

Chapter 1

Elementary differential geometry

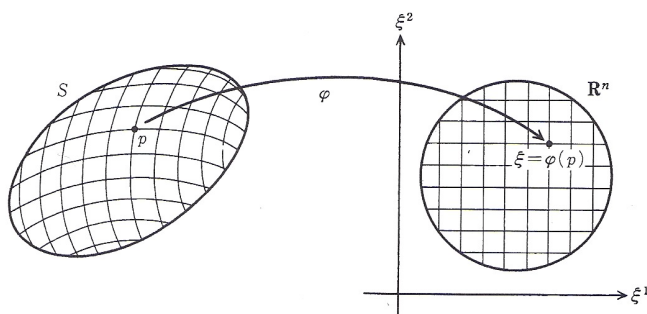
Differential geometry is a mature field of mathematics and has many introductory texts; still, it is not an easy field to master. However, in this book we shall require only the fundamental ideas and methodologies of differential geometry. The main theme of modern differential geometry has been to characterize the global properties of manifolds, and much theory has been developed towards this end. At this time, the field of information geometry (mostly) requires only the theory of the locally characterizable properties of manifolds.

For information geometry the most important aspects of differential geometry are those which allow us to take problems from a variety of fields: statistics, information theory, and control theory; visualize them geometrically; and from this develop novel tools with which to extend and advance these fields. In this chapter we present an introduction to differential geometry from this point of view.

1.1 Differentiable manifolds

A **differentiable manifold** is a mathematical concept denoting a generalization/abstraction of geometric objects such as smooth curves and surfaces in an n -dimensional space. Intuitively, a manifold S is a “set with a coordinate system.” Since S is a set, it has elements. It does not matter what these elements are (these elements are also called the **points** of S .) For example, in this book, we shall introduce manifolds whose points are probability distributions and also those whose points are linear systems. S must also have a **coordinate system**. By this we mean a one-to-one mapping from S (or its subset) to \mathbb{R}^n , which allows us to specify each point in S using a vector of n real numbers (this vector is called the **coordinates** of the corresponding point). We call the natural number n the **dimension** of S , and write $n = \dim S$.

We call a coordinate system that has S as its domain a global coordinate

Figure 1.1: A coordinate system for S .

system. In our analysis below, we shall consider only the case where there exists a global coordinate system. However, in general there are many manifolds which do not have global coordinate systems. Examples of such a manifold include the surface of a sphere and the torus (the surface of a donut). These manifolds have only local coordinate systems. This may be viewed informally in the following way. Consider an open subset U of S , and suppose that U has a coordinate system. This provides a local coordinate system for those points contained in U . For a point not contained in U , consider another open subset V containing that point which also has a coordinate system. Repeat this process until the original set S is covered, so that each point in S is contained in an open subset which has a coordinate system. Then this collection of open subsets of S and their corresponding coordinate systems would allow us to express any point in S using coordinates. However, as mentioned above, in this chapter we shall consider only the case when there exists a global coordinate system. This will suffice to prepare us for the later chapters. Indeed, since in this chapter we principally develop the local theory of manifolds, this assumption does not typically affect the generality of the analysis.

Let S be a manifold and $\varphi : S \rightarrow \mathbb{R}^n$ be a coordinate system for S . Then φ maps each point p in S to n real numbers: $\varphi(p) = [\xi^1(p), \dots, \xi^n(p)] = [\xi^1, \dots, \xi^n]$. These are the coordinates of the point p . Each ξ^i may be viewed as a function $p \rightarrow \xi^i(p)$ which maps a point p to its i^{th} coordinate; we call these n functions $\xi^i : S \rightarrow \mathbb{R}$ ($i = 1, \dots, n$) the **coordinate functions**.¹ We shall write the coordinate system φ in ways such as $\varphi = [\xi^1, \dots, \xi^n] = [\xi^i]$ (Figure 1.1).

Let $\psi = [\rho^i]$ be another coordinate system for S . Then the same point $p \in S$ has both the coordinates $[\xi^i(p)] = [\xi^i] \in \mathbb{R}^n$ with respect to the coordinate system φ , and the coordinates $[\rho^i(p)] = [\rho^i] \in \mathbb{R}^n$ with respect to the coordinate system ψ . The coordinates $[\rho^i]$ may be obtained from $[\xi^i]$ in the following way. First apply the inverse mapping φ^{-1} to $[\xi^i]$; this gives us a point p in S . Then apply ψ to this point; this result is $[\rho^i]$. In other words, we apply the

¹We shall use ξ^i, ρ^i to denote both (the variable representing) the i^{th} coordinate of a point and a coordinate function. This is similar to writing "the function $y = y(x)$."

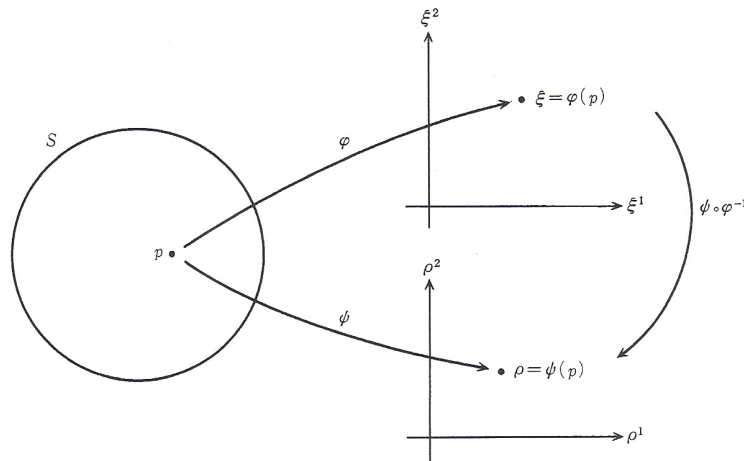


Figure 1.2: Coordinate transformation.

transformation on \mathbb{R}^n given by

$$\psi \circ \varphi^{-1} : [\xi^1, \dots, \xi^n] \mapsto [\rho^1, \dots, \rho^n]. \quad (1.1)$$

This is called the coordinate transformation from $\varphi = [\xi^i]$ to $\psi = [\rho^i]$ (Figure 1.2).

To consider S as a manifold means that one is interested in investigating those properties of S which are invariant under coordinate transformations. In particular, differential geometry analyzes the geometry of objects using differential operators with respect to a variety of functions on S , and it would be problematic if these operators depended fundamentally on the choice of coordinates. Hence it is necessary to restrict the coordinate systems to those which allow smooth transformations between each other.

In order to properly formalize the concepts described above, let us now formally define manifolds for which there exists a global coordinate system.

Let S be a set. If there exists a set of coordinate systems \mathcal{A} for S which satisfies the conditions (i) and (ii) below, we call S (more properly, (S, \mathcal{A})) an n -dimensional C^∞ **differentiable manifold**, or more simply, a **manifold**.

- (i) Each element φ of \mathcal{A} is a one-to-one mapping from S to some open subset of \mathbb{R}^n .
- (ii) For all $\varphi \in \mathcal{A}$, given any one-to-one mapping ψ from S to \mathbb{R}^n , the following holds:

$$\psi \in \mathcal{A} \iff \psi \circ \varphi^{-1} \text{ is a } C^\infty \text{ diffeomorphism.}$$

Here, by a C^∞ **diffeomorphism** we mean that $\psi \circ \varphi^{-1}$ and its inverse $\varphi \circ \psi^{-1}$ are both C^∞ (infinitely many times differentiable). From these conditions, and

given the coordinate transformation described in Equation (1.1), it follows that we may take the partial derivative of the function $\rho^i = \rho^i(\xi^1, \dots, \xi^n)$ with respect to its variable arguments as many times as needed, and that the same holds for $\xi^i = \xi^i(\rho^1, \dots, \rho^n)$. In this book, the condition C^∞ is used a number of times, but in fact it is usually not necessary; it would suffice for the relevant functions to be differentiable some appropriate number of times. Intuitively, then, we may consider C^∞ to simply mean “sufficiently smooth”.

Let S be a manifold and φ be a coordinate system for S . Let U be a subset of S . If the image $\varphi(U)$ is an open subset of \mathbb{R}^n , then we say that U is an open subset of S . From condition (ii) above, we see that this property is invariant over the choice of coordinate system φ . This allows us to consider S as a topological space. For any non-empty open subset U of S , we may restrict φ , the coordinate system of S , to obtain $\varphi|_U$ (the mapping $U \rightarrow \mathbb{R}^n$ obtained by restricting the domain of φ to U), which may be taken as a coordinate system for U . Hence we see that U is a manifold whose dimension is the same as that of S .

Let $f : S \rightarrow \mathbb{R}$ be a function on a manifold S . Then if we select a coordinate system $\varphi = [\xi^i]$ for S , this function may be rewritten as a function of the coordinates; i.e., letting $[\xi^i]$ denote the coordinates of the point p , we have $f(p) = \bar{f}(\xi^1, \dots, \xi^n)$, where $\bar{f} = f \circ \varphi^{-1}$. Note that \bar{f} is a real-valued function whose domain is $\varphi(S)$, an open subset of \mathbb{R}^n . Now suppose that $\bar{f}(\xi^1, \dots, \xi^n)$ is partially differentiable at each point in $\varphi(S)$. Then the partial derivative $\frac{\partial}{\partial \xi^i} \bar{f}(\xi^1, \dots, \xi^n)$ is also a function on $\varphi(S)$. By transforming the domain back to S , we may define the partial derivatives of f to be $\frac{\partial f}{\partial \xi^i} \stackrel{\text{def}}{=} \frac{\partial \bar{f}}{\partial \xi^i} \circ \varphi : S \rightarrow \mathbb{R}$. We write $\left(\frac{\partial f}{\partial \xi^i}\right)_p$ to denote the ‘value of this function at point p (the partial derivative at point p).

When $\bar{f} = f \circ \varphi^{-1}$ is C^∞ , in other words when $\bar{f}(\xi^1, \dots, \xi^n)$ can be partially differentiated with respect to its variables an unbounded number of times, we call f a **C^∞ function** on S . This definition does not depend on the choice of coordinate system φ . The partial derivatives $\frac{\partial f}{\partial \xi^i}$ of a C^∞ function f are also C^∞ functions. We may similarly define the higher-order partial derivatives, e.g. $\frac{\partial^2 f}{\partial \xi^j \partial \xi^i} = \frac{\partial}{\partial \xi^j} \frac{\partial f}{\partial \xi^i}$. These will also be C^∞ . As with the case of C^∞ functions on \mathbb{R}^n , $\frac{\partial^2 f}{\partial \xi^j \partial \xi^i} = \frac{\partial}{\partial \xi^j} \frac{\partial f}{\partial \xi^i}$ holds.

Let us denote the class of C^∞ functions on S by $\mathcal{F}(S)$, or simply \mathcal{F} . For all f and g in \mathcal{F} and a real number c , we define the sum $f + g$ as $(f + g)(p) = f(p) + g(p)$, the scaling cf as $(cf)(p) = cf(p)$, and the product $f \cdot g$ as $(f \cdot g)(p) = f(p) \cdot g(p)$; these functions are also members of \mathcal{F} .

Let $[\xi^i]$ and $[\rho^j]$ be two coordinate systems. Since the coordinate functions ξ^i and ρ^j are clearly C^∞ , the partial derivatives $\frac{\partial \xi^i}{\partial \rho^j}$ and $\frac{\partial \rho^j}{\partial \xi^i}$ are well defined, and they satisfy

$$\sum_{j=1}^n \frac{\partial \xi^i}{\partial \rho^j} \frac{\partial \rho^j}{\partial \xi^k} = \sum_{j=1}^n \frac{\partial \rho^j}{\partial \xi^i} \frac{\partial \xi^i}{\partial \rho^k} = \delta_k^i, \quad (1.2)$$

where δ_k^i is 1 if $k = i$, and 0 otherwise (the Kronecker delta). In addition, for

any C^∞ function f , we have

$$\frac{\partial f}{\partial \rho^j} = \sum_{i=1}^n \frac{\partial \xi^i}{\partial \rho^j} \frac{\partial f}{\partial \xi^i} \quad \text{and} \quad \frac{\partial f}{\partial \xi^i} = \sum_{j=1}^n \frac{\partial \rho^j}{\partial \xi^i} \frac{\partial f}{\partial \rho^j}. \quad (1.3)$$

Note: In this book there often appear equations which contain indices such as i, j, \dots , and are to be summed over those indices that are both super and subscripted. For these equations we shall abbreviate by omitting the summation sign \sum corresponding to these indices. For example, Equations (1.2) and (1.3) above would be written as

$$\frac{\partial \xi^i}{\partial \rho^j} \frac{\partial \rho^j}{\partial \xi^k} = \frac{\partial \rho^i}{\partial \xi^j} \frac{\partial \xi^j}{\partial \rho^k} = \delta_k^i$$

$$\frac{\partial f}{\partial \rho^j} = \frac{\partial \xi^i}{\partial \rho^j} \frac{\partial f}{\partial \xi^i}, \quad \frac{\partial f}{\partial \xi^i} = \frac{\partial \rho^j}{\partial \xi^i} \frac{\partial f}{\partial \rho^j}.$$

We shall also abbreviate $\sum_{i=1}^n \sum_{j=1}^n A_{jk}^{ij} B_i^h$ as $A_{jk}^{ij} B_i^h$. Hence (unless there is ambiguity), whenever there appears such an equation we shall assume that there is an implicit \sum (i.e., there is a summation over the relevant indices). Note therefore that $A_j^i X^j = A_k^i X^k$, for instance, is always true. This notation is known as Einstein's convention.

Let S and Q be manifolds with coordinate systems $\varphi : S \rightarrow \mathbb{R}^n$ and $\psi : Q \rightarrow \mathbb{R}^m$. A mapping $\lambda : S \rightarrow Q$ is said to be C^∞ or smooth if $\psi \circ \lambda \circ \varphi^{-1}$ is a C^∞ mapping from an open subset of \mathbb{R}^n to \mathbb{R}^m . A necessary and sufficient condition for λ to be C^∞ is that $f \circ \lambda \in \mathcal{F}(S)$ for all $f \in \mathcal{F}(Q)$. If a C^∞ mapping λ is a bijection (i.e., one-to-one and $\lambda(S) = Q$) and the inverse λ^{-1} is also C^∞ , then λ is called a C^∞ diffeomorphism from S onto Q .

1.2 Tangent vectors and tangent spaces

The tangent space T_p at a point $p \in S$ of a manifold S is intuitively the vector space obtained by "locally linearizing" S around p . Let $[\xi^i]$ be some coordinate system for S , and let e_i denote the "tangent vector" which goes through point p and is parallel to the i^{th} coordinate curve (coordinate axis). By the i^{th} coordinate curve we mean the curve which is obtained by fixing the values of all ξ^j for $j \neq i$ and varying only the value of ξ^i . The n -dimensional space spanned by the n tangent vectors e_1, \dots, e_n is the tangent space T_p at point p (Figure 1.3). Let p' be a point "very close" to p , and let $[\xi^i]$ and $[\xi^i + d\xi^i]$ (where $d\xi^i$ is an infinitesimal) be the coordinates of p and p' , respectively. Then the segment joining these two points may be described by $\overrightarrow{pp'} = d\xi^i e_i$, an infinitesimal vector in T_p .

Let us make the above concepts more precise. To do so, we must first formally define what we mean by curves and the tangent vector of curves on a

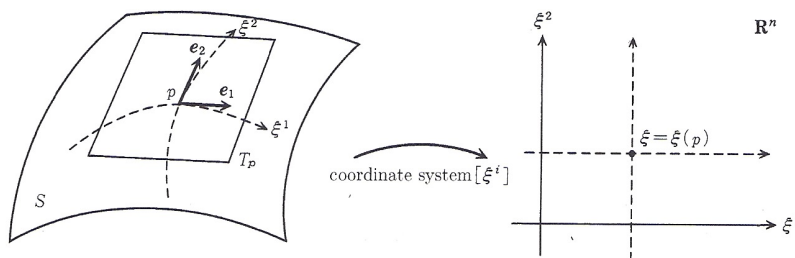


Figure 1.3: Tangent Space

manifold. Consider a one-to-one function $\gamma : I \rightarrow S$ from some interval $I \subset \mathbb{R}$ to S . By defining $\gamma^i(t) \stackrel{\text{def}}{=} \xi^i(\gamma(t))$ we may express the point $\gamma(t)$ ($t \in I$) using coordinates as $\bar{\gamma}(t) = [\gamma^1(t), \dots, \gamma^n(t)]$. If $\bar{\gamma}(t)$ is C^∞ for $t \in I$, we call γ a C^∞ **curve** on S . This definition is independent of coordinate system choice.

Now, given a curve γ and a point $\gamma(a) = p$, let us consider what is meant by the “derivative” of γ at p , or alternatively the “tangent vector” $\left(\frac{d\gamma}{dt}\right)_p = \dot{\gamma}(a)$. When S is simply an open subset of \mathbb{R}^n , or can be embedded smoothly into \mathbb{R}^ℓ ($\ell \geq n$), the range of γ is contained within a single linear space, and hence it suffices to consider the standard derivative

$$\dot{\gamma}(a) = \lim_{h \rightarrow 0} \frac{\gamma(a+h) - \gamma(a)}{h}. \quad (1.4)$$

In general, however, the equation above is not meaningful. On the other hand, if we take a C^∞ function $f \in \mathcal{F}$ on S and consider the value of $f(\gamma(t))$ on the curve, since this is a real-valued function, we may define the derivative $\frac{d}{dt}f(\gamma(t))$ in the usual way. Using coordinates, we have $f(\gamma(t)) = \bar{f}(\bar{\gamma}(t)) = \bar{f}(\gamma^1(t), \dots, \gamma^n(t))$, and the derivatives may be rewritten as

$$\frac{d}{dt}f(\gamma(t)) = \left(\frac{\partial \bar{f}}{\partial \xi^i}\right)_{\bar{\gamma}(t)} \frac{d\gamma^i(t)}{dt} = \left(\frac{\partial f}{\partial \xi^i}\right)_{\gamma(t)} \frac{d\gamma^i(t)}{dt}. \quad (1.5)$$

We call this the directional derivative of f along the curve γ . Let us consider this directional derivative as an expression of the tangent vector of γ . In other words, we take the operator $\mathcal{F} \rightarrow \mathbb{R}$ which maps $f \in \mathcal{F}$ to $\frac{d}{dt}f(\gamma(t))|_{t=a}$, and simply define the tangent vector $\left(\frac{d\gamma}{dt}\right)_p = \dot{\gamma}(a)$ to be this operator. Then we may rewrite Equation (1.5) as

$$\dot{\gamma}(a) = \left(\frac{d\gamma}{dt}\right)_p = \dot{\gamma}^i(a) \left(\frac{\partial}{\partial \xi^i}\right)_p \quad (1.6)$$

($\dot{\gamma}^i(a) = \frac{d}{dt}\gamma^i(t)|_{t=a}$). Here $\left(\frac{\partial}{\partial \xi^i}\right)_p$ is an operator which maps $f \mapsto \left(\frac{\partial f}{\partial \xi^i}\right)_p$. It is possible to show that when the tangent vectors can be defined using Equation (1.4), there is a natural one-to-one correspondence between Equations (1.4)

and (1.6). Hence the definition of tangent vectors as operators may be viewed as a generalization of Equation (1.4).

Since a partial derivative is simply a directional derivative along a coordinate axis, the operator $\left(\frac{\partial}{\partial \xi^i}\right)_p$ is the tangent vector at point p of the i^{th} coordinate curve. The e_i mentioned previously corresponds to this $\left(\frac{\partial}{\partial \xi^i}\right)_p$. From Equation (1.3), we see that

$$\left(\frac{\partial}{\partial \rho^j}\right)_p = \left(\frac{\partial \xi^i}{\partial \rho^j}\right)_p \left(\frac{\partial}{\partial \xi^i}\right)_p \quad \text{and} \quad \left(\frac{\partial}{\partial \xi^i}\right)_p = \left(\frac{\partial \rho^j}{\partial \xi^i}\right)_p \left(\frac{\partial}{\partial \rho^j}\right)_p. \quad (1.7)$$

Consider all curves which pass through the point p . We denote the set of all tangent vectors corresponding to these curves by T_p , or $T_p(S)$. From Equation (1.6), we see that

$$T_p(S) = \left\{ c^i \left(\frac{\partial}{\partial \xi^i}\right)_p \mid [c^1, \dots, c^n] \in \mathbb{R}^n \right\}. \quad (1.8)$$

This forms a linear space, and since the operators $\left\{ \left(\frac{\partial}{\partial \xi^i}\right)_p ; i = 1, \dots, n \right\}$ are clearly linearly independent, the dimension of this space is n ($= \dim S$). We call $T_p(S)$ and its elements the **tangent space** and **tangent vectors**, of S at the point p , respectively. In addition, we call $\left(\frac{\partial}{\partial \xi^i}\right)_p$ the **natural basis** of the coordinate system $[\xi^i]$.

Let $D \in T_p$ be some tangent vector. Then for all $f, g \in \mathcal{F}$ and all $a, b \in \mathbb{R}$, D satisfies the following:

$$\text{[Linearity]} \quad D(af + bg) = aD(f) + bD(g). \quad (1.9)$$

$$\text{[Leibniz's rule]} \quad D(f \cdot g) = f(p)D(g) + g(p)D(f). \quad (1.10)$$

Conversely, it can be shown that any operator $D : \mathcal{F} \rightarrow \mathbb{R}$ satisfying these properties is an element of T_p . Hence, it is possible to define tangent vectors in terms of these properties.

Let $\lambda : S \rightarrow Q$ be a smooth mapping from a manifold S to another manifold Q . Given a tangent vector $D \in T_p(S)$ of S , the mapping $D' : \mathcal{F}(Q) \rightarrow \mathbb{R}$ defined by $D'(f) = D(f \circ \lambda)$ satisfies Equations (1.9) (1.10) with p replaced with $\lambda(p)$, and hence D' belongs to $T_{\lambda(p)}(Q)$. Representing this correspondence as $D' = (d\lambda)_p(D)$, we may define a linear mapping $(d\lambda)_p : T_p(S) \rightarrow T_{\lambda(p)}(Q)$, which is called the **differential** of λ at p . When S and Q are provided with coordinate systems $[\xi^i]$ and $[\rho^j]$ respectively, we have

$$(d\lambda)_p \left(\left(\frac{\partial}{\partial \xi^i}\right)_p \right) = \left(\frac{\partial(\rho^j \circ \lambda)}{\partial \xi^i} \right)_p \left(\frac{\partial}{\partial \rho^j}\right)_{\lambda(p)}. \quad (1.11)$$

Moreover, for any curve $\gamma(t)$ on S passing through the point p it follows that

$$(d\lambda)_p \left(\left(\frac{d\gamma}{dt}\right)_p \right) = \left(\frac{d(\lambda \circ \gamma)}{dt} \right)_{\lambda(p)}. \quad (1.12)$$

1.3 Vector fields and tensor fields

Let $X : p \mapsto X_p$ be a mapping which maps each point p in the manifold S to a tangent vector $X_p \in T_p(S)$. We call such a mapping a **vector field**. For example, if $[\xi^i]$ is a coordinate system, then we may define n vector fields through the mappings $\frac{\partial}{\partial \xi^i} : p \mapsto \left(\frac{\partial}{\partial \xi^i}\right)_p$ ($i = 1, \dots, n$). These are the vector fields formed by the natural basis. Below, we shall write ∂_i to mean $\frac{\partial}{\partial \xi^i}$. In general, given a vector field X , for each point p there exists n real numbers $\{X_p^1, \dots, X_p^n\}$ which uniquely determine $X_p = X_p^i(\partial_i)_p$. Hence we may define the functions $X^i : p \mapsto X_p^i$ on S . We call the n functions $\{X^1, \dots, X^n\}$ the **components** of X with respect to $[\xi^i]$. This allows us to write $X = X^i \partial_i$. If, in addition, we let $[\rho^j]$ be another coordinate system and $X = \tilde{X}^j \tilde{\partial}_j$ ($\tilde{\partial}_j \stackrel{\text{def}}{=} \frac{\partial}{\partial \rho^j}$) be the component expression of X with respect to $[\rho^j]$, then the following hold:

$$\tilde{X}^j = X^i \frac{\partial \rho^j}{\partial \xi^i} \quad \text{and} \quad X^i = \tilde{X}^j \frac{\partial \xi^i}{\partial \rho^j}. \quad (1.13)$$

If the components of a vector field are C^∞ with respect to some coordinate system, then the components are C^∞ with respect to any other. We call such a vector field a C^∞ vector field. Since we consider only C^∞ vector fields in this book, we shall refer to them as simply vector fields. We shall denote this family of vector fields by $\mathcal{T}(S)$, or simply \mathcal{T} . Clearly $\partial_i \in \mathcal{T}$ ($i = 1, \dots, n$).

Now for any $X, Y \in \mathcal{T}$ and any $c \in \mathbb{R}$, the mappings $X + Y : p \mapsto X_p + Y_p$ and $cX : p \mapsto cX_p$ are also members of \mathcal{T} . Hence \mathcal{T} is a linear space. In addition, for any $f \in \mathcal{F}$, the mapping $fX : p \mapsto f(p)X_p$ is a member of \mathcal{T} .

We call $F : V_1 \times V_2 \times \dots \times V_r \rightarrow W$, where V_1, \dots, V_r, W are linear spaces, a **multilinear mapping** if the following property holds. Let $\tilde{F}(v_i)$ denote a mapping of one variable equal to $F(v_1, \dots, v_r)$ where some v_i has been distinguished as the variable, and the other v_j ($j \neq i$) are held constant to some value ($\in V_j$). Then $\tilde{F} : v_i \mapsto \tilde{F}(v_i)$ is a linear mapping from V_i to W .

Now for each point $p \in S$, let $[T_p]_r^0$ denote the family of multilinear mappings of the form $\underbrace{T_p \times \dots \times T_p}_r \rightarrow \mathbb{R}$, and let $[T_p]_r^1$ denote the family of the form

$\underbrace{T_p \times \dots \times T_p}_{r \text{ direct products}} \rightarrow T_p$. We call mappings $A : p \mapsto A_p$ which maps each point

p in S to some element A_p of $[T_p]_r^q$ ($q = 0, 1$) a **tensor field of type (q, r)** on S . The types $(0, r)$ and $(1, r)$ are also respectively called **tensor fields of covariant degree r** and **tensor fields of contravariant degree 1 and covariant degree r** . Vector fields may be considered to be tensor fields of type $(1, 0)$. Although it is possible to define tensor fields of type (q, r) for $q = 2, 3, \dots$, they will not be used in this book. In addition, we shall occasionally refer to tensor fields as simply tensors.

Let A be a tensor field of type (q, r) and X_1, \dots, X_r be r vector fields. Then

we may consider a mapping with domain S of the following form:

$$A(X_1, \dots, X_r) : p \mapsto A_p((X_1)_p, \dots, (X_r)_p). \quad (1.14)$$

When $q = 0$, $A_p((X_1)_p, \dots, (X_r)_p) \in \mathbb{R}$ and hence this mapping is a real-valued function on S . When $q = 1$, $A_p((X_1)_p, \dots, (X_r)_p) \in T_p$, and hence this defines a vector field on S . Given A , if for all C^∞ vector fields $X_1, \dots, X_r \in \mathcal{T}$ the mapping $A(X_1, \dots, X_r)$ is C^∞ (i.e., when $q = 0$ the mapping is in \mathcal{F} , and when $q = 1$ it is in \mathcal{T}), we call A a C^∞ tensor field. Below, we consider only C^∞ tensor fields, and shall simply call them tensor fields.

Consider the tensor field A of type (q, r) to be a mapping $(X_1, \dots, X_r) \mapsto A(X_1, \dots, X_r)$. Then when $q = 0$ we have $A : \underbrace{\mathcal{T} \times \dots \times \mathcal{T}}_{r \text{ direct products}} \rightarrow \mathcal{F}$, and when

$q = 1$ we have $A : \underbrace{\mathcal{T} \times \dots \times \mathcal{T}}_{r \text{ direct products}} \rightarrow \mathcal{T}$. This, in addition to forming a multilinear

mapping, has the following property: for all $f_1, \dots, f_r \in \mathcal{F}$,

$$A(f_1 X_1, \dots, f_r X_r) = f_1 \cdots f_r A(X_1, \dots, X_r).$$

We call this the \mathcal{F} -**multilinearity** of A . Conversely, if the mapping $A : \mathcal{T} \times \dots \times \mathcal{T} \rightarrow \mathcal{F}$, or alternatively $A : \mathcal{T} \times \dots \times \mathcal{T} \rightarrow \mathcal{T}$ is \mathcal{F} -multilinear, then this determines a tensor field $p \mapsto A_p$ satisfying Equation (1.14).

The operation of a tensor field A of type $(0, r)$ on the r basis vector fields $\partial_{i_1}, \dots, \partial_{i_r}$ ($\partial_i \stackrel{\text{def}}{=} \frac{\partial}{\partial \xi^i}$) defines a function. Let us denote this by

$$A(\partial_{i_1}, \dots, \partial_{i_r}) = A_{i_1 \dots i_r}.$$

We call the n^r functions $\{A_{i_1 \dots i_r}\}$ obtained by changing the values of i_1, \dots, i_r the **components** of A with respect to the coordinate system $[\xi^i]$. Let X_1, \dots, X_r be r vector fields; these may be expressed component-wise as $X_j = X_j^i \partial_i$. Then from \mathcal{F} -multilinearity, we have

$$A(X_1, \dots, X_r) = A_{i_1 \dots i_r} X_1^{i_1} \cdots X_r^{i_r}$$

In the case of a tensor field A of type $(1, r)$, $A(\partial_{i_1}, \dots, \partial_{i_r})$ is a vector field, and its component expression is given by

$$A(\partial_{i_1}, \dots, \partial_{i_r}) = A_{i_1 \dots i_r}^k \partial_k.$$

The n^{r+1} functions $\{A_{i_1 \dots i_r}^k\}$ thus defined are called the components of A with respect to $[\xi^i]$. As in the previous case, letting $X_j = X_j^i \partial_i$, the following holds:

$$A(X_1, \dots, X_r) = (A_{i_1 \dots i_r}^k X_1^{i_1} \cdots X_r^{i_r}) \partial_k.$$

Let $[\rho^j]$ be another coordinate system. Using \sim to denote components with respect to $[\rho^j]$, we have

$$\tilde{A}_{j_1 \dots j_r} = A_{i_1 \dots i_r} \left(\frac{\partial \xi^{i_1}}{\partial \rho^{j_1}} \right) \cdots \left(\frac{\partial \xi^{i_r}}{\partial \rho^{j_r}} \right) \quad \text{and} \quad (1.15)$$

$$\tilde{A}_{j_1 \dots j_r}^\ell = A_{i_1 \dots i_r}^k \left(\frac{\partial \xi^{i_1}}{\partial \rho^{j_1}} \right) \cdots \left(\frac{\partial \xi^{i_r}}{\partial \rho^{j_r}} \right) \left(\frac{\partial \rho^\ell}{\partial \xi^k} \right). \quad (1.16)$$

1.4 Submanifolds

Let S and M be manifolds, where M is a subset of S . Let $[\xi^1, \dots, \xi^n] = [\xi^i]$ and $[u^1, \dots, u^m] = [u^a]$ be coordinate systems for S and M , respectively, where $n = \dim S$ and $m = \dim M$. Below, we shall use the indices i, j, k, \dots over $\{1, \dots, n\}$ for S and a, b, c, \dots over $\{1, \dots, m\}$ for M .

We call M a **submanifold** of S if the following conditions (i), (ii), and (iii) hold.

- (i) The restriction $\xi^i|_M$ of each $\xi^i : S \rightarrow \mathbb{R}$ to M , is a C^∞ function on M .
- (ii) Let $B_a^i \stackrel{\text{def}}{=} \left(\frac{\partial \xi^i}{\partial u^a} \right)_p$ (more precisely, $\left(\frac{\partial \xi^i|_M}{\partial u^a} \right)_p$) and $B_a \stackrel{\text{def}}{=} [B_a^1, \dots, B_a^n] \in \mathbb{R}^n$. Then for each point p in M , $\{B_1, \dots, B_m\}$ are linearly independent (hence $m \leq n$).
- (iii) For any open subset W of M , there exists U , an open subset of S , such that $W = M \cap U$.

These conditions are independent of the choice of coordinate systems $[\xi^i]$ and $[u^a]$. Indeed, conditions (i) and (ii) mean that the embedding $\iota : M \rightarrow S$ defined by $\iota(p) = p, \forall p \in M$, is a C^∞ mapping and that its differential $(d\iota)_p$ is nondegenerate at each point p .

An open subset of S , as we noted in §1.1, forms an n -dimensional manifold; in addition, it is also a submanifold of S . We may construct an example of a submanifold of dimension m ($< n$) in the following way. Let $[\xi^i]$ be a coordinate system of S and $\{c^{m+1}, \dots, c^n\}$ be $n - m$ real numbers. Now define

$$M \stackrel{\text{def}}{=} \{p \in S \mid \xi^i(p) = c^i, m+1 \leq i \leq n\}.$$

We assume that $M \neq \emptyset$ (the empty set). Then if we let $u^a \stackrel{\text{def}}{=} \xi^a|_M$ ($1 \leq a \leq m$), M is an m -dimensional manifold with coordinate system $[u^a]$, and hence it is a submanifold of S . The "reverse" of this is also true at least locally. In other words, if M is an m -dimensional submanifold of S , with $[u^a]$ its coordinate system, and $\{c^{m+1}, \dots, c^n\}$ is a set of $n - m$ real numbers, then it is possible to choose U , an open subset of S , and a coordinate system $[\xi^i]$, so that

$$M \cap U = \{p \in U \mid \xi^i(p) = c^i, m+1 \leq i \leq n\},$$

and moreover, $u^a|_{M \cap U} = \xi^a|_{M \cap U}$ ($1 \leq a \leq m$).

If M is a submanifold of S then a curve $\gamma : t \mapsto \gamma(t)$ in M is also a curve in S . Hence letting p be a point on γ , the tangent vector $\left(\frac{d\gamma}{dt} \right)_p$ of γ may be considered both as an element of $T_p(M)$ and as one of $T_p(S)$. Using coordinate systems $[u^a]$ and $[\xi^i]$ for M and S , respectively, and letting $\gamma^a \stackrel{\text{def}}{=} u^a \circ \gamma$ and $\gamma^i \stackrel{\text{def}}{=} \xi^i \circ \gamma$, these tangent vectors may be written as $\left(\frac{d\gamma^a}{dt} \right)_p (\partial_a)_p \in T_p(M)$

and $\left(\frac{d\gamma^i}{dt}\right)_p (\partial_i)_p \in T_p(S)$, where $\partial_a \stackrel{\text{def}}{=} \frac{\partial}{\partial u^a}$ and $\partial_i \stackrel{\text{def}}{=} \frac{\partial}{\partial \xi^i}$. Since

$$\left(\frac{d\gamma^i}{dt}\right)_p = \left(\frac{\partial \xi^i}{\partial u^a}\right)_p \left(\frac{d\gamma^a}{dt}\right)_p, \quad (1.17)$$

from condition (ii) for submanifolds we see that there is a one-to-one correspondence between these tangent vectors. In other words, this correspondence is given by the differential $(d\iota)_p$ of the embedding $\iota: M \rightarrow S$. By considering the corresponding pairs to be equivalent, we may view $T_p(M)$ as a linear subspace of $T_p(S)$. From Equation (1.17) we obtain

$$\left(\frac{\partial}{\partial u^a}\right)_p = \left(\frac{\partial \xi^i}{\partial u^a}\right)_p \left(\frac{\partial}{\partial \xi^i}\right)_p = B_a^i \partial_i. \quad (1.18)$$

This shows that $B_a^i \partial_i$ is the natural basis vector ∂_a of M with respect to coordinate system $[u^a]$ seen as a vector in $T_p(S)$. In addition, this may be interpreted as the equality of the differential operators: for all $f \in \mathcal{F}(S)$, $\left(\frac{\partial f}{\partial u^a}\right)_p = \left(\frac{\partial \xi^i}{\partial u^a}\right)_p \left(\frac{\partial f}{\partial \xi^i}\right)_p$.

1.5 Riemannian metrics

Let S be a manifold. For each point p in S , let us assume that an inner product $\langle \cdot, \cdot \rangle_p$ has been defined on the tangent space $T_p(S)$. In other words, for any tangent vectors $D, D' \in T_p(S)$ we have $\langle D, D' \rangle_p \in \mathbb{R}$, and the following hold.

$$\begin{aligned} \text{[Linearity]} \quad \langle aD + bD', D'' \rangle_p &= a \langle D, D'' \rangle_p + b \langle D', D'' \rangle_p \\ &(\forall a, b \in \mathbb{R}) \end{aligned} \quad (1.19)$$

$$\text{[Symmetry]} \quad \langle D, D' \rangle_p = \langle D', D \rangle_p \quad (1.20)$$

$$\text{[Positive-definiteness]} \quad \text{If } D \neq 0 \text{ then } \langle D, D \rangle_p > 0 \quad (1.21)$$

Note that $\langle \cdot, \cdot \rangle_p \in [T_p(S)]_2^0$ since from Equations (1.19) and (1.20) we see that $\langle \cdot, \cdot \rangle_p$ is a bilinear form. Hence the mapping from points p in S to their inner product on $T_p(S)$, say $g: p \mapsto \langle \cdot, \cdot \rangle_p$, is a tensor field of covariant degree 2. We call this a (C^∞) **Riemannian metric** on S . Such a metric, g , is not naturally determined by the structure of S as a manifold; it is possible to consider an infinite number of Riemannian metrics on S . Given a Riemannian metric g on S , we call S (more precisely (S, g)) a **Riemannian manifold**.

Let $[\xi^i]$ be a coordinate system for S , and let $\partial_i \stackrel{\text{def}}{=} \frac{\partial}{\partial \xi^i}$. Then the components $\{g_{ij}; i, j = 1, \dots, n\}$ ($n = \dim S$) of a Riemannian metric g with respect to $[\xi^i]$ are determined by $g_{ij} = \langle \partial_i, \partial_j \rangle$. This is a C^∞ function which maps each point p in S to $g_{ij}(p) = \langle (\partial_i)_p, (\partial_j)_p \rangle_p$. If we rewrite the tangent vectors $D, D' \in T_p$ in terms of their coordinates as $D = D^i (\partial_i)_p$ and $D' = D'^j (\partial_j)_p$, their inner product may then be written as:

$$\langle D, D' \rangle_p = g_{ij}(p) D^i D'^j.$$

Also, the length $\|D\|$ of the tangent vector D is given by

$$\|D\|^2 = \langle D, D \rangle_p = g_{ij}(p) D^i D^j.$$

If we let $G(p) = [g_{ij}(p)]$ be an $n \times n$ matrix whose $(i, j)^{\text{th}}$ element is $g_{ij}(p)$, we see from Equations (1.20) and (1.21) that this is a positive definite symmetric matrix. Conversely, suppose we are given a coordinate system $[\xi^i]$ for an n -dimensional manifold S , and n^2 C^∞ functions $\{g_{ij}\} (\subset \mathcal{F}(S))$. Then if $G(p) = [g_{ij}(p)]$ is a positive definite symmetric matrix for every point $p \in S$, the corresponding Riemannian metric on S which has g_{ij} as its components with respect to $[\xi^i]$ is uniquely determined. The relationship between these components and the components $\tilde{g}_{k\ell} = \langle \tilde{\partial}_k, \tilde{\partial}_\ell \rangle$ ($\tilde{\partial}_k \stackrel{\text{def}}{=} \frac{\partial}{\partial \rho^k}$) with respect to a different coordinate system $[\rho^k]$ is given by the following transformations of covariant tensor fields of order 2 (refer to Equation (1.15):

$$\tilde{g}_{k\ell} = g_{ij} \left(\frac{\partial \xi^i}{\partial \rho^k} \right) \left(\frac{\partial \xi^j}{\partial \rho^\ell} \right) \quad \text{and} \quad g_{ij} = \tilde{g}_{k\ell} \left(\frac{\partial \rho^k}{\partial \xi^i} \right) \left(\frac{\partial \rho^\ell}{\partial \xi^j} \right). \quad (1.22)$$

Let $g^{ij}(p)$ be the $(i, j)^{\text{th}}$ component of the inverse $G(p)^{-1}$ of $G(p) = [g_{ij}(p)]$ (this inverse is also positive definite symmetric). Now define the function $g^{ij} : p \mapsto g^{ij}(p)$ on S . Then

$$g_{ij} g^{jk} = \delta_i^k = \begin{cases} 1 & (k = i) \\ 0 & (k \neq i) \end{cases}, \quad (1.23)$$

and the relationship between this inverse and $\tilde{G}(p)^{-1} = [\tilde{g}^{k\ell}(p)]$, which is the inverse of $\tilde{G}(p) = [\tilde{g}_{k\ell}(p)]$, is given by the following.

$$\tilde{g}^{k\ell} = g^{ij} \left(\frac{\partial \rho^k}{\partial \xi^i} \right) \left(\frac{\partial \rho^\ell}{\partial \xi^j} \right) \quad \text{and} \quad g^{ij} = \tilde{g}^{k\ell} \left(\frac{\partial \xi^i}{\partial \rho^k} \right) \left(\frac{\partial \xi^j}{\partial \rho^\ell} \right). \quad (1.24)$$

Let $\gamma : [a, b] \rightarrow S$ be a curve in the Riemannian manifold S . We define its length $\|\gamma\|$ to be

$$\|\gamma\| \stackrel{\text{def}}{=} \int_a^b \left\| \frac{d\gamma}{dt} \right\| dt = \int_a^b \sqrt{g_{ij} \dot{\gamma}^i \dot{\gamma}^j} dt, \quad (1.25)$$

where $\dot{\gamma}^i$ is the derivative of $\gamma^i \stackrel{\text{def}}{=} \xi^i \circ \gamma$ (see Equation (1.6).)

Let M be a submanifold of a Riemannian manifold S . As noted in §1.4, for each point $p \in M$, we may view $T_p(M)$ as a linear subspace of $T_p(S)$, and hence an inner product $g(p) = \langle \cdot, \cdot \rangle_p$ on $T_p(S)$ naturally defines an inner product on $T_p(M)$. Then, letting $g|_M(p)$ denote this inner product, $g|_M : p \mapsto g|_M(p)$ is a Riemannian metric on M . Given a coordinate system $[u^a]$ on M , we see from Equation (1.18) that the components of $g|_M$, $\{g_{ab}\}$ satisfy

$$g_{ab} = \left\langle \frac{\partial}{\partial u^a}, \frac{\partial}{\partial u^b} \right\rangle = g_{ij} \left(\frac{\partial \xi^i}{\partial u^a} \right) \left(\frac{\partial \xi^j}{\partial u^b} \right). \quad (1.26)$$

1.6 Affine connections and covariant derivatives

Let S be an n -dimensional manifold. If S is an open subset of \mathbb{R}^n , then by defining the tangent vector of a curve γ according to Equation (1.4), the tangent space $T_p = T_p(S)$ at each point $p \in S$ may be considered equivalent to \mathbb{R}^n . This means that for p and q not equal, there is still a natural correspondence between T_p and T_q . For a general manifold S , however, T_p and T_q are entirely different spaces when $p \neq q$. Hence, to consider relationships between T_p and T_q , we must somehow augment the structure of S as a manifold. Affine connections are such a structural augmentation.

Intuitively, defining an affine connection on a manifold S means that for each point p in S and its "neighbor" p' , we define a linear one-to-one mapping between T_p and $T_{p'}$. Here we call p' a neighbor of p if, given a coordinate system $[\xi^i]$ of S , the difference between the coordinates of p and p' , $d\xi^i \stackrel{\text{def}}{=} \xi^i(p') - \xi^i(p)$, when construed as a first-order infinitesimal, is sufficiently small that we may ignore the second-order infinitesimals $(d\xi^i)(d\xi^j)$. Below we shall introduce the notion of affine connections in an intuitive manner using infinitesimals. (It is possible to formalize this discussion by using fiber bundles.)

As shown in Figure 1.4, in order to establish a linear mapping $\Pi_{p,p'}$ between T_p and $T_{p'}$ we must specify, for each $j \in \{1, \dots, n\}$, how to express $\Pi_{p,p'}((\partial_j)_p)$ in terms of a linear combination of $\{(\partial_1)_{p'}, \dots, (\partial_n)_{p'}\}$ ($\partial_j \stackrel{\text{def}}{=} \frac{\partial}{\partial \xi^j}$). Let us assume that the difference between $\Pi_{p,p'}((\partial_j)_p)$ and $(\partial_j)_{p'}$ is an infinitesimal, and that it may be expressed as a linear combination of $\{d\xi^1, \dots, d\xi^n\}$. Then we have

$$\Pi_{p,p'}((\partial_j)_p) = (\partial_j)_{p'} - d\xi^i (\Gamma_{ij}^k)_p (\partial_k)_{p'}, \quad (1.27)$$

where $\{(\Gamma_{ij}^k)_p; i, j, k = 1, \dots, n\}$ are n^3 real numbers which depend on the point p .

If for each pair of neighboring points p and p' in S , there is defined a linear mapping $\Pi_{p,p'} : T_p \rightarrow T_{p'}$ of the form described in Equation (1.27), and if the n^3 functions $\Gamma_{ij}^k : p \mapsto (\Gamma_{ij}^k)_p$ are all C^∞ , then we say that we have introduced an **affine connection** on S . In addition, we call $\{\Gamma_{ij}^k\}$ the **connection coefficients** of the affine connection with respect to the coordinate system $[\xi^i]$. Note that the only constraint on the connection coefficients are that they be C^∞ , and that therefore affine connections have this degree of freedom. Below, we often refer to affine connections as simply connections.

Let $[\rho^r] = [\rho^1, \dots, \rho^n]$ be a coordinate system distinct from $[\xi^i]$, and let $\tilde{\partial}_r \stackrel{\text{def}}{=} \frac{\partial}{\partial \rho^r} = \frac{\partial \xi^i}{\partial \rho^r} \partial_i$. From Equation (1.27) and the linearity of $\Pi_{p,p'}$ we have

$$\Pi_{p,p'}((\tilde{\partial}_s)_p) = \left(\frac{\partial \xi^j}{\partial \rho^s} \right)_p \{(\partial_j)_{p'} - d\xi^i (\Gamma_{ij}^k)_p (\partial_k)_{p'}\}.$$

By substituting into the right hand side of this equation

$$\left(\frac{\partial \xi^j}{\partial \rho^s} \right)_{p'} = \left(\frac{\partial \xi^j}{\partial \rho^s} \right)_p + \left(\frac{\partial^2 \xi^j}{\partial \rho^r \partial \rho^s} \right)_p d\rho^r \quad \text{and}$$

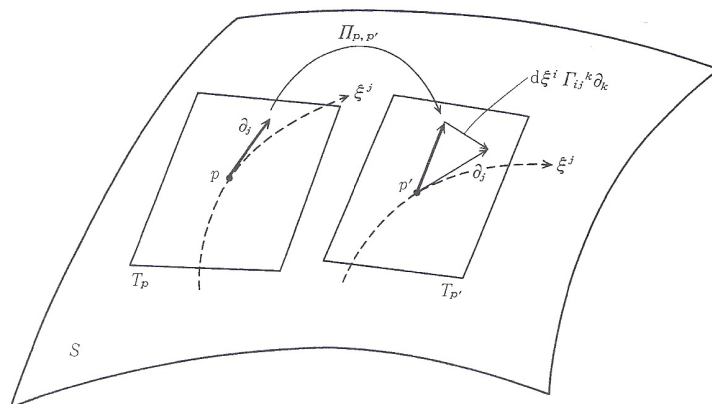


Figure 1.4: Affine connection (an infinitesimal translation)

$$d\xi^i = \left(\frac{\partial \xi^i}{\partial \rho^r} \right)_p d\rho^r \quad \left(d\rho^r \stackrel{\text{def}}{=} \rho^r(p') - \rho^r(p) \right),$$

and ignoring second order infinitesimals, we obtain

$$\Pi_{p,p'}((\tilde{\partial}_s)_p) = (\tilde{\partial}_s)_{p'} - d\rho^r (\tilde{\Gamma}_{rs}^t)_p (\tilde{\partial}_t)_{p'}, \quad (1.28)$$

where $(\tilde{\Gamma}_{rs}^t)_p$ is the value of the function

$$\tilde{\Gamma}_{rs}^t = \left\{ \Gamma_{ij}^k \frac{\partial \xi^i}{\partial \rho^r} \frac{\partial \xi^j}{\partial \rho^s} + \frac{\partial^2 \xi^k}{\partial \rho^r \partial \rho^s} \right\} \frac{\partial \rho^t}{\partial \xi^k} \quad (1.29)$$

at the point p . Note that Equations (1.27) and (1.28) are of the same form. Furthermore, if the functions Γ_{ij}^k are C^∞ for all (i, j, k) then so are the functions $\tilde{\Gamma}_{rs}^t$ for all (r, s, t) . In other words, the notion of affine connections is independent of the choice of coordinate system. Their connection coefficients, however, are related according to Equation (1.29).

An affine connection determines, for neighboring points p and p' , a correspondence between T_p and $T_{p'}$. By connecting a sequence of such correspondences, we may find, for non-neighboring points p and q , a correspondence between T_p and T_q . This correspondence depends, however, on the curve γ by which one connects p and q . Let us define the notion of “translating tangent vectors along a curve” in the following way.

Let $\gamma: [a, b] \rightarrow S$, where $\gamma(a) = p$ and $\gamma(b) = q$, be a curve which connects points p and q in S . We call a mapping from each point $\gamma(t)$ to a tangent vector $X(t) \in T_{\gamma(t)}$, say $X: t \mapsto X(t)$, a **vector field along γ** . Given such a vector field X , if, for all $t \in [a, b]$ and the corresponding infinitesimal dt , the corresponding tangent vectors are linearly related as specified by the connection, i.e., if

$$X(t + dt) = \Pi_{\gamma(t), \gamma(t+dt)}(X(t)), \quad (1.30)$$

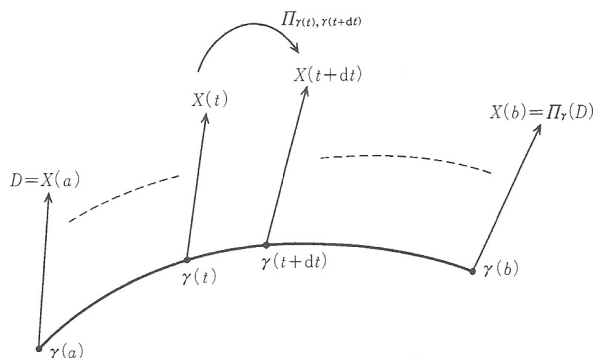


Figure 1.5: Translation of a tangent vector along a curve

then we say that X is **parallel** along γ (Figure 1.5).

Let us rewrite the equation above with respect to the coordinate system $[\xi^i]$. Letting $\partial_i = \frac{\partial}{\partial \xi^i}$, we have $X(t) = X^i(t)(\partial_i)_{\gamma(t)}$. From Equation (1.27) we have that

$$\Pi_{\gamma(t), \gamma(t+dt)}(X(t)) = \{X^k(t) - dt \dot{\gamma}^i(t) X^j(t) (\Gamma_{ij}^k)_{\gamma(t)}\} (\partial_k)_{\gamma(t+dt)}, \quad (1.31)$$

where $\gamma^i \stackrel{\text{def}}{=} \xi^i \circ \gamma$, and $\dot{\gamma}^i(t)$ is its derivative with respect to t . Now since in addition, $X(t+dt) = X^i(t+dt)(\partial_i)_{\gamma(t+dt)}$, substituting this into Equation (1.30) we obtain

$$\dot{X}^k(t) + \dot{\gamma}^i(t) X^j(t) (\Gamma_{ij}^k)_{\gamma(t)} = 0, \quad (1.32)$$

where $\dot{X}^k(t) \stackrel{\text{def}}{=} \frac{dX^k(t)}{dt} = \frac{X^k(t+dt) - X^k(t)}{dt}$. Equation (1.32) is an ordinary linear differential equation on $X^1(t), \dots, X^n(t)$, and hence given an initial (boundary) condition there exists a unique solution. From this, given $D \in T_{\gamma(a)} = T_p$, we see that there exists a unique parallel vector field along γ such that $X(a) = D$. Then letting $\Pi_\gamma(D)$ denote the vector $X(b) \in T_{\gamma(b)} = T_q$ determined by D , we see that Π_γ is a linear isomorphism from T_p to T_q . We call Π_γ the **parallel translation along γ** .

Let $\gamma : [a, b] \rightarrow S$ be a curve and X be a vector field along γ . In general, $X(t)$ and $X(t+h)$ lie in different tangent spaces and hence it is not possible to consider the derivative $\frac{dX(t)}{dt} = \lim_{h \rightarrow 0} \frac{X(t+h) - X(t)}{h}$. However, if an affine connection is given on S , then the parallel translation of $X(t+h) \in T_{\gamma(t+h)}$ to the space $T_{\gamma(t)}$ along γ gives us $X_t(t+h) = \Pi_{\gamma(t+h), \gamma(t)}(X(t+h))$, and using this we may consider within $T_{\gamma(t)}$ the quantity $\lim_{h \rightarrow 0} \frac{X_t(t+h) - X(t)}{h}$. We call this the **covariant derivative** of $X(t)$, and denote it by $\frac{\delta X(t)}{dt}$. In other words, instead of $dX(t) = X(t+dt) - X(t)$, we use

$$\delta X(t) = \Pi_{\gamma(t+dt), \gamma(t)}(X(t+dt)) - X(t) \quad (1.33)$$

(see Figure 1.6).

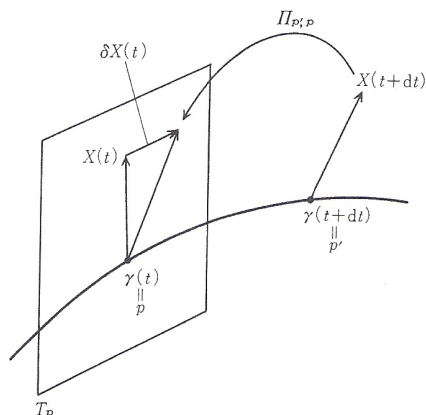


Figure 1.6: The covariant derivative along a curve.

Rewriting $X(t)$ as $X^j(t)(\partial_j)_{\gamma(t)}$, we have

$$\Pi_{\gamma(t+dt), \gamma(t)}(X(t+dt)) = \{X^k(t+dt) + dt\dot{\gamma}^i(t)X^j(t)(\Gamma_{ij}^k)_{\gamma(t)}\}(\partial_k)_{\gamma(t)}, \quad (1.34)$$

and substituting this into Equation (1.33), we obtain

$$\frac{\delta X(t)}{dt} = \left\{ \dot{X}^k(t) + \dot{\gamma}^i(t)X^j(t)(\Gamma_{ij}^k)_{\gamma(t)} \right\} (\partial_k)_{\gamma(t)}. \quad (1.35)$$

This also forms a vector field along γ . In addition, we see that the parallel translation condition in Equation (1.32) may now be written simply as $\frac{\delta X}{dt} = 0$.

In this way, using an affine connection it is possible to define the infinitesimal δX and the derivative $\frac{\delta X}{dt}$ of a vector field $X(t)$ along a curve. Extending this to "the directional derivative of a vector field $X = X^i\partial_i \in \mathcal{T}$ on S along a tangent vector $D = D^i(\partial_i)_p \in T_p$ " is straightforward as follows: consider a curve whose tangent vector at the point p is D , and by taking the covariant derivative of X along this curve we obtain

$$\nabla_D X = D^i \{ (\partial_i X^k)_p + X^j_p (\Gamma_{ij}^k)_p \} (\partial_k)_p \in T_p(S). \quad (1.36)$$

In fact, letting $X_\gamma : t \mapsto X_{\gamma(t)}$ for an arbitrary curve γ , we have from Equations (1.35) and (1.36) that

$$\frac{\delta X_{\gamma(t)}}{dt} = \nabla_{\dot{\gamma}(t)} X. \quad (1.37)$$

We may also define for each $X, Y \in \mathcal{T}(S)$ the vector field $\nabla_X Y \in \mathcal{T}(S)$ by $(\nabla_X Y)_p = \nabla_{X_p} Y \in T_p(S)$. We call this the **covariant derivative** of Y with respect to X . Given $X = X^i\partial_i$ and $Y = Y^j\partial_j$, we may write

$$\nabla_X Y = X^i \{ \partial_i Y^k + Y^j \Gamma_{ij}^k \} \partial_k. \quad (1.38)$$

In particular, when $X = \partial_i$ and $Y = \partial_j$, we obtain the component expression of the covariant derivative

$$\nabla_{\partial_i} \partial_j = \Gamma_{ij}^k \partial_k. \quad (1.39)$$

This may be construed as the vector field which expresses the change in the basis vector ∂_j as it is moved in the direction of ∂_i .

The operator $\nabla : \mathcal{T} \times \mathcal{T} \rightarrow \mathcal{T}$ which maps (X, Y) to $\nabla_X Y$ satisfies the following properties: for arbitrary $X, Y, Z \in \mathcal{T}$ and $f \in \mathcal{F}$ (: the set of C^∞ functions on S),

$$(i) \quad \nabla_{X+Y} Z = \nabla_X Z + \nabla_Y Z.$$

$$(ii) \quad \nabla_X (Y + Z) = \nabla_X Y + \nabla_X Z.$$

$$(iii) \quad \nabla_{fX} Y = f \nabla_X Y.$$

$$(iv) \quad \nabla_X (fY) = f \nabla_X Y + (Xf)Y.$$

Here, Xf denotes the function $p \mapsto X_p f$ ($\in \mathcal{F}$). Note that $\nabla_X Y$ is \mathcal{F} -linear with respect to X , but not with respect to Y , and hence ∇ is not a tensor field.

In fact, it is possible to consider the conditions (i)-(iv) as the defining properties of affine connections. In other words, we may define an affine connection on S to be a mapping $\nabla : \mathcal{T}(S) \times \mathcal{T}(S) \rightarrow \mathcal{T}(S)$ which satisfies conditions (i)-(iv). In addition, we may define the connection coefficients $\{\Gamma_{ij}^k\}$ of ∇ with respect to some coordinate system $[\xi^i]$ to be the n^3 functions determined by Equation (1.39). Then it is possible to prove Equations (1.38) and (1.29) from conditions (i)-(iv). It is also possible to reverse the derivation in Equations (1.32)-(1.37) to arrive at the definitions of $\frac{\delta X(t)}{dt}$ and Π_γ from that of ∇ . This method would make the use of both infinitesimals and fiber bundles unnecessary. In this book, we shall often refer to the "connection ∇ ".

Finally, we note that the totality of affine connections on a manifold forms an affine space. In other words, for any affine connections ∇ and ∇' and for any real number $\alpha \in \mathbb{R}$, the affine combination $\alpha \nabla + (1 - \alpha) \nabla'$ defines another affine connection. Note also that the difference of two affine connections is a tensor field of type (1, 2).

1.7 Flatness

Let $X \in \mathcal{T}(S)$ be a vector field on S . If for any curve γ on S , $X_\gamma : t \mapsto X_{\gamma(t)}$ is parallel along γ (with respect to the connection ∇), we say that X is **parallel** on S (with respect to ∇ .) In this case, for any curve γ which connects points p and q , $X_q = \Pi_\gamma(X_p)$ holds. A necessary and sufficient condition for an $X = X^i \partial_i$ to be parallel is that $\nabla_Y X = 0$ for all $Y \in \mathcal{T}(S)$, or equivalently that

$$\partial_i X^k + X^j \Gamma_{ij}^k = 0. \quad (1.40)$$

Note that nonzero parallel vector fields do not exist in general.

Let $[\xi^i]$ be a coordinate system of S , and suppose that with respect to this coordinate system the n basis vector fields $\partial_i = \frac{\partial}{\partial \xi^i}$ ($i = 1, \dots, n$) are all parallel on S . Then we call $[\xi^i]$ an **affine coordinate system** for ∇ . This condition is equivalent both to $\nabla_{\partial_i} \partial_j = 0$ and also to the condition that the connection coefficients $\{\Gamma_{ij}^k\}$ of ∇ with respect to $[\xi^i]$ are all identically 0.

Given some connection, a corresponding affine coordinate system does not in general exist. If an affine coordinate system exists for connection ∇ , we say that ∇ is **flat**, or alternatively that S is flat with respect to ∇ . Let $[\xi^i]$ be an affine coordinate system. Then with respect to a different coordinate system $[\rho^r]$, we see from Equation (1.29) that the connection coefficients $\{\tilde{\Gamma}_{rs}^t\}$ may be written as $\tilde{\Gamma}_{rs}^t = \frac{\partial^2 \xi^k}{\partial \rho^r \partial \rho^s} \frac{\partial \rho^t}{\partial \xi^k}$. Hence a necessary and sufficient condition for $[\rho^r]$ to be another affine coordinate system is that $\frac{\partial^2 \xi^k}{\partial \rho^r \partial \rho^s} = 0$. This is equivalent to the condition that there exist an $n \times n$ matrix A and an n -dimensional vector B such that

$$\xi(p) = A\rho(p) + B \quad (\forall p \in S) \quad (1.41)$$

($\xi(p) = [\xi^i(p)]$ and $\rho(p) = [\rho^r(p)]$.) We call a transformation of the form described in Equation (1.41) an **affine transformation** (when $B = 0$, this is simply a linear transformation). In addition, we see that this transformation is regular, i.e., one-to-one, and that A is a regular matrix. The collection of such regular affine transformations form a group, and affine coordinate systems have this degree of freedom.

Let ∇ be a connection on S . Then for vector fields $X, Y, Z \in \mathcal{T}$, if we define

$$R(X, Y)Z \stackrel{\text{def}}{=} \nabla_X(\nabla_Y Z) - \nabla_Y(\nabla_X Z) - \nabla_{[X, Y]}Z \quad \text{and} \quad (1.42)$$

$$T(X, Y) \stackrel{\text{def}}{=} \nabla_X Y - \nabla_Y X - [X, Y], \quad (1.43)$$

then these are also vector fields ($\in \mathcal{T}$). Here, letting $X = X^i \partial_i$ and $Y = Y^i \partial_i$, we have defined $[X, Y]$ to be the vector field

$$[X, Y] = (X^j \partial_j Y^i - Y^j \partial_j X^i) \partial_i$$

(this does not depend on the choice of coordinate system). The mappings $R : \mathcal{T} \times \mathcal{T} \times \mathcal{T} \rightarrow \mathcal{T}$ and $T : \mathcal{T} \times \mathcal{T} \rightarrow \mathcal{T}$ as defined above are both \mathcal{F} -multilinear. Hence R and T are respectively tensor fields of types (1, 3) and (1, 2). We call R the **Riemann-Christoffel curvature tensor (field)** of ∇ , or more simply the **curvature tensor (field)**, and T the **torsion tensor (field)** of ∇ . The component expressions of R and T with respect to coordinate system $[\xi^i]$ are given by

$$R(\partial_i, \partial_j) \partial_k = R_{ijk}^\ell \partial_\ell \quad \text{and} \quad T(\partial_i, \partial_j) = T_{ij}^k \partial_k \quad (1.44)$$

($\partial_i \stackrel{\text{def}}{=} \frac{\partial}{\partial \xi^i}$), and these may be computed in the following way:

$$R_{ijk}^\ell = \partial_i \Gamma_{jk}^\ell - \partial_j \Gamma_{ik}^\ell + \Gamma_{ih}^\ell \Gamma_{jk}^h - \Gamma_{jh}^\ell \Gamma_{ik}^h \quad \text{and} \quad (1.45)$$

$$T_{ij}^k = \Gamma_{ij}^k - \Gamma_{ji}^k. \quad (1.46)$$

If $[\xi^i]$ is an affine coordinate system for ∇ , then clearly $R_{ijk}^\ell = 0$ and $T_{ij}^k = 0$. In fact, in this case the components of R and T , since they are tensors, are always all 0 with respect to any coordinate system. In other words, if ∇ is flat, then $R = 0$ and $T = 0$. Conversely, if $R = 0$ and $T = 0$, it is known that ∇ is locally flat in the following sense: for each point $p \in S$, there exists a neighborhood U of p such that ∇ is flat on U . A proof will be found in standard textbooks of differential geometry.

In general, when $T = 0$ (i.e., $\Gamma_{ij}^k = \Gamma_{ji}^k$) holds, ∇ is called a **symmetric connection** or **torsion-free connection**. The connections having appeared so far in information geometry are mostly symmetric connections. However, the incorporation of torsion into the framework of information geometry, which would relate it to such fields as quantum mechanics (noncommutative probability theory) and systems theory, is an interesting topic for the future. We will make an attempt in this direction in §7.3.

If a connection is flat, then parallel translation does not depend on the curve selected to connect the two points. In particular, the n basis vector fields $\partial_i = \frac{\partial}{\partial \xi^i}$ ($i = 1, \dots, n$) of an affine coordinate system $[\xi^i]$ are parallel vector fields, and hence $\Pi_\gamma((\partial_i)_p) = (\partial_i)_q$ regardless of the curve γ used to connect the points p and q . In addition, if the components X^i of a vector field $X = X^i \partial_i$ are all constant on S , then X is parallel, and $\Pi_\gamma(X_p) = X_q$.

In general, if parallel translation does not depend on curve choice, or in other words if there are n linearly independent parallel vector fields on S then $R = 0$, and in addition, when S is simply connected (i.e., when arbitrary closed loops may be continuously contracted to a single point) it is known that the converse also holds.² There exist, however, connections for which $R = 0$ and $T \neq 0$. When this is the case, although parallel translation does not depend on the curve selected, there does not exist an affine coordinate system. Such spaces, called spaces of distant parallelism, were introduced by Einstein within the context of unified field theory, and also serve a major role within the theory of non-Riemannian plasticity. Another example will be shown in §7.3.

From Equations (1.45) and (1.46) we see that in general $R_{ijk}^\ell = -R_{jik}^\ell$ and $T_{ij}^k = -T_{ji}^k$. Hence, in the particular case when S is 1-dimensional, $R = 0$ and $T = 0$ necessarily hold, and therefore S is flat.

1.8 Autoparallel submanifolds

Let S be an n -dimensional manifold and M be an m -dimensional submanifold of S . Let $[\xi^i]$ and $[u^a]$ be coordinate systems for S and M , respectively, and let $\partial_i = \frac{\partial}{\partial \xi^i}$ and $\partial_a = \frac{\partial}{\partial u^a}$. Suppose also that ∇ is an affine connection on S and that $\{\Gamma_{ij}^k\}$ are the connection coefficients of ∇ with respect to $[\xi^i]$. Now letting $X = X^a \partial_a$ and $Y = Y^a \partial_a \in \mathcal{T}(M)$ be vector fields on M , we may consider $\nabla_{X_p} Y$, the “directional derivative of Y along X_p ”, as we did in Equation (1.36). However, even though in general $\nabla_{X_p} Y$ is a tangent vector of S ($\in T_p(S)$), it

²There are those who define “flat” to denote this case.

is not necessarily a tangent vector of M ($\in T_p(M)$). If we let $\nabla_X Y$ denote the mapping from points p in M to $\nabla_{X_p} Y \in T_p(S)$, then using identities such as $\partial_a = (\partial_a \xi^i) \partial_i$ we have

$$\nabla_X Y = X^a (\partial_a Y^b) \partial_b + X^a Y^b \{ (\partial_a \xi^i) (\partial_b \xi^j) \Gamma_{ij}^k + \partial_a \partial_b \xi^k \} \partial_k. \quad (1.47)$$

In particular, letting $X = \partial_a$ and $Y = \partial_b$, we obtain

$$\nabla_{\partial_a} \partial_b = \{ (\partial_a \xi^i) (\partial_b \xi^j) \Gamma_{ij}^k + \partial_a \partial_b \xi^k \} \partial_k. \quad (1.48)$$

Note also that Equation (1.47) may be written as

$$\nabla_X Y = X^a (\partial_a Y^b) \partial_b + X^a Y^b \nabla_{\partial_a} \partial_b \quad (1.49)$$

As we mentioned above, for $X, Y \in \mathcal{T}(M)$, $(\nabla_X Y)_p = \nabla_{X_p} Y$ is an element of $T_p(S)$, but not necessarily one of $T_p(M)$, i.e., in general, $\nabla_X Y \notin \mathcal{T}(M)$. If, however,

$$\nabla_X Y \in \mathcal{T}(M) \quad \text{for} \quad \forall X, Y \in \mathcal{T}(M), \quad (1.50)$$

then ∇ determines a covariant derivative on M . In fact, when this is the case the conditions (i)-(iv) from §1.6 hold for all $X, Y, Z \in \mathcal{T}(M)$ and all $f \in \mathcal{F}(M)$, and ∇ is an affine connection on M . If we use this connection to define a parallel translation $\Pi_\gamma^M : T_{\gamma(a)}(M) \rightarrow T_{\gamma(b)}(M)$ on M along the curves $\gamma : [a, b] \rightarrow M$, then this translation coincides exactly with the parallel translation $\Pi_\gamma : T_{\gamma(a)}(S) \rightarrow T_{\gamma(b)}(S)$ on S restricted to the tangent spaces of M , using the original connection on S . In other words

$$\Pi_\gamma^M = \Pi_\gamma|_{T_{\gamma(a)}(M)}. \quad (1.51)$$

If a submanifold M of S satisfies Equation (1.50), we say that M is **autoparallel** with respect to ∇ . In particular, open subsets of S are autoparallel. From Equation (1.49) we see that a necessary and sufficient condition for M to be autoparallel is that $\nabla_{\partial_a} \partial_b \in \mathcal{T}(M)$ holds for all a, b . This, in turn, is equivalent to there existing m^3 functions $\{\Gamma_{ab}^c\} (\in \mathcal{F}(M))$ which satisfy

$$\nabla_{\partial_a} \partial_b = \Gamma_{ab}^c \partial_c. \quad (1.52)$$

These $\{\Gamma_{ab}^c\}$ form the connection coefficients of ∇ with respect to $[u^a]$. Using Equation (1.48) we may rewrite Equation (1.52) in the following way:

$$\Gamma_{ab}^c \partial_c \xi^k = (\partial_a \xi^i) (\partial_b \xi^j) \Gamma_{ij}^k + \partial_a \partial_b \xi^k. \quad (1.53)$$

We can also see that M is autoparallel in S if and only if M is closed with respect to the parallel translation on S in the following sense: for every curve $\gamma : [a, b] \rightarrow M$ in M and for every tangent vector D of M at $\gamma(a)$, the result $\Pi_\gamma(D)$ of the parallel translation $\Pi_\gamma : T_{\gamma(a)}(S) \rightarrow T_{\gamma(b)}(S)$ belongs to the tangent space of M at $\gamma(b)$.

1-dimensional autoparallel submanifolds are called **autoparallel curves** or **geodesics**. For a curve $\gamma : t \mapsto \gamma(t)$, the condition in Equation (1.52) may be rewritten using Equation (1.37) as

$$\frac{\delta}{dt} \frac{d\gamma}{dt} = \Gamma(t) \frac{d\gamma}{dt}, \quad (1.54)$$

where $\Gamma : t \mapsto \Gamma(t)$ is a C^∞ function. As we noted at the end of §1.7, connections on 1-dimensional manifolds are necessarily flat, and hence by substituting into Equation (1.54) a suitable one-to-one transformation (change of variable) of t , we may obtain $\Gamma(t) \equiv 0$. We call such a t an **affine parameter** of γ . In this case Equation (1.54) reduces to

$$\frac{\delta}{dt} \frac{d\gamma}{dt} = 0, \quad (1.55)$$

and implies that $\frac{d\gamma}{dt}$ is parallel along γ . It is possible to define geodesics using Equation (1.55). Rewriting Equation (1.55) using the coordinate system $[\xi^i]$ and the corresponding representation $\gamma^i = \xi^i \circ \gamma$, we obtain

$$\ddot{\gamma}^i(t) + \dot{\gamma}^i(t) \dot{\gamma}^j(t) (\Gamma_{ij}^k)_{\gamma(t)} = 0. \quad (1.56)$$

Let M be an autoparallel submanifold of S . If the torsion tensor of S is 0, then the torsion tensor of M is also 0. This is clear from Equations (1.46) and (1.53). The same holds for the curvature tensor. The latter fact may be derived using Equations (1.45) and (1.53), but it is in fact immediate from the analysis of parallel translation as follows: from Equation (1.51) we see that if the choice of curve does not affect parallel translation in S , then it similarly does not in M . Note that, in the case when parallel translation does not depend on curve choice, a necessary and sufficient condition for a submanifold M to be autoparallel in S is that there exist $m = (\dim M)$ linearly independent vector fields on M which are parallel with respect to the connection on S .

Consider the case when S is flat with respect to ∇ . Then by the argument above autoparallel submanifolds of S are also flat. Hence without loss of generality we may assume that $[\xi^i]$ and $[u^a]$ are affine coordinate systems in Equation (1.53), the condition for a submanifold M of S to be autoparallel. Equation (1.53) then reduces to $\partial_a \partial_b \xi^k = 0$. This condition is equivalent to there existing an $n \times m$ matrix A and an n -dimensional vector B which satisfies

$$\xi(p) = Au(p) + B \quad (\forall p \in M) \quad (1.57)$$

($\xi(p) = [\xi^i(p)]$ and $u(p) = [u^a(p)]$.) In general, a subspace of \mathbb{R}^n which may be expressed as $\{Au + B | u \in \mathbb{R}^m\}$ is called an **affine subspace** of \mathbb{R}^n (; when $B = 0$ we have a linear subspace). We summarize the discussion above in the following theorem.

Theorem 1.1 *If S is flat, then a necessary and sufficient condition for a submanifold M to be autoparallel is that M is expressed as an affine subspace (or*

an open subset of an affine subspace) of S with respect to an affine coordinate system. In particular, geodesics may be expressed using linear equations (as a line or a segment) with respect to affine coordinate systems. In addition, if M is autoparallel, then it is also flat.

1.9 Projection of connections and embedding curvature

If M is a submanifold of S which is not autoparallel with respect to ∇ on S , then there is no natural connection on M which may be derived from ∇ . However, if there is for each point p a mapping π_p from $T_p(S)$ to $T_p(M)$, then we may use this to define a connection on M . Assume that $\pi_p : T_p(S) \rightarrow T_p(M)$ is a linear mapping and that $\pi_p(D) = D$ for every $D \in T_p(M)$, and that the relation $p \mapsto \pi_p$ is C^∞ . Now suppose, for each $X, Y \in \mathcal{T}(M)$, we define $\nabla_X^{(\pi)} \in \mathcal{T}(M)$ in the following way:

$$(\nabla_X^{(\pi)} Y)_p = \pi_p((\nabla_X Y)_p) \quad (\forall p \in M). \quad (1.58)$$

Then $\nabla^{(\pi)}$ is a connection on M . In particular, if a Riemannian metric $g = \langle \cdot, \cdot \rangle$ is given on S , we may take as π_p the orthogonal projection with respect to g . This is defined to be that which satisfies, for all $D \in T_p(S)$ and all $D' \in T_p(M)$,

$$\langle \pi_p(D), D' \rangle_p = \langle D, D' \rangle_p. \quad (1.59)$$

We call such $\nabla^{(\pi)}$ the **projection of ∇ onto M with respect to g** .

If S has a coordinate system $[\xi^i]$, then the connection coefficients $\{\Gamma_{ij}^k\}$ of ∇ are determined by Equation (1.39). If S also has a Riemannian metric g , then we may define n^3 additional functions $\{\Gamma_{ij,k}\}$ in the following way:

$$\Gamma_{ij,k} \stackrel{\text{def}}{=} \langle \nabla_{\partial_i} \partial_j, \partial_k \rangle = \Gamma_{ij}^h g_{hk}. \quad (1.60)$$

The quantities $\{\Gamma_{ij,k}\}$, like $\{\Gamma_{ij}^k\}$, may be considered as a different component expression of the same ∇ . With respect to a different coordinate system $[\rho^r]$ for S , these may be written as follows ($\tilde{\partial}_r \stackrel{\text{def}}{=} \frac{\partial}{\partial \rho^r}$):

$$\tilde{\Gamma}_{rs,t} \stackrel{\text{def}}{=} \langle \nabla_{\tilde{\partial}_r} \tilde{\partial}_s, \tilde{\partial}_t \rangle = \left(\frac{\partial \xi^i}{\partial \rho^r} \frac{\partial \xi^j}{\partial \rho^s} \Gamma_{ij,k} + \frac{\partial^2 \xi^h}{\partial \rho^r \partial \rho^s} g_{hk} \right) \frac{\partial \xi^k}{\partial \rho^t}. \quad (1.61)$$

Similarly, for the projection $\nabla^{(\pi)}$ of ∇ onto M , we may define, given a coordinate system $[u^a]$ for M , $\Gamma_{ab,c}^{(\pi)} \stackrel{\text{def}}{=} \langle \nabla_{\partial_a}^{(\pi)} \partial_b, \partial_c \rangle$ ($\partial_a \stackrel{\text{def}}{=} \frac{\partial}{\partial u^a}$). Using Equations (1.58), (1.59) and (1.48) we may rewrite this as

$$\Gamma_{ab,c}^{(\pi)} = \langle \Gamma_{\partial_a} \partial_b, \partial_c \rangle = \{ (\partial_a \xi^i) (\partial_b \xi^j) \Gamma_{ij,k} + (\partial_a \partial_b \xi^j) g_{jk} \} (\partial_c \xi^k). \quad (1.62)$$

The connection coefficients of $\nabla^{(\pi)}$ are then given by $\Gamma_{ab}^{(\pi)d} = \Gamma_{ab,c}^{(\pi)} g^{cd}$. From this, we see that if ∇ is symmetric, then so is $\nabla^{(\pi)}$.

Now let

$$H(X, Y) \stackrel{\text{def}}{=} \nabla_X Y - \nabla_X^{(\pi)} Y \quad (1.63)$$

for $X, Y \in \mathcal{T}(M)$. Then $(H(X, Y))_p = (\nabla_X Y)_p - \pi_p((\nabla_X Y)_p)$ is the orthogonal projection of $(\nabla_X Y)_p$ onto $[T_p(M)]^\perp$, the orthocomplement of $T_p(M)$. Given this, note that the autoparallel condition for M in Equation (1.50) is equivalent to stating that $H(X, Y) = 0$ holds for all $X, Y \in \mathcal{T}(M)$, and that this, in turn, is equivalent to simply stating that $H = 0$. Intuitively, H may be considered as measuring the degree to which M is “not autoparallel” or “curved” in S . In addition, since $H(X, Y)$ is $\mathcal{F}(M)$ -linear with respect to both X and Y (i.e., is $\mathcal{F}(M)$ -bilinear), H may be considered as “a kind of” tensor field, even though $H(X, Y)$ is not a vector field on M in general. We call such an H an **embedding curvature** of the submanifold $M (\subset S)$ with respect to ∇ .

Since M has $\nabla^{(\pi)}$ as a connection, we may use this to compute its Riemann-Christoffel curvature $R^{(\pi)}$. This $R^{(\pi)}$ expresses the “inherent curvature” of M itself, while the embedding curvature H expresses the curvature of the arrangement of M within S . As we noted in §1.8, if R , the Riemann-Christoffel curvature of S , is 0, and if, in addition, $H = 0$ (i.e., M is autoparallel), then $R^{(\pi)} = 0$ also. However, $R^{(\pi)} = 0$ does not entail $H = 0$. For example, consider a cylinder surface M embedded within a 3-dimensional Euclidean space. The 2-dimensional geometry on the surface of this cylinder is Euclidean, and $R^{(\pi)} = 0$. However, within the 3-dimensional space it is curved, and hence H is not 0. It is important to distinguish these two notions of curvature.

For each point p in S , let $\{(\partial_a)_p; 1 \leq a \leq m\}$ ($m = \dim M$) be a basis for $T_p(M)$, and let $\{(\partial_\kappa)_p; m+1 \leq \kappa \leq n\}$ ($n = \dim S$) be a basis for $[T_p(M)]^\perp$. Then we may define the $m^2(n-m)$ functions $\{H_{ab\kappa}\}$ in the following way:

$$H_{ab\kappa} \stackrel{\text{def}}{=} \langle H(\partial_a, \partial_b), \partial_\kappa \rangle = \langle \nabla_{\partial_a} \partial_b, \partial_\kappa \rangle. \quad (1.64)$$

It follows from the properties of tensors that $H = 0 \Leftrightarrow H_{ab\kappa} = 0$ ($\forall a, b, \kappa$).

1.10 Riemannian connection

Let ∇ be an affine connection on a Riemannian manifold $(S, g = \langle \cdot, \cdot \rangle)$, and suppose ∇ satisfies, for all vector fields $X, Y, Z \in \mathcal{T}(S)$,

$$Z \langle X, Y \rangle = \langle \nabla_Z X, Y \rangle + \langle X, \nabla_Z Y \rangle. \quad (1.65)$$

Then we say that ∇ is a **metric connection** with respect to g . Using the coordinate expressions of g and ∇ we may rewrite this condition as follows:

$$\partial_k g_{ij} = \Gamma_{ki,j} + \Gamma_{kj,i}. \quad (1.66)$$

Let us show that, under a metric connection, the parallel translation of two vectors leaves their inner product unchanged. Consider a curve $\gamma : t \mapsto \gamma(t)$

on S and two vector fields X and Y along γ . Letting $\frac{\delta X}{dt}$ and $\frac{\delta Y}{dt}$ respectively denote the covariant derivatives of X and Y with respect to ∇ , we see from Equation (1.65) that

$$\frac{d}{dt} \langle X(t), Y(t) \rangle = \left\langle \frac{\delta X(t)}{dt}, Y(t) \right\rangle + \left\langle X(t), \frac{\delta Y(t)}{dt} \right\rangle. \quad (1.67)$$

Now if X and Y are both parallel on γ (i.e., $\frac{\delta X}{dt} = \frac{\delta Y}{dt} = 0$), then the right hand side of the equation above is 0, and hence $\langle X(t), Y(t) \rangle$ does not depend on t and is constant. The parallel translation Π_γ along γ , then, is a metric isomorphism which preserves inner products. In other words, letting p and q be the boundary points of γ , for any two tangent vectors $D_1, D_2 \in T_p$ the following holds:

$$\langle \Pi_\gamma(D_1), \Pi_\gamma(D_2) \rangle_q = \langle D_1, D_2 \rangle_p. \quad (1.68)$$

We call a connection which is both metric and symmetric the **Riemannian connection** or the **Levi-Civita connection** with respect to g . For a given g , such a connection exists uniquely. In fact, combining Equation (1.66) with the requirement that $\Gamma_{ij,k} = \Gamma_{ji,k}$, we have

$$\Gamma_{ij,k} = \frac{1}{2} (\partial_i g_{jk} + \partial_j g_{ki} - \partial_k g_{ij}). \quad (1.69)$$

The geodesics with respect to the Riemannian connection ∇ are known to (locally) coincide with the shortest curve joining two points (where we measure length according to Equation (1.25).) In addition, if we consider the case when ∇ is flat and there exists an affine coordinate system $[\xi^i]$, we find that since $\partial_i = \frac{\partial}{\partial \xi^i}$ is parallel on S , $\langle \partial_i, \partial_j \rangle$ is constant on S . Since affine coordinate systems have a degree of freedom as given in Equation (1.41), we see in particular that there exists an affine coordinate system which satisfies

$$\langle \partial_i, \partial_j \rangle = \delta_{ij}. \quad (1.70)$$

A coordinate system which satisfies the equation above is called a **Euclidean coordinate system** (with respect to g). Hence the Riemannian connection is flat if and only if there exists a Euclidean coordinate system.

In most differential geometry textbooks, only Riemannian connections are introduced on Riemannian manifolds. Non-metric connections are not even discussed. However, when considering families of probability distributions as manifolds, we find that the natural connections which one would introduce are non-metric (see §2.3). As we shall discuss in Chapter 3, this leads us to the novel notion of dual connections.