

# **A Gentle Introduction to Tensors**

Boaz Porat  
Department of Electrical Engineering  
Technion – Israel Institute of Technology  
[boaz@ee.technion.ac.il](mailto:boaz@ee.technion.ac.il)

May 27, 2014

## Opening Remarks

This document was written for the benefits of Engineering students, Electrical Engineering students in particular, who are curious about physics and would like to know more about it, whether from sheer intellectual desire or because one's awareness that physics is the key to our understanding of the world around us. Of course, anybody who is interested and has some college background may find this material useful. In the future, I hope to write more documents of the same kind. I chose tensors as a first topic for two reasons. First, tensors appear everywhere in physics, including classical mechanics, relativistic mechanics, electrodynamics, particle physics, and more. Second, tensor theory, at the most elementary level, requires only linear algebra and some calculus as prerequisites. Proceeding a small step further, tensor theory requires background in multivariate calculus. For a deeper understanding, knowledge of manifolds and some point-set topology is required. Accordingly, we divide the material into three chapters. The first chapter discusses constant tensors and constant linear transformations. Tensors and transformations are inseparable. To put it succinctly, tensors are geometrical objects over vector spaces, whose coordinates obey certain laws of transformation under change of basis. Vectors are simple and well-known examples of tensors, but there is much more to tensor theory than vectors. The second chapter discusses tensor fields and curvilinear coordinates. It is this chapter that provides the foundations for tensor applications in physics. The third chapter extends tensor theory to spaces other than vector spaces, namely manifolds. This chapter is more advanced than the first two, but all necessary mathematics is included and no additional formal mathematical background is required beyond what is required for the second chapter.

I have used the coordinate approach to tensors, as opposed to the formal geometrical approach. Although this approach is a bit old fashioned, I still find it the easier to comprehend on first learning, especially if the learner is not a student of mathematics or physics.

All vector spaces discussed in this document are over the field  $\mathbb{R}$  of real numbers. We will not mention this every time but assume it implicitly.

I would appreciate feedback, comments, corrections, and criticisms. Please e-mail to [boaz@ee.technion.ac.il](mailto:boaz@ee.technion.ac.il).

# Chapter 1

## Constant Tensors and Constant Linear Transformations

### 1.1 Plane Vectors

Let us begin with the simplest possible setup: that of plane vectors. We think of a plane vector as an arrow having direction and length, as shown in Figure 1.1.

The length of a physical vector must have physical units; for example: distance is measured in meter, velocity in meter/second, force in Newton, electric field in Volt/meter, and so on. The length of a "mathematical vector" is a pure number. Length is absolute, but direction must be measured relative to some (possibly arbitrarily chosen) reference direction, and has units of radians (or, less conveniently, degrees). Direction is usually assumed positive in counterclockwise rotation from the reference direction.

Vectors, by definition, are free to move parallel to themselves anywhere in the plane and they remain invariant under such moves (such a move is called *translation*).

Vectors are abstract objects, but they may be manipulated numerically and algebraically by expressing them in bases. Recall that a basis in a plane is

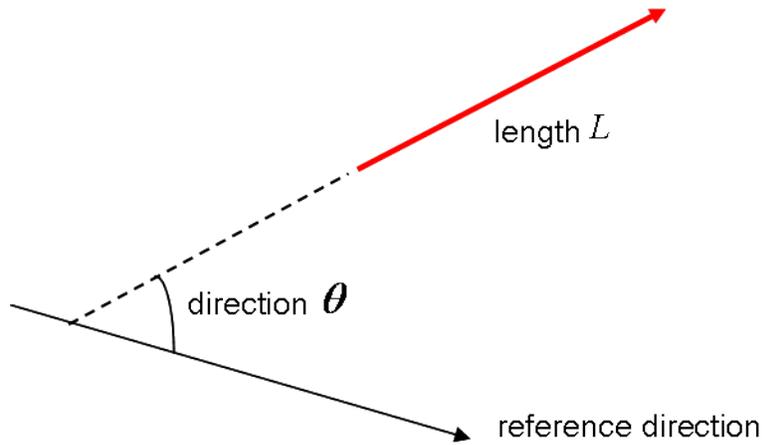


Figure 1.1: A plane vector having length and direction

a pair of non-zero and non-collinear vectors  $(\mathbf{e}_1, \mathbf{e}_2)$ . When drawing a basis, it is customary to translate  $\mathbf{e}_1$  and  $\mathbf{e}_2$  until their tails touch, as is shown in Figure 1.2.

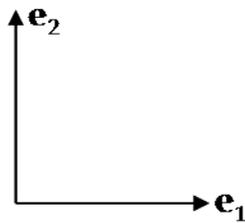


Figure 1.2: A basis in the plane

The basis depicted in Figure 1.2 happens to be *orthonormal*; that is, the two vectors are perpendicular and both have unity length. However, a basis need not be orthonormal. Figure 1.3 shows another basis  $(\tilde{\mathbf{e}}_1, \tilde{\mathbf{e}}_2)$ , whose vectors are neither perpendicular nor having equal length.

Let  $\mathbf{x}$  be an arbitrary plane vector and let  $(\mathbf{e}_1, \mathbf{e}_2)$  be some basis in the plane. Then  $\mathbf{x}$  can be expressed in a unique manner as a linear combination of the basis vectors; that is,

$$\mathbf{x} = \mathbf{e}_1 x^1 + \mathbf{e}_2 x^2 \quad (1.1)$$

The two real numbers  $(x^1, x^2)$  are called the *coordinates* of  $\mathbf{x}$  in the basis

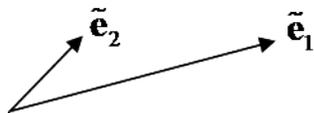


Figure 1.3: Another basis in the plane

$(\mathbf{e}_1, \mathbf{e}_2)$ . The following are worth noting:

1. Vectors are set in bold font whereas coordinates are set in italic font.
2. The basis vectors are numbered by subscripts whereas the coordinates are numbered by superscripts. This distinction will be important later.
3. In products such as  $\mathbf{e}_1 x^1$  we place the vector on the left and the scalar on the right. In most linear algebra books the two are reversed — the scalar is on the left of the vector. The reason for our convention will become clear later, but for now it should be kept in mind.

Recalling notations from vector-matrix algebra, we may express (1.1) as

$$\mathbf{x} = [ \mathbf{e}_1 \quad \mathbf{e}_2 ] \begin{bmatrix} x^1 \\ x^2 \end{bmatrix} \quad (1.2)$$

For now we will use row vectors to store basis vectors and column vectors to store coordinates. Later we will abandon expressions such as (1.2) in favor of more compact and more general notations.

## 1.2 Transformation of Bases

Consider two bases  $(\mathbf{e}_1, \mathbf{e}_2)$ , which we will henceforth call the *old basis*, and  $(\tilde{\mathbf{e}}_1, \tilde{\mathbf{e}}_2)$ , which we will call the *new basis*. See, for example, Figure 1.4, in which we have brought the two bases to a common origin.

Since  $(\mathbf{e}_1, \mathbf{e}_2)$  is a basis, each of the vectors  $(\tilde{\mathbf{e}}_1, \tilde{\mathbf{e}}_2)$  can be uniquely expressed as a linear combination of  $(\mathbf{e}_1, \mathbf{e}_2)$ , similarly to (1.1):

$$\begin{aligned} \tilde{\mathbf{e}}_1 &= \mathbf{e}_1 S_1^1 + \mathbf{e}_2 S_2^1 \\ \tilde{\mathbf{e}}_2 &= \mathbf{e}_1 S_1^2 + \mathbf{e}_2 S_2^2 \end{aligned} \quad (1.3)$$

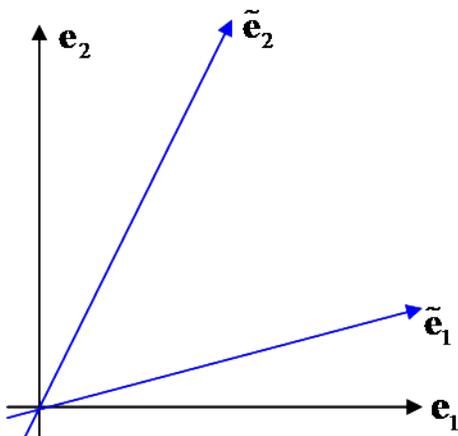


Figure 1.4: Two bases in the plane

Equation (1.3) is the *basis transformation formula* from  $(\mathbf{e}_1, \mathbf{e}_2)$  to  $(\tilde{\mathbf{e}}_1, \tilde{\mathbf{e}}_2)$ . The 4-parameter object  $\{S_i^j, 1 \leq i, j \leq 2\}$  is called the *direct transformation* from the old basis to the new basis. We may also write (1.3) in vector-matrix notation:

$$[\tilde{\mathbf{e}}_1 \quad \tilde{\mathbf{e}}_2] = [\mathbf{e}_1 \quad \mathbf{e}_2] \begin{bmatrix} S_1^1 & S_2^1 \\ S_1^2 & S_2^2 \end{bmatrix} = [\mathbf{e}_1 \quad \mathbf{e}_2] S \quad (1.4)$$

The matrix  $S$  is the *direct transformation matrix* from the old basis to the new basis. This matrix is uniquely defined by the two bases. Note that the rows of  $S$  appear as superscripts and the columns appear as subscripts; remember this convention for later.

A special case occurs when the new basis is identical with the new basis. In this case, the transformation matrix becomes the identity matrix  $I$ , where  $I_i^i = 1$  and  $I_i^j = 0$  for  $i \neq j$ .

Since  $(\tilde{\mathbf{e}}_1, \tilde{\mathbf{e}}_2)$  is a basis, each of the vectors  $(\mathbf{e}_1, \mathbf{e}_2)$  may be expressed as a linear combination of  $(\tilde{\mathbf{e}}_1, \tilde{\mathbf{e}}_2)$ . Hence the transformation  $S$  is performe invertible and we can write

$$\begin{aligned} [\mathbf{e}_1 \quad \mathbf{e}_2] &= [\tilde{\mathbf{e}}_1 \quad \tilde{\mathbf{e}}_2] \begin{bmatrix} S_1^1 & S_2^1 \\ S_1^2 & S_2^2 \end{bmatrix}^{-1} = [\tilde{\mathbf{e}}_1 \quad \tilde{\mathbf{e}}_2] \begin{bmatrix} T_1^1 & T_2^1 \\ T_1^2 & T_2^2 \end{bmatrix} \\ &= [\tilde{\mathbf{e}}_1 \quad \tilde{\mathbf{e}}_2] T \end{aligned} \quad (1.5)$$

where  $T = S^{-1}$  or, equivalently,  $ST = TS = I$ . The object  $\{T_i^j, 1 \leq i, j \leq 2\}$  is the *inverse transformation* and  $T$  is the *inverse transformation matrix*.

In summary, with each pair of bases there are associated two transformations. Once we agree which of the two bases is labeled old and which is labeled new, there is a unique direct transformation (from the old to the new) and a unique inverse transformation (from the new to the old). The two transformations are the inverses of each other.

### 1.3 Coordinate Transformation of Vectors

Equation (1.2) expresses a vector  $\mathbf{x}$  in terms of coordinates relative to a given basis  $(\mathbf{e}_1, \mathbf{e}_2)$ . If a second basis  $(\tilde{\mathbf{e}}_1, \tilde{\mathbf{e}}_2)$  is given, then  $\mathbf{x}$  may be expressed relative to this basis using a similar formula

$$\mathbf{x} = \tilde{\mathbf{e}}_1 \tilde{x}^1 + \tilde{\mathbf{e}}_2 \tilde{x}^2 = \begin{bmatrix} \tilde{\mathbf{e}}_1 & \tilde{\mathbf{e}}_2 \end{bmatrix} \begin{bmatrix} \tilde{x}^1 \\ \tilde{x}^2 \end{bmatrix} \quad (1.6)$$

The coordinates  $(\tilde{x}^1, \tilde{x}^2)$  differ from  $(x^1, x^2)$ , but the vector  $\mathbf{x}$  is the same. The situation is depicted in Figure 1.5. The vector  $\mathbf{x}$  is shown in red. The basis  $(\mathbf{e}_1, \mathbf{e}_2)$  and its associated coordinates  $(x^1, x^2)$  are shown in black; the basis  $(\tilde{\mathbf{e}}_1, \tilde{\mathbf{e}}_2)$  and its associated coordinates  $(\tilde{x}^1, \tilde{x}^2)$  are shown in blue.

We now pose the following question: how are the coordinates  $(\tilde{x}^1, \tilde{x}^2)$  related to  $(x^1, x^2)$ ? To answer this question, recall the transformation formulas between the two bases and perform the following calculation:

$$\begin{bmatrix} \tilde{\mathbf{e}}_1 & \tilde{\mathbf{e}}_2 \end{bmatrix} \begin{bmatrix} \tilde{x}^1 \\ \tilde{x}^2 \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 \end{bmatrix} S \begin{bmatrix} \tilde{x}^1 \\ \tilde{x}^2 \end{bmatrix} = \mathbf{x} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 \end{bmatrix} \begin{bmatrix} x^1 \\ x^2 \end{bmatrix} \quad (1.7)$$

Since (1.7) must hold identically for an arbitrary vector  $\mathbf{x}$ , we are led to conclude that

$$S \begin{bmatrix} \tilde{x}^1 \\ \tilde{x}^2 \end{bmatrix} = \begin{bmatrix} x^1 \\ x^2 \end{bmatrix} \quad (1.8)$$

Or, equivalently,

$$\begin{bmatrix} \tilde{x}^1 \\ \tilde{x}^2 \end{bmatrix} = S^{-1} \begin{bmatrix} x^1 \\ x^2 \end{bmatrix} = T \begin{bmatrix} x^1 \\ x^2 \end{bmatrix} \quad (1.9)$$

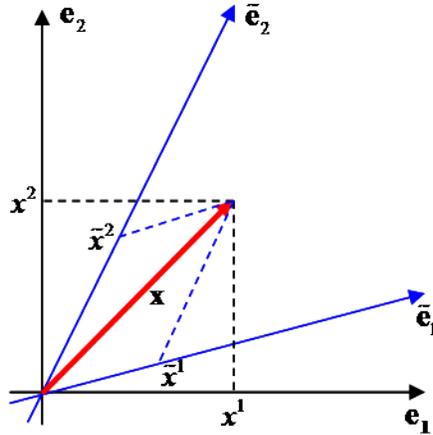


Figure 1.5: A vector in two bases

We have thus arrived at a somewhat surprising conclusion: *the coordinates of a vector when passing from an old basis to a new basis are transformed via the inverse of the transformation from the old basis to the new.*

As an example, the direct transformation between the bases in Figure 1.4 is

$$S = \begin{bmatrix} 1 & 0.5 \\ 0.25 & 1 \end{bmatrix}$$

The inverse transformation is

$$T = 0.875 \begin{bmatrix} 1 & -0.5 \\ -0.25 & 1 \end{bmatrix}$$

Examination of Figure 1.4 confirms this result, at least qualitatively.

The result obtained in the section is important and should be memorized: When a basis is transformed using a direct transformation, the coordinates of an arbitrary vector are transformed using the inverse transformation. For this reasons, vectors are said to be *contravariant* ("they vary in a contrary manner", in a way of speaking).

## 1.4 Generalization to Higher-Dimensional Vector Spaces

We assume that you have studied a course a linear algebra; therefore you are familiar with general (abstract) finite-dimensional vector spaces. In particular, an  $n$ -dimensional vector space possesses a set of  $n$  linearly independent vectors, but no set of  $n + 1$  linearly independent vectors. A basis for an  $n$ -dimensional vector space  $\mathcal{V}$  is any ordered set of linearly independent vectors  $(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n)$ . An arbitrary vector  $\mathbf{x}$  in  $\mathcal{V}$  can be expressed as a linear combination of the basis vectors:

$$\mathbf{x} = \sum_{i=1}^n \mathbf{e}_i x^i \quad (1.10)$$

The real numbers in (1.10) are called linear coordinates. We will refer to them simply as coordinates, until we need to distinguish them from curvilinear coordinates in Chapter 2. Note again our preferred convention of writing the vector on the left of the scalar. If a second basis  $(\tilde{\mathbf{e}}_1, \tilde{\mathbf{e}}_2, \dots, \tilde{\mathbf{e}}_n)$  is given, there exist unique transformations  $S$  and  $T$  such that

$$\tilde{\mathbf{e}}_i = \sum_{j=1}^n \mathbf{e}_j S_i^j, \quad \mathbf{e}_i = \sum_{j=1}^n \tilde{\mathbf{e}}_j T_i^j, \quad T = S^{-1} \quad (1.11)$$

The coordinates of  $\mathbf{x}$  in the new basis are related to those in the old basis according to the transformation law

$$\tilde{x}^i = \sum_{j=1}^n T_j^i x^j \quad (1.12)$$

Equation (1.12) is derived in exactly the same way as (1.9). Thus, vectors in an  $n$ -dimensional space are *contravariant*.

Note that the rows of  $S$  appear as superscripts and the columns appear as subscripts. This convention is important and should be kept in mind.

We remark that orthonormality of the bases is nowhere required or even mentioned. Moreover, we have not even defined the concept of vectors being

orthogonal or normal, although you may know these definitions from previous studies; we will return to this topic later. All that is needed here are the concepts of linear dependence/independence, finite dimension, basis, and transformation of bases.

## 1.5 Interlude: Implicit Summation Notation

Looking at the equations derived in the preceding section, we observe the frequent use of the summation symbol  $\sum_{j=1}^n$ . When it comes to tensors, the equations will contain more and more summations and will become cumbersome and potentially confusing. Albert Einstein (yes, *the* Einstein) has noted that, when the range of summation is understood from the context (as is usually the case), the appearance of the summation symbol is redundant. Einstein thus proposed to eliminate the summation symbol and the summation range, and assume that the reader will mentally add them. This has come to be known as Einstein's notation, or implicit summation notation. Thus, equations (1.10), (1.11), (1.12) can be written as

$$\mathbf{x} = \mathbf{e}_j x^j, \quad \tilde{\mathbf{e}}_i = \mathbf{e}_j S_i^j, \quad \mathbf{e}_i = \tilde{\mathbf{e}}_j T_i^j, \quad \tilde{x}^i = T_j^i x^j, \quad x^i = S_j^i \tilde{x}^j \quad (1.13)$$

Einstein's notation rules can be summarized as follows:

1. Whenever the same index appears twice in an expression, once as a superscript and once as a subscript, summation over the range of that index is implied.
2. The range of summation is to be understood from the context; in case of doubt, the summation symbol should appear explicitly.
3. It is illegal for a summation index to appear more than twice in an expression, once as a superscript and once as a subscript.
4. The index used in implicit summation is arbitrary and may be replaced by any other index that is not already in use. For example,  $j$  in (1.13) may be replaced by  $k$  with no change in the meaning of the expressions. Such indices are called *dummy* or *bound*.

5. Indices other than dummy indices may appear any number of times and are *free*, in the same sense as common algebraic variables. For example,  $i$  in (1.13) is a free index. A free index may be replaced by any other index that is not already in use, provided that replacement is consistent throughout the equation. For example, replacing all  $i$ 's by  $m$ 's in (1.13) will not change the meaning.
6. The last rule is not in common use and we include it ad-hoc here, for a reason to be explained soon: Attempt to place the symbol carrying the summation index as subscript on the *left* of the symbol carrying the summation index as superscript. For example, write  $a_j^i x^j$  or  $x_j a_i^j$ , but avoid  $x^j a_j^i$  or  $a_i^j x_j$ .

Einstein's notation takes some getting used to, but then it becomes natural and convenient. As an example, let us use it for multiplying two matrices. Let  $a_j^i$  and  $b_m^k$  stand for the square matrices  $A$  and  $B$ , and recall that superscripts denote row indices and subscripts denote column indices. Then  $c_k^i = a_j^i b_k^j$  stands for the product  $C = AB$ , and  $d_j^i = b_k^i a_j^k$  stands the product  $D = BA$ . You may write  $d_j^i = a_j^k b_k^i$  and this will also be a legal expression for  $D$ , but the untrained reader may interpret this as  $C$ . Rule 6 helps avoiding such potential confusion because the order of terms in Einstein's notation then agrees with the order of operands in matrix multiplication. Sometimes, however, we must abandon rule 6 because other factors take precedence. For example, rule 6 cannot be used in the expression  $a_k^{ij} b_{mj}^k$ .

To conclude this interlude, we introduce the *Kronecker* delta symbol, defined as

$$\delta_j^i = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad (1.14)$$

The Kronecker delta is useful in many circumstances. Often it is used as a function of two integer variables. On other occasions it stands for the identity matrix  $I$ . In connection with Einstein's notation it may be used as a coordinate selector, in the following sense:

$$\delta_j^i x^j = x^i \quad (1.15)$$

It may be useful to contemplate equation (1.15) for a minute or so and be sure its meaning is understood. We will encounter further uses of the Kronecker delta later.

## 1.6 Covectors

Let  $\mathcal{V}$  be an  $n$ -dimensional space and  $(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n)$  a basis for  $\mathcal{V}$ . As explained in Appendix A, to  $\mathcal{V}$  there corresponds a dual space  $\mathcal{V}^*$  and to  $(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n)$  there corresponds a dual basis  $(\mathbf{f}^1, \mathbf{f}^2, \dots, \mathbf{f}^n)$ . The members of  $\mathcal{V}^*$  are called *dual vectors* or *covectors*. A covector  $\mathbf{y}$  can be expressed as a linear combination of the basis members  $\mathbf{y} = y_i \mathbf{f}^i$ ; note the use of implicit summation and that superscripts and subscripts are used in an opposite manner to their use in vectors.

Let  $S$  be the change-of-basis transformation from the basis  $(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n)$  to the basis  $(\tilde{\mathbf{e}}_1, \tilde{\mathbf{e}}_2, \dots, \tilde{\mathbf{e}}_n)$ . What is the corresponding transformation from the dual basis  $(\mathbf{f}^1, \mathbf{f}^2, \dots, \mathbf{f}^n)$  to the dual basis  $(\tilde{\mathbf{f}}^1, \tilde{\mathbf{f}}^2, \dots, \tilde{\mathbf{f}}^n)$ ? To answer this question, recall the definition of the members of a dual basis as "coordinate selectors" to find

$$\tilde{\mathbf{f}}^i(\tilde{\mathbf{x}}) = \tilde{x}^i = T_j^i x^j = T_j^i \mathbf{f}^j(\mathbf{x}) \quad (1.16)$$

Since this equality holds for all  $\mathbf{x} \in \mathcal{V}$ , necessarily

$$\tilde{\mathbf{f}}^i = T_j^i \mathbf{f}^j \quad (1.17)$$

We conclude that the members of the dual basis are transformed by change of basis using the inverse transformation  $T$ . It follows, as in the case of vectors, that the coordinates of covectors are transformed by change of basis using the direct transformation  $S$ :

$$\tilde{y}_i = y_j S_i^j \quad (1.18)$$

So, in summary, covectors behave opposite to the behavior of vectors under change of basis. Vector bases are transformed using  $S$  and vector coordinates are transformed using  $T$ . Covector bases are transformed using  $T$  and covector coordinates are transformed using  $S$ . Consequently, covectors are said to be *covariant* whereas vectors, as we recall, are *contravariant*.

Let us exemplify covectors and their covariant behavior by introducing functions of vectors. Consider a function  $f(\mathbf{x})$  which assigns to every vector a scalar value  $y = f(\mathbf{x})$ . Note that we are dealing with plane vectors again. The gradient of  $f(\mathbf{x})$  consists of the two partial derivatives  $(\partial f / \partial x^1, \partial f / \partial x^2)$ . In math (or physics) courses you probably learned that the gradient of a function is a vector. Let us examine the behavior of the gradient under a change

of basis. We want to express the partial derivatives of  $f(\mathbf{x})$  with respect to the coordinates of the new basis. Using the chain rule for partial derivatives we obtain

$$\begin{aligned}\frac{\partial f}{\partial \tilde{x}^1} &= \frac{\partial f}{\partial x^1} \frac{\partial x^1}{\partial \tilde{x}^1} + \frac{\partial f}{\partial x^2} \frac{\partial x^2}{\partial \tilde{x}^1} \\ \frac{\partial f}{\partial \tilde{x}^2} &= \frac{\partial f}{\partial x^1} \frac{\partial x^1}{\partial \tilde{x}^2} + \frac{\partial f}{\partial x^2} \frac{\partial x^2}{\partial \tilde{x}^2}\end{aligned}\tag{1.19}$$

Or, in Einstein's notation:

$$\frac{\partial f}{\partial \tilde{x}^i} = \frac{\partial f}{\partial x^j} \frac{\partial x^j}{\partial \tilde{x}^i}\tag{1.20}$$

But from (1.13) we know that

$$x^j = S_i^j \tilde{x}^i \quad \Rightarrow \quad \frac{\partial x^j}{\partial \tilde{x}^i} = S_i^j\tag{1.21}$$

If we try to substitute (1.21) in (1.20) we will find that things do not quite work well, because the index  $j$  will appear twice as a superscript, once in  $\partial x^j$  and once in  $S_i^j$ . Therefore, such substitution will result in an expression that does not conform to Einstein's notation. To remedy this problem, we introduce a new notational device for the partial derivatives, as follows:

$$\nabla_k f = \frac{\partial f}{\partial x^k}, \quad \tilde{\nabla}_k f = \frac{\partial f}{\partial \tilde{x}^k}\tag{1.22}$$

Then we can combine (1.20), (1.21), and (1.22) to find

$$\tilde{\nabla}_i f = (\nabla_j f) S_i^j\tag{1.23}$$

It is now evident that  $\nabla_i f$  cannot be regarded as vector, but as a covector, because it is transformed under a change of basis using the direct transformation, rather than the inverse transformation. We may further generalize and remove the function  $f$ , leaving only  $\nabla_i$ . The "covector"  $\nabla_i$  is, in fact, a *covariant operator*. It has no numeric value on its own, but when applied to a scalar-valued function of a vector, it produces a covector.  $\nabla_i$  is the gradient operator, expressed in coordinates relative to a given basis. The gradient operator is also called *del* or *nabla*. The preceding discussion applies to finite-dimensional vector spaces of any dimension. You are probably familiar with its use in classical physics, in the 3-dimensional physical space.

## 1.7 Linear Operators on Vector Spaces

Having covered vectors and covectors, and their laws of transformation under change of basis, we are now ready to introduce new geometric objects. A linear operator on an  $n$ -dimensional vector space  $\mathcal{V}$  is a function  $f : \mathcal{V} \rightarrow \mathcal{V}$  which is additive and homogeneous. We thus require that

$$f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y}), \quad f(a\mathbf{x}) = af(\mathbf{x}) \quad (1.24)$$

A linear operator acts on the coordinates of a vector in a linear way; namely, each coordinate of the result is a linear combination of the coordinates of the argument. So, if  $\mathbf{y} = f(\mathbf{x})$ , then

$$y^i = f_j^i x^j \quad (1.25)$$

**Remark:** Although the notation  $f_j^i$  resembles the notations  $S_i^j, T_i^j$  used for the direct and inverse basis transformations, there is subtle but important difference in our interpretation of these notations. The objects  $S_i^j, T_i^j$  are nothing to us but square arrays of real numbers and their definitions depend on a specific choice of bases, so they both depend on *two* bases (the old and the new). By contrast, we interpret  $f_j^i$  as a *basis-independent* geometric object, whose numerical representation depends on a single chosen basis.

Let us explore the transformation law for  $f_j^i$  when changing from a basis  $\mathbf{e}_i$  to a basis  $\tilde{\mathbf{e}}_i$ . We find from the contravariance of  $x^i$  and  $y^i$  that

$$\tilde{y}^k = T_i^k y^i = T_i^k f_j^i x^j = T_i^k f_j^i S_m^j \tilde{x}^m \quad (1.26)$$

Hence we conclude that, in order for the relation (1.25) to hold in the new basis, we must have

$$\tilde{f}_m^k = T_i^k f_j^i S_m^j \quad (1.27)$$

Expression (1.27) is the transformation law for linear operators. As we see, the transformation involves both the direct and the inverse transformations. Therefore a linear operator is contravariant in one index and covariant in the second index. The transformation (1.27) can also be expressed in matrix form as  $\tilde{F} = TFS = S^{-1}FS$ , which is the way it is presented in linear algebra.

## 1.8 Tensors

Vectors, covectors, and linear operators are all special cases of tensors. We will not attempt to define tensors in abstract terms, but settle for a coordinate-based definition, as follows.

A *tensor of type* (or *valency*)  $(r, s)$  over an  $n$ -dimensional vector space is an object consisting of  $n^{r+s}$  coordinates, denoted by the generic symbol  $a_{j_1 \dots j_s}^{i_1 \dots i_r}$ , and obeying the following change-of-basis transformation law:

$$\tilde{a}_{j_1 \dots j_s}^{i_1 \dots i_r} = T_{k_1}^{i_1} \dots T_{k_r}^{i_r} a_{m_1 \dots m_s}^{k_1 \dots k_r} S_{j_1}^{m_1} \dots S_{j_s}^{m_s} \quad (1.28)$$

Let us spend some time discussing this equation. According to Einstein's summation notation, summation must be performed  $r + s$  times, using the indices  $k_1 \dots k_r, m_1 \dots m_s$ . At what order should these summations be performed? The answer is: it does not matter — any order would yield the same result. To see this, apply the distributive and commutative laws of real numbers, open all parentheses in (1.28) (you don't see the parentheses, but they are there!), and think of (1.28) as a flat multiple sum

$$\tilde{a}_{j_1 \dots j_s}^{i_1 \dots i_r} = \sum_{k_1=1}^n \dots \sum_{k_r=1}^n \sum_{m_1=1}^n \dots \sum_{m_s=1}^n T_{k_1}^{i_1} \dots T_{k_r}^{i_r} a_{m_1 \dots m_s}^{k_1 \dots k_r} S_{j_1}^{m_1} \dots S_{j_s}^{m_s} \quad (1.29)$$

The coordinates  $i_1 \dots i_r$  are the *contravariant coordinates* and the coordinates  $j_1 \dots j_s$  are the *covariant coordinates*, in an obvious generalization of the terminology used for vectors and covectors. You should remember that contravariant coordinates appear as superscripts and are transformed using  $T$ ; covariant coordinates appear as subscripts and are transformed using  $S$ .

We know from linear algebra that a vector is an abstract object, which should be distinguished from its coordinate representation. Whereas the coordinates depend on the choice of basis, the vector itself is invariant. The same applies to tensors. Tensors are abstract objects and what we see in (1.28) is only a law of transformation of the coordinates of the tensor, while the tensor itself is invariant.

One way to think of a tensor is as storage (say a computer memory) having capacity of  $n^{r+s}$  addresses, each address holding a real number. Think of

the address as an integer represented in radix  $n$  and call an  $n$ -radix digit "enit" (derived from "bit"). Then  $k_1 \dots k_r m_1 \dots m_s$  may be thought of as an address consisting of  $r + s$  enits. However, the  $r$  leftmost enits and the  $s$  rightmost enits are used differently in specifying the address: the former are used as superscripts and the latter as subscripts. Thus  $k_1 \dots k_r$  is the contravariant part of the address and  $m_1 \dots m_s$  is the covariant part of the address. The contents of the storage, which is to say the coordinates, depend on the basis. Under change of basis, the transformation law (1.28) applies.

Looking back at what we presented in the preceding sections, we now realize that a contravariant vector is in fact a  $(1, 0)$ -tensor and a covariant vector is a  $(0, 1)$ -tensor. A linear operator on a vector space is a  $(1, 1)$ -tensor. A scalar can be regarded as a  $(0, 0)$ -tensor, since  $n^0 = 1$ ; therefore a  $(0, 0)$ -tensor has a single address and stores a single value.

## 1.9 Operations on Tensors

Having defined the general concept of tensor over an  $n$ -dimensional vector space, let us now introduce the basic arithmetic operations involving tensors.

### 1.9.1 Addition

Two tensors of the same type can be added term-by-term. The expression

$$c_{j_1 \dots j_s}^{i_1 \dots i_r} = a_{j_1 \dots j_s}^{i_1 \dots i_r} + b_{j_1 \dots j_s}^{i_1 \dots i_r} \quad (1.30)$$

means that each coordinate on the left hand holds the sum of the corresponding coordinates on the right side. We can write tensor addition symbolically as  $c = a + b$ . Tensor addition is obviously commutative. Furthermore, it is straightforward to verify that the change-of-basis transformation law holds for  $c$ , hence  $c$  is indeed a tensor.

#### Remarks:

1. Tensors of different ranks cannot be added.

2. The tensor 0 can be defined as a tensor of any rank whose coordinates are all 0. Therefore, 0 is not a uniquely defined tensor, but if we accept this vague definition, then  $a + 0 = 0 + a = a$  holds for any tensor  $a$ .

## 1.9.2 Multiplication by a scalar

Each of the coordinates of a tensor can be multiplied by a given scalar to yield a new tensor of the same type; this is expressed as follows:

$$c_{j_1 \dots j_s}^{i_1 \dots i_r} = x a_{j_1 \dots j_s}^{i_1 \dots i_r} \quad (1.31)$$

We can write tensor multiplication by a scalar symbolically as  $c = xa$ . It is straightforward to verify that the change-of-basis transformation law holds for  $c$ , hence  $c$  is indeed a tensor. It is also easy to check that

$$x(a + b) = xa + xb, \quad (x + y)a = xa + ya, \quad (xy)a = x(ya) = y(xa) \quad (1.32)$$

## 1.9.3 The Tensor Product

Let  $a$  be an  $(r, s)$ -tensor and  $b$  a  $(p, q)$ -tensor. We write the coordinates of the first tensor as  $a_{j_1 \dots j_s}^{i_1 \dots i_r}$  and those of the second tensor as  $b_{j_{s+1} \dots j_{s+q}}^{i_{r+1} \dots i_{r+p}}$ . Note that all indices are distinct within and across tensors. The tensor product  $c = a \otimes b$  is defined as the  $(r + p, s + q)$ -tensor having the coordinates

$$c_{j_1 \dots j_s j_{s+1} \dots j_{s+q}}^{i_1 \dots i_r i_{r+1} \dots i_{r+p}} = a_{j_1 \dots j_s}^{i_1 \dots i_r} b_{j_{s+1} \dots j_{s+q}}^{i_{r+1} \dots i_{r+p}} \quad (1.33)$$

Let us elaborate more on this definition. Recall the storage interpretation of tensors, explained in Sec. 1.8. The tensor  $c$  is allocated  $n^{r+p+s+q}$  addresses. The  $r+p$  contravariant enits of the address  $i_1 \dots i_r i_{r+1} \dots i_{r+p}$  are subdivided into  $r$  leftmost enits and  $p$  rightmost enits. Similarly, the  $s+q$  covariant enits of the address  $j_1 \dots j_s j_{s+1} \dots j_{s+q}$  are subdivided into  $s$  leftmost enits and  $q$  rightmost enits. The rule for filling the addresses of  $c$  with contents is as follows. Access the number in  $a$  having the address from the contravariant  $r$ -part and covariant  $s$ -part and call it  $x$ . Access the number in  $b$  having the address from the contravariant  $p$ -part and covariant  $q$ -part and call it  $y$ . Then the corresponding value of the  $c$  address is  $xy$ .

That the tensor product is indeed a tensor is almost self-evident and is left to the reader as an exercise. Simply write the change-of-basis transformation to each factor, examine the resulting expression and compare it with the change-of-basis transformation needed for  $c$  to be a tensor.

The coordinates of the tensor product  $a \otimes b$  comprise all possible products of coordinates of  $a$  and coordinates of  $b$ . Tensor product is not commutative; that is,  $a \otimes b \neq b \otimes a$ . The reason is that, although all possible products of coordinates of the two tensors appear in both products, they appear at different locations (or different addresses, if you wish). Let us exemplify this in the case where both tensors are  $(1, 0)$ , say  $a^i$  and  $b^k$ . Let  $c^{ik} = a^i b^k$  and  $d^{ki} = b^k a^i$ . Let  $g$  and  $h$  denote specific enits such that  $g \neq h$ ; for example,  $g = 1, h = 2$ . Let  $u = a^g, v = a^h, x = b^g, y = b^h$  (these are specific values, not general variables!) Then  $c^{gh} = uy$ , but  $d^{gh} = xv$ , so  $c^{gh} \neq d^{gh}$  in general.

The tensor product is *bilinear*, in the following sense:

$$(xa + yb) \otimes c = x(a \otimes c) + y(b \otimes c), \quad c \otimes (xa + yb) = x(c \otimes a) + y(c \otimes b) \quad (1.34)$$

We can extend the tensor product to any finite number of tensors in an obvious manner. If  $\{a(i), 1 \leq i \leq m\}$  are tensors, then

$$a(1) \otimes a(2) \otimes \dots \otimes a(m)$$

denotes their tensor product. This tensor product is *multilinear* in the following sense: For every  $k$  we have

$$\begin{aligned} a(1) \otimes \dots [xa(k) + yb] \otimes \dots a(m) = \\ x[a(1) \otimes \dots a(k) \otimes \dots a(m)] + y[a(1) \otimes \dots b \otimes \dots a(m)] \end{aligned} \quad (1.35)$$

## 1.9.4 Contraction

Let  $a$  be an  $(r, s)$ -tensor. Choose the contravariant index at any position, say the  $i$ -th position and rename it by a new symbol, say  $k$ . Next choose a covariant index at any position, say the  $j$ -th position and rename it by the same symbol  $k$ . Let

$$b_{h_1 \dots h_{j-1} h_{j+1} \dots h_s}^{g_1 \dots g_{i-1} g_{i+1} \dots g_r} = a_{h_1 \dots h_{j-1} k h_{j+1} \dots h_s}^{g_1 \dots g_{i-1} k g_{i+1} \dots g_r} \quad (1.36)$$

Note how the index  $k$  disappears through the implied summation and the resulting tensor has type  $(r - 1, s - 1)$ . The operation (1.36) is called *contraction*. For a general  $(r, s)$ -tensor there are  $rs$  possible contractions, one for each pair of contravariant and covariant indices.

To see that (1.36) indeed defines a tensor, consider first a simple case of a  $(1, 1)$  tensor  $a_h^g$ . Then  $a_g^g$  is the trace of  $a$ . Applying a change-of-basis transformation to  $a$  and then computing the trace gives

$$T_g^m a_h^g S_m^h = a_h^g \delta_g^h = a_g^g \quad (1.37)$$

Hence the trace is invariant under change of basis. Now replace  $a_h^g$  by a general tensor and repeat (1.37) for a chosen pair of indices to find

$$T_{g_i}^k a_{h_1 \dots h_{j-1} h_{j+1} \dots h_s}^{g_1 \dots g_{i-1} g_{i+1} \dots g_r} S_k^{h_j} = a_{h_1 \dots h_{j-1} g_i h_{j+1} \dots h_s}^{g_1 \dots g_{i-1} g_{i+1} \dots g_r} = b_{h_1 \dots h_{j-1} h_{j+1} \dots h_s}^{g_1 \dots g_{i-1} g_{i+1} \dots g_r} \quad (1.38)$$

Now, applying the transformations associated with the remaining indices would complete the change-of-basis transformation for (1.36), resulting in  $\tilde{b}_{h_1 \dots h_{j-1} h_{j+1} \dots h_s}^{g_1 \dots g_{i-1} g_{i+1} \dots g_r}$  as required.

In view of (1.37), contraction is a generalized trace operation. Often, implied summation can be viewed as outer product followed by contraction. Consider, for example, the following operation between a covector and a vector:

$$c = a_i b^i \quad (1.39)$$

The result of (1.39) is a scalar, called the *scalar product* of  $a_j$  and  $b^i$ . If we form the outer product  $d_j^i = a_j b^i$  and then contract the upper and lower indexes of  $d_j^i$ , we obtain the scalar product (1.39). It follows immediately that scalar product is invariant under change of basis. The scalar product has important applications in quantum mechanics, as well as in other fields, but is less general than the inner product, to be introduced soon.

## 1.10 Tensor Spaces as Vector Spaces; New Notations

Consider the set  $\mathcal{T}(r, s)$  of all  $(r, s)$ -tensors, including the zero tensor  $0$ . Equipped with the addition operation of two tensors and the multiplication

operation of a scalar by a tensor, this set becomes a vector space of dimension  $n^{r+s}$ . Let  $(\mathbf{e}_1, \dots, \mathbf{e}_n)$  be a basis for  $\mathcal{V}$  and  $(\mathbf{f}^1, \dots, \mathbf{f}^n)$  a basis for  $\mathcal{V}^*$ . Then all tensor products  $\mathbf{e}_{i_1} \otimes \dots \otimes \mathbf{e}_{i_r} \otimes \mathbf{f}^{j_1} \otimes \dots \otimes \mathbf{f}^{j_s}$ , whose number is  $n^{r+s}$ , constitute a basis for  $\mathcal{T}(r, s)$ . To save space, we will henceforth use the abridged notation  $\mathbf{e}_{i_1} \dots \mathbf{e}_{i_r} \mathbf{f}^{j_1} \dots \mathbf{f}^{j_s}$  for the basis tensors (i.e., the  $\otimes$  operation between the basis vectors is implied).

We can now express an  $(r, s)$ -tensor in a way similar to vectors:

$$a = a_{j_1 \dots j_s}^{i_1 \dots i_r} \mathbf{e}_{i_1} \dots \mathbf{e}_{i_r} \mathbf{f}^{j_1} \dots \mathbf{f}^{j_s} \quad (1.40)$$

Expression (1.40) is to be understood, as usual, to imply summation over all indices. This expression is a formal, unambiguous way to write a tensor.

Expression (1.40) brings out a problem that is both fundamental and notational. We implicitly assumed that, when writing a basis tensor

$$\mathbf{e}_{i_1} \dots \mathbf{e}_{i_r} \mathbf{f}^{j_1} \dots \mathbf{f}^{j_s}$$

we place the contravariant vectors first, followed by the covariant vectors. But why not use the reverse order? Moreover, why not mix the two sets? Consider, for example, the space  $\mathcal{T}(2, 1)$ . Are the basis vectors  $\mathbf{e}_i \mathbf{e}_j \mathbf{f}^k$ , or perhaps  $\mathbf{f}^k \mathbf{e}_i \mathbf{e}_j$  or even  $\mathbf{e}_i \mathbf{f}^k \mathbf{e}_j$ ? Since a tensor product is not commutative, all three possibilities are different and all three are equally valid. It follows that  $\mathcal{T}(2, 1)$  is not a well-defined object and can stand for three different tensor spaces (although the three are isomorphic). In general, there are  $\binom{r+s}{r}$  different spaces  $\mathcal{T}(r, s)$ , corresponding to the different ways of mixing covariant and contravariant basis vectors. We can always distinguish between the different possibilities if we replace the generic  $\mathcal{T}(r, s)$  by a specific product involving  $\mathcal{V}$  and  $\mathcal{V}^*$  in the required order. For example, in the case of  $\mathcal{T}(2, 1)$ , we may use one of the following three notations:

$$\mathcal{V} \otimes \mathcal{V} \otimes \mathcal{V}^*, \quad \mathcal{V} \otimes \mathcal{V}^* \otimes \mathcal{V}, \quad \mathcal{V}^* \otimes \mathcal{V} \otimes \mathcal{V}$$

However, this works only for specific  $r$  and  $s$ , and a specific order of contravariance/covariance.

The full notation (1.40) disambiguates the tensor; for example, in the case of  $(2, 1)$ -tensors, we understand  $a_k^{ij} \mathbf{e}_i \mathbf{e}_j \mathbf{f}^k$ ,  $a_k^{ij} \mathbf{f}^k \mathbf{e}_i \mathbf{e}_j$ , and  $a_k^{ij} \mathbf{e}_i \mathbf{f}^k \mathbf{e}_j$  as being different. But what about the simplified coordinate notation  $a_k^{ij}$  that we have

used until now? In fact, many authors distinguish between  $a_{k}^{ij}$ ,  $a_k^{ij}$ , and  $a_k^{i j}$ . The coordinates of general tensors are often written as  $a_{j_1 \dots j_s}^{i_1 \dots i_r}$  when the contravariant basis vectors are meant to come first. Authors who use such notations may also use  $S_j^i$  for transformation matrices. Another common notation is  $\Lambda_{\mu'}^{\mu}$  for the direct transformation matrix and  $\Lambda_{\mu}^{\mu'}$  for its inverse. Here the same symbol  $\Lambda$  is used for both transformations and one must pay attention to the location of primes to distinguish between the two transformations. Such notations quickly become awkward and difficult to follow<sup>1</sup>. We have chosen to adopt a pragmatic approach. We will continue to use  $a_{j_1 \dots j_s}^{i_1 \dots i_r}$  whenever the distinction between the  $\binom{r+s}{r}$  possibilities is unimportant; such was the case until now. In the future, we will adopt the notation  $a_{j_1 \dots j_s}^{i_1 \dots i_r}$  and its mixed variations only when distinction is significant and ambiguity may arise otherwise.

## 1.11 Inner Products, The Gram Matrix, and Metric Tensors

The material in section is extremely important, albeit not difficult. Be sure to understand it fully and review several times if necessary.

### 1.11.1 Inner Products

Little progress can be made in the application of tensors to physics problems without imposing additional structure on the underlying vector space. The defining axioms of vector space include no operations that act on vectors and produce scalars. Here we introduce one such function, arguably the most important one: the inner product. Let  $\mathbf{u}$  and  $\mathbf{v}$  be vectors in a vector space  $\mathcal{V}$  and denote by  $\mathbf{u} \cdot \mathbf{v}$  a function acting on  $\mathbf{u}$  and  $\mathbf{v}$  and producing a scalar  $a = \mathbf{u} \cdot \mathbf{v}$ , such that the following property holds:

---

<sup>1</sup>And also a nightmare to typeset.

1. Bilinearity:

$$(a\mathbf{u}_1 + b\mathbf{u}_2) \cdot \mathbf{v} = a(\mathbf{u}_1 \cdot \mathbf{v}) + b(\mathbf{u}_2 \cdot \mathbf{v})$$

$$\mathbf{u} \cdot (a\mathbf{v}_1 + b\mathbf{v}_2) = a(\mathbf{u} \cdot \mathbf{v}_1) + b(\mathbf{u} \cdot \mathbf{v}_2)$$

for all  $\mathbf{u}, \mathbf{u}_1, \mathbf{u}_2, \mathbf{v}, \mathbf{v}_1, \mathbf{v}_2 \in \mathcal{V}$  and  $a, b \in \mathbb{R}$ .

Such a function is called an *inner product* on the space  $\mathcal{V}$ . If, in addition, the following property holds:

2. Symmetry:

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$$

for all  $\mathbf{u}, \mathbf{v} \in \mathcal{V}$ , then the inner product is called *symmetric*.

If, furthermore, the following property holds:

3. Nondegeneracy:

$$\mathbf{u} \cdot \mathbf{x} = 0 \text{ for all } \mathbf{u} \in \mathcal{V} \Rightarrow \mathbf{x} = \mathbf{0}$$

then the inner product is called *nondegenerate*.

A vector space equipped with an inner product is called an *inner product space*. We will only consider symmetric nondegenerate inner products, without mentioning this explicitly every time.

When we substitute  $\mathbf{u} = \mathbf{v}$  in the inner product, the resulting scalar-valued function  $\mathbf{u} \cdot \mathbf{u}$  of the vector  $\mathbf{u}$  is called the quadratic form induced by the inner product. A quadratic form satisfying  $\mathbf{u} \cdot \mathbf{u} > 0$  for all  $\mathbf{u} \neq \mathbf{0}$  is called *positive*. An inner product space whose associated quadratic form is positive is called *Euclidean*.

An inner product space admits the concept of orthogonality. Vectors  $\mathbf{u}$  and  $\mathbf{v}$  are *orthogonal* if  $\mathbf{u} \cdot \mathbf{v} = 0$ . The notation  $\mathbf{u} \perp \mathbf{v}$  is used to signify orthogonality of two vectors. If the space is Euclidean, then the concept of length is also defined. The *length* (or *Euclidean norm*) of  $\mathbf{u}$  is the nonnegative scalar  $\|\mathbf{u}\| = \sqrt{\mathbf{u} \cdot \mathbf{u}}$ .

A basis  $(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n)$  in an  $n$ -dimensional inner product space is *orthogonal* if  $\mathbf{e}_i \perp \mathbf{e}_j$  for all  $i \neq j$ . If the space is Euclidean, a basis is *orthonormal* if it is orthogonal and, additionally,  $\|\mathbf{e}_i\| = 1$  for all  $i$ . Every inner product space possesses an orthogonal basis and every Euclidean space possesses an orthonormal basis. These are well known results in linear algebra and we will not prove them here.

### 1.11.2 The Gram Matrix

Let us express the inner product in some basis  $(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n)$ . Let  $\mathbf{u} = \mathbf{e}_i u^i$  and  $\mathbf{v} = \mathbf{e}_j v^j$  be some vectors. Then, using the bilinearity of the inner product, we find that

$$\mathbf{u} \cdot \mathbf{v} = (\mathbf{e}_i \cdot \mathbf{e}_j) u^i v^j \quad (1.41)$$

The entity  $\{\mathbf{e}_i \cdot \mathbf{e}_j, 1 \leq i, j \leq n\}$ , consisting of  $n^2$  numbers, is usually considered as an  $n \times n$  matrix in linear algebra, and is called the *Gram matrix* of the basis. Let us denote this matrix by  $G$ . By symmetry of the inner product,  $G$  is symmetric. As we now prove, this matrix is nonsingular.

**Theorem 1.** *The Gram matrix  $G$  is nonsingular.*

*Proof.* Let us assume that there exists a vector  $\mathbf{x}$  such that

$$(\mathbf{e}_i \cdot \mathbf{e}_j) x^j = (\mathbf{e}_i \cdot \mathbf{x}) = 0.$$

Since this holds for all  $i$ ,  $\mathbf{x}$  is orthogonal to every member of the basis. It follows that  $\mathbf{x}$  is orthogonal to every vector in the space. But then it follows from the nondegeneracy of the inner product that  $\mathbf{x} = 0$ . Since the only vector in the null space of  $G$  is the zero vector,  $G$  is nonsingular.  $\square$

We now examine the behavior of  $G$  under change of basis. Consider a new basis  $(\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n)$ , related to the old basis  $(\mathbf{e}_1, \dots, \mathbf{e}_n)$  through the transformation  $S$ . Then,

$$\tilde{\mathbf{e}}_i \cdot \tilde{\mathbf{e}}_j = (\mathbf{e}_k S_i^k) \cdot (\mathbf{e}_m S_j^m) = (\mathbf{e}_k \cdot \mathbf{e}_m) S_i^k S_j^m \quad (1.42)$$

Or, in matrix form,

$$\tilde{G} = S'GS \quad (1.43)$$

where  $S'$  is the transpose of  $S$ . Matrices  $G$  and  $\tilde{G}$  related by (1.43) with nonsingular  $S$  are called *congruent*. Thus a change of basis of a vector space provides a congruence relation between their corresponding Gram matrices.

The celebrated *Sylvester law of inertia* asserts the following:

**Theorem 2** (Sylvester). *Every real symmetric matrix  $G$  is congruent to a diagonal matrix  $\Lambda$  whose entries have values  $+1$ ,  $-1$ , or  $0$ . The matrix  $\Lambda$  is unique for all matrices congruent to  $G$  (up to the ordering of the diagonal entries of  $\Lambda$ ).*

If  $n$  is the dimension of  $G$  then, by Sylvester's theorem,  $n$  can be expressed as  $n = n^+ + n^- + n^0$ , according to the numbers of  $+1$ ,  $-1$ , and  $0$  along the diagonal of  $\Lambda$ . The triplet  $(n^+, n^-, n^0)$  is called the *signature* of  $G$ . Sylvester's theorem tells us that the signature is invariant under congruence relation. If  $G$  is nonsingular, as in our case, then necessarily  $n^0 = 0$  and  $n = n^+ + n^-$ . If  $G$  is positive, then  $n^- = n^0 = 0$  and  $n = n^+$ . This happens if and only if the space is Euclidean.

### 1.11.3 The Metric Tensor

Examination of (1.42) reveals that  $(\mathbf{e}_k \cdot \mathbf{e}_m)$  is transformed like a  $(0, 2)$ -tensor under change of basis. Defining  $g_{ij} = \mathbf{e}_i \cdot \mathbf{e}_j$ , we have

$$\mathbf{u} \cdot \mathbf{v} = g_{ij} u^i v^j \quad (1.44)$$

The  $(0, 2)$ -tensor  $g_{ij}$  is called the *metric tensor* of the inner product space. Like all tensors, it is a geometric object, invariant under change-of-basis transformations. By Sylvester's theorem, there exists a basis which makes the metric diagonal and reveals the signature of the space. This signature is uniquely defined by the definition of the inner product. It immediately follows from (1.44) that the inner product is invariant under change of basis. This is not surprising, since the definition of inner product does not depend on a basis.

Since, by our assumption,  $G$  is nonsingular, it possesses an inverse  $G^{-1}$ . The entries of  $G^{-1}$  may be viewed as the coordinates of a  $(2, 0)$ -tensor, called the *dual metric tensor*, and usually denoted by  $g^{ij}$ . It follows immediately that

$$g_{jk} g^{ki} = \delta_j^i \quad (1.45)$$

We may summarize the preceding discussion by the following theorem:

**Theorem 3.** *For every finite-dimensional inner product space there exists a unique symmetric nonsingular  $(0, 2)$ -tensor  $g_{ij}$  such that  $\mathbf{u} \cdot \mathbf{v} = g_{ij} u^i v^j$  for any pair of vectors  $\mathbf{u}$  and  $\mathbf{v}$ . Conversely, if  $g_{ij}$  is a symmetric nonsingular  $(0, 2)$ -tensor on a finite-dimensional vector space, then an inner product  $\mathbf{u} \cdot \mathbf{v}$  is uniquely defined such that  $\mathbf{u} \cdot \mathbf{v} = g_{ij} u^i v^j$  for any pair of vectors  $\mathbf{u}$  and  $\mathbf{v}$ .*

If the vector space is Euclidean and the basis is orthonormal, then the coordinates of the metric tensor in this basis are, by definition,  $g_{ij} = \delta_{ij}$ , where  $\delta_{ij}$  is defined in a similar manner to the Kronecker delta  $\delta_j^i$ . In this case the inner product is simply given by  $\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^n u^i v^i$ . This is the inner-product expression usually encountered in linear algebra courses. Our discussion makes it clear that this expression holds only when the space is Euclidean and the basis is orthonormal.

### 1.11.4 Example: The Minkowski Space

We will now provide an example of non-Euclidean inner product space, known as the Minkowski space. This is a 4-dimensional inner product vector space possessing an orthogonal basis  $(\mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$  and a metric tensor whose coordinates in this orthogonal basis are

$$g_{ij} = \begin{cases} -1, & i = j = 0 \\ 1, & i = j = 1, 2, 3 \\ 0, & i \neq j \end{cases} \quad (1.46)$$

The metric of this space has signature  $n^+ = 3, n^- = 1$ . Some authors<sup>2</sup> use the negative of (1.46) as the definition of Minkowski space, in which case  $n^+ = 1, n^- = 3$ .

The Minkowski space is clearly non-Euclidean; indeed, this space underlies relativity theory, so it is the space in which our universe exists! It is common, in relativity theory, to number the dimensions starting at 0 (rather than 1). The index 0 is associated with  $ct$ , time multiplied by the speed of light. The remaining indices are associated with the usual space coordinates  $x, y, z$ . However, relativity theory convention for the coordinates is  $(x^0, x^1, x^2, x^3)$ , in agreement with tensor notation (be careful not to confuse contravariant indices with powers!)

Let  $\mathbf{x}$  be a vector in the Minkowski space, expressed in the time-space basis of the space. Then we have

$$\mathbf{x} \cdot \mathbf{x} = -(x^0)^2 + \sum_{i=1}^3 (x^i)^2 \quad (1.47)$$

---

<sup>2</sup>This inconsistency has plagued relativity theory for a century and continues to do so.

Clearly,  $\mathbf{x} \cdot \mathbf{x}$  is not always nonnegative because the inner product of the Minkowski space is not positive. The following terminology is in use, depending on the sign of  $\mathbf{x} \cdot \mathbf{x}$ :

$$\mathbf{x} \cdot \mathbf{x} \begin{cases} < 0 : \textit{timelike} \\ = 0 : \textit{lightlike} \\ > 0 : \textit{spacelike} \end{cases}$$

There is much more to tell about relativity theory, but this is not the place to do so.

## 1.12 Lowering and Raising of Tensors

Let  $a_{j_1 \dots j_s}^{i_1 \dots i_r}$  be the coordinates of the  $(r, s)$ -tensor  $a$  in some basis and  $g_{ij}$  be the metric tensor in this basis. Let us form the tensor product  $g_{pq} a_{j_1 \dots j_s}^{i_1 \dots i_r}$ . This tensor has type  $(r, s+2)$ . Now choose one of the contravariant coordinates of  $a$ , say  $i_k$ . Replace  $i_k$  by  $q$  and perform contraction with respect to  $q$ . Then  $q$  will disappear and we will be left with a tensor of type  $(r-1, s+1)$

$$b_{pj_1 \dots j_s}^{i_1 \dots i_{k-1} i_{k+1} \dots i_r} = g_{pq} a_{j_1 \dots j_s}^{i_1 \dots i_{k-1} q i_{k+1} \dots i_r} \quad (1.48)$$

This operation is called *lowering*. Lowering acts to decrease the contravariance valency by 1 and increase the covariant valency by 1. There are  $r$  possible lowerings, depending on the choice of  $k$ . Note that the new covariant index of the result,  $p$  in (1.48), is placed in the first position. If different placement is necessary, some ad-hoc notation must be used.

Raising is the dual of lowering. We start with the dual metric tensor  $g^{pq}$  and form the tensor product  $g^{pq} a_{j_1 \dots j_s}^{i_1 \dots i_r}$ . We choose an index  $j_k$ , replace  $j_k$  by  $q$  and perform contraction with respect to  $q$ , obtaining

$$c_{j_1 \dots j_{k-1} j_{k+1} \dots j_s}^{pi_1 \dots i_r} = g^{pq} a_{j_1 \dots j_{k-1} q j_{k+1} \dots j_s}^{i_1 \dots i_r} \quad (1.49)$$

This operation is called *raising*. Raising acts to increase the contravariance valency by 1 and decrease the covariant valency by 1, so the resulting tensor has type  $(r+1, s-1)$ . There are  $s$  possible raisings, depending on the choice of  $k$ . Note that the new contravariant index of the result,  $p$  in (1.49),

is placed in the first position. If different placement is necessary, some ad-hoc notation must be used.

A common use of lowering and raising is in moving between vectors and covectors. If  $v^i$  is a vector in some basis, we define its corresponding covector  $v_i$  through the relationships

$$v_i = g_{ik}v^k, \quad v^i = g^{ik}v_k \quad (1.50)$$

These relationships establish a natural isomorphism between the given vector space  $\mathcal{V}$  and its dual space of covectors  $\mathcal{V}^*$ .

## 1.13 The Levi-Civita Symbol and Related Topics

### 1.13.1 Permutations and Parity

A *permutation* of the integers  $(12 \dots n)$  is simply a rearrangement of this set of integers in a different order, say  $(i_1i_2 \dots i_n)$ . As you surely know, there are  $n!$  different permutations of  $n$  integers.

Suppose we want to undo the permutation  $(i_1i_2 \dots i_n)$  by some mechanical procedure (e.g., a procedure which may be programmed and executed by a computer). One way of doing this is as follows. Look for the number 1 among  $(i_1i_2 \dots i_n)$ . If  $i_1 = 1$ , do nothing; otherwise if  $i_j = 1$ , interchange  $i_j$  and  $i_1$ . Such an operation is called *transposition* and in this case it will bring 1 to the first position and will move  $i_1$  to where the 1 was. Next repeat for the number 2. If  $i_2 = 2$ , do nothing; otherwise if  $i_j = 2$ , interchange  $i_j$  and  $i_2$ . Repeat the same procedure until all numbers from 1 to  $n$  have been moved to their natural locations.

The total number of transpositions in the preceding procedure determines the *parity* of the permutation  $(i_1i_2 \dots i_n)$ . If the number is even, the parity is *even*; if the number is odd, the parity is *odd*.

### 1.13.2 The Levi-Civita Symbol

The *Levi-Civita Symbol*  $\epsilon_{i_1 i_2 \dots i_n}$  is a function of  $n$  indices, each taking values from 1 to  $n$ . It is therefore fully defined by  $n^n$  values, one for each choice of indices. The definition of the Levi-Civita symbol is as follows.

$$\epsilon_{i_1 i_2 \dots i_n} = \begin{cases} 1, & i_1 i_2 \dots i_n \text{ is an even permutation of } 12 \dots n \\ -1, & i_1 i_2 \dots i_n \text{ is an odd permutation of } 12 \dots n \\ 0, & i_1 i_2 \dots i_n \text{ is not a permutation of } 12 \dots n \end{cases} \quad (1.51)$$

As we see,  $\epsilon_{i_1 i_2 \dots i_n}$  is 1 in  $n!/2$  cases out of  $n^n$ , is  $-1$  in  $n!/2$  cases, and is 0 in all other cases.

Let  $A$  be an  $n \times n$  matrix. Using the Levi-Civita symbol, the determinant of  $A$  can be expressed as

$$\det A = \epsilon_{i_1 i_2 \dots i_n} A_1^{i_1} A_2^{i_2} \dots A_n^{i_n} \quad (1.52)$$

with implied summation over all indices. We are used to seeing the determinant definition as a sum of  $n!$  terms whereas (1.52) contains  $n^n$  terms. However, only  $n!$  of these are nonzero, and each nonzero term is equal to a product of  $n$  elements of  $A$  taken from all rows and all columns, with positive or negative sign depending on the parity of the corresponding row permutation relative to the order of columns. Thus (1.52) is nothing but the definition of a determinant.

### 1.13.3 The Volume Tensor/Pseudotensor

Although the Levi-Civita symbol is written in tensor notation, we refrained from calling it so. And, indeed, let us check the behavior of the Levi-Civita symbol under change of basis and see if it transformed like a tensor. If  $\epsilon_{i_1 i_2 \dots i_n}$  were a tensor, it would have to be an  $(0, n)$ -tensor. As such, its transformation law would have to be

$$\tilde{\epsilon}_{j_1 j_2 \dots j_n} = \epsilon_{i_1 i_2 \dots i_n} S_{j_1}^{i_1} S_{j_2}^{i_2} \dots S_{j_n}^{i_n} \quad (1.53)$$

However, comparing the right side of (1.53) with (1.52), we see that it is equal to  $\det S$  if  $j_1 j_2 \dots j_n$  is an even permutation of  $i_1 i_2 \dots i_n$ , to  $-(\det S)$  if

it is an odd permutation, and to 0 if it is not a permutation at all. It follows that

$$\tilde{\epsilon}_{j_1 j_2 \dots j_n} = (\det S) \epsilon_{j_1 j_2 \dots j_n} \quad (1.54)$$

The expression (1.54) indicates why  $\epsilon_{i_1 i_2 \dots i_n}$  cannot be a tensor. If it were, the values of its coordinates would be in the set  $\{1, -1, 0\}$  in some preferred basis and would have different values in all other bases. But, in defining  $\epsilon_{i_1 i_2 \dots i_n}$  we have not used any preferred basis, so we must assume that its values are in the set  $\{1, -1, 0\}$  in any basis. Expression (1.54) contradicts this and hence  $\epsilon_{i_1 i_2 \dots i_n}$  is not a tensor.

Let us try correcting the problem described above by a small modification. It follows from (1.43) that the determinant of the Gram matrix undergoes the following change under the transformation  $S$ :

$$\det \tilde{G} = (\det S)^2 (\det G) \quad (1.55)$$

Note that  $\det G$  may be positive or negative<sup>3</sup>, but the sign of  $\det G$  is invariant under change of basis, because  $(\det S)^2$  is positive. Let us thus define a new object:

$$\omega_{i_1 i_2 \dots i_n} = \sqrt{|\det G|} \epsilon_{i_1 i_2 \dots i_n} \quad (1.56)$$

The quantity under the square root on the right side of (1.56) is the absolute value of the determinant of  $G$ . Combining (1.53)–(1.56), we see that  $\omega_{i_1 i_2 \dots i_n}$  is transformed as follows:

$$\tilde{\omega}_{j_1 j_2 \dots j_n} = \pm \omega_{i_1 i_2 \dots i_n} S_{j_1}^{i_1} S_{j_2}^{i_2} \dots S_{j_n}^{i_n} \quad (1.57)$$

where the sign on the right side of (1.57) is the same as the sign of  $\det S$ .

The conclusion from (1.57) is that  $\omega_{i_1 i_2 \dots i_n}$  is “almost” a tensor, except for possible sign change. As long as all transformations  $S$  in the context of the application are guaranteed to have positive determinant,  $\omega_{i_1 i_2 \dots i_n}$  is a tensor. In the general case, we refer to  $\omega_{i_1 i_2 \dots i_n}$  as a *pseudotensor*. We warn, however, that this terminology is not standard.

The quantity  $\sqrt{|\det G|}$  is equal to the volume of the parallelepiped formed by the basis vectors in the case of a 3-dimensional Euclidean space. Therefore, we call  $\omega_{i_1 i_2 \dots i_n}$  the *volume pseudotensor*, or the *volume tensor* in the special case of positive  $\det S$ . Again, this name is not standard.

---

<sup>3</sup>For example,  $\det G$  is positive in Euclidean spaces but is negative in the Minkowski space.

## 1.14 Symmetry and Antisymmetry

In the following discussion,  $a_{j_1 j_2 \dots j_s}$  is a  $(0, s)$ -tensor, but can also be a symbol such as the Levi-Civita symbol or a pseudotensor such as the volume tensor.

We say that  $a_{j_1 j_2 \dots j_s}$  is *symmetric* with respect to a pair of indices  $p$  and  $q$  if

$$a_{j_1 \dots p \dots q \dots j_s} = a_{j_1 \dots q \dots p \dots j_s} \quad (1.58)$$

We say that  $a_{j_1 j_2 \dots j_s}$  is *antisymmetric* with respect to a pair of indices  $p$  and  $q$  if

$$a_{j_1 \dots p \dots q \dots j_s} = -a_{j_1 \dots q \dots p \dots j_s} \quad (1.59)$$

We note that each of (1.58) and (1.59) involves transposition of  $p$  and  $q$ ; thus, symmetry and antisymmetry are defined by the behavior of the coordinates under transpositions.

A tensor (or symbol or pseudotensor) is called *completely symmetric* if it exhibits symmetry under all possible transpositions; it is called *completely antisymmetric* if it exhibits antisymmetry under all possible transpositions. Sometimes the word “completely” is omitted but we keep it for clarity.

The Levi-Civita symbol provides an example of a completely antisymmetric symbol, and the same is true for the volume pseudotensor. It holds for any completely antisymmetric tensor that  $a_{j_1 \dots p \dots p \dots j_s} = 0$ . In other words, a completely antisymmetric tensor may have nonzero coordinates only when all indices are different.

A tensor  $a_{j_1 j_2 \dots j_s}$  is completely symmetric if and only if

$$a_{k_1 k_2 \dots k_s} = a_{j_1 j_2 \dots j_s}$$

for any permutation  $k_1 k_2 \dots k_s$ . The proof of this statement is as follows. If  $a_{j_1 j_2 \dots j_s}$  is completely symmetric and  $k_1 k_2 \dots k_s$  is a permutation of  $j_1 j_2 \dots j_s$  then, as we saw, this permutation can be decomposed into a sequence of transpositions. Since symmetry holds at each step, it holds for the entire permutation. Conversely, if  $a_{j_1 j_2 \dots j_s}$  is invariant under any permutation of its indices, then it is invariant under all transpositions as a special case and is therefore symmetric.

A tensor  $a_{j_1 j_2 \dots j_s}$  is completely antisymmetric if and only if

$$a_{k_1 k_2 \dots k_s} = \pm a_{j_1 j_2 \dots j_s}$$

where the sign is positive for any even permutation  $k_1 k_2 \dots k_s$  and is negative for any odd permutation. The proof is similar to the preceding proof for the symmetric case.

The *symmetric part* of a tensor  $a_{j_1 j_2 \dots j_s}$  with respect to a pair of indices  $p$  and  $q$  is defined by

$$a_{j_1 \dots (p| \dots |q) \dots j_s} = 0.5(a_{j_1 \dots p \dots q \dots j_s} + a_{j_1 \dots q \dots p \dots j_s}) \quad (1.60)$$

The *antisymmetric part* of a tensor  $a_{j_1 j_2 \dots j_s}$  with respect to a pair of indices  $p$  and  $q$  is defined by

$$a_{j_1 \dots [p| \dots |q] \dots j_s} = 0.5(a_{j_1 \dots p \dots q \dots j_s} - a_{j_1 \dots q \dots p \dots j_s}) \quad (1.61)$$

This notation is nonstandard and different authors use different notations. The bar lines are used to isolate the indices of the transposition from the intervening indices. If  $p$  and  $q$  are adjacent indices, the bar lines are omitted. For example,  $(pq)$  appears in the symmetric part and  $[pq]$  appears in the antisymmetric part if  $p$  and  $q$  are adjacent.

The tensor  $a_{j_1 \dots p \dots q \dots j_s}$  is the sum of its symmetric and antisymmetric parts:

$$a_{j_1 \dots p \dots q \dots j_s} = a_{j_1 \dots (p| \dots |q) \dots j_s} + a_{j_1 \dots [p| \dots |q] \dots j_s} \quad (1.62)$$

This follows from (1.60) and (1.61) in an obvious manner.

Complete symmetrization and complete antisymmetrization are also of importance. We will use verbal definitions, which are simple, whereas the mathematical formulas are somewhat awkward. The *complete symmetrization* of  $a_{j_1 j_2 \dots j_s}$ , denoted by  $a_{(j_1 j_2 \dots j_s)}$ , is defined as follows. We form all  $s!$  permutations of indices  $a_{k_1 k_2 \dots k_s}$  and compute their sum. We then divide by  $s!$ . The *complete antisymmetrization* of  $a_{j_1 j_2 \dots j_s}$ , denoted by  $a_{[j_1 j_2 \dots j_s]}$ , is defined as follows. We form all  $s!$  permutations of indices  $a_{k_1 k_2 \dots k_s}$  and compute their alternating sum. By “alternating” we mean that even permutations are added and odd permutations subtracted. We then divide by  $s!$ . It is easy to prove that  $a_{(j_1 j_2 \dots j_s)}$  is completely symmetric and  $a_{[j_1 j_2 \dots j_s]}$  is completely antisymmetric.

Partial symmetrizations and antisymmetrizations are also useful in certain applications. Their definitions are easy to understand and we discuss them only briefly. We may select any subset of  $j_1 j_2 \dots j_s$  and perform complete symmetrization or antisymmetrization using this subset and leaving the remaining indices intact. If the subset of indices is consecutive, the notation is self explanatory. For example, if  $a_{ijkm}$  is a tensor, then  $a_{i(jkm)}$  is partial symmetrization with respect to  $ijk$  and  $a_{[ijk]m}$  is partial antisymmetrization with respect to  $ijk$ . To compute  $a_{i(jkm)}$ , we add all  $a_{iuvw}$ , where  $uvw$  is a permutation of  $ijk$ , and divide by 6. To compute  $a_{[ijk]m}$ , we compute the alternating sum of all  $a_{uvw}$ , where  $uvw$  is a permutation of  $ijk$ , and divide by 6.

Partial symmetrizations and antisymmetrizations for non-consecutive subsets of indices are computed in the same way as for consecutive subsets; their notations become awkward and will not be discussed here.

The preceding material carries over without change to contravariant tensors  $a^{i_1 i_2 \dots i_r}$ . In the case of mixed tensors, one has to be careful. In most applications, either the covariant coordinates or the contravariant coordinates are manipulated, but not both at the same time. Under this limitation, the preceding definitions and properties are applicable in the same way.

We finally remark that antisymmetrizations are of great importance in relativity theory whereas symmetrizations are seldom used.

## 1.15 Summary

In this chapter we introduced tensors in the simplest setting—that of common vector spaces. Using this approach enabled us to stay within the realm of linear algebra, with very little need for calculus. Even in this simple framework, there is a need for understanding dual vectors (or covectors) and dual bases. Once this difficulty is overcome, the road is clear for tensors of any rank. By defining tensors as storage for coordinates, we were able (hopefully) to gain immediate intuition as to their laws of transformation under change of basis. We learned that a general tensor possesses a number of covariant coordinates and a number of contravariant coordinates. The former

transform under the direct change-of-basis matrix and the latter transform under its inverse.

The elementary operations on tensors include common vector operations (addition and multiplication by a scalar) as well as operations that are unique to tensors. Among the latter, the most important is the tensor product. The second is contraction. By combining tensor products and contractions we can form almost any algebraic tensor operation of interest.

Vector spaces may be equipped with an inner product. We considered inner products that are symmetric and nondegenerate, but not necessarily positive. An inner product space permits the definition of the important Gram matrix of a basis. The Gram matrix leads naturally to the metric tensor and to the operations of raising and lowering.

The remaining sections of this chapter deal with somewhat more advanced subjects—the Levi-Civita symbol, the volume pseudotensor, and symmetrization and antisymmetrization of tensors. These may be skipped on first reading, but in the long run they are important.

We define vector spaces to be *flat*, because all geometric objects are fixed, although their coordinates vary depending on the basis. Applications in physics, as well as many branches of mathematics, require more complex objects, in particular ones which involve functions and calculus. The remaining chapters deal with tensors in more general settings and, in particular, introduce tensor calculus and some special tensors of importance.

# Chapter 2

## Tensor Fields and Tensor Calculus

### 2.1 Tensor Fields

In this chapter we will consider tensors that vary from point to point in space. We therefore change our viewpoint on the underlying vector space  $\mathcal{V}$ . Rather than an abstract space, we will think of  $\mathcal{V}$  as a real physical space, which can be the usual 3-dimensional Euclidean Newtonian space or the 4-dimensional Minkowski space. The former is associated with classical physics whereas the latter constitutes the framework for relativistic physics. To  $\mathcal{V}$  we attach a fixed origin and a *reference basis*  $(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n)$ . Each point in  $\mathcal{V}$  has a radius vector  $\mathbf{r}$  with respect to the origin. The coordinates of  $\mathbf{r}$  can be expressed as linear coordinates in terms of the reference basis, as we learned in Chapter 1, say  $(x^1, x^2, \dots, x^n)$ . However, they can also be expressed in other ways, rather than through bases. For example, you are probably familiar with cylindrical and spherical coordinates in a 3-dimensional space. We may express a radius vector in cylindrical coordinates as  $\mathbf{r}(\rho, \varphi, z)$  or in spherical coordinates as  $\mathbf{r}(r, \theta, \varphi)$ . In both cases the arguments are not linear coordinates relative to any basis. These are special cases of curvilinear coordinates, which we will study in great detail in the sequel.

To each point in the space  $\mathcal{V}$  we will assign a tensor  $a(\mathbf{r})$ . These tensors have the same rank  $(r, s)$  throughout the space; they are geometrical objects over the space  $\mathcal{V}$  by their very definition. Initially we define their coordinates relative to the reference basis and later consider change of basis. Thus,  $a_{j_1 \dots j_s}^{i_1 \dots i_r}(\mathbf{r})$  denotes the coordinates of a space-dependent tensor with respect to the reference basis  $(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n)$ . Such an object is called a *tensor field* over  $\mathcal{V}$ .

The simplest way to think of  $a_{j_1 \dots j_s}^{i_1 \dots i_r}(\mathbf{r})$  is as a collection of  $n^{r+s}$  functions of  $\mathbf{r}$ . However, at every fixed  $\mathbf{r}$ , the values of the functions must obey the change-of-basis transformation laws defined in Chapter 1. Specifically, the covariant coordinates must transform according to the direct transformation and the contravariant coordinates must transform according to the inverse transformation. Additionally, regarded as functions of  $\mathbf{r}$ ,  $a_{j_1 \dots j_s}^{i_1 \dots i_r}(\mathbf{r})$  are usually assumed to satisfy certain continuity and differentiability conditions, to be discussed later.

## 2.2 The Gradient Operator in Linear Coordinates

We have already met the gradient operator  $\nabla_k$  in Sec. 1.6, applied to a scalar field and expressed as a covariant operator in linear coordinates. We wish to generalize the gradient operator to tensor fields. As long as we restrict ourselves to linear coordinates, this is not difficult. Let  $a_{j_1 \dots j_s}^{i_1 \dots i_r}(\mathbf{r})$  be a tensor field. Upon expressing the radius vector  $\mathbf{r}$  in terms of the reference basis (i.e.  $\mathbf{r} = \mathbf{e}_i x^i$ ), the tensor  $a_{j_1 \dots j_s}^{i_1 \dots i_r}(\mathbf{r})$  becomes a function of the coordinates  $x^i$ . We may now differentiate the tensor with respect to a particular coordinate  $x^p$ :

$$\nabla_p a_{j_1 \dots j_s}^{i_1 \dots i_r}(\mathbf{r}) = a_{j_1 \dots j_s; p}^{i_1 \dots i_r}(\mathbf{r}) = \frac{\partial a_{j_1 \dots j_s}^{i_1 \dots i_r}(\mathbf{r})}{\partial x^p} \quad (2.1)$$

There are several novelties in equation (2.1) that we should note. First, the right side expresses the fact that each component of the tensor, contravariant and covariant, is differentiated separately and each provides  $n$  partial derivatives, one for each  $p$ . Second, the result depends of the choice of basis, since the partial derivatives are with respect to the coordinates in the

basis. Third, the resulting object is a tensor field of type  $(r, s + 1)$ , with an additional covariant component; this fact is far from being obvious and must be proved. Fourth, (2.1) introduces a new notation, with the new covariant component appearing as the last one and separated by a semicolon. Several different notations are used in the literature but we will adhere to this one.

Before we can discuss whether  $\nabla_p a_{j_1 \dots j_s}^{i_1 \dots i_r}(\mathbf{r})$  is a tensor, we must clarify what being a tensor means in this case. Two possibilities must be considered. In the first, we may change the coordinates of the tensor  $a_{j_1 \dots j_s}^{i_1 \dots i_r}(\mathbf{r})$  to a new basis, thereby yielding  $\tilde{a}_{j_1 \dots j_s}^{i_1 \dots i_r}(\mathbf{r})$ , but continue differentiating with respect to  $x^p$ . In the second, we may change the coordinates of  $a_{j_1 \dots j_s}^{i_1 \dots i_r}(\mathbf{r})$  to  $\tilde{a}_{j_1 \dots j_s}^{i_1 \dots i_r}(\mathbf{r})$  and differentiate with respect to the new coordinates  $\tilde{x}^p$ . Whereas both definitions are plausible, we adopt the latter one and define

$$\tilde{\nabla}_p a_{j_1 \dots j_s}^{i_1 \dots i_r}(\mathbf{r}) = \tilde{a}_{j_1 \dots j_s; p}^{i_1 \dots i_r}(\mathbf{r}) = \frac{\partial \tilde{a}_{j_1 \dots j_s}^{i_1 \dots i_r}(\mathbf{r})}{\partial \tilde{x}^p} \quad (2.2)$$

Now, that the definition of  $\nabla_p$  is clear, let us prove

**Theorem 4.**  $\nabla_p$ , as defined in (2.1) and (2.2), is a tensor.

*Proof.* The change-of-basis transformation of  $a_{j_1 \dots j_s}^{i_1 \dots i_r}(\mathbf{r})$  gives

$$\tilde{a}_{j_1 \dots j_s}^{i_1 \dots i_r}(\mathbf{r}) = T_{k_1}^{i_1} \dots T_{k_r}^{i_r} a_{m_1 \dots m_s}^{k_1 \dots k_r}(\mathbf{r}) S_{j_1}^{m_1} \dots S_{j_s}^{m_s} \quad (2.3)$$

Next we differentiate (2.3) with respect to some coordinate  $\tilde{x}^p$ . In doing so, we must remember that (2.3) is, in fact, a multiple sum, so we must differentiate each term and then perform the sum. However, in doing so, only  $a_{j_1 \dots j_s}^{i_1 \dots i_r}(\mathbf{r})$  is differentiated because the matrices  $S$  and  $T$  are constant. Therefore,

$$\frac{\partial \tilde{a}_{j_1 \dots j_s}^{i_1 \dots i_r}(\mathbf{r})}{\partial \tilde{x}^p} = T_{k_1}^{i_1} \dots T_{k_r}^{i_r} \frac{\partial a_{m_1 \dots m_s}^{k_1 \dots k_r}(\mathbf{r})}{\partial \tilde{x}^p} S_{j_1}^{m_1} \dots S_{j_s}^{m_s} \quad (2.4)$$

Now, using the chain rule for partial differentiation of  $a_{j_1 \dots j_s}^{i_1 \dots i_r}(\mathbf{r})$  on the right side of (2.4), we find that

$$\frac{\partial a_{m_1 \dots m_s}^{k_1 \dots k_r}(\mathbf{r})}{\partial \tilde{x}^p} = \frac{\partial a_{m_1 \dots m_s}^{k_1 \dots k_r}(\mathbf{r})}{\partial x^q} \frac{\partial x^q}{\partial \tilde{x}^p} = \frac{\partial a_{m_1 \dots m_s}^{k_1 \dots k_r}(\mathbf{r})}{\partial x^q} S_p^q \quad (2.5)$$

with implied summation over  $q$  on the right side of (2.5). Substitution of (2.5) in (2.4) gives

$$\frac{\partial \tilde{a}_{j_1 \dots j_s}^{i_1 \dots i_r}(\mathbf{r})}{\partial \tilde{x}^p} = T_{k_1}^{i_1} \dots T_{k_r}^{i_r} \frac{a_{m_1 \dots m_s}^{k_1 \dots k_r}(\mathbf{r})}{\partial x^q} S_{j_1}^{m_1} \dots S_{j_s}^{m_s} S_p^q \quad (2.6)$$

In our new notation, equation (2.6) reads

$$\tilde{a}_{j_1 \dots j_s; p}^{i_1 \dots i_r}(\mathbf{r}) = T_{k_1}^{i_1} \dots T_{k_r}^{i_r} a_{m_1 \dots m_s; q}^{k_1 \dots k_r}(\mathbf{r}) S_{j_1}^{m_1} \dots S_{j_s}^{m_s} S_p^q \quad (2.7)$$

This is precisely the change-of-basis formula for an  $(r, s + 1)$ -tensor, as was claimed.  $\square$

## 2.3 Curvilinear Coordinates

Let us assume that we are given  $n$  functions of the coordinates of the reference basis, to be denoted by  $y^i(x^1, \dots, x^n)$ ,  $1 \leq i \leq n$ . These functions are assumed to be continuous and to possess continuous partial derivatives on a certain region in the space<sup>1</sup>. Additionally, they are assumed to be invertible, and the inverse functions  $x^i(y^1, \dots, y^n)$ ,  $1 \leq i \leq n$  are also assumed to be continuous and to possess continuous partial derivatives on the same region. Such functions  $y^i(x^1, \dots, x^n)$  are called *curvilinear coordinates*. “Curvi” implies nonlinearity, and “linear” implies that the functions can be locally linearized in the vicinity of each point  $\mathbf{r}$ , as we shall see soon.

Consider the partial derivatives of the radius vector  $\mathbf{r}$  with respect to the curvilinear coordinates. We define

$$\mathbf{E}_i = \frac{\partial \mathbf{r}}{\partial y^i} = \frac{\partial}{\partial y^i} (\mathbf{e}_j x^j) = \mathbf{e}_j \left( \frac{\partial x^j}{\partial y^i} \right) \quad (2.8)$$

This derivation is correct because  $\mathbf{e}_j$  are fixed vectors; hence their derivatives are identically zero. The  $n^2$  partial derivatives  $\partial x^j / \partial y^i$  form a square matrix

---

<sup>1</sup>Technically, this region must be an open set, but you may ignore this if you are not familiar with the notion of an open set. In the next chapter we will be more precise about this point.

called the *Jacobian matrix*. This matrix is nonsingular since  $y^i(x^1, \dots, x^n)$ ,  $1 \leq i \leq n$  are invertible. Let us denote

$$S_i^j = \frac{\partial x^j}{\partial y^i}, \quad T_k^m = \frac{\partial y^m}{\partial x^k} \quad (2.9)$$

Then

$$T_j^m S_i^j = \delta_i^m, \quad S_m^i T_k^m = \delta_k^i \quad (2.10)$$

The vectors  $(\mathbf{E}_1, \dots, \mathbf{E}_n)$  are called the *tangent vectors* of the curvilinear coordinates at the point  $\mathbf{r}$ . The vector space spanned by the tangent vector is called the *tangent space* at the point  $\mathbf{r}$ . The tangent space varies from point to point, unless all functions  $y^i(x^1, \dots, x^n)$  are linear. The space is  $n$ -dimensional in general, unless one (or more) of  $\mathbf{E}_i$  is zero<sup>2</sup>.

To distinguish such a basis from an ordinary fixed basis, we use uppercase bold letters for the basis members. If  $\mathbf{r} + d\mathbf{r}$  is a point close to  $\mathbf{r}$ , then the differential vector  $d\mathbf{r}$  can be conveniently expressed in terms of the local basis as  $d\mathbf{r} = \mathbf{E}_i dy^i$ . We note that  $d\mathbf{r}$  is also called the *line element*.

Equation (2.8) can be written as

$$\mathbf{E}_i = \mathbf{e}_j S_i^j, \quad \mathbf{e}_k = \mathbf{E}_m T_k^m \quad (2.11)$$

Expressions (2.11) have the same form as an ordinary change-of-basis transformation as can be seen by comparing them with (1.13). We should keep in mind, however, that (2.11) is local at each  $\mathbf{r}$  whereas (1.13) is global on the space.

## 2.4 The Affine Connections

When each of the tangent vectors  $\mathbf{E}_i$  is differentiated with respect to each curvilinear component  $y^j$ , we obtain  $n^2$  new vectors  $\partial \mathbf{E}_i / \partial y^j$ . Each such

---

<sup>2</sup>If some of the  $\mathbf{E}_i$  are zero at an isolated point, we may define the tangent space at that point by continuity. However, if this happens on an interval, we have a case of irreparable degeneracy of the tangent space.

vector may be expressed in terms of the local basis  $(\mathbf{E}_1, \dots, \mathbf{E}_n)$ . Doing so yields  $n^2$  linear expressions

$$\frac{\partial \mathbf{E}_i}{\partial y^j} = \mathbf{E}_k \Gamma_{ij}^k \quad (2.12)$$

where implicit summation over  $k$  is understood as usual. The  $n^3$  coefficients  $\Gamma_{ij}^k$  in (2.12) are called the *affine connections* or the *Christoffel symbols of the second kind*. Although the notation  $\Gamma_{ij}^k$  may imply that the affine connections constitute a tensor, this is in fact not the case.

At this point it becomes convenient to use a notation commonly used in multivariate calculus: whenever we differentiate a function  $f(y^1, \dots, y^n)$  with respect to  $y^p$ , we write  $f_{,p} = \partial f / \partial y^p$  for the derivative. The function  $f$  may have other indices. For example, (2.12) can be written as

$$\mathbf{E}_{i,j} = \mathbf{E}_k \Gamma_{ij}^k \quad (2.13)$$

In general, an index following a comma will denote partial derivative with respect to the corresponding  $y$  coordinate. Second derivatives will contain two indices after the comma, etc. Note that partial derivatives with respect to  $x^p$  will be denoted explicitly as  $\partial f / \partial x^p$ , to avoid ambiguity.

### 2.4.1 Formulas for the Affine Connections

We now derive several explicit formulas for the affine connections, which will be useful later. First, differentiation of (2.11) and substitution in (2.12) gives

$$\mathbf{E}_{i,j} = (\mathbf{e}_k S_i^k)_{,j} = \mathbf{e}_k S_{i,j}^k = \mathbf{E}_m \Gamma_{ij}^m = \mathbf{e}_k S_m^k \Gamma_{ij}^m \quad (2.14)$$

Since  $(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n)$  is a basis, the expansion of any vector in terms of this basis is unique. It hence follows from (2.14) that

$$S_{i,j}^k = S_m^k \Gamma_{ij}^m \quad (2.15)$$

We may now apply the inverse transformation to (2.15), to find

$$T_k^p S_{i,j}^k = T_k^p S_m^k \Gamma_{ij}^m = \delta_m^p \Gamma_{ij}^m = \Gamma_{ij}^p \quad (2.16)$$

We thus obtain the first explicit formula for the affine connections:

$$\Gamma_{ij}^p = T_k^p S_{i,j}^k = T_k^p x_{,ij}^k \quad (2.17)$$

We therefore deduce that the affine connections are symmetric in their lower indices.

A second formula for the affine connections is derived as follows. Upon differentiating (2.10), we get

$$(T_k^p S_i^k)_{,j} = T_{k,j}^p S_i^k + T_k^p S_{i,j}^k = 0 \quad \Rightarrow \quad T_k^p S_{i,j}^k = -T_{k,j}^p S_i^k \quad (2.18)$$

Substitution of (2.18) in (2.17) gives

$$\Gamma_{ij}^p = -T_{k,j}^p S_i^k \quad (2.19)$$

The third explicit formula expresses the affine connections in terms of the metric tensor  $g_{ij} = \mathbf{E}_i \cdot \mathbf{E}_j$  associated with the basis  $(\mathbf{E}_1, \dots, \mathbf{E}_n)$ :

$$\Gamma_{ij}^k = 0.5g^{km}(g_{mi,j} + g_{mj,i} - g_{ij,m}) \quad (2.20)$$

We prove (2.20) by direct calculation, as follows:

$$\begin{aligned} g_{mi,j} &= (\mathbf{E}_m \cdot \mathbf{E}_i)_{,j} = \mathbf{E}_{m,j} \cdot \mathbf{E}_i + \mathbf{E}_m \cdot \mathbf{E}_{i,j} \\ &= (\mathbf{E}_p \Gamma_{mj}^p) \cdot \mathbf{E}_i + \mathbf{E}_m \cdot (\mathbf{E}_p \Gamma_{ij}^p) = g_{pi} \Gamma_{mj}^p + g_{mp} \Gamma_{ij}^p \end{aligned} \quad (2.21)$$

In the same way we find for the two other terms

$$g_{mj,i} = g_{pj} \Gamma_{mi}^p + g_{mp} \Gamma_{ji}^p, \quad g_{ij,m} = g_{pj} \Gamma_{im}^p + g_{ip} \Gamma_{jm}^p \quad (2.22)$$

Substitution of (2.21) and (2.22) in (2.20) and using the symmetry of the affine connection and the metric tensor with respect to the lower indices, we find

$$\begin{aligned} &0.5g^{km} (g_{pi} \Gamma_{mj}^p + g_{mp} \Gamma_{ij}^p + g_{pj} \Gamma_{mi}^p + g_{mp} \Gamma_{ji}^p - g_{pj} \Gamma_{im}^p - g_{ip} \Gamma_{jm}^p) \\ &= g^{km} g_{mp} \Gamma_{ij}^p = \delta_p^k \Gamma_{ij}^p = \Gamma_{ij}^k \end{aligned} \quad (2.23)$$

Hence the proof is complete.

## 2.4.2 Example

We illustrate the derivation of the affine connections for a two-dimensional vector space with Cartesian coordinates  $(x, y)$  and curvilinear polar coordinates  $(r, \varphi)$ . related by

$$x = r \cos \varphi, \quad y = r \sin \varphi$$

This transformation is invertible at all points, except  $r = 0$ . The Jacobian matrix and its inverse are

$$S = \begin{bmatrix} \cos \varphi & -r \sin \varphi \\ \sin \varphi & r \cos \varphi \end{bmatrix}, \quad T = \begin{bmatrix} \cos \varphi & \sin \varphi \\ -r^{-1} \sin \varphi & r^{-1} \cos \varphi \end{bmatrix}$$

The partial derivatives of the Jacobian matrix are

$$S_{,r} = \begin{bmatrix} 0 & -\sin \varphi \\ 0 & \cos \varphi \end{bmatrix}, \quad S_{,\varphi} = \begin{bmatrix} -\sin \varphi & -r \cos \varphi \\ \cos \varphi & -r \sin \varphi \end{bmatrix}$$

Multiplying these matrices by  $T$  gives

$$TS_{,r} = \begin{bmatrix} 0 & 0 \\ 0 & r^{-1} \end{bmatrix}, \quad TS_{,\varphi} = \begin{bmatrix} 0 & -r \\ r^{-1} & 0 \end{bmatrix}$$

Finally,

$$\Gamma_{rr}^r = 0, \quad \Gamma_{r\varphi}^r = \Gamma_{\varphi r}^r = 0, \quad \Gamma_{\varphi\varphi}^r = -r, \quad \Gamma_{rr}^\varphi = 0, \quad \Gamma_{r\varphi}^\varphi = \Gamma_{\varphi r}^\varphi = r^{-1}, \quad \Gamma_{\varphi\varphi}^\varphi = 0$$

## 2.5 Differentiation of Tensor Fields in Curvilinear Coordinates

We now turn our attention to the problem of differentiating tensor fields in curvilinear coordinates, which differentiation generalizes the tensor derivative  $a_{j_1 \dots j_s; p}^{i_1 \dots i_r}(\mathbf{r})$  introduced in Sec. 2.2. In the following we omit the explicit dependence on  $\mathbf{r}$  for simplicity, but we keep in mind that  $\mathbf{r}$  may be expressed in terms of either the linear coordinates  $x^i$  or the curvilinear coordinates  $y^i$ , depending on the context; cf. Sec. 2.3.

For the purpose of the material on tensor fields in curvilinear coordinates, we will use the following convention henceforth. Tensor coordinates relative to the reference basis will continue to be denoted by lower case letters. For tensors in curvilinear coordinates we will use upper case letters, but without the tilde. Thus, if  $a_{m_1 \dots m_s}^{k_1 \dots k_r}$  are the coordinates of the tensor  $a$  in the reference basis  $\mathbf{e}_i$ , expressed as functions of the linear coordinates  $x^i$ , then  $A_{m_1 \dots m_s}^{k_1 \dots k_r}$  will denote the coordinates of the same tensor in the tangent basis  $\mathbf{E}_i$ , expressed as functions of the curvilinear coordinates  $y^i$ . This convention will facilitate convenient visual identification of tensors in curvilinear coordinates, which tensors are heavily used in subsequent material.

Recall (2.7), which was shown to hold for change of basis in linear coordinates. Let us require that (2.7) hold in curvilinear coordinates as well; that is,

$$A_{j_1 \dots j_s; p}^{i_1 \dots i_r} = T_{k_1}^{i_1} \dots T_{k_r}^{i_r} a_{m_1 \dots m_s; q}^{k_1 \dots k_r} S_{j_1}^{m_1} \dots S_{j_s}^{m_s} S_p^q \quad (2.24)$$

where  $S, T$  are the transformations (2.9). We wish to derive an expression for  $A_{j_1 \dots j_s; p}^{i_1 \dots i_r}$  which will make (2.24) valid. In order to simplify the derivation and make it more transparent we will start with the special case of a (1, 1)-tensor. In this case we know that

$$a_m^k = S_i^k A_j^i T_m^j \quad (2.25)$$

Therefore,

$$a_{m,p}^k = S_{i,p}^k A_j^i T_m^j + S_i^k A_{j,p}^i T_m^j + S_i^k A_j^i T_{m,p}^j \quad (2.26)$$

Substitute (2.15) and (2.18):

$$\frac{\partial a_m^k}{\partial x^q} \frac{\partial x^q}{\partial y^p} = \frac{\partial a_m^k}{\partial x^q} S_p^q = S_h^k \Gamma_{ip}^h A_j^i T_m^j + S_i^k A_{j,p}^i T_m^j - S_i^k A_j^i \Gamma_{hp}^j T_m^h \quad (2.27)$$

This is the same as

$$T_k^i \frac{\partial a_m^k}{\partial x^q} S_j^m S_p^q = \Gamma_{hp}^i A_j^h + A_{j,p}^i - A_h^i \Gamma_{jp}^h \quad (2.28)$$

or

$$T_k^i a_{m;q}^k S_j^m S_p^q = A_{j,p}^i + \Gamma_{hp}^i A_j^h - A_h^i \Gamma_{jp}^h \quad (2.29)$$

Finally, comparing the left side of (2.29) with the right side of (2.24), we find that

$$A_{j;p}^i = A_{j,p}^i + \Gamma_{hp}^i A_j^h - A_h^i \Gamma_{jp}^h \quad (2.30)$$

Expression (2.30) is called the *covariant derivative* of the (1,1)-tensor  $A_j^i$  in curvilinear coordinates. As we saw,  $A_{j;p}^i$  is a (1,2)-tensor, related to the (1,2)-tensor  $a_{m;q}^k$  via the change-of-basis formula (2.24).

We now have two kinds of derivatives in curvilinear coordinates: the conventional derivative, denoted by  $X_{,p}$  (where  $X$  stands for any object depending on  $y^i$ ), and the covariant derivative  $X_{;p}$ , defined only for tensors  $X$ . The latter is equal to the former plus additional terms; these are called the *affinities*.

The covariant derivative of a general tensor in curvilinear coordinates is derived in a similar manner, yielding the expression

$$A_{j_1 \dots j_s ; p}^{i_1 \dots i_r} = A_{j_1 \dots j_s , p}^{i_1 \dots i_r} + \sum_{u=1}^r \Gamma_{h_u p}^{i_u} A_{j_1 \dots j_s}^{i_1 \dots i_{u-1} h_u i_{u+1} \dots i_r} - \sum_{u=1}^s A_{j_1 \dots j_{u-1} h_u j_{u+1} \dots j_s}^{i_1 \dots i_r} \Gamma_{j_u p}^{h_u} \quad (2.31)$$

The special cases of the covariant derivatives of vectors and covectors are worthwhile writing explicitly:

$$A_{;p}^i = A_{,p}^i + \Gamma_{hp}^i A^h, \quad A_{j;p} = A_{j,p} - A_h \Gamma_{jp}^h \quad (2.32)$$

The metric tensor  $g_{ij}$  has the following useful property:

**Theorem 5.** *The covariant derivative of  $g_{ij}$  is identically zero; that is,*

$$g_{ij;p} = 0$$

*Proof.* It follows as a special case of (2.31) that

$$g_{ij;p} = g_{ij,p} - g_{kj} \Gamma_{ip}^k - g_{ik} \Gamma_{jp}^k \quad (2.33)$$

Now substitute the affine connections expression (2.20) in (2.33), then carry out all cancelations, ending up in 0, as claimed.  $\square$

Let  $T^{ij}$  be a symmetric (2, 0)-tensor. The *covariant divergence* of  $T^{ij}$  is defined as

$$\nabla_{;i} T^{ij} = T^{ij}_{;i} = T^{ij}_{,i} + \Gamma_{ki}^i T^{kj} + \Gamma_{ik}^j T^{ik} \quad (2.34)$$

The covariant divergence is a contravariant vector whose free index is  $j$  in this definition. The covariant divergence is important in physics applications because it is typically associated with conservation laws.

## 2.6 The Second Covariant Derivative

We continue our discussion of the preceding section by differentiating tensor fields in curvilinear coordinates a second time. However, we do not discuss the general case but limit ourselves to the second covariant derivative of a (contravariant) vector. The general case can be derived using the procedure derived below, but the special case we consider is sufficient for drawing some interesting conclusions.

The second covariant derivative of  $A^i$ , denoted by  $A^i_{;pq}$ , is the first covariant derivative of  $A^i_{;p}$ . This is, in turn, a (1, 1)-tensor; therefore its first covariant derivative can be found from (2.30):

$$A^i_{;pq} = (A^i_{;p})_{;q} + \Gamma^i_{hq} A^h_{;p} - A^i_{;h} \Gamma^h_{pq} \quad (2.35)$$

Let us expand each of the first two terms on the right side of (2.35) (but not the third, for reasons to be seen later), using  $A^i_{;p}$  from (2.32):

$$(A^i_{;p})_{;q} = (A^i_{;p} + \Gamma^i_{hp} A^h)_{;q} = A^i_{;pq} + \Gamma^i_{hp,q} A^h + \Gamma^i_{hp} A^h_{;q} \quad (2.36)$$

$$\Gamma^i_{hq} A^h_{;p} = \Gamma^i_{hq} A^h_{;p} + \Gamma^i_{hq} \Gamma^h_{kp} A^k \quad (2.37)$$

Substitution of (2.36) and (2.37) in (2.35) gives

$$A^i_{;pq} = A^i_{;pq} + \Gamma^i_{hp,q} A^h + \Gamma^i_{hp} A^h_{;q} + \Gamma^i_{hq} A^h_{;p} + \Gamma^i_{hq} \Gamma^h_{kp} A^k - A^i_{;h} \Gamma^h_{pq} \quad (2.38)$$

Expression (2.38) provides the desired result for the second covariant derivative of a vector in curvilinear coordinates. In the next section we will continue to explore this result.

## 2.7 The Riemann Curvature Tensor

As we know from multivariate calculus, second derivatives possess the symmetry  $X_{;pq} = X_{;qp}$  if the second derivatives exist and are continuous. Let us explore if this property holds for second covariant derivatives. To find out,

let us compute  $A^i_{;pq} - A^i_{;qp}$  and check whether the difference is identically zero. Using (2.38) and canceling terms that are equal due to symmetries gives

$$\begin{aligned} A^i_{;pq} - A^i_{;qp} &= \Gamma^i_{hp,q} A^h + \Gamma^i_{kq} \Gamma^k_{hp} A^h - \Gamma^i_{hq,p} A^h - \Gamma^i_{kp} \Gamma^k_{hq} A^h \\ &= (\Gamma^i_{hp,q} + \Gamma^i_{kq} \Gamma^k_{hp} - \Gamma^i_{hq,p} - \Gamma^i_{kp} \Gamma^k_{hq}) A^h \end{aligned} \quad (2.39)$$

Let us introduce a new symbol:

$$R^i_{hqp} = \Gamma^i_{hp,q} + \Gamma^i_{kq} \Gamma^k_{hp} - \Gamma^i_{hq,p} - \Gamma^i_{kp} \Gamma^k_{hq} \quad (2.40)$$

Equation (2.39) then becomes

$$A^i_{;pq} - A^i_{;qp} = R^i_{hqp} A^h \quad (2.41)$$

An immediate consequence of this result is that the second covariant derivative is not symmetric in general, except when  $R^i_{hqp}$  is identically zero. Later we will see when this happens.

**Theorem 6.**  $R^i_{hqp}$  is a tensor.

*Proof.* The left side of (2.41) is a difference of tensors and is therefore a tensor. Denoting this tensor temporarily as  $\delta A^i_{pq}$ , we have that  $\delta A^i_{pq} = R^i_{hqp} A^h$ , where both  $\delta A^i_{pq}$  and  $A^h$  are tensors. In this equality,  $A^h$  is arbitrary,  $R^i_{hqp}$  is given, and  $\delta A^i_{pq}$  is the resulting tensor. Now make a change-of-basis transformation. If  $R^i_{hqp}$  is a tensor, then the corresponding equality in the new basis will be  $\delta \tilde{A}^i_{pq} = \tilde{R}^i_{hqp} \tilde{A}^h$ , where all tildes denote the new basis coordinates of the tensors. If  $R^i_{hqp}$  is not a tensor, then the corresponding equality in the new basis will be  $\delta \tilde{A}^i_{pq} = \hat{R}^i_{hqp} \tilde{A}^h$ , where  $\hat{R}^i_{hqp}$  is an unknown array of numbers. Therefore, in the latter case,

$$(\tilde{R}^i_{hqp} - \hat{R}^i_{hqp}) \tilde{A}^h = 0$$

Since this equality holds for all  $\tilde{A}^h$  it holds, in particular, for  $\tilde{A}^k = 1$  and  $\tilde{A}^h = 0$ ,  $h \neq k$ . It therefore follows that

$$\tilde{R}^i_{kqp} - \hat{R}^i_{kqp} = 0$$

Finally, this holds for all  $i, k, p, q$ , so  $\hat{R}^i_{kqp} = \tilde{R}^i_{kqp}$  and it follows that  $R^i_{hqp}$  is a tensor.  $\square$

The tensor  $R_{hqp}^i$  is called the *Riemann curvature tensor*. The Riemann curvature tensor plays a central role in differential geometry and the theory of manifolds, as well as in applications of these theories. Here we do not provide enough material on these subjects to really appreciate the importance of the Riemann tensor. We will settle for discussing some of its mathematical properties and then present some tensor related to the Riemann tensor.

The first thing to note is that  $R_{hqp}^i$  depend only on the affine connections  $\Gamma_{ij}^k$ . These, in turn, depend only on the metric  $g_{ij}$ , as seen from (2.20). It follows that *The Riemann tensor depends only on the metric*. We should remember, however, that in curvilinear coordinates the metric varies from point to point and the Riemann tensor varies with it. Only when the curvilinear coordinates are linear is the metric constant. In this case the affine connections are identically zero and so is the Riemann tensor.

The Riemann tensor is antisymmetric in the indices  $q$  and  $p$ ; that is,

$$R_{hqp}^i = -R_{h pq}^i \quad (2.42)$$

This follows directly from the definition (2.40).

The Riemann tensor satisfies the so-called *First Bianchi identity*:

$$R_{hqp}^i + R_{qph}^i + R_{phq}^i = 0 \quad (2.43)$$

This identity also follows by substitution of (2.40) and performing the cancellations. Note that the lower indices of the three terms in (2.43) are cyclic permutations of one another.

Other symmetries of the Riemann tensor are more subtle; to make them explicit, we introduce the purely covariant Riemann tensor, obtained by lowering the contravariant index of  $R_{hqp}^i$ :

$$R_{ihqp} = g_{ij} R_{hqp}^j \quad (2.44)$$

The covariant Riemann tensor possesses the symmetries

$$R_{ihqp} = -R_{ihpq} = -R_{hiqp} = R_{qpih} \quad (2.45)$$

The first equality in (2.45) follows from the corresponding anti-symmetry of (2.42). The other two are established in Appendix B.

## 2.8 Some Special Tensors

The space-geometry tensors of general relativity are, in a figure of speech, children of the Riemann tensor. The Ricci<sup>3</sup> tensor and the Einstein tensor, in particular, are derived from the Riemann tensor as we will now present.

The *Ricci tensor* is the result of contracting the contravariant and last covariant indices of the Riemann tensor:

$$R_{ij} = R_{ijk}^k \quad (2.46)$$

The Ricci tensor is symmetric, as results from the following chain of equalities:

$$R_{ij} = R_{ijk}^k = g^{ku} R_{uijk} = g^{ku} R_{jkui} = g^{ku} R_{kjiu} = R_{jiu}^u = R_{ji} \quad (2.47)$$

In (2.47) we used the symmetries and anti-symmetries of the covariant Riemann tensor, given in (2.44). Be sure to understand which symmetry is used at each step of (2.47).

The *curvature scalar* is the full contraction of the Ricci tensor:

$$R = g^{ij} R_{ij} \quad (2.48)$$

The *Einstein tensor* is defined by

$$E_{ij} = R_{ij} - 0.5g_{ij}R \quad (2.49)$$

Einstein's tensor is symmetric, since both  $R_{ij}$  and  $g_{ij}$  are symmetric. Einstein's tensor can also be expressed in mixed and contravariant forms:

$$E_j^i = R_j^i - 0.5\delta_j^i R, \quad E^{ij} = R^{ij} - 0.5g^{ij}R \quad (2.50)$$

---

<sup>3</sup>Gregorio Ricci-Curbastro and Tullio Levi-Civita developed tensor theory in the late 19th century, following in the footsteps of Gauss and Riemann. Levi-Civita helped Einstein understand tensors, thereby facilitating Einstein's development of general relativity. Two Jews—one a brilliant mathematician and the other greatest physicist ever—in an increasingly hostile and antisemitic Europe.

## 2.9 Summary

In this section we extended tensor theory from constant tensors in constant bases to tensor fields; that is, tensors that vary from point to point in space. The space itself is still a vector space with an inner product (positive or not); therefore tensors can still be expressed relative to a fixed (reference) basis. However, they may also be expressed relative to bases that vary from point to point. We introduced the concept of curvilinear coordinates, which may be global on the entire space or local to a subset of the space.

Calculus demands the ability to differentiate functions. We distinguished between two kinds of derivatives: the usual partial derivative, as defined in calculus; and the covariant derivative. The latter has an important property not shared by the former—it is a bona-fide tensor and therefore transforms properly under change of basis.

The definition of covariant derivative relies on the extension of the concept of metric to local bases on curvilinear coordinates, and on the affine connections. A second covariant derivative can also be defined. It turns out that the second covariant derivative, unlike a usual second partial derivative, depends on the order of the coordinates with respect to which we differentiate; in other words, it is not commutative. An important consequence of this lack of commutativity is the ability to define the curvature tensor. This tensor, also called the Riemann tensor, leads to the Ricci tensor and to the Einstein tensor; the latter plays a fundamental role in general relativity.

# Chapter 3

## Tensors on Manifolds

### 3.1 Introduction

The space  $\mathbb{R}^n$  is one of mathematics' greatest success stories: it is at the same time the star of linear algebra, geometry, and analysis. What makes  $\mathbb{R}^n$  such a fertile ground for mathematical theories? I am not sure whether there is an official answer, but I suspect that if there is one, it is: Pythagoras' theorem; or, more accurately, the  $n$ -dimensional generalization thereof:

$$d^2 = \sum_{i=1}^n (x^i)^2 \tag{3.1}$$

where  $d$  is the hypotenuse and  $x^i$  the sides of an  $n$ -dimensional triangle. This being said, there is more to mathematics than Euclidean spaces. To mention but one of many examples, consider the the surface of a sphere. It follows immediately from (3.1) that the surface of a sphere is just the set of all points of constant distance  $d$  from the origin. But, although the surface of a sphere is intimately related to the Euclidean space in which it is embedded, it does not at all look like a Euclidean space of any dimension.

The surface of a sphere is an example of a *smooth topological manifold*. We can loosely define such an object as a set patched up from subsets, each of which is “like” a Euclidean space. “Like” in this description means that each

patch can be mapped in a one-to-one way to a patch of a Euclidean space such that the map and its inverse are continuous. Let us exemplify this informal description by considering the surface of a sphere in  $\mathbb{R}^3$  centered at the origin and having radius  $d = 1$ . Denote the surface of the sphere by  $S$  and the three cartesian coordinates by  $(x, y, z)$ . Then the two sets

$$O_1 = \{(x, y, z) \in S, z \neq 1\} \text{ and } O_2 = \{(x, y, z) \in S, z \neq -1\}$$

together cover  $S$ . Note that these sets considerably overlap; the first contains all points of  $S$  except the north pole and the second contains all points but the south pole. Let  $(u, v)$  be cartesian coordinates on  $\mathbb{R}^2$  and define

$$u = \frac{2x}{1-z}, \quad v = \frac{2y}{1-z}$$

These functions, called the *stereographic projection*, map  $O_1$  onto  $\mathbb{R}^2$  continuously. The inverse functions

$$z = \frac{0.25(u^2 + v^2) - 1}{0.25(u^2 + v^2) + 1}, \quad x = 0.5u(1-z), \quad y = 0.5v(1-z)$$

are continuous on  $\mathbb{R}^2$  onto  $O_1$ . In a similar way  $O_2$  can be mapped onto  $\mathbb{R}^2$ ; simply replace  $1 - z$  by  $1 + z$  in the above. The surface of a sphere in three dimensions is therefore a two-dimensional manifold.

This chapter extends tensor calculus from curvilinear coordinates in Euclidean spaces to tensor fields on manifolds. This chapter uses a higher level of abstraction and requires greater mathematical maturity on part of the reader beyond what was needed in Chapter 2. Learning the material in this chapter may thus be considerably more difficult than the material in the preceding chapters. On the other hand, tensors on manifolds are essential for understanding general relativity, which is why this chapter is included here.

The mathematical background presented in Section 3.2 contains several parts. Sets and functions should be part of the education of all engineering students. Hopefully, this material will not be new to you. The Euclidean space  $\mathbb{R}^n$  is usually studied in the course of multivariate calculus, but its topological properties, as presented here, are not always emphasized. Introducing the topological properties of  $\mathbb{R}^n$  provides us with an opportunity to introduce general topological spaces, at the most elementary level. The minimal necessary background on continuity and homeomorphism is then introduced.

Section 3.3 introduces manifolds. This is the most challenging section; reading and re-reading will probably be needed by most students. By comparison, the remaining sections should be easier. Section 3.4, on tensors and tensor fields, is a relatively straightforward extension of the material in Chapter 2. Sections 3.5, 3.6, and 3.7 on curves, curvature and geodesics are optional, and rely mostly on standard multivariate calculus.

## 3.2 Mathematical Background

### 3.2.1 Sets and Functions

Recall the definition of **smooth manifold** as a ...

We assume that you know the concept of an abstract set, at least at the intuitive level. We remind that the notation  $x \in A$  means that  $x$  *belongs* to the set  $A$  or, equivalently, that  $x$  is a *member* of the set  $A$ . We also say that  $A$  *contains*  $x$ . The notation  $x \notin A$  means that  $x$  does not belong to  $A$ .

If  $A$  is a set and  $B$  is another set such that, for all  $x \in B$  it holds true that  $x \in A$ , then  $B$  is a *subset* of  $A$ , denoted as  $B \subseteq A$ . We also say that  $B$  is *included* in  $A$  and that  $A$  *includes*  $B$ . For two sets  $A$  and  $B$ , the notation  $A \setminus B$  stands for the set of all  $x \in A$  and  $x \notin B$ . We sometime call this set (informally) “ $A$  minus  $B$ ”. If, as a special case,  $B \subseteq A$ , then  $A \setminus B$  is called the *complement* of  $B$  in  $A$ .

The *union*  $A \cup B$  of sets  $A$  and  $B$  is the set of all  $x$  such that  $x \in A$  or  $x \in B$  or both. The *intersection*  $A \cap B$  of  $A$  and  $B$  is the set of all  $x$  such that  $x \in A$  and  $x \in B$ . If there is no such  $x$ , the two sets are said to be *disjoint* and the intersection is the *empty set*  $\emptyset$ .

Unions and intersections are not limited to two sets, or even to a finite collection of sets. If  $I$  is an arbitrary set—finite, countably infinite, or even uncountably infinite—we may assign to each member  $i \in I$  a set  $A_i$ , thus obtaining a collection  $\{A_i, i \in I\}$ . Then the union  $\bigcup_{i \in I} A_i$  is the set of all  $x$  such that  $x \in A_i$  for some  $i \in I$  and the intersection  $\bigcap_{i \in I} A_i$  is the set of all

$x$  such that  $x \in A_i$  for all  $i \in I$ .

A function  $f$  from a domain  $X$  to a codomain  $Y$ , denoted  $f: X \rightarrow Y$ , is a set of ordered pairs  $(x, y)$  where  $x$  is present in exactly one pair for each  $x \in X$ . On the other hand, any  $y \in Y$  may appear once, or multiple times, or none at all. We commonly write  $y = f(x)$  to denote that  $(x, y)$  belongs to  $f$ ; with this notation, the pair  $(x, f(x))$  is trivially in  $f$ .

A function is called *injective* (or *one-to-one*) if each  $y \in Y$  appears at one pair  $(x, y)$  at most; equivalently,  $f(x_1) \neq f(x_2)$ , unless  $x_1 = x_2$ . A function is called *surjective* (or *onto*) if each  $y \in Y$  appears at least in one pair  $(x, y)$ ; equivalently, for every  $y \in Y$  there exists  $x \in X$  such that  $y = f(x)$ . A function is called *bijective* (or *one-one correspondence*) if each  $y \in Y$  appears in exactly one pair. Clearly, a function is bijective if and only if it is both injective and surjective.

The *range* of a function is the subset of all  $y \in Y$  for which  $y = f(x)$  for some  $x \in X$ . The range of a surjective function is therefore identical with its codomain. Every function becomes surjective if we redefine its codomain as its range. On the other hand, a function that is not injective cannot be made injective in a similar manner.

If  $f: X \rightarrow Y$  is a bijective function, its *inverse*  $f^{-1}$  is defined as follows. First, the domain of  $f^{-1}$  is the codomain of  $f$  and vice versa; that is,  $f^{-1}: Y \rightarrow X$ . Second,  $f^{-1}$  consists of all pairs  $(y, x)$  corresponding to the pairs  $(x, y)$  of  $f$ ; equivalently,  $f^{-1}(f(x)) = x$  for all  $x \in X$ .

Let  $A$  be a subset of the domain  $X$  of a function. Then the *image* of  $A$ , denoted  $f[A]$ , is the subset of  $Y$  consisting of all points  $y \in Y$  such that  $y = f(x)$  for some  $x \in A$ . It is clear from the definition that  $f[X]$  is identical with the range of the function.

Let  $B$  be a subset of the codomain of a function. Then the *inverse image* of  $B$ , denoted  $f^{-1}[B]$ , is the subset of  $X$  consisting of all points  $x \in X$  such that  $y = f(x)$  for some  $y \in B$ . It is clear from the definition that  $f^{-1}[Y] = X$ .

The image and inverse image satisfy the following union and intersection

properties:

$$f[A \cup B] = f[A] \cup f[B] \quad (3.2a)$$

$$f[A \cap B] \subseteq f[A] \cap f[B] \quad (3.2b)$$

$$f^{-1}[A \cup B] = f^{-1}[A] \cup f^{-1}[B] \quad (3.2c)$$

$$f^{-1}[A \cap B] = f^{-1}[A] \cap f^{-1}[B] \quad (3.2d)$$

However, if  $f$  is injective, then (3.2b) changes to

$$f[A \cap B] = f[A] \cap f[B] \quad (3.3)$$

Let  $X$ ,  $Y$ , and  $Z$  be three sets and let  $f: X \rightarrow Y$  and  $g: Y \rightarrow Z$ . Then the *composition*  $g \circ f: X \rightarrow Z$  is the set of all pairs  $(x, z)$  such that, for some  $y \in Y$ ,  $y = f(x)$  and  $z = g(y)$ . We may omit  $y$  and write  $z = g(f(x))$  to denote  $g \circ f$  when applied to  $x$ . It is obvious that, in the definition of composition, the domain of  $g$  can be restricted to the range  $f[X]$  of  $f$  without affecting the composition.

### 3.2.2 The Topological Structure of $\mathbb{R}^n$

The space  $\mathbb{R}^n$  consists of all  $n$ -tuples of real numbers  $(x^1, \dots, x^n)$ . Each such  $n$ -tuple is denoted as a vector, for example  $\mathbf{x}$ . Points and vectors are synonymous in  $\mathbb{R}^n$ . The numbers  $x^i$  are called the *coordinates* of  $\mathbf{x}$ .

A *distance function* is defined for all pairs of vectors:

$$|\mathbf{x} - \mathbf{y}| = \left[ \sum_{i=1}^n (x^i - y^i)^2 \right]^{1/2} \quad (3.4)$$

The distance functions has three fundamental properties:

**D1** The distance is zero if  $\mathbf{x} = \mathbf{y}$  and positive if  $\mathbf{x} \neq \mathbf{y}$ .

**D2** The distance is symmetric in  $\mathbf{x}$  and  $\mathbf{y}$ .

**D3** The distance satisfies the *triangle inequality* for all  $\mathbf{x}$ ,  $\mathbf{y}$ ,  $\mathbf{z}$ :

$$|\mathbf{x} - \mathbf{y}| \leq |\mathbf{x} - \mathbf{z}| + |\mathbf{z} - \mathbf{x}| \quad (3.5)$$

Let  $\mathbf{x}_0$  be a point in  $\mathbb{R}^n$  and  $d$  a positive number. The set

$$B(\mathbf{x}_0, d) = \{\mathbf{y} : |\mathbf{x}_0 - \mathbf{y}| < d\} \quad (3.6)$$

is called an *open ball* centered at  $\mathbf{x}_0$  and having radius  $d$ .

A subset  $O$  of  $\mathbb{R}^n$  is *open* if it is a union (finite, countable, or uncountable) of open balls. Open sets have three fundamental properties, as follows.

**T1** The empty set  $\emptyset$  and  $\mathbb{R}^n$  are open sets.

**T2** The union of any number of open sets (finite, countable, or uncountable) is an open set.

**T3** The intersection of two open sets is an open set.

Property **T1** is true because of the following. By convention, an empty union is empty. Thus  $\bigcup_{i \in \emptyset} O_i = \emptyset$ ; therefore  $\emptyset$  is open. The entire space  $\mathbb{R}^n$  is open because it is the union of all open balls.

Property **T2** is almost self evident. Given any collection of open sets, then each is a union of open balls, thus their union is a union of unions of open balls, which is itself a union of open balls, which is open.

Property **T3** is more difficult to prove. We break the proof into three steps, as follows.

Step One. Consider the intersection of two open balls, say

$$A = B(\mathbf{x}_1, d_1) \cap B(\mathbf{x}_2, d_2)$$

Let  $\mathbf{y}$  be a point in  $A$ . We aim to find an open ball  $B(\mathbf{y}, \varepsilon)$  such that  $B(\mathbf{y}, \varepsilon) \subseteq B(\mathbf{x}_1, d_1)$  and  $B(\mathbf{y}, \varepsilon) \subseteq B(\mathbf{x}_2, d_2)$ . It will then follow that  $B(\mathbf{y}, \varepsilon) \subseteq A$ . Define

$$\varepsilon_1 = d_1 - |\mathbf{y} - \mathbf{x}_1|, \quad \varepsilon_2 = d_2 - |\mathbf{y} - \mathbf{x}_2|, \quad \varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$$

Since  $\mathbf{y} \in B(\mathbf{x}_1, d_1)$ , it follows that  $\varepsilon_1 > 0$ . Similarly,  $\varepsilon_2 > 0$ . Therefore  $\varepsilon > 0$  and hence  $B(\mathbf{y}, \varepsilon)$  is indeed an open ball. Let us now show that  $B(\mathbf{y}, \varepsilon) \subseteq B(\mathbf{x}_1, d_1)$ . If  $\mathbf{z} \in B(\mathbf{y}, \varepsilon)$ ; then

$$|\mathbf{z} - \mathbf{y}| < \varepsilon = d_1 - |\mathbf{y} - \mathbf{x}_1|$$

so by the triangle inequality

$$|\mathbf{z} - \mathbf{x}_1| \leq |\mathbf{z} - \mathbf{y}| + |\mathbf{y} - \mathbf{x}_1| < d_1 - |\mathbf{y} - \mathbf{x}_1| + |\mathbf{y} - \mathbf{x}_1| = d_1$$

This proves that  $\mathbf{z} \in B(\mathbf{x}_1, d_1)$ . Since  $\mathbf{z}$  is an arbitrary point in  $B(\mathbf{y}, \varepsilon)$ , it follows that  $B(\mathbf{y}, \varepsilon) \subseteq B(\mathbf{x}_1, d_1)$ . In the same we show that  $B(\mathbf{y}, \varepsilon) \subseteq B(\mathbf{x}_2, d_2)$  and conclude that  $B(\mathbf{y}, \varepsilon) \subseteq A$ .

Step Two. So far we proved that, given a point  $\mathbf{y}$  in the intersection  $A$  of two open balls, there exists an open ball  $B(\mathbf{y}, \varepsilon)$  that contains  $\mathbf{y}$  and is included in  $A$ . The union  $\bigcup_{\mathbf{y} \in A} B(\mathbf{y}, \varepsilon)$  therefore both includes and is included in  $A$ , so it is equal to  $A$ . It follows that  $A$  is an open set<sup>1</sup>.

Step Three. To complete the proof, consider two open sets  $O_1, O_2$ . Each is a union of open balls and hence  $U_1 \cap U_2$  is the union of all intersections of the open balls comprising  $O_1$  with those comprising  $O_2$  (this follows from De Morgan's distributive laws for unions and intersections of sets). As we have just shown, each such intersection is an open set; therefore  $U_1 \cap U_2$  is an open set and the proof is complete.

Generalization. The intersection of a finite number of open sets is open. This follows from **T3** using induction on the number of sets.

### 3.2.3 General Topological Spaces

Borrowing from definitions stated in the preceding subsection, we now define a *topological space* as a set  $S$ , equipped with a collection  $\mathcal{T}$  of subsets of  $S$ , such that axioms **T1**, **T2** and **T3** are satisfied. The member sets of  $\mathcal{T}$  are the *open sets* of the topology.

The space  $\mathbb{R}^n$ , together with the collection of all open sets, is an example of a topological space. It is called the *usual topology* on  $\mathbb{R}^n$ . This, in fact, was the model on which general topology was founded.

---

<sup>1</sup>If you have not seen this kind of mathematical proof, you must be surprised and hopefully amazed. This is the power of infinity in action. It is Georg Cantor's theory of infinity that revolutionized mathematics; the revolution in physics followed only shortly after.

Two simple examples of topological spaces are:

- Any set  $S$  with  $\mathcal{T} = \{\emptyset, S\}$ . This topology has only two open sets and is called the *indiscrete topology* on  $S$ .
- Any set  $S$  with  $\mathcal{T}$  containing all subsets of  $S$ . This is called the *discrete topology* on  $S$ .

Many more examples can be given, but since our aim is not to teach topology for its own sake, we will not provide more examples.

We call a set  $C$  *closed* if its complement  $S \setminus C$  is open. It follows easily from this definition that  $\emptyset$  and  $S$  are closed, that an arbitrary intersection of closed sets is closed, and that the union of two closed sets is closed. In the usual topology on  $\mathbb{R}^n$ , the *closed ball*

$$C(\mathbf{x}_0, d) = \{\mathbf{y} : |\mathbf{x}_0 - \mathbf{y}| \leq d\} \quad (3.7)$$

is closed. For a proof, note that the union of all open balls in  $\mathbb{R}^n$  which are disjoint from  $C(\mathbf{x}_0, d)$  is open and is equal to  $\mathbb{R}^n \setminus C(\mathbf{x}_0, d)$

If  $\mathbf{x}$  is a point and  $O$  is an open set containing  $\mathbf{x}$ , then  $O$  is said to be an *open neighborhood* of  $\mathbf{x}$ . If  $\mathbf{x}$  is a point and  $N$  is a set containing  $x$  (not necessarily open), then  $N$  is a *neighborhood* of  $\mathbf{x}$  if there exists an open neighborhood  $O$  of  $x$  such that  $O \subseteq N$ . Thus, a neighborhood of  $x$  is any set including an open neighborhood of  $x$ .

### 3.2.4 More on $\mathbb{R}^n$

The usual topology on  $\mathbb{R}^n$  has many interesting properties that arise frequently in advanced calculus on  $\mathbb{R}^n$ . We will mention some of them here; most are not essential to understanding of the material in this chapter but are good to know for completeness.

**Hausdorff** If  $\mathbf{x}_1$  and  $\mathbf{x}_2$  are two different points, then there exist open neighborhoods  $O_1$  and  $O_2$  of  $\mathbf{x}_1$  and  $\mathbf{x}_2$  such that  $O_1 \cap O_2 = \emptyset$ . A topological space having this property is called a *Hausdorff space*<sup>2</sup>. To

---

<sup>2</sup>Named after Felix Hausdorff, one of the founders of the abstract theory of topological spaces.

prove the Hausdorff property for  $\mathbb{R}^n$ , let  $d = |\mathbf{x}_1 - \mathbf{x}_2|$  and then let  $O_1 = B(\mathbf{x}_1, d/3)$ ,  $O_2 = B(\mathbf{x}_2, d/3)$ . It is straightforward to prove that  $O_1 \cap O_2 = \emptyset$ .

**Separability** A subset  $A$  of a topological space is *dense* if every open set contains a point of  $A$ . A topological space is *separable* if it contains a countable dense set. It is well known that the rational numbers are dense in  $\mathbb{R}$  (we will not prove this here). It then follows that the points  $\mathbf{x}$  whose coordinates are rational are dense in  $\mathbb{R}^n$ . The set of rational numbers is known to be countable. It then follows that the set of  $n$ -tuples of rational numbers is countable. Therefore,  $\mathbb{R}^n$  equipped with the usual topology is separable.

**Second Countability** A *base* for a topology is a collection  $\mathcal{B}$  of open sets such that every open set is a union of members of  $\mathcal{B}$ . The definition of the usual topology on  $\mathbb{R}^n$  automatically makes the collection of all open balls a base for the topology. A topological space is *second countable* if it has a countable base [first countable is another property, which we will not discuss here]. The collection  $\mathcal{B}$  of all open balls whose centers have rational coordinates and whose radii are rational is a countable base for usual topology on  $\mathbb{R}^n$ . Here is a sketch of the proof. Given an open set  $O$ , take all points  $\mathbf{x}$  of  $O$  having rational coordinates. For each  $\mathbf{x}$ , take all open balls  $B(\mathbf{x}, r)$  with rational  $r$  such that  $B(\mathbf{x}, r) \subseteq O$ . Then  $\bigcup_{\mathbf{x}, r} B(\mathbf{x}, r) = O$ .

All these definitions are applicable to any topological space, not just to  $\mathbb{R}^n$ .

### 3.2.5 Continuity and Homeomorphisms

A function on a topological space  $X$  to a topological space  $Y$  is *continuous at a point*  $x$  if, for any open neighborhood  $V$  of  $f(x)$ , there is an open neighborhood  $U$  of  $x$  such that  $f[U] \subseteq V$ . Compare this definition with the usual conventional “epsilon-delta” definition for functions on and to  $\mathbb{R}$  and you will see that the “epsilon-delta” definition follows from the general definition, when applied to the usual topology on  $\mathbb{R}$ .

A function is continuous on an open set  $O \subseteq X$  if it is continuous at all points of  $O$ . In particular, a function is continuous on  $X$  if it is continuous at all

points of  $X$ . The following theorem is of interest: A function is continuous on  $X$  if and only if, for any open set  $V$  in  $Y$ , the inverse image  $U = f^{-1}[V]$  is an open set in  $X$ . The proof is a simple exercise which you may want to carry out, to test your understanding of images, inverse images, and open sets.

Let  $X$  and  $Y$  be two topological spaces and assume that there exists a bijective function  $f$  on  $X$  onto  $Y$  such that both  $f$  and  $f^{-1}$  are continuous. Then the two spaces are *homeomorphic* and  $f$  is called *homeomorphism*. You may think of two spaces as being homeomorphic if they are made of infinitely flexible rubber and one can be obtained from the other by arbitrary stretching, squeezing, or bending, but no tearing or punching holes. The two spaces “look the same” in the sense that an arbitrary set  $O$  in  $X$  is open if and only if  $f[O]$  is open in  $Y$ .

Let  $X$  and  $Y$  be two topological spaces. Let  $x$  be a point in  $X$  and assume that there exist an open neighborhood  $O_x$  of  $x$  and an injective function  $f_x$  on  $O_x$  onto  $f_x[O_x] \subseteq Y$  such that  $f_x$  is continuous at all points of  $O_x$  and  $f_x^{-1}$  is continuous at all points of  $f_x[O_x]$ . Assume further that this condition holds for all points  $x \in X$  (note that  $O_x$  and  $f_x$  exist for each  $x$  individually and not globally). Then  $Y$  is *locally homeomorphic* to  $X$ .

Let us now proceed to functions  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ . We will occasionally refer to such functions as *numerical*. A numerical function is *continuous* if it is continuous with respect to the usual topology, as defined above. Equivalently,  $f$  is continuous if for every  $\mathbf{x} \in \mathbb{R}^n$  and every open ball  $B_{\mathbf{y}}$  containing  $\mathbf{y} = f(\mathbf{x})$ , there exists an open ball  $B_{\mathbf{x}}$  containing  $\mathbf{x}$  such that  $f[B_{\mathbf{x}}] \subseteq B_{\mathbf{y}}$ .

We assume you are familiar with the concept of differentiability, as taught in calculus courses. A function is of *class  $C^k$*  if all its partial derivatives up to order  $k$  exist and are continuous. A function is  $C^\infty$  if it is of class  $C^k$  for all  $k$ . A function of class  $C^\infty$  is also called *smooth*, and we will henceforth use “smooth” consistently for  $C^\infty$ . A function is *analytic* if it can be expressed as an infinite power series about each point and the power series converges absolutely in an open neighborhood of the point. An analytic function is always smooth, but not conversely.

## 3.3 Manifolds

### 3.3.1 Definition of a Manifold

A *smooth topological manifold of dimension  $n$*  is a set  $M$  satisfying the following axioms:

- M1**  $M$  is a topological space whose topology is Hausdorff and second countable.
- M2** There is a fixed collection of open sets  $\mathcal{O} = \{O_i, i \in I\}$  on  $M$  that *covers*  $M$ ; that is,  $\bigcup_{i \in I} O_i = M$ .
- M3** For each  $O_i \in \mathcal{O}$  there is an injective function  $\psi_i: O_i \rightarrow \mathbb{R}^n$  such that  $\psi_i$ , when its range is restricted to  $\psi_i[O_i]$ , is a homeomorphism between  $O_i$  and  $\psi_i[O_i]$ . The pair  $(O_i, \psi_i)$  is called a *chart* and the collection of all charts is called an *atlas*.
- M4** Two charts  $(O_i, \psi_i)$  and  $(O_j, \psi_j)$  are said to be *compatible* if either  $O_i \cap O_j = \emptyset$  or  $O_i \cap O_j = U \neq \emptyset$  and the function  $\psi_i \circ \psi_j^{-1}: \psi_j[U] \rightarrow \psi_i[U]$  is smooth. Then every pair of charts in the atlas is compatible.
- M5** The atlas is *maximal* in the following sense: If  $(O, \psi)$  is a chart that is compatible with every chart  $(O_i, \psi_i)$  in the atlas, then  $(O, \psi)$  is in the atlas.

The following comments should clarify this definition.

1. Imposing a topology on  $M$  is the first step in establishing similarity between the local structure of  $M$  and that of  $\mathbb{R}^n$ , since we aim to use the usual topology on  $\mathbb{R}^n$  as a “reference” topology. The Hausdorff and second countability requirements are technical and will not be used here. Nonetheless, they are needed to establish some deeper aspects of theory, outside our scope.
2. The number of sets in  $\mathcal{O}$  needed in most applications of the theory is finite and fairly small. For example, two charts are sufficient to cover the surface of a two-dimensional sphere, as we explained in Section 3.1. Nonetheless, axiom **M5**, which is included for technical reasons, expands  $\mathcal{O}$  beyond necessity

3. It follows from axiom **M3** that  $\psi_i[O_i]$  is open in  $\mathbb{R}^n$ .
4. It is the dimension  $n$  of  $\mathbb{R}^n$  in axiom **M3** that makes us call the manifold  $n$ -dimensional.
5. A chart is also called a *coordinate system*. Given a chart  $(O, \psi)$ , the set  $O$  is called the *chart set* and the function  $\psi$  is called the *chart function*.
6. Note that the set  $U$  in axiom **M4** is open, being an intersection of open sets. Therefore, and since homeomorphisms carry open sets to open sets,  $\psi_i[U]$  and  $\psi_j[U]$  are open in  $\mathbb{R}^n$ .
7. Pay attention to the construction in axiom **M4**. The idea is to be able to work with smooth functions on subsets of  $\mathbb{R}^n$  into  $\mathbb{R}^n$ . Since the points of  $M$  are not numerical ( $M$  is an abstract set rather than a concrete numerical set), we must find a way to “go through”  $M$ . The function  $\psi_i \circ \psi_j^{-1}$  answers this need. Since  $\psi_j^{-1} : \psi_j[U] \rightarrow U$  and  $\psi_i : U \rightarrow \psi_i[U]$ , the composition is indeed on  $\psi_j[U]$  into  $\psi_i[U]$ , both of which are open subsets of  $\mathbb{R}^n$ .
8. Axiom **M5** serves to make the definition of a manifold unique, since otherwise we could choose many different atlases satisfying axioms **M2** through **M4**. This is a matter of convenience rather than necessity, since it eliminates the need to explicitly state which atlas we are referring to when working with a manifold.

### 3.3.2 Smooth Functions on Manifolds

We want to be able to define a smooth function on a manifold  $M$  to  $\mathbb{R}^m$ , where  $m$  may be equal to or different from  $n$ . Since  $M$  is not numerical, we cannot do this directly using definitions from calculus. However, we may use the same device used in axiom **M4**. Let  $f : M \rightarrow \mathbb{R}^m$  be a function satisfying the following requirement:

**MF1** For every  $(O_i, \psi_i)$  in the atlas, the function  $f \circ \psi_i^{-1} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is smooth.

Then  $f$  is defined to be smooth. Note that a single  $(O_i, \psi_i)$  makes  $f \circ \psi_i^{-1}$  “go through”  $O_i$  only and hence is not sufficient to characterize  $f$  on the entire manifold. However, since **MF1** holds for every chart in the atlas,  $f$  is, in fact,

characterized on  $M$  in its entirety. Analytic and  $C^k$  functions on manifolds to  $\mathbb{R}^m$  can be defined similarly.

A particular class of smooth functions is given by the coordinate functions of a manifold. These are defined as follows. First define the functions

$$\xi_k : \mathbb{R}^n \rightarrow \mathbb{R}, \quad \xi_k(\mathbf{x}) = x^k \quad (3.8)$$

The function  $\xi_k$  selects the  $k^{\text{th}}$  coordinate of a vector. This is clearly a smooth function. It now follows from **M4** and **MF1** that the composition  $\xi_k \circ \psi_i : O_i \rightarrow \mathbb{R}$  is smooth for all  $i$  and all  $k$ . The functions  $(\xi_1 \circ \psi_i, \dots, \xi_n \circ \psi_i)$  are the *coordinate functions* of the chart  $(O_i, \psi_i)$ . Of course, the different charts of the atlas have different sets of coordinate functions.

### 3.3.3 Derivatives on Manifolds

We want to define derivatives of functions on manifolds. Again, we cannot use calculus directly since a general manifold is abstract and its points are not numerical. Nonetheless, we can define a differentiation operator by imposing the fundamental properties of a derivative on such an operator, namely linearity and the rule of differentiation of a product. We will now explain this idea in precise terms.

Let  $M$  be a smooth manifold and let  $\mathcal{F}$  be the set of all smooth functions  $f : M \rightarrow \mathbb{R}$ , as defined in the preceding subsection. Note that the target space here is the real line  $\mathbb{R}$  rather than a general Euclidean space. Let  $f$  and  $g$  be two functions in  $\mathcal{F}$ . Then  $\alpha f + \beta g$  clearly belongs to  $\mathcal{F}$  for any two real numbers  $\alpha$  and  $\beta$ . It is, perhaps, less obvious that  $fg$  also belongs to  $\mathcal{F}$ . Note that  $fg$  denotes *pointwise multiplication*; that is, the value of  $fg$  at  $p \in M$  is equal to  $f(p)g(p)$ . Once this is understood, it is easy to verify that  $fg$  is smooth by recalling **MF1** and the fact that the product of two numerical smooth functions is a smooth function.

Now let  $p$  be a fixed point in  $M$  and define an operator  $d_p : \mathcal{F} \rightarrow \mathbb{R}$  to be a *derivative operator* at  $p$  if the following axioms hold:

**MD1** Linearity:

$$d_p(\alpha f + \beta g) = \alpha d_p(f) + \beta d_p(g) \text{ for all } f, g \in \mathcal{F} \text{ and all real } \alpha, \beta \quad (3.9)$$

**MD2** Product rule:

$$d_p(fg) = fd_p(g) + gd_p(f) \text{ for all } f, g \in \mathcal{F} \quad (3.10)$$

We emphasize again that  $f$ ,  $g$ ,  $d_p(f)$ , and  $d_p(g)$  in (3.9), (3.10) are understood to be evaluated at  $p$ , therefore they are all real numbers.

As an example of the preceding definition, let us prove that the derivative of a constant function is zero. Suppose that  $f(p) = c \neq 0$  for all  $p \in M$ . Let us compute  $d_p(f^2)$  it two ways, once from (3.9) and once from (3.10). Using (3.9) with  $f(p) = c$  gives

$$d_p(f^2) = d_p(cf) = cd_p(f)$$

On the other hand, using (3.10) with  $f(p) = c$  gives

$$d_p(f^2) = 2cd_p(f)$$

These two equalities agree only if  $d_p(f) = 0$ .

### 3.3.4 Directional Derivatives Along Cartesian Coordinates

Let us continue to explore ways of differentiation related to smooth functions on manifolds. As before, let  $p \in M$  be a fixed point,  $f \in \mathcal{F}$  a given function, and  $(O_p, \psi_p)$  a fixed chart such that  $p \in O_p$ . By the definition of a smooth function on a manifold, the function  $f \circ \psi_p^{-1} : \psi_p[O] \rightarrow \mathbb{R}$  is smooth on an open subset of  $\mathbb{R}^n$  to  $\mathbb{R}$ . For such a function, the conventional partial derivatives with respect to the cartesian coordinates  $x^k$  in  $\mathbb{R}^n$ :

$$\frac{\partial}{\partial x^k}(f \circ \psi_p^{-1})$$

are well defined as functions on an open subset of  $\mathbb{R}^n$  to  $\mathbb{R}$ .

Now comes a point that calls for special attention. Let us substitute  $f(p)$  for  $f$  and  $\psi(p)$  for  $\psi$  in the partial derivatives. Doing so will result in a real number. Now let us repeat for *all* functions  $f \in \mathcal{F}$ . This will result in an operator  $\partial/\partial x^k|_p : \mathcal{F} \rightarrow \mathbb{R}$ . Moreover, being a bona-fide partial derivative, it

satisfies (3.9) and (3.10) automatically. Therefore,  $\partial/\partial x^k|_p$  is a derivative operator on  $M$ , as defined in the preceding subsection.

The operators  $(\partial/\partial x^k|_p, 1 \leq k \leq n)$  are called the *directional derivatives along the coordinates*. They will play an important role in the remainder of this chapter.

### 3.3.5 Tangent Vectors and Tangent Spaces

The definition of an abstract derivative operator **MD1**, **MD2** leaves room for many different derivative operators on the same manifold. Given any two derivative operators  $d_{p,1}$  and  $d_{p,2}$ , we can define their linear sum:

$$(\alpha_1 d_{p,1} + \alpha_2 d_{p,2})(f) = \alpha_1 d_{p,1}(f) + \alpha_2 d_{p,2}(f) \quad (3.11)$$

Showing that (3.11) satisfies (3.9) and (3.10) is straightforward. Also, we may define the zero operator as one that assigns the value 0 to any function  $f$ . Therefore, the collection of all derivative operators at a point  $p$ , as defined by **MD1** and **MD2**, is a vector space. We denote this space by  $\mathcal{D}_p$  and call it the *tangent space* of the manifold at  $p$ . The elements of  $\mathcal{D}_p$ , namely the derivative operators, are called *tangent vectors*.

The main property of the the tangent space of a manifold at a point is given by the following theorem

**Theorem 7.** *The tangent space  $\mathcal{D}_p$  is  $n$ -dimensional and the directional derivatives  $(\partial/\partial x^k|_p, 1 \leq k \leq n)$  constitute a basis for the space.*

We will prove only one part, namely that the directional derivatives are independent. The other part, that every derivative operator  $d$  can be expressed as a linear sum of the directional derivatives, will be omitted<sup>3</sup>. Assume that

$$v^k \frac{\partial}{\partial x^k} \Big|_p = 0$$

---

<sup>3</sup>Although I strive to include, or at list provide hints for, proofs of all claims in this document, I decided to make an exception here. The omitted proof is rather long and technical. Perhaps I will add it in the future.

for some  $n$  numbers  $v^k$ . Then, by the definition of the directional derivatives,

$$v^k \frac{\partial}{\partial x^k} (f \circ \psi^{-1}) \Big|_p = 0$$

This equality cannot hold identically for *all* smooth functions at a specific point  $p$ , unless all  $v^k$  are zero.

We can now compare the present construction of tangent spaces to that in Chapter 2. Recall the definition (2.8) of tangent vectors as partial derivatives of the position vector  $\mathbf{r}$  with respect to the curvilinear coordinates. In Chapter 2 we started with a fixed  $n$ -dimensional vector space, and all constructions were in this space. In particular, the tangent vectors and the tangent space were concrete objects, defined via numerical coordinates. Here we defined a tangent vector as an abstract operator on an abstract space, and imposed certain properties in order to make the collection of tangent vectors a linear space. The abstract construction led eventually to numerically defined basis—that of  $\partial/\partial x^k|_p$ , the directional derivatives along the coordinates. We will call this basis the *coordinates basis* for short.

At this point, the road is open to the development of tensor theory on manifolds, parallel to tensor theory on curvilinear coordinates. This endeavor will be taken up in the next section.

### 3.4 Tensors and Tensor Fields on Manifolds

Before we start developing the material in this section, we make a comment on notations. In Chapter 2 we made a distinction between lowercase symbols, which were defined on the fixed underlying vector space, and uppercase symbols, which were defined on local tangent bases. Now we have no underlying vector space and all objects are on manifolds. Therefore there will be no need for distinction and we will mix lowercase and uppercase symbols as we find convenient. Also, there are no curvilinear coordinates here, only cartesian coordinates on open sets of  $\mathbb{R}^n$  that are locally homeomorphic to the charts of the manifold. Therefore, we will use mostly the letter  $x$  for coordinates, unless convenience calls for a different letter.

### 3.4.1 Coordinate Transformations

Suppose that a point  $p \in M$  belongs to two charts  $(O, \psi)$  and  $(\tilde{O}, \tilde{\psi})$ . The definition of tangent vectors and tangent spaces is independent of the chart, so there is only one tangent space at  $p$ ; moreover, a specific tangent vector at  $p$  is uniquely defined, independently of the chart. However, the two different charts have different bases  $(\partial/\partial x^1, \dots, \partial/\partial x^n)$  and  $(\partial/\partial \tilde{x}^1, \dots, \partial/\partial \tilde{x}^n)$  at  $p$ . These bases are related by the usual chain rule

$$\frac{\partial}{\partial \tilde{x}^i} = \frac{\partial}{\partial x^j} \frac{\partial x^j}{\partial \tilde{x}^i}, \quad \frac{\partial}{\partial x^i} = \frac{\partial}{\partial \tilde{x}^j} \frac{\partial \tilde{x}^j}{\partial x^i} \quad (3.12)$$

where the partial derivatives are defined on the numerical function  $\psi \circ \tilde{\psi}^{-1}$  and its inverse, and are understood to be evaluated at  $p$ .

If  $\mathbf{v}$  is a tangent vector at  $p$ , then  $\mathbf{v}$  can be expressed in terms of its coordinates in the two bases:

$$\mathbf{v} = v^i \frac{\partial}{\partial x^i} = \tilde{v}^j \frac{\partial}{\partial \tilde{x}^j} \quad (3.13)$$

from which we deduce the change-of-basis transformation rules

$$v^i = \frac{\partial x^i}{\partial \tilde{x}^j} \tilde{v}^j, \quad \tilde{v}^i = \frac{\partial \tilde{x}^i}{\partial x^j} v^j \quad (3.14)$$

Comparing these formulas with (1.11) and (1.12), we see that with

$$S_j^i = \frac{\partial x^i}{\partial \tilde{x}^j}, \quad T_j^i = \frac{\partial \tilde{x}^i}{\partial x^j} \quad (3.15)$$

the transformation rules for vectors on manifolds become identical to the transformation rules for conventional vectors on an  $n$ -dimensional space. We should bear in mind, however, the essential difference between the two: (1.11) and (1.12) apply globally on the underlying space, whereas (3.14) holds for a given point of the manifold. Consequently, the matrices  $S_j^i$  and  $T_j^i$  vary from point to point, as they do in the case of curvilinear coordinates.

### 3.4.2 Cotangent Spaces

We have denoted by  $\mathcal{D}_p$  the tangent space at  $p$ . As we learned in Chapter 1, we can assign to  $\mathcal{D}_p$  a dual space  $\mathcal{D}_p^*$ , as defined in Appendix A. This space is

called the *cotangent space* at  $p$  and we will refer to its elements as covectors or dual vectors, as in Chapter 1. A common notation for the covectors comprising the dual basis is  $(dx^1, dx^2, \dots, dx^n)$ . The change-of-basis transformation for dual bases is

$$d\tilde{x}^i = \frac{\partial \tilde{x}^i}{\partial x^j} dx^j, \quad dx^i = \frac{\partial x^i}{\partial \tilde{x}^j} d\tilde{x}^j \quad (3.16)$$

and the change-of-basis rules for the coordinates of a covector  $\mathbf{w}$  are

$$w_i = \tilde{w}_j \frac{\partial \tilde{x}^j}{\partial x^i}, \quad \tilde{w}_i = w_j \frac{\partial x^j}{\partial \tilde{x}^i} \quad (3.17)$$

### 3.4.3 Vector Fields

A *vector field*  $d$  on a manifold is a collection  $\{d_p : p \in M\}$  of derivative operators; that is, an assignment of a tangent vector to each point in the manifold.

We want to construct a meaningful definition of smoothness of a vector field. We recall that, for each  $p$ ,  $d_p(f) \in \mathbb{R}$ . Therefore, for a fixed  $f$ ,  $\{d_p(f) : p \in M\}$  defines a function  $d_f : M \rightarrow \mathbb{R}$ . Pay attention to the reversal of roles of subscripts:  $d_p$  is a function on  $\mathcal{F} \rightarrow \mathbb{R}$  for a fixed  $p$  and we denote its value on  $f \in \mathcal{F}$  by  $d_p(f)$ . On the other hand,  $d_f$  is a function on  $M \rightarrow \mathbb{R}$  and we denote its value on  $p \in M$  by  $d_f(p)$ . This notation takes some time to get used to, but it is important, so the time is well spent.

We can now define the notion of a smooth vector field. A vector field  $d$  is *smooth* if  $d_f$  is smooth on  $M$  for every smooth function  $f$  on  $M$ .

Recall the definition of coordinate bases:

$$\left. \frac{\partial}{\partial x^k} \Big|_p = \left\{ \frac{\partial}{\partial x^k} (f \circ \psi_p^{-1}) \Big|_p : f \in \mathcal{F} \right\} \quad (3.18)$$

As before, we can interchange the roles of  $p$  and  $f$ :

$$\left. \frac{\partial}{\partial x^k} \Big|_f = \left\{ \frac{\partial}{\partial x^k} (f \circ \psi_p^{-1}) \Big|_f : p \in O_p \right\} \quad (3.19)$$

Note that (3.19) can be defined only on the chart set  $O_p$  corresponding to the chart function  $\psi_p$ . The numerical functions  $\partial(f \circ \psi_p^{-1})/\partial x^k$  are obviously

smooth on  $\psi_p$ . When we move to a different chart, we must differentiate with respect to the new coordinates, so all we can say is that the vector field  $\partial/\partial x^k$  can be defined on each chart separately and is smooth on each chart.

It is now easy to conclude that, since the fields of coordinate bases are smooth, a general vector field  $v$  is smooth if and only if its coordinates  $v^k$  (which are functions  $M \rightarrow \mathbb{R}$ ) are smooth on every chart.

### 3.4.4 Tensors and Tensor Fields

We will construct tensors in the same way we did in Chapter 1, which is essentially a definition based on numerical coordinates in a selected basis, together with rules of transformation under change of basis. We define an  $(r, s)$ -tensor field as a collection of  $n^{r+s}$  smooth functions on  $M$ , denoted as  $a_{j_1 \dots j_s}^{i_1 \dots i_r}$  and obeying the transformation law

$$\tilde{a}_{j_1 \dots j_s}^{i_1 \dots i_r} = \frac{\partial \tilde{x}^{i_1}}{\partial x^{k_1}} \cdots \frac{\partial \tilde{x}^{i_r}}{\partial x^{k_r}} a_{m_1 \dots m_s}^{k_1 \dots k_r} \frac{\partial x^{m_1}}{\partial \tilde{x}^{j_1}} \cdots \frac{\partial x^{m_s}}{\partial \tilde{x}^{j_s}} \quad (3.20)$$

Compare this definition with (1.28) and you will see that the two are essentially identical, except for the difference in interpretation. Whereas (1.28) defines constant tensors over a vector space, (3.20) defines a tensor field over a manifold. The adjectives “covariant” and “contravariant” are used here in the same way as for constant tensors.

We can use ordinary partial differentiation on tensor fields:

$$a_{j_1 \dots j_s, p}^{i_1 \dots i_r} = \frac{\partial a_{j_1 \dots j_s}^{i_1 \dots i_r}}{\partial x^p} \quad (3.21)$$

The result is not a tensor field, however, since it does not satisfy the transformation law (3.20). To obtain a derivative tensor field, we will need the covariant derivative, as we saw in Chapter 2 and as we will see again later in this chapter.

The special case of a  $(1, 0)$ -tensor field gives rise to a vector field, as defined in Subsection 3.4.3. The special case of a  $(0, 1)$ -tensor field gives rise to a *covector field*.

### 3.4.5 The Metric Tensor

A *metric tensor* on a manifold  $M$  is a smooth  $(0, 2)$ -tensor field  $g_{ij}$  satisfying the following axioms

**MM1**  $g_{ij}$  is symmetric.

**MM2**  $g_{ij}$  is nondegenerate, as defined in Section 1.11.

**MM3** The signature  $\Lambda$  of  $g_{ij}$ , as defined in Theorem 2 (Sylvester's theorem of inertia) is constant on the entire manifold.

These three axioms allow us to write  $g_{ij}$  as

$$g_{ij} = S_i^k S_j^m \Lambda_{km} \quad (3.22)$$

where  $S$  is a nonsingular matrix and  $\Lambda$  is diagonal with a constant pattern of  $\pm 1$ 's along the diagonal. Both  $g_{ij}$  and  $S$  vary from point to point on the manifold, but  $\Lambda$  does not.

The inverse, or dual, metric tensor  $g^{ij}$  is given by

$$g^{ij} = T_k^i T_m^j \Lambda^{km} \quad (3.23)$$

where  $T$  is the inverse of  $S$ ; that is,  $T_k^i S_j^k = S_k^i T_j^k = \delta_j^i$ . It is easy to verify that, since  $\Lambda$  is its own inverse,

$$g_{jk} g^{ki} = g^{ik} g_{kj} = \delta_j^i \quad (3.24)$$

The metric tensor can be expressed in full form, including its basis covectors, as in (1.40),

$$ds^2 = g_{ij} dx^i dx^j \quad (3.25)$$

The notation  $ds^2$ , although it is merely symbolic and should not be understood as the square of a real number, is called the (square of the) *line element*.

The transformation law of the metric tensor under change of basis is the same as any  $(0, 2)$ -tensor:

$$\tilde{g}_{ij} = \frac{\partial x^p}{\partial \tilde{x}^i} \frac{\partial x^q}{\partial \tilde{x}^j} g_{pq} = \left( \frac{\partial x^p}{\partial \tilde{x}^i} S_p^k \right) \left( \frac{\partial x^q}{\partial \tilde{x}^j} S_q^m \right) \Lambda_{km} \quad (3.26)$$

### 3.4.6 Interlude: A Device

A simple device, which we now introduce, will enable us to save many pages of definitions and derivations. Let us conjure up an  $n$ -dimensional inner product space  $\mathcal{V}$ , together with a basis  $(\mathbf{e}_1, \dots, \mathbf{e}_n)$  whose Gram matrix is

$$\mathbf{e}_k \cdot \mathbf{e}_m = \Lambda_{km} \quad (3.27)$$

Now, as we did in Chapter 2, let us define a local basis at each point of the manifold:

$$\mathbf{E}_i = S_i^k \mathbf{e}_k \quad (3.28)$$

We then find that

$$\mathbf{E}_i \cdot \mathbf{E}_j = S_i^k S_j^m (\mathbf{e}_k \cdot \mathbf{e}_m) = S_i^k S_j^m \Lambda_{km} = g_{ij} \quad (3.29)$$

We thus find ourselves in exactly the same framework as in Chapter 2. Instead of using the metric tensor directly, we may use the hypothetical bases  $(\mathbf{e}_1, \dots, \mathbf{e}_n)$  and  $(\mathbf{E}_1, \dots, \mathbf{E}_n)$  and base all definitions and derivations on properties of these bases.

We emphasize that this device is possible only because of axiom **MM3** of the metric tensor. The constancy of  $\Lambda$  facilitates the use of a fixed inner product space with a fixed signature for the entire manifold. It should also be clear that  $\mathcal{V}$  and the bases  $\mathbf{e}_i$  and  $\mathbf{E}_i$  are not part of the definition of a manifold. They are artificial devices, introduced only for their technical usefulness, as we will see in the next subsection.

### 3.4.7 (Almost) All Work Done

Once the device employed in the preceding subsection is understood, all the material in Chapter 2 from Section 2.4 onwards applies to tensors on manifolds, with no changes. In particular, affine connections, covariant derivatives, and the special tensors, are defined and used as in Chapter 2. All you need to remember is that:

- The coordinates  $x^k$  on some chart  $(O, \psi)$  take the place of the curvilinear coordinates  $y^k$ .

- We have not dealt with the difficulty of patching charts and thereby facilitating working on the manifold as a whole. Although such patching can be made rigorous, it is quite technical and is outside the scope of this document. Here we must keep in mind that a single fixed chart is assumed, although this can be any chart in the atlas.

The “almost” in the title of this subsection implies that we are not quite finished yet. Additional material, both interesting and important, is provided in the next section.

### 3.5 Curves and Parallel Transport

A *smooth curve* on a manifold  $M$  is a smooth function  $\gamma: \mathbb{R} \rightarrow M$ , or, in some cases, from a closed interval in  $\mathbb{R}$  to  $M$ . Smoothness is defined in a natural way, namely, by requiring that  $f \circ \gamma: \mathbb{R} \rightarrow \mathbb{R}$  be smooth for all  $f \in \mathcal{F}$ . Expressing the curve explicitly as  $\gamma(t)$ , we may visualize the curve as a line lying in  $M$  and progressing as  $t$  increases. If  $\gamma$  is injective, the curve does not cross or touch itself at any point.

The *tangent vector* at a point  $p(t)$  on the curve is defined as

$$\tau(f) = \frac{d(f \circ \gamma)}{dt} \tag{3.30}$$

for every  $f \in \mathcal{F}$ . This definition makes  $\tau(f)$  is a derivative operator; that is, a member of  $\mathcal{D}_p$ , hence a vector at  $p$ .

The tangent vector can be expressed in terms of the coordinates basis:

$$\tau = \frac{dx^k}{dt} \frac{\partial}{\partial x^k} = \tau^k \frac{\partial}{\partial x^k} \tag{3.31}$$

where  $\tau^k = dx^k/dt$  are the coordinates of the tangent vector in the coordinates basis. Note that  $x^k(t)$  are the coordinates of the curve in the coordinates basis.

Let  $\gamma(t)$  be a curve and  $\tau^k(t)$  its tangent vector. Let  $v^i(t)$  be a function  $\mathbb{R} \rightarrow \mathcal{D}$ ; that is, a collection of vectors defined on all points of the curve.

Then  $v^i(t)$  is *parallel transported* on the curve if

$$\tau^k(t)v_{;k}^i(t) = 0 \quad \text{for all } t \text{ and all } i \quad (3.32)$$

where, as we recall,  $v_{;k}^i$  is the covariant derivative of  $v^i$  along the  $k^{\text{th}}$  coordinate. The left side of equation (3.32) is thus a contraction of a vector by a mixed (1, 1)-tensor, resulting in a vector.

The definition (3.32) is, perhaps, not very illuminating and makes it hard to understand what “parallel” means in this context. Let us therefore bring (3.32) to a form that will make it more transparent. First, note that the requirement that the left side of (3.32) be zero for all  $i$  translates to

$$\tau^k(t)v_{;k}^i(t)\mathbf{E}_i = 0 \quad \text{for all } t \quad (3.33)$$

Let us substitute for  $\tau^k(t)$ ,  $v_{;k}^i$  and  $\mathbf{E}_i$  their definitions:

$$\frac{dx^k}{dt} \left( \frac{\partial v^i}{\partial x^k} + \Gamma_{km}^i v^m \right) S_i^p \mathbf{e}_p = \frac{dx^k}{dt} \left( \frac{\partial v^i}{\partial x^k} S_i^p + \Gamma_{km}^i S_i^p v^m \right) \mathbf{e}_p = 0 \quad (3.34)$$

Next we substitute (2.15):

$$\frac{dx^k}{dt} \left( \frac{\partial v^i}{\partial x^k} S_i^p + \frac{\partial S_m^p}{\partial x^k} v^m \right) \mathbf{e}_p = \frac{dx^k}{dt} \frac{\partial(v^i S_i^p)}{\partial x^k} \mathbf{e}_p = 0 \quad (3.35)$$

We now observe that  $v^i S_i^p = V^p$ , where  $V^p$  are the coordinates of the vector  $v$  in the hypothetical reference vector space  $\mathcal{V}$ . Therefore,

$$\frac{\partial(V^p \mathbf{e}_p)}{\partial x^k} \frac{dx^k}{dt} = \frac{d(V^p \mathbf{e}_p)}{dt} = 0 \quad (3.36)$$

The conclusion from this derivation is: A parallel transported vector remains unchanged—that is, remains parallel to itself—when viewed as an abstract geometrical object in the hypothetical reference space  $\mathcal{V}$ .

What have we learned from the preceding derivation? On the practical level, certainly not much. On the intuitive level, we must exercise our imagination and visualize the vector transported parallel to itself in some hypothetical space along a curve that exists on a “real” manifold. Of course, both  $M$  and  $\mathcal{V}$  are ultimately abstract, but the former a defined mathematical object,

whereas the latter is an artificial device. If this digression was not helpful, you may simply disregard it.

Returning to the defining equation of parallel transport, we may write it as

$$\frac{dv^i}{dt} + \Gamma_{km}^i \frac{dx^k}{dt} v^m = 0 \quad \text{for all } t \text{ and all } i \quad (3.37)$$

Assume that  $v^i$  is known at point  $p$  on the curve and we assign  $t = 0$  to this point. Then (3.37) is a coupled set of  $n$  nonlinear differential equations in the unknown functions  $v^i(t)$ . The functions  $dx(t)/dt$  are known, being the derivatives of the curve's coordinates. The affine connections are also known, since they depend only on the metric. Since  $\Gamma_{km}^i$  and  $dx(t)/dt$  are smooth functions, the differential equations have a unique solution, although the solution may not be easy to find. It follows that knowledge of  $v^i$  on a single point uniquely defines the parallel transported vector  $v^i(t)$  on the entire curve. In view of the interpretation given above to the notion of parallel transport, this conclusion should come at no surprise. You may think of the curve as “moving under the feet of the vector” rather than the vector moving along the curve.

## 3.6 Curvature

Let  $\gamma(t)$  be a curve on a manifold defined for  $0 \leq t \leq 1$  and such that  $\gamma(0) = \gamma(1)$ . Such a curve is called *closed* for an obvious reason. Let  $v^k(0)$  be a fixed vector at  $\gamma(0)$  and perform parallel transport on  $\gamma$ , as defined by Equation (3.37). What would  $v^k(1)$  be? You may think intuitively that  $v^k(1) = v^k(0)$ , since the vector is back to its initial point. If indeed you think so, then your intuition have failed you this time. It is a surprising and important fact that  $v^k(1) \neq v^k(0)$  in general. This follows from the curvature of the manifold, and we analyze this phenomenon in mathematical terms in this section.

Let us consider a simple example first. Place an arrow at the North Pole of the earth, tangent to the surface and pointing south in the direction of the Greenwich meridian (longitude  $0^\circ$ ). Now begin to move south along the meridian with the arrow pointing south all the time, until you hit the equator.

At this point start moving east along the equator with the arrow continuing pointing south. When you get to longitude  $90^\circ$  (you will be somewhere in the Indian Ocean, but never mind that), start moving north along the meridian with the arrow still pointing south. When you get to the north pole, the arrow will be pointing south parallel to the  $90^\circ$  meridian, so it will be rotated  $90^\circ$  relative to the direction of the arrow when you started! Although we have cheated slightly in this example (the curve in this example is not smooth, having three corners), the general behavior is indeed true.

Let us now do the mathematics. We first need an auxiliary result, derived in the next subsection. Then we will be ready for deriving the main result.

### 3.6.1 Preparation: The Area of a Closed Curve

We want to compute the area of a closed planar curve  $\gamma(t)$  surrounding the origin, as shown in Figure 3.1. The curve is parameterized such that  $\gamma(0) = \gamma(1)$ .

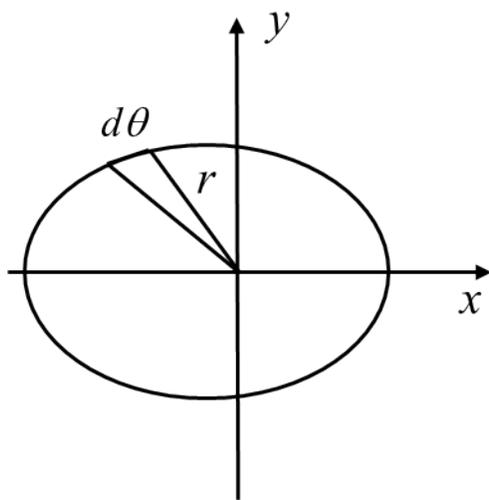


Figure 3.1: The Area of a Closed Curve

The area of the triangle spanning the angle between  $\theta$  and  $\theta + d\theta$  is

$$dA = 0.5r^2d\theta \quad (3.38)$$

Therefore, the total area enclosed by the curve is

$$A = 0.5 \int_0^{2\pi} r^2 d\theta = 0.5 \int_0^1 r^2(t) \frac{d\theta}{dt} dt \quad (3.39)$$

Let us convert to cartesian coordinates:

$$\begin{aligned} r^2 &= x^2 + y^2 \\ \theta(t) &= \arctan \frac{y}{x} \\ \frac{d\theta}{dt} &= \frac{1}{x^2 + y^2} \left( \frac{dy}{dt} x - \frac{dx}{dt} y \right) \end{aligned} \quad (3.40)$$

Now substitute in (3.39):

$$A = 0.5 \int_0^1 \left( \frac{dy}{dt} x - \frac{dx}{dt} y \right) dt \quad (3.41)$$

Next observe the following:

$$0.5 \int_0^1 \left( \frac{dy}{dt} x + \frac{dx}{dt} y \right) dt = 0.5 \int_0^1 \frac{d(xy)}{dt} dt = 0.5[x(1)y(1) - x(0)y(0)] = 0 \quad (3.42)$$

The last equality follows from the fact that the curve is closed. Finally, adding and subtracting (3.41) and (3.42) gives

$$A = \int_0^1 \frac{dy(t)}{dt} x(t) dt = - \int_0^1 \frac{dx(t)}{dt} y(t) dt \quad (3.43)$$

### 3.6.2 Approximate Solution of the Parallel Transport Equations

Choose a point on the manifold having coordinates  $x_0^k$ . Construct a closed curve

$$x(t) = x_0 + \varepsilon\gamma(t) \quad (3.44)$$

such that  $\gamma(t)$  is a fixed curve and  $\varepsilon$  is a small scalar. You may visualize the curve as a small loop surrounding  $x_0$ . Our aim is to approximate the parallel transport equations (3.37) to order  $\varepsilon^2$ ; that is, to neglect terms of higher order in  $\varepsilon$ .

Beginning with  $x(t)$ , we see from (3.44) that

$$\frac{dx^k(t)}{dt} = \varepsilon \frac{d\gamma^k(t)}{dt} \quad (3.45)$$

Looking at (3.37), we see that, since (3.45) already contains a factor  $\varepsilon$ , we do not need more than a first order expansion of  $\Gamma_{km}^i$  in  $\varepsilon$ . Since  $\Gamma_{km}^i$  depends on  $\varepsilon$  only through its dependence on the coordinates, the first-order expansion is

$$\Gamma_{km}^i[x(t)] = \Gamma_{km}^i[x_0 + \varepsilon\gamma(t)] \approx \Gamma_{km}^i(x_0) + \varepsilon\Gamma_{km,p}^i(x_0)\gamma^p(t) \quad (3.46)$$

The third element in (3.37) is the unknown variable itself, which we expand in a second-order series:

$$\begin{aligned} v^i(t) &\approx v_0^i + \varepsilon v_1^i(t) + \varepsilon^2 v_2^i(t) \\ \frac{dv^i(t)}{dt} &\approx \varepsilon \frac{dv_1^i(t)}{dt} + \varepsilon^2 \frac{dv_2^i(t)}{dt} \end{aligned} \quad (3.47)$$

We can now substitute (3.45), (3.46), (3.47) in (3.37):

$$\begin{aligned} &\varepsilon \frac{dv_1^i(t)}{dt} + \varepsilon^2 \frac{dv_2^i(t)}{dt} \\ &+ \varepsilon \frac{d\gamma^k(t)}{dt} [\Gamma_{km}^i(x_0) + \varepsilon\Gamma_{km,p}^i(x_0)\gamma^p(t)] [v_0^m + \varepsilon v_1^m(t) + \varepsilon^2 v_2^m(t)] \approx 0 \end{aligned} \quad (3.48)$$

The first-order term in  $\varepsilon$  gives

$$\frac{dv_1^i(t)}{dt} + \frac{d\gamma^k(t)}{dt} \Gamma_{km}^i(x_0) v_0^m = 0 \quad (3.49)$$

This is easily integrated to yield

$$v_1^i(t) + \gamma^k(t) \Gamma_{km}^i(x_0) v_0^m = C \quad (3.50)$$

Since  $\gamma^k(t) \leftarrow 0$  must give  $v_1^i(t) \leftarrow v_0^i$  by definition, the constant of integration  $C$  is zero. Therefore,

$$v_1^i(t) + \gamma^k(t) \Gamma_{km}^i(x_0) v_0^m = 0 \quad (3.51)$$

and thus

$$v_1^i(1) - v_1^i(0) = -[\gamma^k(1) - \gamma^k(0)]\Gamma_{km}^i(x_0)v_0^m = 0 \quad (3.52)$$

We conclude that there is no first-order effect of parallel transport on a closed curve.

Let us now proceed to the second-order term in  $\varepsilon$ . We again find from (3.48)

$$\frac{dv_2^i(t)}{dt} + \frac{d\gamma^k(t)}{dt}[\Gamma_{km}^i(x_0)v_1^m(t) + \Gamma_{km,p}^i(x_0)\gamma^p(t)v_0^m] = 0 \quad (3.53)$$

We want to substitute (3.51) in (3.53). To facilitate this substitution, we must first make several dummy index replacements in both (3.51) and (3.53), after which (3.53) becomes

$$\frac{dv_2^i(t)}{dt} + [\Gamma_{kj,\ell}^i(x_0) - \Gamma_{km}^i(x_0)\Gamma_{j\ell}^m(x_0)]v_0^j \frac{d\gamma^k(t)}{dt} \gamma^\ell(t) = 0 \quad (3.54)$$

This differential equation can be integrated to give

$$\begin{aligned} v_2^i(1) - v_2^i(0) &= -[\Gamma_{kj,\ell}^i(x_0) - \Gamma_{km}^i(x_0)\Gamma_{j\ell}^m(x_0)]v_0^j \int_0^1 \frac{d\gamma^k(t)}{dt} \gamma^\ell(t) dt \\ &= [\Gamma_{kj,\ell}^i(x_0) - \Gamma_{km}^i(x_0)\Gamma_{j\ell}^m(x_0)]v_0^j A^{k\ell} \end{aligned} \quad (3.55)$$

The integral, which we denote by  $A^{k\ell}$ , is recognized to be the area of the projection of the curve  $\gamma(t)$  on the plane  $k$ - $\ell$ ; cf. (3.43).

We can give (3.55) a more symmetrical form. First, since  $k$  and  $\ell$  are dummy indices, we may interchange them, without affecting the result. Therefore, (3.55) is equal to

$$v_2^i(1) - v_2^i(0) = [\Gamma_{\ell j,k}^i(x_0) - \Gamma_{\ell m}^i(x_0)\Gamma_{jk}^m(x_0)]v_0^j A^{\ell k} \quad (3.56)$$

However,  $A^{\ell k}$  is easily seen to be antisymmetric, therefore

$$v_2^i(1) - v_2^i(0) = -[\Gamma_{\ell j,k}^i(x_0) - \Gamma_{\ell m}^i(x_0)\Gamma_{jk}^m(x_0)]v_0^j A^{k\ell} \quad (3.57)$$

Now add (3.55) and (3.57) and divide by 2 to obtain

$$\begin{aligned} v_2^i(1) - v_2^i(0) &= \\ 0.5[\Gamma_{kj,\ell}^i(x_0) + \Gamma_{\ell m}^i(x_0)\Gamma_{jk}^m(x_0) - \Gamma_{\ell j,k}^i(x_0) - \Gamma_{km}^i(x_0)\Gamma_{j\ell}^m(x_0)]v_0^j A^{k\ell} \end{aligned} \quad (3.58)$$

We now recall the definition of the Riemann curvature tensor (2.40) and recognize the quantity in the brackets in (3.58) as  $R_{k\ell j}^i$ . Therefore, finally,

$$v^i(1) - v^i(0) \approx 0.5\varepsilon^2 R_{k\ell j}^i A^{k\ell} v_0^j \quad (3.59)$$

We arrived at the result promised in the beginning of this section and, at the same time, established an interesting interpretation of the Riemann curvature tensor. When a vector is parallel transported along an infinitesimal closed curve in a manifold, there is a second-order, nonzero difference between the vectors at the end of the curve and at the beginning of the curve. This difference is proportional to the Riemann tensor at the central point of the loop and also to the area of the loop, as expressed by  $\varepsilon^2 A^{k\ell}$ . Another way of expressing this result is: When a given vector is parallel transported from a point  $p$  to a point  $q$  on a manifold, the resulting vector is not uniquely determined by the points  $p$  and  $q$ , but depends on the chosen path between the points.

## 3.7 Geodesics and Line Length

### 3.7.1 Geodesics

A *geodesic* in a manifold is a curve having the property that its tangent vector is parallel transported along the curve. Returning to (3.37), the differential equation for the coordinates of a parallel transported vector, and substituting the tangent vector in place of the transported vector, yields the equation

$$\frac{d^2 x^i}{dt^2} + \Gamma_{km}^i \frac{dx^k}{dt} \frac{dx^m}{dt} = 0 \quad \text{for all } t \text{ and all } i \quad (3.60)$$

This is known as the *geodesic equation*. It is a set of  $n$  second-order, nonlinear, coupled differential equations in the unknown functions  $x^i(t)$ , which are the coordinates of the geodesic. One must remember that the affine connections  $\Gamma_{km}^i$  are not constant coefficients but functions of the  $x^i(t)$ , because the affine connections are not constant in general on a curved space.

Assuming that the metric  $g_{ij}$  is smooth on the manifold, the affine connections are smooth. Then, given  $x^i(0)$  and  $dx^i(0)/dt$  at point  $p$  on the manifold,

the geodesic equation has a unique solution. This solution can be found analytically in very simple cases, or numerically when no analytic solution exists.

### 3.7.2 Length in a Euclidean Spaces

In an  $n$ -dimensional Euclidean space, the length of a curve  $\gamma(t)$  having cartesian coordinates  $x^i(t)$ ,  $0 \leq t \leq t_f$  is given by the integral

$$S = \int_0^{t_f} \left[ \sum_{k=1}^n (dx^k)^2 \right]^{1/2} = \int_0^{t_f} \left[ \sum_{k=1}^n \left( \frac{dx^k(t)}{dt} \right)^2 \right]^{1/2} dt \quad (3.61)$$

It is well known that the curve of shortest length between two points is a straight line. By definition, a straight line has coordinates  $x^i(t) = a^i + b^i t$ , where  $a^i$ ,  $b^i$  are constants. We note that a straight line satisfies the geodesic equation (3.60), because the affine connections are identically zero on a flat space and  $d^2x^i(t)/dt^2 = 0$  for a straight line. So, the curve of shortest length between two points on a Euclidean space satisfies the geodesic equation. We wish to find out whether this result can be generalized to manifolds.

While we are still in a Euclidean space, let us introduce the concept of natural parameterization. First define the *partial length* of a curve as

$$s(t) = \int_0^t \left[ \sum_{k=1}^n \left( \frac{dx^k(u)}{du} \right)^2 \right]^{1/2} du \quad (3.62)$$

We then make a variable change

$$t = t(s) \quad (3.63)$$

and redefine the functions  $x^i(s)$  and  $dx^i(s)$  accordingly. We distinguish between functions of  $t$  and functions of  $s$  only by the argument in parentheses and do not assign different symbols to the functions themselves. This is obviously an abuse of notation, but is convenient and hopefully will not lead to confusion. Note also that (3.64) requires the inversion of the function  $s(t)$ . This may be computationally difficult but we are ignoring such difficulties in the present discussion.

The parameterization  $x^i(s)$  is called the *natural parameterization* of the curve. With the natural parameterization, (3.62) becomes a triviality

$$\int_0^s du = s \quad (3.64)$$

### 3.7.3 Length in a Manifold

Suppose that the metric is known to be positive in a region of the manifold or on the entire manifold; that is,  $g_{ij} \geq 0$ . We can then define the length of a curve by generalizing (3.61) to

$$S = \int_0^{t_f} [g_{ij}(x^k(t)) dx^i(t) dx^j(t)]^{1/2} = \int_0^{t_f} \left[ g_{ij}(x^k(t)) \frac{dx^i(t)}{dt} \frac{dx^j(t)}{dt} \right]^{1/2} dt \quad (3.65)$$

Note that (3.66) cannot be proved and must be taken as a definition. A similar definition can be made if the metric is known to be negative, by simply replacing  $g_{ij}$  by  $-g_{ij}$  in the definition<sup>4</sup>. Note also that the metric does not depend on  $t$  directly, but via its dependence on the coordinates  $x^i(t)$ .

As in the case of Euclidean spaces, we may define the partial length

$$s(t) = \int_0^t \left[ g_{ij}(x^k(u)) \frac{dx^i(u)}{du} \frac{dx^j(u)}{du} \right]^{1/2} du \quad (3.66)$$

Then we can define the natural parameterization in the same way as for Euclidean spaces, via the variable change (3.64).

Let us now pose the following question: Given two points on the manifold, what is the curve of minimum length connecting these points? A partial answer is given as follows: The curve of minimum length satisfies the geodesic equation, provided the curve is parameterized in the natural parameterization. Note that this condition is necessary, but not sufficient, for the curve to have minimum length.

The proof of the necessary condition is given in Appendix C; it is not difficult, but somewhat lengthy and may be skipped.

---

<sup>4</sup>This happens in relativity theory, in timelike regions of the space.

# Appendix A

## Dual Vector Spaces

The subject of dual vector spaces is not usually covered in linear algebra courses, but is necessary for tensors. We will therefore provide a brief introduction to this subject in this appendix.

Let  $\mathcal{V}$  be an  $n$ -dimensional vector space over  $\mathbb{R}$ . A function  $f : \mathcal{V} \rightarrow \mathbb{R}$  is called a *functional*. A functional is linear if it satisfies the usual linearity conditions, namely

$$f(\mathbf{x}_1 + \mathbf{x}_2) = f(\mathbf{x}_1) + f(\mathbf{x}_2), \quad f(a\mathbf{x}) = af(\mathbf{x}) \quad (\text{A.1})$$

The *zero functional* maps every  $\mathbf{x} \in \mathcal{V}$  to the number 0; this functional is obviously linear.

The sum of two linear functionals is a linear functional and the product of a linear functional by a scalar is a linear functional. Sums of linear functionals and products by scalars obey all the properties of vectors in a vector space. The proofs of these statements are straightforward. Therefore, the set of all linear functionals on  $\mathcal{V}$  is a vector space, called the *dual space* of  $\mathcal{V}$  and denoted by  $\mathcal{V}^*$ . The elements of  $\mathcal{V}^*$  are called *dual vectors* or *covectors*.

We will use bold font for dual vectors; usually there will be no confusion with vectors, but in case of ambiguity will mention explicitly whether the symbol stands for a vector or a dual vector. If  $\mathbf{y}$  is a dual vector and  $\mathbf{x}$  is a vector,

we will denote

$$\langle \mathbf{y}, \mathbf{x} \rangle = \mathbf{y}(\mathbf{x})$$

Let  $(\mathbf{e}_1, \dots, \mathbf{e}_n)$  be a basis for  $\mathcal{V}$ . We aim to prove that  $\mathcal{V}^*$  is also an  $n$ -dimensional vector space over  $\mathbb{R}$  and find a basis for this space. Let  $\mathbf{x}$  be some vector in  $\mathcal{V}$  and express  $\mathbf{x}$  in terms of the basis,  $\mathbf{x} = \sum_{i=1}^n x^i \mathbf{e}_i$ . Define the functional  $\mathbf{f}^i$  in  $\mathcal{V}^*$  as follows:

$$\langle \mathbf{f}^i, \mathbf{x} \rangle = x^i \tag{A.2}$$

Thus  $\mathbf{f}^i$  selects the  $i$ -th coordinate of  $\mathbf{x}$  when  $\mathbf{x}$  is expressed in terms of the basis.

**Theorem 8.** *The space  $\mathcal{V}^*$  is  $n$ -dimensional and the set  $(\mathbf{f}^1, \dots, \mathbf{f}^n)$  is a basis for  $\mathcal{V}^*$ .*

*Proof.* We first show that  $(\mathbf{f}^1, \dots, \mathbf{f}^n)$  are independent. Let  $\mathbf{g} = \sum_{i=1}^n a_i \mathbf{f}^i$  and assume that  $\mathbf{g}$  is the zero functional for some  $(a_1, \dots, a_n)$ . Applying  $\mathbf{g}$  to any vector  $\mathbf{f}$ , we must have  $\langle \mathbf{g}, \mathbf{x} \rangle = 0$ . Thus  $\sum_{i=1}^n a_i x^i = 0$  for all  $(x_1, \dots, x_n)$ , which immediately implies  $a^i = 0$  for all  $i$ .

Next we show that every linear functional can be expressed as a linear combination of  $(\mathbf{f}^1, \dots, \mathbf{f}^n)$ . Let  $\mathbf{g} \in \mathcal{V}^*$  and define the  $n$  scalars  $g_i = \langle \mathbf{g}, \mathbf{e}_i \rangle$ . We will prove that  $\mathbf{g}$  is given by  $\mathbf{g} = \sum_{i=1}^n g_i \mathbf{f}^i$ . For an arbitrary vector  $\mathbf{x}$ ,

$$\langle \mathbf{g}, \mathbf{x} \rangle = \left\langle \mathbf{g}, \sum_{i=1}^n \mathbf{e}_i x^i \right\rangle = \sum_{i=1}^n \langle \mathbf{g}, \mathbf{e}_i \rangle x^i = \sum_{i=1}^n g_i x^i = \sum_{i=1}^n g_i \langle \mathbf{f}^i, \mathbf{x} \rangle \tag{A.3}$$

Since this holds identically for all  $\mathbf{x}$ , it follows that  $\mathbf{g} = \sum_{i=1}^n g_i \mathbf{f}^i$  and the proof is complete.  $\square$

The basis  $(\mathbf{f}^1, \dots, \mathbf{f}^n)$  is called the *dual basis* of  $(\mathbf{e}_1, \dots, \mathbf{e}_n)$ . It is worthwhile noting that

$$\langle \mathbf{f}^i, \mathbf{e}_j \rangle = \delta_j^i \tag{A.4}$$

This is true because, when  $\mathbf{e}_j$  is expressed in terms of the basis  $(\mathbf{e}_1, \dots, \mathbf{e}_n)$ , its  $j$ -th coordinate is 1 and all other coordinates are 0.

# Appendix B

## Derivation of the Symmetries of the Covariant Riemann Tensor

In this appendix we derive an expression for the covariant Riemann curvature tensor, which makes the symmetries of this tensor transparent. First we define the lowered affine connection by

$$\Gamma_{ijk} = 0.5(g_{ij,k} + g_{ik,j} - g_{jk,i}) \quad (\text{B.1})$$

Compare with (2.20); as we see,  $\Gamma_{ijk}$  does not include the metric tensor factor that appears in  $\Gamma_{jk}^i$

The following identity is verified by direct substitution of (B.1) in the right side:

$$g_{ij,k} = \Gamma_{ijk} + \Gamma_{jik} \quad (\text{B.2})$$

To find the covariant Riemann tensor, we must compute each of the four terms in (2.40) and lower the contravariant index. Let us begin with the first term  $g_{iu}\Gamma_{hp,q}^u$ . We use the product rule for derivative and then substitute (B.2) to find

$$g_{iu}\Gamma_{hp,q}^u = (g_{iu}\Gamma_{hp}^u)_{,q} - g_{iu,q}\Gamma_{hp}^u = \Gamma_{ihp,q} - (\Gamma_{iuq} + \Gamma_{uiq})\Gamma_{hp}^u \quad (\text{B.3})$$

The second term of (2.40) is simpler:

$$g_{iu}\Gamma_{kq}^u\Gamma_{hp}^k = \Gamma_{ikq}\Gamma_{hp}^k = \Gamma_{iuq}\Gamma_{hp}^u \quad (\text{B.4})$$

Adding (B.3) and (B.4) gives

$$g_{iu}(\Gamma_{hp,q}^u + \Gamma_{kq}^u \Gamma_{hp}^k) = \Gamma_{ihp,q} - \Gamma_{uiq} \Gamma_{hp}^u = \Gamma_{ihp,q} - g^{ut} \Gamma_{uiq} \Gamma_{thp} \quad (\text{B.5})$$

The sum of the third and fourth terms of (2.40) is obtained from (B.5) upon interchanging  $p$  and  $q$ :

$$g_{iu}(\Gamma_{hq,p}^u + \Gamma_{kp}^u \Gamma_{hq}^k) = \Gamma_{ihq,p} - g^{ut} \Gamma_{uip} \Gamma_{thq} \quad (\text{B.6})$$

Since the indices  $u$  and  $t$  in (B.6) are dummy (they are summation indices) and since  $g^{ut}$  is symmetric in  $u$  and  $t$ , we may interchange them and rewrite (B.6) as

$$g_{iu}(\Gamma_{hq,p}^u + \Gamma_{kp}^u \Gamma_{hq}^k) = \Gamma_{ihq,p} - g^{ut} \Gamma_{tip} \Gamma_{uhq} \quad (\text{B.7})$$

Adding (B.5) and (B.7) gives the desired formula for  $R_{ihqp}$ :

$$\begin{aligned} R_{ihqp} &= g_{iu} R_{hqp}^i = g_{iu}(\Gamma_{hp,q}^u + \Gamma_{kq}^u \Gamma_{hp}^k) - g_{iu}(\Gamma_{hq,p}^u + \Gamma_{kp}^u \Gamma_{hq}^k) \\ &= (\Gamma_{ihp,q} - \Gamma_{ihq,p}) + g^{ut}(\Gamma_{tip} \Gamma_{uhq} - \Gamma_{thp} \Gamma_{uiq}) \end{aligned} \quad (\text{B.8})$$

It is easy to see that the second term on the right side of (B.8) is anti-symmetric in  $i$  and  $u$ , and it remains to check the first term. We have

$$\begin{aligned} (\Gamma_{ihp,q} - \Gamma_{ihq,p}) &= 0.5(g_{ih,pq} + g_{ip,hq} - g_{hp,iq}) - 0.5(g_{ih,qp} + g_{iq,hp} - g_{hq,ip}) \\ &= 0.5(g_{ip,hq} - g_{hp,iq}) - 0.5(g_{iq,hp} - g_{hq,ip}) \end{aligned} \quad (\text{B.9})$$

The anti-symmetry of  $(\Gamma_{ihp,q} - \Gamma_{ihq,p})$  in  $i$  and  $u$  is now obvious, so in summary

$$R_{hiqp} = -R_{ihqp} \quad (\text{B.10})$$

The covariant Riemann tensor exhibits the symmetry

$$R_{ihqp} = R_{qpih} \quad (\text{B.11})$$

This symmetry can also be read from (B.8) and (B.9).

# Appendix C

## Proof that the Curve of Minimum Length Satisfies the Geodesic Equation

Let  $S$  be the length of the minimum-length curve, as expressed by `mlength4`. Let us move from the curve  $x(t)$  of minimum length to a neighboring curve  $x(t) + \varepsilon y(t)$ , where  $\varepsilon$  is a small real number and  $y(t)$  is smooth and satisfies  $y(0) = y(t_f) = 0$ , but is otherwise arbitrary. We will derive an expression for the length of the new curve. We emphasize that  $t$  is the argument of the natural parameterization of the minimum-length curve, but not that of  $y(t)$  in general.

The length that we wish to calculate is given by

$$\int_0^{t_f} [F(x(t) + \varepsilon y(t), \dot{x}(t) + \varepsilon \dot{y}(t))]^{1/2} dt \quad (\text{C.1})$$

The scalar function  $F$  will result from a  $(2, 2)$ -tensor  $F_{pq}^{ij}$  upon performing the contraction  $F_{ij}^{ij}$ . For the sake of clarity, let us keep the  $(2, 2)$ -tensor for the time being. Also, to make the derivation easier to typeset and read, we use the shorthand notation  $\dot{x}^i(t) = dx^i(t)/dt$ . So, we define

$$F_{pq}^{ij}(x(t), \dot{x}(t)) = g_{pq}(x(t)) \dot{x}^i(t) \dot{x}^j(t) \quad (\text{C.2})$$

and then

$$\begin{aligned}
& F_{pq}^{ij}(x(t) + \varepsilon y(t), \dot{x}(t) + \varepsilon \dot{y}(t)) = \\
& g_{pq}(x(t) + \varepsilon y(t))[\dot{x}^i(t) + \varepsilon \dot{y}^j(t)][\dot{x}^j(t) + \varepsilon \dot{y}^j(t)] = \\
& F_{pq}^{ij}(x(t), \dot{x}(t)) + \varepsilon \Delta F_{pq}^{ij}(x(t), \dot{x}(t)) + O(\varepsilon^2)
\end{aligned} \tag{C.3}$$

where  $\Delta F_{pq}^{ij}(x(t), \dot{x}(t))$  is the linear term in  $\varepsilon$  and  $O(\varepsilon^2)$  is a remainder term; that is, a term bounded in magnitude by some multiple of  $\varepsilon^2$ .

To compute an explicit expression for  $\Delta F_{pq}^{ij}(x(t), \dot{x}(t))$ , let us first approximate  $g_{pq}$  up to first order in  $\varepsilon$ :

$$g_{pq}(x(t) + \varepsilon y(t)) = g_{pq}(x(t)) + \varepsilon g_{pq,k}(x(t))y^k(t) + O(\varepsilon^2) \tag{C.4}$$

Therefore,

$$\begin{aligned}
\Delta F_{pq}^{ij}(x(t), \dot{x}(t)) &= g_{pq,k}(x(t))y^k(t)\dot{x}^i(t)\dot{x}^j(t) \\
&+ g_{pq}(x(t))\dot{x}^i(t)\dot{y}^j(t) + g_{pq}(x(t))\dot{x}^j(t)\dot{y}^i(t)
\end{aligned} \tag{C.5}$$

Consider the following identity:

$$\begin{aligned}
\frac{d}{dt}[g_{pq}(x(t))\dot{x}^i(t)y^j(t)] &= g_{pq,\ell}(x(t))\dot{x}^\ell(t)\dot{x}^i(t)y^j(t) \\
&+ g_{pq}(x(t))\ddot{x}^i(t)y^j(t) + g_{pq}(x(t))\dot{x}^i(t)\dot{y}^j(t)
\end{aligned} \tag{C.6}$$

from which we can write

$$\begin{aligned}
g_{pq}(x(t))\dot{x}^i(t)\dot{y}^j(t) &= \frac{d}{dt}[g_{pq}(x(t))\dot{x}^i(t)y^j(t)] \\
&- g_{pq,\ell}(x(t))\dot{x}^\ell(t)\dot{x}^i(t)y^j(t) - g_{pq}(x(t))\ddot{x}^i(t)y^j(t)
\end{aligned} \tag{C.7}$$

and similarly

$$\begin{aligned}
g_{pq}(x(t))\dot{x}^j(t)\dot{y}^i(t) &= \frac{d}{dt}[g_{pq}(x(t))\dot{x}^j(t)y^i(t)] \\
&- g_{pq,\ell}(x(t))\dot{x}^\ell(t)\dot{x}^j(t)y^i(t) - g_{pq}(x(t))\ddot{x}^j(t)y^i(t)
\end{aligned} \tag{C.8}$$

We can now substitute (C.7) and (C.8) in (C.5) and get

$$\begin{aligned}
\Delta F_{pq}^{ij}(x(t), \dot{x}(t)) &= g_{pq,k}(x(t))y^k(t)\dot{x}^i(t)\dot{x}^j(t) \\
&+ \frac{d}{dt}[g_{pq}(x(t))\dot{x}^i(t)y^j(t)] + \frac{d}{dt}[g_{pq}(x(t))\dot{x}^j(t)y^i(t)] \\
&- g_{pq,\ell}(x(t))\dot{x}^\ell(t)\dot{x}^i(t)y^j(t) - g_{pq,\ell}(x(t))\dot{x}^\ell(t)\dot{x}^j(t)y^i(t) \\
&- g_{pq}(x(t))\ddot{x}^i(t)y^j(t) - g_{pq}(x(t))\ddot{x}^j(t)y^i(t)
\end{aligned} \tag{C.9}$$

We are now in a position to perform the contraction from (2,2)-tensors to scalars:

$$F(x(t), \dot{x}(t)) = g_{ij}(x(t))\dot{x}^i(t)\dot{x}^j(t) = 1 \quad (\text{C.10})$$

and

$$\begin{aligned} \Delta F(x(t), \dot{x}(t)) &= g_{ij,k}(x(t))y^k(t)\dot{x}^i(t)\dot{x}^j(t) \\ &+ \frac{d}{dt}[g_{ij}(x(t))\dot{x}^i(t)y^j(t)] + \frac{d}{dt}[g_{ij}(x(t))\dot{x}^j(t)y^i(t)] \\ &- g_{ij,\ell}(x(t))\dot{x}^\ell(t)\dot{x}^i(t)y^j(t) - g_{ij,\ell}(x(t))\dot{x}^\ell(t)\dot{x}^j(t)y^i(t) \\ &- g_{ij}(x(t))\ddot{x}^i(t)y^j(t) - g_{ij}(x(t))\ddot{x}^j(t)y^i(t) \end{aligned} \quad (\text{C.11})$$

Taking advantage of the symmetry of  $g_{ij}$  and relabeling indices as necessary, we find

$$\begin{aligned} \Delta F(x(t), \dot{x}(t)) &= 2\frac{d}{dt}[g_{ij}(x(t))\dot{x}^i(t)y^j(t)] \\ &+ g_{ij,k}(x(t))y^k(t)\dot{x}^i(t)\dot{x}^j(t) - g_{ik,j}(x(t))\dot{x}^j(t)\dot{x}^i(t)y^k(t) \\ &- g_{kj,i}(x(t))\dot{x}^i(t)\dot{x}^j(t)y^k(t) - 2g_{ik}(x(t))\ddot{x}^i(t)y^k(t) \end{aligned} \quad (\text{C.12})$$

We now recall the formula (2.20) for the affine connection to express (C.12) as

$$\begin{aligned} \Delta F(x(t), \dot{x}(t)) &= 2\frac{d}{dt}[g_{ij}(x(t))\dot{x}^i(t)y^j(t)] \\ &- 2[g_{ik}\ddot{x}^i(t) + g_{k\ell}\Gamma_{ij}^\ell\dot{x}^i(t)\dot{x}^j(t)]y^k(t) \end{aligned} \quad (\text{C.13})$$

Adding the zero-order term and the first-order term and approximating the square-root up to first order gives

$$\begin{aligned} [F(x(t), \dot{x}(t)) + \varepsilon\Delta F(x(t), \dot{x}(t)) + O(\varepsilon^2)]^{1/2} &= \\ [1 + \varepsilon\Delta F(x(t), \dot{x}(t)) + O(\varepsilon^2)]^{1/2} &= 1 + 0.5\varepsilon\Delta F(x(t), \dot{x}(t)) + O(\varepsilon^2) \end{aligned} \quad (\text{C.14})$$

where the number 1 results from the fact that we are using the natural parameterization for the minimum-length solution.

We need one more simplifying result:

$$\int_0^{t_f} \frac{d}{dt}[g_{ij}(x(t))\dot{x}^i(t)y^j(t)]dt = g_{ij}(x(t))\dot{x}^i(t)y^j(t)\Big|_0^{t_f} = 0 \quad (\text{C.15})$$

which follows from the fact that  $y^j(0) = y^j(t_f) = 0$ .

We can now perform the integration:

$$\begin{aligned} & \int_0^{t_f} [F(x(t), \dot{x}(t)) + \varepsilon \Delta F(x(t), \dot{x}(t)) + O(\varepsilon^2)]^{1/2} dt = \\ & S - 2\varepsilon \int_0^{t_f} [g_{ik} \ddot{x}^i(t) + g_{kl} \Gamma_{ij}^\ell \dot{x}^i(t) \dot{x}^j(t)] y^k(t) dt + O(\varepsilon^2) \end{aligned} \quad (\text{C.16})$$

Now, finally, comes the main point of the preceding derivation. For small enough  $\varepsilon$ , only the first-order term determines whether the right side of (C.16) is less than, equal to, or greater than  $S$ . Since we are free to choose a positive or negative  $\varepsilon$ , we can force the first-order term to be positive or negative, unless the integral is identically zero. It follows that a necessary condition for  $S$  to be the minimum length is that

$$\int_0^{t_f} [g_{ik} \ddot{x}^i(t) + g_{kl} \Gamma_{ij}^\ell \dot{x}^i(t) \dot{x}^j(t)] y^k(t) dt = 0 \quad (\text{C.17})$$

But, since  $y^k(t)$  are arbitrary functions, the necessary condition is, in fact,

$$g_{ik} \ddot{x}^i(t) + g_{kl} \Gamma_{ij}^\ell \dot{x}^i(t) \dot{x}^j(t) = 0 \quad (\text{C.18})$$

It remains to bring (C.18) to the final desired form. Multiply by  $g^{mk}$  to obtain

$$g^{mk} g_{ik} \ddot{x}^i(t) + g^{mk} g_{kl} \Gamma_{ij}^\ell \dot{x}^i(t) \dot{x}^j(t) = 0 \quad (\text{C.19})$$

Now perform the contraction:

$$\ddot{x}^m(t) + \Gamma_{ij}^m \dot{x}^i(t) \dot{x}^j(t) = 0 \quad (\text{C.20})$$

Finally, revert from the dot notation to the  $d/dt$  notation:

$$\frac{d^2 x^m(t)}{dt^2} + \Gamma_{ij}^m \frac{dx^i(t)}{dt} \frac{dx^j(t)}{dt} = 0 \quad (\text{C.21})$$

Equation (C.21) is the geodesic equation.