

# Chapter 1

## Describing the Physical World: Vectors & Tensors

It is now well established that all matter consists of elementary particles<sup>1</sup> that interact through mutual attraction or repulsion. In our everyday life, however, we do not see the elemental nature of matter, we see continuous regions of material such as the air in a room, the water in a glass or the wood of a desk. Even in a gas, the number of individual molecules in macroscopic volumes is enormous  $\approx 10^{19} \text{ cm}^{-3}$ . Hence, when describing the behaviour of matter on scales of microns or above it is simply not practical to solve equations for each individual molecule. Instead, theoretical models are derived using average quantities that represent the state of the material.

In general, these average quantities will vary with position in space, but an important concept to bear in mind is that the fundamental behaviour of matter should not depend on the particular coordinate system chosen to represent the position. The consequences of this almost trivial observation are far-reaching and we shall find that it dictates the form of many of our governing equations. We shall always consider our space to be three-dimensional and Euclidean<sup>2</sup> and we describe position in space by a *position vector*,  $\mathbf{r}$ , which runs from a specific point, the origin, to our chosen location. The exact coordinate system chosen will depend on the problem under consideration; ideally it should make the problem as “easy” as possible.

### 1.1 Vectors

A vector is a geometric quantity that has a “magnitude and direction”. A more mathematically precise, but less intuitive, definition is that a vector is an element of a vector space. Many physical quantities are naturally described in terms of vectors, e.g. position, velocity and acceleration, force. The invariance of material behaviour under changes in coordinates means that if a vector represents a physical quantity then it must not vary if we change our coordinate system. Imagine drawing a line that connects two points in two-dimensional (Euclidean) plane, that line remains unchanged whether we describe it as “ $x$  units across and  $y$  units up from the origin” or “ $r$  units from the origin in the  $\theta$  direction”. Thus, a vector is an object that exists independent of any coordinate system, but if we wish to describe it we must choose a specific coordinate system and its representation in that coordinate system (its components) will depend on the specific coordinates chosen.

---

<sup>1</sup>The exact nature of the most basic unit is, of course, still debated, but the fundamental discrete nature of matter is not.

<sup>2</sup>We won't worry about relativistic effects at all.

### 1.1.1 Cartesian components and summation convention

The fact that we have a Euclidean space means that we can **always** choose a Cartesian coordinate system with fixed orthonormal base vectors,  $\mathbf{e}_1 = \mathbf{i}$ ,  $\mathbf{e}_2 = \mathbf{j}$  and  $\mathbf{e}_3 = \mathbf{k}$ . For a compact notation, it is much more convenient to use the numbered subscripts rather than different symbols to distinguish the base vectors. Any vector quantity  $\mathbf{a}$  can be written as a sum of its components in the direction of the base vectors

$$\mathbf{a} = a^1 \mathbf{e}_1 + a^2 \mathbf{e}_2 + a^3 \mathbf{e}_3; \quad (1.1)$$

and the vector  $\mathbf{a}$  can be represented via its components  $(a^1, a^2, a^3)$ ; and so,  $\mathbf{e}_1 = (1, 0, 0)$ ,  $\mathbf{e}_2 = (0, 1, 0)$  and  $\mathbf{e}_3 = (0, 0, 1)$ . We will often represent the components of vectors using an index, *i.e.*  $(a^1, a^2, a^3)$  is equivalent to  $a^I$ , where  $I \in \{1, 2, 3\}$ . In addition, we use the Einstein summation convention in which any index that appears twice represents a sum over all values of that index

$$\mathbf{a} = a^J \mathbf{e}_J = \sum_{J=1}^3 a^J \mathbf{e}_J. \quad (1.2)$$

Note that we can change the (dummy) summation index without affecting the result

$$a^J \mathbf{e}_J = \sum_{J=1}^3 a^J \mathbf{e}_J = \sum_{K=1}^3 a^K \mathbf{e}_K = a^K \mathbf{e}_K.$$

The summation is ambiguous if an index appears more than twice and such terms are not allowed. For clarity later, an upper case index is used for objects in a Cartesian (or, in fact, any orthonormal) coordinate system and, in general, we will insist that summation can only occur over a raised index and a lowered index for reasons that will hopefully become clear shortly.

It is important to recognise that the components of a vector  $a^I$  do not actually make sense unless we know the base vectors as well. In isolation the components give you distances but not direction, which is only half the story.

### 1.1.2 Curvilinear coordinate systems

For a complete theoretical development, we shall consider general coordinate systems<sup>3</sup>. Unfortunately the use of general coordinate systems introduces considerable complexity because the lines on which individual coordinates are constant are not necessarily straight lines nor are they necessarily orthogonal to one another. A consequence is that the base vectors in general coordinate systems are not orthonormal and vary as functions of position.

### 1.1.3 Tangent (covariant base) vectors

The position vector  $\mathbf{r} = x^K \mathbf{e}_K$ , is a function of the Cartesian coordinates,  $x^K$ , where  $x^1 = x$ ,  $x^2 = y$  and  $x^3 = z$ . Note that the Cartesian base vectors can be recovered by differentiating the position vector with respect to the appropriate coordinate

$$\frac{\partial \mathbf{r}}{\partial x^K} = \mathbf{e}_K. \quad (1.3)$$

---

<sup>3</sup>For our purposes a coordinate system to be a set of independent scalar variables that can be used to describe any position in the entire Euclidean space.

In other words the derivative of position with respect to a coordinate returns a vector tangent to the coordinate direction; a statement that is true for any coordinate system. For a general coordinate system,  $\xi^i$ , we can write the position vector as

$$\mathbf{r}(\xi^1, \xi^2, \xi^3) = x^K(\xi^i) \mathbf{e}_K, \quad (1.4)$$

because the Cartesian base vectors are fixed. Here the notation  $x^K(\xi^i)$  means that the Cartesian coordinates can be written as functions of the general coordinates, e.g. in plane polars  $x(r, \theta) = r \cos \theta$ ,  $y(r, \theta) = r \sin \theta$ , see Example 1.1. Note that equation (1.4) is the first time in which we use upper and lower case indices to distinguish between the Cartesian and general coordinate systems. A tangent vector in the  $\xi^1$  direction,  $\mathbf{t}_1$ , is the difference between two position vectors associated with a small (infinitesimal) change in the  $\xi^1$  coordinate

$$\mathbf{t}_1(\xi^i) = \mathbf{r}(\xi^i + d\xi^1) - \mathbf{r}(\xi^i), \quad (1.5)$$

where  $d\xi^1$  represents the small change in the  $\xi^1$  coordinate direction, see Figure 1.1.

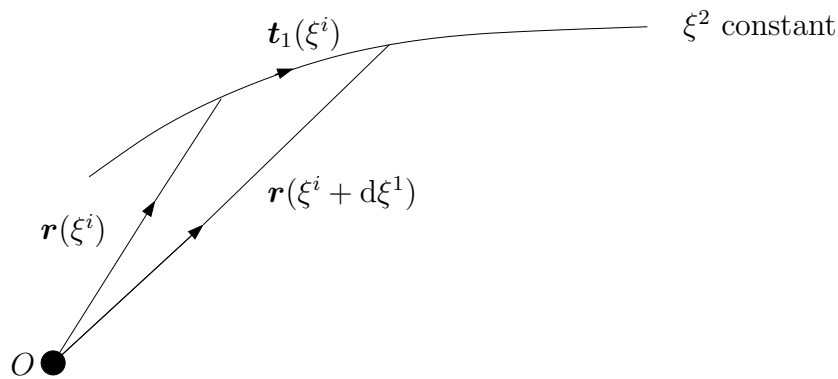


Figure 1.1: Sketch illustrating the tangent vector  $\mathbf{t}_1(\xi)$  corresponding to a small change  $d\xi^1$  in the coordinate  $\xi^1$ . The tangent lies along a line of constant  $\xi^2$  in two dimensions, or a plane of constant  $\xi^2$  and  $\xi^3$  in three dimensions.

Assuming that  $\mathbf{r}$  is differentiable and Taylor expanding the first term in (1.5) demonstrates that

$$\mathbf{t}_1 = \mathbf{r}(\xi^i) + \frac{\partial \mathbf{r}}{\partial \xi^1} d\xi^1 - \mathbf{r}(\xi^i) + \mathcal{O}((d\xi^i)^2),$$

which yields

$$\mathbf{t}_1 = \frac{\partial \mathbf{r}}{\partial \xi^1} d\xi^1,$$

if we neglect the (small) quadratic and higher-order terms. Note that exactly the same argument can be applied to increments in the  $\xi^2$  and  $\xi^3$  directions and because  $d\xi^i$  are scalar lengths, it follows that

$$\mathbf{g}_i = \frac{\partial \mathbf{r}}{\partial \xi^i}$$

is also a tangent vector in the  $\xi^i$  direction, as claimed. Hence, using equation (1.4), we can compute tangent vectors in the general coordinate directions via

$$\mathbf{g}_i = \frac{\partial \mathbf{r}}{\partial \xi^i} = \frac{\partial x^K}{\partial \xi^i} \mathbf{e}_K. \quad (1.6)$$

We can interpret equation (1.6) as defining a local linear transformation between the Cartesian base vectors and our new tangent vectors  $\mathbf{g}_i$ . The transformation is linear because  $\mathbf{g}_i$  is a linear combination of the vectors  $\mathbf{e}_K$ , which should not be a surprise because we explicitly neglected the quadratic and higher terms in the Taylor expansion. The transformation is local because, in general, the coefficients will change with position. The coefficients of the transformation can be written as entries in a matrix,  $\mathbf{M}$ , in which case equation (1.6) becomes

$$\begin{pmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \\ \mathbf{g}_3 \end{pmatrix} = \overbrace{\begin{pmatrix} \frac{\partial x^1}{\partial \xi^1} & \frac{\partial x^2}{\partial \xi^1} & \frac{\partial x^3}{\partial \xi^1} \\ \frac{\partial x^1}{\partial \xi^2} & \frac{\partial x^2}{\partial \xi^2} & \frac{\partial x^3}{\partial \xi^2} \\ \frac{\partial x^1}{\partial \xi^3} & \frac{\partial x^2}{\partial \xi^3} & \frac{\partial x^3}{\partial \xi^3} \end{pmatrix}}^{\mathbf{M}} \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \mathbf{e}_3 \end{pmatrix}. \quad (1.7)$$

Provided that the transformation is non-singular (the determinant of the matrix  $\mathbf{M}$  is non-zero) the tangent vectors will also be a basis of the space and they are called *covariant* base vectors because the transformation preserves the tangency of the vectors to the coordinates. In general, the covariant base vectors are neither orthogonal nor of unit length. It is also important to note that the covariant base vectors will usually be functions of position.

**Example 1.1. Finding the covariant base vectors for plane polar coordinates**

A plane polar coordinate system is defined by the two coordinates  $\xi^1 = r$ ,  $\xi^2 = \theta$  such that

$$x = x^1 = r \cos \theta \quad \text{and} \quad y = x^2 = r \sin \theta.$$

Find the covariant base vectors.

**Solution 1.1.** The position vector is given by

$$\mathbf{r} = x^1 \mathbf{e}_1 + x^2 \mathbf{e}_2 = r \cos \theta \mathbf{e}_1 + r \sin \theta \mathbf{e}_2 = \xi^1 \cos \xi^2 \mathbf{e}_1 + \xi^1 \sin \xi^2 \mathbf{e}_2,$$

and using the definition (1.6) gives

$$\mathbf{g}_1 = \frac{\partial \mathbf{r}}{\partial \xi^1} = \cos \xi^2 \mathbf{e}_1 + \sin \xi^2 \mathbf{e}_2, \quad \text{and} \quad \mathbf{g}_2 = \frac{\partial \mathbf{r}}{\partial \xi^2} = -\xi^1 \sin \xi^2 \mathbf{e}_1 + \xi^1 \cos \xi^2 \mathbf{e}_2.$$

Note that  $\mathbf{g}_1$  is a unit vector,

$$|\mathbf{g}_1| = \sqrt{\mathbf{g}_1 \cdot \mathbf{g}_1} = \sqrt{\cos^2 \xi^2 + \sin^2 \xi^2} = 1,$$

but  $\mathbf{g}_2$  is not,  $|\mathbf{g}_2| = \xi^1$ . The vectors are orthogonal  $\mathbf{g}_1 \cdot \mathbf{g}_2 = 0$  and can be related to the standard orthonormal polar base vectors via  $\mathbf{g}_1 = \mathbf{e}_r$  and  $\mathbf{g}_2 = r \mathbf{e}_\theta$ .

### 1.1.4 Contravariant base vectors

The fact that the covariant basis is *not* necessarily orthonormal makes life somewhat awkward. For orthonormal systems we are used to the fact that when  $\mathbf{a} = a^K \mathbf{e}_K$ , then unique components can be obtained via a dot product<sup>4</sup>.

$$\mathbf{a} \cdot \mathbf{e}_I = a^K \mathbf{e}_K \cdot \mathbf{e}_I = a^I, \quad (1.8)$$

---

<sup>4</sup>The dot or scalar product is an operation on two vectors that returns a unique scalar: the product of the lengths of the two vectors and the cosine of the angle between them. In the present context we only need to know that for orthogonal vectors the dot product is zero and the dot product of a unit vector with itself is one, see §1.3.1 for further discussion.

where the last equality is a consequence of the orthonormality. In our covariant basis, we would write  $\mathbf{a} = a^k \mathbf{g}_k$ , so that

$$\mathbf{a} \cdot \mathbf{g}_i = a^k \mathbf{g}_k \cdot \mathbf{g}_i, \quad (1.9)$$

but no further simplification can be made. Equations (1.9) are a linear system of simultaneous equations that must be solved in order to find the values of  $a^k$ , which is considerably more effort than using an explicit formula such as (1.8). The explicit formula (1.8) arises because the matrix  $\mathbf{e}_K \cdot \mathbf{e}_I$  is diagonal, which means that the equations decouple and are no longer simultaneous.

We can, however, recover most of the nice properties of an orthonormal coordinate system if we define another set of base vectors that are each orthogonal to two of the covariant base vectors and have unit length when projected in the direction of the remaining covariant base vector. In other words, the new vectors  $\mathbf{g}^i$  are **defined** such that

$$\mathbf{g}^i \cdot \mathbf{g}_j = \delta_j^i \equiv \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{otherwise,} \end{cases} \quad (1.10)$$

where the object  $\delta_j^i$  is known as the Kronecker delta. In orthonormal coordinate systems the two sets of base vectors coincide; for example, in our global Cartesian coordinates  $\mathbf{e}^I \equiv \mathbf{e}_I$ .

We can decompose  $\mathbf{g}^i$  into its components in the Cartesian basis  $\mathbf{g}^i = g_K^i \mathbf{e}^K$ , where we have used the raised index on the base vectors for consistency with our summation convention. Note that  $g_K^i$  is thus defined to be the  $K$ -th Cartesian component of the  $i$ -th contravariant base vector. From the definition (1.10) and (1.6)

$$(g_K^i \mathbf{e}^K) \cdot \left( \frac{\partial x^L}{\partial \xi^j} \mathbf{e}_L \right) = g_K^i \frac{\partial x^L}{\partial \xi^j} \mathbf{e}^K \cdot \mathbf{e}_L = g_K^i \frac{\partial x^L}{\partial \xi^j} \delta_L^K = g_L^i \frac{\partial x^L}{\partial \xi^j} = \delta_j^i. \quad (1.11)$$

Note that we have used the ‘‘index-switching’’ property of the Kronecker delta to write  $g_K^i \delta_L^K = g_L^i$ , which can be verified by writing out all terms explicitly.

Multiplying both side of equation (1.11) by  $\frac{\partial \xi^j}{\partial x^K}$  yields

$$g_L^i \frac{\partial x^L}{\partial \xi^j} \frac{\partial \xi^j}{\partial x^K} = \delta_j^i \frac{\partial \xi^j}{\partial x^K} = \frac{\partial \xi^i}{\partial x^K};$$

and from the chain rule

$$\frac{\partial x^L}{\partial \xi^j} \frac{\partial \xi^j}{\partial x^K} = \frac{\partial x^L}{\partial x^K} = \delta_K^L,$$

because the Cartesian coordinates are independent. Hence,

$$g_L^i \delta_K^L = g_K^i = \frac{\partial \xi^i}{\partial x^K},$$

and so the new set of base vectors are

$$\mathbf{g}^i = \frac{\partial \xi^i}{\partial x^K} \mathbf{e}^K. \quad (1.12)$$

The equation (1.12) defines a local linear transformation between the Cartesian base vectors and the vectors  $\mathbf{g}^i$ . In a matrix representation, equation (1.12) is

$$\begin{pmatrix} \mathbf{g}^1 \\ \mathbf{g}^2 \\ \mathbf{g}^3 \end{pmatrix} = \overbrace{\begin{pmatrix} \frac{\partial \xi^1}{\partial x^1} & \frac{\partial \xi^1}{\partial x^2} & \frac{\partial \xi^1}{\partial x^3} \\ \frac{\partial \xi^2}{\partial x^1} & \frac{\partial \xi^2}{\partial x^2} & \frac{\partial \xi^2}{\partial x^3} \\ \frac{\partial \xi^3}{\partial x^1} & \frac{\partial \xi^3}{\partial x^2} & \frac{\partial \xi^3}{\partial x^3} \end{pmatrix}}^{\mathbf{M}^{-T}} \begin{pmatrix} \mathbf{e}^1 \\ \mathbf{e}^2 \\ \mathbf{e}^3 \end{pmatrix}, \quad (1.13)$$

and we see that the new transformation is the inverse transpose<sup>5</sup> of the linear transformation that defines the covariant base vectors (1.6). For this reason, the vectors  $\mathbf{g}^i$  are called *contravariant* base vectors.

**Example 1.2. Finding the contravariant base vectors for plane polar coordinates**

For the plane polar coordinate system defined in Example 1.1, find the contravariant base vectors.

**Solution 1.2.** The contravariant base vectors are defined by equation (1.12) and in order to use that equation directly, we must express our polar coordinates as functions of the Cartesian coordinates

$$r = \xi^1 = \sqrt{x^1x^1 + x^2x^2}, \quad \text{and} \quad \tan \theta = \tan \xi^2 = \frac{x^2}{x^1},$$

and then we can compute

$$\frac{\partial \xi^1}{\partial x^1} = \cos \xi^2, \quad \frac{\partial \xi^1}{\partial x^2} = \sin \xi^2, \quad \frac{\partial \xi^2}{\partial x^1} = -\frac{\sin \xi^2}{\xi^1} \quad \text{and} \quad \frac{\partial \xi^2}{\partial x^2} = \frac{\cos \xi^2}{\xi^1}$$

Thus, using the transformation (1.13), we have

$$\mathbf{g}^1 = \frac{\partial \xi^1}{\partial x^1} \mathbf{e}^1 + \frac{\partial \xi^1}{\partial x^2} \mathbf{e}^2 = \cos \xi^2 \mathbf{e}_1 + \sin \xi^2 \mathbf{e}_2 = \mathbf{g}_1,$$

where we have used the fact that  $\mathbf{e}_I = \mathbf{e}^I$ , and also

$$\mathbf{g}^2 = \frac{\partial \xi^2}{\partial x^1} \mathbf{e}^1 + \frac{\partial \xi^2}{\partial x^2} \mathbf{e}^2 = -\frac{\sin \xi^2}{\xi^1} \mathbf{e}_1 + \frac{\cos \xi^2}{\xi^1} \mathbf{e}_2 = \frac{1}{(\xi^1)^2} \mathbf{g}_2.$$

We can now easily verify that  $\mathbf{g}^i \cdot \mathbf{g}_j = \delta_j^i$ .

An alternative (and often easier) approach is to find the contravariant base vectors by finding the inverse transpose of the matrix  $\mathbf{M}$  that defines the covariant base vectors and using equation (1.13).

**1.1.5 Components of vectors in covariant and contravariant bases**

We can find the components of a vector  $\mathbf{a}$  in the covariant basis by taking the dot product with the appropriate contravariant base vectors

$$\mathbf{a} = a^k \mathbf{g}_k, \quad \text{where} \quad a^i = \mathbf{a} \cdot \mathbf{g}^i \quad (= a^k \mathbf{g}_k \cdot \mathbf{g}^i = a^k \delta_k^i = a^i.) \quad (1.14)$$

Similarly components of the vector  $\mathbf{a}$  in the contravariant basis are given by taking the dot product with the appropriate covariant base vectors

$$\mathbf{a} = a_k \mathbf{g}^k, \quad \text{where} \quad a_i = \mathbf{a} \cdot \mathbf{g}_i \quad (= a_k \mathbf{g}^k \cdot \mathbf{g}_i = a_k \delta_i^k = a_i.) \quad (1.15)$$

---

<sup>5</sup>That the inverse matrix is given by

$$\mathbf{M}^{-1} = \begin{pmatrix} \frac{\partial \xi^1}{\partial x^1} & \frac{\partial \xi^2}{\partial x^1} & \frac{\partial \xi^3}{\partial x^1} \\ \frac{\partial \xi^1}{\partial x^2} & \frac{\partial \xi^2}{\partial x^2} & \frac{\partial \xi^3}{\partial x^2} \\ \frac{\partial \xi^1}{\partial x^3} & \frac{\partial \xi^2}{\partial x^3} & \frac{\partial \xi^3}{\partial x^3} \end{pmatrix},$$

can be confirmed by checking that  $\mathbf{M}\mathbf{M}^{-1} = \mathbf{M}^{-1}\mathbf{M} = \mathbf{I}$ , the identity matrix. Alternatively, the relationship follows directly from equation (1.11) written in matrix form.

In fact, we can obtain the components of a general vector in either the covariant or contravariant basis directly from the Cartesian coordinates. If  $\mathbf{a} = a^K \mathbf{e}_K = a_K \mathbf{e}^K$ , then the components in the covariant basis associated with the curvilinear coordinates  $\xi^i$  are

$$a^i = \mathbf{a} \cdot \mathbf{g}^i = a^K \mathbf{e}_K \cdot \mathbf{g}^i = a^K \frac{\partial \xi^i}{\partial x^J} \mathbf{e}^J \cdot \mathbf{e}_K = a^K \frac{\partial \xi^i}{\partial x^J} \delta_K^J = a^K \frac{\partial \xi^i}{\partial x^K},$$

a contravariant transform. Similarly, the components of the vector in the contravariant basis may be obtained by covariant transform from the Cartesian components and so

$$a_i = \frac{\partial x^K}{\partial \xi^i} a^K \tag{1.16a}$$

and

$$a^i = \frac{\partial \xi^i}{\partial x^K} a^K. \tag{1.16b}$$

### 1.1.6 Invariance of vectors: (significance of index position)

Having established the need for two different types of transformations in curvilinear coordinate systems, we are now in a position to consider the significance of the raised and lowered indices in our summation convention. We shall insist that for an index to be lowered the object must transform covariantly under a change in coordinates and for an index to be raised the object must transform contravariantly under a change in coordinates<sup>6</sup>. An important exception to this rule are the coordinates themselves:  $\xi^i$  represents the three scalar coordinates, e.g. in spherical polar coordinates  $\xi^1 = r$ ,  $\xi^2 = \theta$  and  $\xi^3 = \phi$ ;  $\xi^i$  are not the components of a vector and do not obey contravariant transformation rules. Equation (1.16a) demonstrates that the components of a vector in the contravariant basis are indeed covariant, justifying the lowered index, and equation (1.16b) provides similar justification for the contravariance of components in the covariant basis. We shall now demonstrate that these transformation properties also follow directly from the requirement that a physical vector should be independent of the coordinate system.

Consider a vector  $\mathbf{a}$ , which can be written in the covariant or contravariant basis

$$\mathbf{a} = a^i \mathbf{g}_i = a_i \mathbf{g}^i. \tag{1.17}$$

We now consider a change in coordinates from  $\xi^i$  to another general coordinate system  $\chi^{\bar{i}}$ . It will be of vital importance later on to know which index corresponds to which coordinate system so we have chosen to add an overbar to the index to distinguish components associated with the two coordinate systems,  $\xi^i$  and  $\chi^{\bar{i}}$ . The covariant base vectors associated with  $\chi^{\bar{i}}$  are then

$$\mathbf{g}_{\bar{i}} \equiv \frac{\partial \mathbf{r}}{\partial \chi^{\bar{i}}} = \frac{\partial x^J}{\partial \chi^{\bar{i}}} \mathbf{e}_J = \frac{\partial \xi^k}{\partial \chi^{\bar{i}}} \frac{\partial x^J}{\partial \xi^k} \mathbf{e}_J = \frac{\partial \xi^k}{\partial \chi^{\bar{i}}} \mathbf{g}_k; \tag{1.18}$$

and the transformation between  $\mathbf{g}_{\bar{i}}$  and  $\mathbf{g}_k$  is of the same (covariant) type as that between  $\mathbf{g}_i$  and  $\mathbf{e}_K$  in equation (1.6). The transformation is covariant because the “new” coordinate is the independent

---

<sup>6</sup>The logic for the choice of index location is the position of the generalised coordinate in the partial derivative defining the transformation:

$$\mathbf{g}_i = \frac{\partial x^K}{\partial \xi^i} \mathbf{e}_K \quad (\text{lowered index}), \quad \mathbf{g}^i = \frac{\partial \xi^i}{\partial x^K} \mathbf{e}_K \quad (\text{raised index}).$$

variable in the partial derivative (it appears in the denominator). In our new basis, the vector  $\mathbf{a} = a^{\bar{i}} \mathbf{g}_{\bar{i}}$  and because  $\mathbf{a}$  must remain invariant

$$\mathbf{a} = a^{\bar{i}} \mathbf{g}_{\bar{i}} = a^i \mathbf{g}_i.$$

Using the transformation of the base vectors (1.18) to replace  $\mathbf{g}_{\bar{i}}$  gives

$$a^{\bar{i}} \frac{\partial \xi^k}{\partial \chi^{\bar{i}}} \mathbf{g}_k = a^i \mathbf{g}_i = a^k \mathbf{g}_k \quad \Rightarrow \quad a^k = \frac{\partial \xi^k}{\partial \chi^{\bar{i}}} a^{\bar{i}}. \quad (1.19)$$

Hence, the components of the vector must transform contravariantly because multiplying both sides of equation (1.19) by the inverse transpose transformation  $\partial \chi^{\bar{j}} / \partial \xi^k$  gives

$$a^{\bar{i}} = \frac{\partial \chi^{\bar{i}}}{\partial \xi^k} a^k. \quad (1.20)$$

This transformation is contravariant because the “new” coordinate is the dependent variable in the partial derivative (it appears in the numerator).

A similar approach can be used to show that the components in the contravariant basis must transform covariantly in order to ensure that the vector remains invariant. Thus, the use of our summation convention ensures that the summed quantities remain invariant under coordinate transformations, which will be essential when deriving coordinate-independent physical laws.

## Interpretation

The fact that base vectors and vector components must transform differently for the vector to remain invariant is actually quite obvious. Consider a one-dimensional Euclidean space in which  $a = a^1 g_1$ . If the base vector is rescaled<sup>7</sup> by a factor  $\lambda$  so that  $g_{\bar{1}} = \lambda g_1$  then to compensate the component must be rescaled by the factor  $1/\lambda$ :  $a^{\bar{1}} = \frac{1}{\lambda} a^1$ . Note that for a  $1 \times 1$  transformation matrix with entry  $\lambda$ , the inverse transpose is  $1/\lambda$ .

### 1.1.7 Orthonormal coordinates

If the coordinates are orthonormal then, by construction, there is no distinction between the covariant and contravariant basis,  $\mathbf{g}_i = \mathbf{g}^i$ . Using equations (1.6) and (1.12), we see that

$$\mathbf{g}_i = \frac{\partial x^K}{\partial \xi^i} \mathbf{e}_K = \mathbf{g}^i = \frac{\partial \xi^i}{\partial x^K} \mathbf{e}_K,$$

and so

$$\frac{\partial x^K}{\partial \xi^i} = \frac{\partial \xi^i}{\partial x^K}. \quad (1.21)$$

Hence, the covariant and contravariant transformations are identical in orthonormal coordinate systems, which means that there is no need to distinguish between raised and lowered indices. This simplification is adopted in many textbooks and the convention is to use only lowered indices. When working with orthonormal coordinates we will also adopt this convention for simplicity, but we must always make sure that we know when the coordinate system is orthonormal. It is for this reason that we have adopted the convention that upper case indices are used for orthonormal coordinates.

---

<sup>7</sup>In one dimension all we can do is rescale the length, although the scaling can vary with position.



If the coordinate system is not known to be orthonormal, we will use lower case indices and must distinguish between the covariant and contravariant transformations.

Condition (1.21) implies that

$$\frac{\partial x^K}{\partial \xi^j} \frac{\partial x^K}{\partial \xi^i} = \frac{\partial x^K}{\partial \xi^j} \frac{\partial \xi^i}{\partial x^K} = \delta_j^i. \quad (1.22)$$

In the matrix representation, equation (1.22) is

$$\mathbf{M}\mathbf{M}^T = \mathbf{I} \quad \Rightarrow \quad \mathbf{M}^T\mathbf{M} = \mathbf{I},$$

where  $\mathbf{I}$  is the identity matrix. In other words the components of the transformation form an orthogonal matrix. It follows that (all) orthonormal coordinates can only be generated by an orthogonal transformation from the reference Cartesians. This should not be a big surprise: any other transform will change the angles between the base vectors or their relative lengths which destroys orthonormality. The argument is entirely reversible: if either the covariant or contravariant transform is orthogonal then the two transforms are identical and the new coordinate system is orthonormal.

### An aside

Further intuition for the reason why the covariant and contravariant transformations are identical when the coordinate transform is orthogonal can be obtained as follows. Imagine that we have a general linear transformation represented as a matrix  $\mathbf{M}$  the acts on vectors such that components in the fixed Cartesian coordinate system  $\mathbf{p} = p^K \mathbf{e}_K$  transform as follows

$$\tilde{p}^K = M_J^K p^J.$$

Note that the index  $K$  **does not** have an overbar because  $\tilde{p}^K$  is a component in the fixed Cartesian coordinate system,  $\mathbf{e}_K$ . The transformation can, of course, also be applied to the base vectors of the fixed Cartesian coordinate system  $\mathbf{e}_I$ ,

$$[\tilde{\mathbf{e}}_I]^K = M_J^K [\mathbf{e}_I]^J,$$

where  $[\ ]^K$  indicates the  $K$ -th component of the base vector. Now,  $[\mathbf{e}_I]^J = \delta_I^J$  and it follows that

$$[\tilde{\mathbf{e}}_I]^K = M_I^K,$$

which allows us to define the operation of the matrix components on the base vectors directly because

$$\tilde{\mathbf{e}}_I = [\tilde{\mathbf{e}}_I]^K \mathbf{e}_K = M_I^K \mathbf{e}_K. \quad (1.23a)$$

Thus the operation of the transformation on the components is the transpose of its operation on the base vectors<sup>8</sup>. We could write the new base vectors as  $\mathbf{e}_{\tilde{I}} = M_{\tilde{I}}^K \mathbf{e}_K$  to be consistent with our previous notation, but this will probably lead to more confusion in the current exposition.

Now consider a vector  $\mathbf{a}$  that must remain invariant under our transformation. Let the vector  $\tilde{\mathbf{a}}'$  be the vector with the same numerical values of its components as  $\mathbf{a}$  but with transformed base vectors, i.e.  $\tilde{\mathbf{a}}' = a^K \tilde{\mathbf{e}}_K$ . Thus, the vector  $\tilde{\mathbf{a}}'$  will be a transformed version of  $\mathbf{a}$ . In order to ensure that the vector remains unchanged under transformation we must apply the appropriate inverse

---

<sup>8</sup>This statement also applies to general bases.

transformation to  $\tilde{\mathbf{a}}'$  relative to the new base vectors,  $\tilde{\mathbf{e}}_I$ . In other words, the transformation of the coordinates must be

$$\tilde{a}^K = [M^{-1}]_J^K \tilde{a}'^J = [M^{-1}]_J^K a^J, \tag{1.23b}$$

where we have used the fact that  $\tilde{a}'^J = a^J$  by definition. Using the two transformation equations (1.23a,b) we see that

$$\tilde{a}^K \tilde{\mathbf{e}}_K = [M^{-1}]_J^K a^J M_K^L \mathbf{e}_L = [M^{-1}]_J^K M_K^L a^J \mathbf{e}_L = \delta_J^L a^J \mathbf{e}_L = a^J \mathbf{e}_J,$$

as required.

Thus, we have the two results: (i) a general property of linear transformations is that the matrix representation of the transformation of vector components is the transpose of the matrix representation of the transformation of base vectors; (ii) in order to remain invariant the transform of the components of the vector must actually undergo the inverse of the coordinate transformation. Thus, the transformations of the base vectors and the coordinates coincide when the inverse transform is equal to its transpose, *i.e.* when the transform is orthogonal.

If that all seems a bit abstract, then hopefully the following specific example will help make the ideas a little more concrete.

**Example 1.3. Equivalence of covariant and contravariant transformations under affine transformations**

Consider a two-dimensional Cartesian coordinate system with base vectors  $\mathbf{e}_I$ . A new coordinate system with base vectors  $\mathbf{e}_{\bar{I}}$  is obtained by rotation through an angle  $\theta$  in the anticlockwise direction about the origin. Derive the transformations for the base vectors and components of a general vector and show that they are the same.

**Solution 1.3.** The original and rotated bases are shown in Figure 1.2(a) from which we determine that the new base vectors are given by

$$\mathbf{e}_{\bar{1}} = \cos\theta \mathbf{e}_1 + \sin\theta \mathbf{e}_2 \quad \text{and} \quad \mathbf{e}_{\bar{2}} = -\sin\theta \mathbf{e}_1 + \cos\theta \mathbf{e}_2.$$

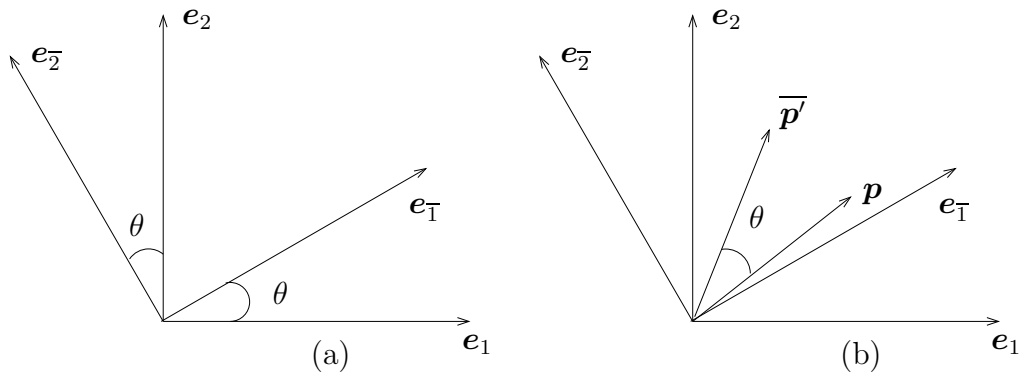


Figure 1.2: (a) The base vectors  $\mathbf{e}_{\bar{I}}$  are the Cartesian base vectors  $\mathbf{e}_I$  rotated through an angle  $\theta$  about the origin. (b) If the coordinates of the position vector  $\mathbf{p}$  are unchanged it is also rotated by  $\theta$  to  $\bar{\mathbf{p}}$ .

Consider a position vector  $\mathbf{p} = p^I \mathbf{e}_I$  in the original basis. If we leave the coordinates unchanged then the new vector  $\overline{\mathbf{p}} = p^{\overline{I}} \mathbf{e}_{\overline{I}}$  is the original vector rotated by  $\theta$ , see Figure 1.2(b).

We must therefore rotate the position vector  $\overline{\mathbf{p}}$  through an angle  $-\theta$  relative to the fixed basis  $\mathbf{e}_{\overline{I}}$ , but this is actually equivalent to a positive rotation of the base vectors. Hence the transforms for the components of vector and the base vectors are the same.

$$p^{\overline{1}} = \cos \theta p^1 + \sin \theta p^2 \quad \text{and} \quad p^{\overline{2}} = -\sin \theta p^1 + \cos \theta p^2.$$

## 1.2 Tensors

Tensors are geometric objects that have magnitude and zero, one or many associated directions, but are linear in character. A more mathematically precise definition is to say that a tensor is multilinear map or alternatively an element of a tensor product of vector spaces, which is somewhat tautological and really not helpful at this point. The order (or degree or rank) of a tensor is the number of associated directions and so a scalar is a tensor of order zero and a vector is a tensor of order one. Many quantities in continuum mechanics such as strain, stress, diffusivity and conductivity are naturally expressed as tensors of order two. We have already seen an example of a tensor in our discussion of vectors: linear transformations from one set of vectors to another, e.g. the transformation from Cartesian to covariant base vectors, are second-order tensors. If the vectors represent physical objects, then they must not depend on the coordinate representation chosen. Hence, the linear transformation must also be independent of coordinates because the same vectors must always transform in the same way. We can write our linear transformation in a coordinate-independent manner as

$$\mathbf{a} = \mathcal{M}(\mathbf{b}), \tag{1.24}$$

and the transformation  $\mathcal{M}$  is a tensor of order two. In order to describe  $\mathcal{M}$  precisely we must pick a specific coordinate system for each vector in equation (1.24). In the global Cartesian basis, equation (1.24) becomes

$$a_I \mathbf{e}_I = \mathcal{M}(b_J \mathbf{e}_J) = b_J \mathcal{M}(\mathbf{e}_J), \tag{1.25}$$

because it is a linear transformation. We now take the dot product with  $\mathbf{e}_K$  to obtain

$$a_I \mathbf{e}_K \cdot \mathbf{e}_I = b_J \mathbf{e}_K \cdot \mathcal{M}(\mathbf{e}_J) \quad \Rightarrow \quad a_K = b_J \mathbf{e}_K \cdot \mathcal{M}(\mathbf{e}_J),$$

where the dot product is written on the left to indicate that we are taking the dot product after the linear transformation has operated on the base vector  $\mathbf{e}_J$ . Hence, we can write the operation of the transformation on the components in the form

$$a_I = M_{IJ} b_J, \tag{1.26}$$

where  $M_{IJ} = \mathbf{e}_I \cdot \mathcal{M}(\mathbf{e}_J)$ . Equation (1.26) can be written in a matrix form to aid calculation

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}.$$

The quantity  $M_{IJ}$  represents the component of the transformed vectors in the  $I$ -th Cartesian direction if the original vector is of unit length in the  $J$ -th direction. Hence, the quantity  $M_{IJ}$  is meaningless without knowing the coordinate system associated with both  $I$  and  $J$ .

In fact, there is no need to choose the same coordinate system for  $I$  and  $J$ . If we write the vector  $\mathbf{a}$  in the covariant basis, equation (1.25) becomes

$$a^i \mathbf{g}_i = b_J \mathcal{M}(\mathbf{e}_J).$$

Taking the dot product with the appropriate contravariant base vector gives

$$a^k = b_J \mathbf{g}^k \cdot \mathcal{M}(\mathbf{e}_J) = b_J \frac{\partial \xi^k}{\partial x^K} \mathbf{e}_K \cdot \mathcal{M}(\mathbf{e}_J),$$

which means that

$$a^k = M_J^k b_J = \frac{\partial \xi^k}{\partial x^K} M_{KJ} b_J \quad \Rightarrow \quad M_J^k = \frac{\partial \xi^k}{\partial x^K} M_{KJ}.$$

In other words the components of each (column) vector corresponding to a fixed second index in a coordinate representation of  $\mathcal{M}$  must obey a contravariant transformation if the associated basis undergoes a covariant transform, *i.e.* the behaviour is exactly the same as for the components of a vector.

If we now also represent the vector  $\mathbf{b}$  in the covariant basis, equation (1.25) becomes

$$a^i \mathbf{g}_i = b^j \mathcal{M}(\mathbf{g}_j).$$

Taking the dot product with the appropriate contravariant base vector gives

$$a^k = b^j \mathbf{g}^k \cdot \mathcal{M}(\mathbf{g}_j) = b^j \frac{\partial \xi^k}{\partial x^K} \mathbf{e}_K \cdot \mathcal{M} \left( \frac{\partial x^J}{\partial \xi^j} \mathbf{e}_J \right) = b^j \frac{\partial \xi^k}{\partial x^K} \frac{\partial x^J}{\partial \xi^j} \mathbf{e}_K \cdot \mathcal{M}(\mathbf{e}_J),$$

on using the linearity of the transformation. Hence,

$$a^k = M_j^k b^j = \frac{\partial \xi^k}{\partial x^K} M_{KJ} \frac{\partial x^J}{\partial \xi^j} b^j \quad \Rightarrow \quad M_j^k = \frac{\partial \xi^k}{\partial x^K} M_{KJ} \frac{\partial x^J}{\partial \xi^j},$$

and the components of each (row) vector associated with a fixed first index in a coordinate representation of  $\mathcal{M}$  undergo a covariant transformation when the associated basis undergoes a covariant transform, *i.e.* the “opposite behaviour” to the components of a vector. The difference in behaviour between the two indices of the components of the linear transformation arises because one index corresponds to the basis of the “input” vector, whereas the other corresponds to the basis of the “output” vector. There is a sum over the second (input) index and the components of the vector  $\mathbf{b}$  and in order for this sum to remain invariant the transform associated with the second index must be the opposite to the components of the vector  $\mathbf{b}$ , in other words the same as the transformation of the base vectors of that vector.

The obvious relationships between components can easily be deduced when we represent our vectors in the contravariant basis,

$$a^i = M^{ij} b_j, \quad a_i = M_i^j b_j, \quad a_i = M_{ij} b^j. \quad (1.27)$$

Many books term  $M^{ij}$  a contravariant second-order tensor;  $M_{ik}$  a covariant second-order tensor and  $M_i^j$  a mixed second-order tensor, but they are simply representations of the *same* coordinate-independent object in different bases. Another more modern notation is to say that  $M^{ij}$  is a type (2, 0) tensor,  $M_{iJ}$  is type (0, 2) and  $M_j^i$  is a type (1, 1) tensor, which allows the distinction between mixed tensors or orders greater than two.

### 1.2.1 Invariance of second-order tensors

Let us now consider a general change of coordinates from  $\xi^i$  to  $\chi^{\bar{i}}$ . Given that

$$a^i = M_j^i b^j, \quad (1.28a)$$

we wish to find an expression for  $M_j^{\bar{i}}$  such that<sup>9</sup>

$$a^{\bar{i}} = M_j^{\bar{i}} b^{\bar{j}}. \quad (1.28b)$$

Using the transformation rules for the components of vectors (1.19) it follows that (1.28a) becomes

$$\frac{\partial \xi^i}{\partial \chi^{\bar{n}}} a^{\bar{n}} = M_j^i \frac{\partial \xi^j}{\partial \chi^{\bar{n}}} b^{\bar{n}}.$$

We now multiply both sides by  $\partial \chi^{\bar{m}} / \partial \xi^i$  to obtain

$$\frac{\partial \chi^{\bar{m}}}{\partial \xi^i} \frac{\partial \xi^i}{\partial \chi^{\bar{n}}} a^{\bar{n}} = \delta_n^{\bar{m}} a^{\bar{n}} = a^{\bar{m}} = \frac{\partial \chi^{\bar{m}}}{\partial \xi^i} M_j^i \frac{\partial \xi^j}{\partial \chi^{\bar{n}}} b^{\bar{n}}.$$

Comparing this expression to equation (1.28b) it follows that

$$M_j^{\bar{i}} = \frac{\partial \chi^{\bar{i}}}{\partial \xi^n} M_m^n \frac{\partial \xi^m}{\partial \chi^{\bar{j}}},$$

and thus we see that covariant components must transform covariantly and contravariant components must transform contravariantly in order for the invariance properties to hold. Similarly, it can be shown that

$$M^{\bar{i}\bar{j}} = \frac{\partial \chi^{\bar{i}}}{\partial \xi^n} \frac{\partial \chi^{\bar{j}}}{\partial \xi^m} M^{nm}, \quad \text{and} \quad M_{\bar{i}\bar{j}} = \frac{\partial \xi^n}{\partial \chi^{\bar{i}}} \frac{\partial \xi^m}{\partial \chi^{\bar{j}}} M_{nm}. \quad (1.29)$$

An alternative definition of tensors is to require that they are sets of index quantities (multi-dimensional arrays) that obey these transformation laws under a change of coordinates.

The transformations can be expressed in matrix form, but we must distinguish between the covariant and contravariant cases. We shall write  $\mathbf{M}^b$  to indicate a matrix where all components transform covariantly and  $\mathbf{M}^\sharp$  for the contravariant case. We define the transformation matrix  $\mathbf{F}$  to have the components

$$\mathbf{F} = \begin{pmatrix} \frac{\partial \chi^{\bar{1}}}{\partial \xi^1} & \frac{\partial \chi^{\bar{1}}}{\partial \xi^2} & \frac{\partial \chi^{\bar{1}}}{\partial \xi^3} \\ \frac{\partial \chi^{\bar{2}}}{\partial \xi^1} & \frac{\partial \chi^{\bar{2}}}{\partial \xi^2} & \frac{\partial \chi^{\bar{2}}}{\partial \xi^3} \\ \frac{\partial \chi^{\bar{3}}}{\partial \xi^1} & \frac{\partial \chi^{\bar{3}}}{\partial \xi^2} & \frac{\partial \chi^{\bar{3}}}{\partial \xi^3} \end{pmatrix}, \quad \text{or} \quad F_j^{\bar{i}} = \frac{\partial \chi^{\bar{i}}}{\partial \xi^j};$$

and then from the chain rule and independence of coordinates

$$\mathbf{F}^{-1} = \begin{pmatrix} \frac{\partial \xi^1}{\partial \chi^{\bar{1}}} & \frac{\partial \xi^1}{\partial \chi^{\bar{2}}} & \frac{\partial \xi^1}{\partial \chi^{\bar{3}}} \\ \frac{\partial \xi^2}{\partial \chi^{\bar{1}}} & \frac{\partial \xi^2}{\partial \chi^{\bar{2}}} & \frac{\partial \xi^2}{\partial \chi^{\bar{3}}} \\ \frac{\partial \xi^3}{\partial \chi^{\bar{1}}} & \frac{\partial \xi^3}{\partial \chi^{\bar{2}}} & \frac{\partial \xi^3}{\partial \chi^{\bar{3}}} \end{pmatrix}, \quad \text{or} \quad [F^{-1}]_j^{\bar{i}} = \frac{\partial \xi^i}{\partial \chi^{\bar{j}}}.$$

If  $\overline{\mathbf{M}}^\sharp$  and  $\overline{\mathbf{M}}^b$  represent the matrices of transformed components then the transformation laws (1.29) become

$$\overline{\mathbf{M}}^\sharp = \mathbf{F} \mathbf{M}^\sharp \mathbf{F}^T, \quad \text{and} \quad \overline{\mathbf{M}}^b = \mathbf{F}^{-T} \mathbf{M}^b \mathbf{F}^{-1}. \quad (1.30)$$

---

<sup>9</sup>This is a place where the use of overbars makes the notation look cluttered, but clarifies precisely which coordinate system is associated with each index. This notation also allows the representation of components in the two different coordinate systems, so-called two-point tensors, e. g.  $M_j^i$ , which will be useful.

## 1.2.2 Cartesian tensors

If we restrict attention to orthonormal coordinate systems, then the transformation between coordinate systems must be orthogonal<sup>10</sup> and we do not need to distinguish between covariant and contravariant behaviour. Consider the transformation from our Cartesian basis  $\mathbf{e}_I$  to another orthonormal basis  $\mathbf{e}_{\bar{I}}$ . The transformation rules for components of a tensor of order two become

$$M_{\bar{I}\bar{J}} = \frac{\partial x^N}{\partial x^{\bar{I}}} \frac{\partial x^M}{\partial x^{\bar{J}}} M_{NM}.$$

The transformation between components of two vectors in the different bases are given by

$$a_{\bar{I}} = \frac{\partial x^{\bar{I}}}{\partial x^K} a_K = \frac{\partial x^K}{\partial x^{\bar{I}}} a_K,$$

which can be written the form

$$a_{\bar{I}} = Q_{\bar{I}K} a_K, \quad \text{where} \quad Q_{\bar{I}K} = \frac{\partial x^{\bar{I}}}{\partial x^K} = \frac{\partial x^K}{\partial x^{\bar{I}}},$$

and the components  $Q_{\bar{I}K}$  form an orthogonal matrix. Hence the transformation property of a (Cartesian) tensor of order two can be written as

$$M_{\bar{I}\bar{J}} = Q_{\bar{I}N} M_{NM} Q_{\bar{J}M}$$

or in matrix form

$$\bar{\mathbf{M}} = \mathbf{Q} \mathbf{M} \mathbf{Q}^T. \tag{1.31}$$

In many textbooks, equation (1.31) is defined to be the transformation rule satisfied by a (Cartesian) tensor of order two.

## 1.2.3 Tensors vs matrices

There is a natural relationship between tensors and matrices because, as we have seen, we can write the components of a second-order tensor in a particular coordinate system as a matrix. It is often helpful to think of a tensor as a matrix when working with it, but the two concepts are distinct.

A summary of all the above is that a tensor is a geometric object that does not depend on any particular coordinate system and expresses a linear relationship between other geometric objects.

## 1.3 Products of vectors: scalar, vector and tensor

### 1.3.1 Scalar product

We have already used the scalar or dot product of two vectors, and the discussion here is included only for completeness. The scalar product is the product of two vectors that returns a unique scalar: the product of the lengths of the vectors and the cosine of the angle between them. Thus far, we have only used the dot product to define orthonormal sets of vectors.

If we represent two vectors  $\mathbf{a}$  and  $\mathbf{b}$  in the co- and contravariant bases, then

$$\mathbf{a} \cdot \mathbf{b} = (a_i \mathbf{g}^i) \cdot (b^j \mathbf{g}_j),$$

---

<sup>10</sup>Although the required orthogonal transformation may vary with position, as is the case in plane polar coordinates.

and so

$$\mathbf{a} \cdot \mathbf{b} = a_i b^j \mathbf{g}^i \cdot \mathbf{g}_j = a_i b^j \delta_j^i = a_i b^i.$$

An alternative decomposition demonstrates that

$$\mathbf{a} \cdot \mathbf{b} = a^i b_i,$$

and we note that the scalar product is invariant under coordinate transformation, as expected. In orthonormal coordinate systems, there is no distinction between co and contravariant bases and so  $\mathbf{a} \cdot \mathbf{b} = a_K b_K$ .

### 1.3.2 Vector product

The vector or cross product is a product of two vectors that returns a unique vector that is orthogonal to both vectors. In orthonormal coordinate systems, the vector product is defined by

$$\mathbf{a} \times \mathbf{b} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \times \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} a_2 b_3 - b_2 a_3 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{pmatrix}.$$

In order to represent the vector product with index notation it is convenient to define a quantity known as the alternating, or Levi-Civita, symbol  $e^{IJK}$ . In orthonormal coordinate systems the components of  $e^{IJK}$  are defined by

$$e^{IJK} = e_{IJK} = \begin{cases} 0 & \text{when any two indices are equal;} \\ +1 & \text{when } I, J, K \text{ is an even permutation of } 1, 2, 3; \\ -1 & \text{when } I, J, K \text{ is an odd permutation of } 1, 2, 3. \end{cases} \quad (1.32)$$

e.g.

$$e^{112} = e^{122} = 0, \quad e^{123} = e^{312} = e^{231} = 1, \quad e^{213} = e^{132} = e^{321} = -1.$$

Strictly speaking  $e^{IJK}$  thus defined is not a tensor because if the handedness of the coordinate system changes then the sign of the entries in  $e^{IJK}$  should change in order for it to respect the appropriate invariance properties; such objects are sometimes called pseudo-tensors. We could ensure that  $e^{IJK}$  is a tensor by restricting our definition to right-handed (or left-handed) orthonormal systems, which will be the approach taken in later chapters.

The vector product of two vectors  $\mathbf{a}$  and  $\mathbf{b}$  in orthonormal coordinates is

$$[\mathbf{a} \times \mathbf{b}]_I = e_{IJK} a_J b_K, \quad (1.33)$$

which can be confirmed by writing out all the components. In addition, the relationship between the Cartesian base vectors  $\mathbf{e}_I$  can be expressed as a vector product using the alternating tensor

$$\mathbf{e}_I \times \mathbf{e}_J = e_{IJK} \mathbf{e}_K. \quad (1.34)$$

Let us now consider the case of general coordinates: the cross product between covariant base vectors is given by

$$\mathbf{g}_i \times \mathbf{g}_j = \frac{\partial x^I}{\partial \xi^i} \mathbf{e}_I \times \frac{\partial x^J}{\partial \xi^j} \mathbf{e}_J = \frac{\partial x^I}{\partial \xi^i} \frac{\partial x^J}{\partial \xi^j} \mathbf{e}_I \times \mathbf{e}_J = \frac{\partial x^I}{\partial \xi^i} \frac{\partial x^J}{\partial \xi^j} e_{IJK} \mathbf{e}_K.$$

The expression on the right-hand side corresponds to the first two indices of the alternating tensor undergoing a covariant transformation so that

$$\mathbf{g}_i \times \mathbf{g}_j = \epsilon_{ijk} \mathbf{e}_K.$$

If we now transform the third index covariantly we must transform the base vector contravariantly so that

$$\mathbf{g}_i \times \mathbf{g}_j = \epsilon_{ijk} \frac{\partial \xi^k}{\partial x^K} \mathbf{e}_K = \epsilon_{ijk} \mathbf{g}^k; \quad (1.35)$$

where  $\epsilon_{ijk} \equiv \frac{\partial x^I}{\partial \xi^i} \frac{\partial x^J}{\partial \xi^j} \frac{\partial x^K}{\partial \xi^k} e_{IJK}$ . A similar argument shows that

$$\mathbf{g}^i \times \mathbf{g}^j = \epsilon^{ijk} \mathbf{g}_k,$$

where  $\epsilon^{ijk} \equiv \frac{\partial \xi^i}{\partial x^I} \frac{\partial \xi^j}{\partial x^J} \frac{\partial \xi^k}{\partial x^K} e^{IJK}$ .

If we decompose the vectors  $\mathbf{a}$  and  $\mathbf{b}$  into the contravariant basis we have

$$\mathbf{a} \times \mathbf{b} = (a_i \mathbf{g}^i) \times (b_j \mathbf{g}^j) = a_i b_j \mathbf{g}^i \times \mathbf{g}^j = a_i b_j \epsilon^{ijk} \mathbf{g}_k.$$

Thus, if we decompose the vector product into the covariant basis we have the following expression for the components

$$[\mathbf{a} \times \mathbf{b}]^k = \epsilon^{ijk} a_i b_j, \quad \text{or} \quad [\mathbf{a} \times \mathbf{b}]^i = \epsilon^{ijk} a_j b_k.$$

### 1.3.3 Tensor product

The tensor product is a product of two vectors that returns a second-order tensor. It can be motivated by the following discussion. Recall that equation (1.24) can be written in the form

$$a_i = M_{ij} b^j, \quad \text{where} \quad M_{ij} = \mathbf{g}_i \cdot \mathcal{M}(\mathbf{g}_j).$$

The components  $M_{ij}$  correspond to the representation of tensor with respect to a basis, but which basis? We shall **define** the basis to be that formed from the tensor product of pairs of base vectors:  $\mathbf{g}^i \otimes \mathbf{g}^j$ , where the symbol  $\otimes$  is used to denote the tensor product. Hence, we can represent a tensor in the different forms

$$\mathcal{M} = M_{ij} \mathbf{g}^i \otimes \mathbf{g}^j = M_{IJ} \mathbf{e}_I \otimes \mathbf{e}_J = M_{ij}^i \mathbf{g}_i \otimes \mathbf{g}^j \dots,$$

which is analogous to representing vectors in the different forms

$$\mathbf{a} = a_I \mathbf{e}_I = a_i \mathbf{g}^i = a^i \mathbf{g}_i.$$

Returning to equation (1.24), we have

$$\mathbf{a} = \mathcal{M}(\mathbf{b}) \quad \Rightarrow \quad \mathbf{a} = (M_{ij} \mathbf{g}^i \otimes \mathbf{g}^j)(\mathbf{b}),$$

and because  $M_{ij}$  are just coefficients it follows that  $\mathbf{g}^i \otimes \mathbf{g}^j$  are themselves tensors of second order<sup>11</sup>. Decomposing  $\mathbf{a}$  and  $\mathbf{b}$  into the contravariant and covariant bases respectively gives

$$a_i \mathbf{g}^i = (M_{ij} \mathbf{g}^i \otimes \mathbf{g}^j)(b^n \mathbf{g}_n) = M_{ij} b^n (\mathbf{g}^i \otimes \mathbf{g}^j)(\mathbf{g}_n), \quad (1.36)$$

---

<sup>11</sup>You should think carefully to convince yourself that this is true.



but from equation (1.27)

$$a_i = M_{ij} b^j \quad \Rightarrow \quad a_i \mathbf{g}^i = M_{ij} b^j \mathbf{g}^i = M_{ij} b^n \delta_n^j \mathbf{g}^i. \quad (1.37)$$

Equating (1.36) and (1.37), it follows that the tensor product acting on the base vectors is

$$(\mathbf{g}^i \otimes \mathbf{g}^j)(\mathbf{g}_n) = \delta_n^j \mathbf{g}^i = (\mathbf{g}^j \cdot \mathbf{g}_n) \mathbf{g}^i. \quad (1.38)$$

Equation (1.38) motivates an alternative definition of the tensor product, which in the case of the product of two vectors is also called the dyadic or outer product. The dyadic product of two vectors  $\mathbf{a}$  and  $\mathbf{b}$  is a second-order tensor that is defined through its action on an arbitrary vector  $\mathbf{v}$

$$(\mathbf{a} \otimes \mathbf{b})(\mathbf{v}) = (\mathbf{b} \cdot \mathbf{v})\mathbf{a}. \quad (1.39)$$

Consider the tensor  $\mathcal{T} = \mathbf{a} \otimes \mathbf{b}$ ; the components of that tensor are given by

$$T_{ij} = \mathbf{g}_i \cdot \mathcal{T}(\mathbf{g}_j) = \mathbf{g}_i \cdot (\mathbf{b} \cdot \mathbf{g}_j)\mathbf{a} = (\mathbf{g}_i \cdot \mathbf{a})(\mathbf{b} \cdot \mathbf{g}_j) = a_i b_j,$$

which gives a third (equivalent) definition of the tensor product. Alternatively, we can write

$$T_j^i = a^i b_j \quad \text{or} \quad T^{ij} = a^i b^j \quad \text{or} \quad T_i^j = a_i b^j. \quad (1.40)$$

Note that we can write

$$\mathcal{T} = \mathbf{a} \otimes \mathbf{b} = a_i \mathbf{g}^i \otimes b_j \mathbf{g}^j = a_i b_j \mathbf{g}^i \otimes \mathbf{g}^j = T_{ij} \mathbf{g}^i \otimes \mathbf{g}^j,$$

again demonstrating that  $T_{ij}$  are the components of the tensor  $\mathcal{T}$  in the basis  $\mathbf{g}^i \otimes \mathbf{g}^j$ .

The tensor product can easily be extended to tensors of arbitrary order via its component-wise definition. e.g.

$$T_{mn}^i \quad ^{jk} = A_{mn}^i B^{jk}$$

## 1.4 The metric tensor

We can reinterpret the Kronecker delta as a tensor with components in the global Cartesian coordinate system, so that

$$\delta_{IJ} = \delta^{IJ} = \delta_J^I \equiv \begin{cases} 1, & \text{if } I = J, \\ 0, & \text{otherwise.} \end{cases}$$

If we transform the Kronecker delta into the general coordinate system  $\xi^i$ , then we obtain the so-called metric tensors

$$g_{ij} \equiv \frac{\partial x^I}{\partial \xi^i} \frac{\partial x^J}{\partial \xi^j} \delta_{IJ} = \frac{\partial x^I}{\partial \xi^i} \frac{\partial x^I}{\partial \xi^j} = \mathbf{g}_i \cdot \mathbf{g}_j, \quad (1.41a)$$

$$g^{ij} \equiv \frac{\partial \xi^i}{\partial x^I} \frac{\partial \xi^j}{\partial x^J} \delta^{IJ} = \frac{\partial \xi^i}{\partial x^I} \frac{\partial \xi^j}{\partial x^I} = \mathbf{g}^i \cdot \mathbf{g}^j, \quad (1.41b)$$

$$g_j^i \equiv \frac{\partial \xi^i}{\partial x^I} \frac{\partial x^J}{\partial \xi^j} \delta_J^I = \frac{\partial \xi^i}{\partial x^I} \frac{\partial x^I}{\partial \xi^j} = \delta_j^i. \quad (1.41c)$$

We note that all the tensors are symmetric  $g_{ij} = g_{ji}$  and that the mixed metric tensor  $g_j^i$  is invariant and equal to  $\delta_j^i$  in all coordinate systems. From equations (1.41a,b) we deduce that

$$g^{ij} g_{jk} = \frac{\partial \xi^i}{\partial x^I} \frac{\partial \xi^j}{\partial x^I} \frac{\partial x^J}{\partial \xi^j} \frac{\partial x^J}{\partial \xi^k} = \frac{\partial \xi^i}{\partial x^I} \delta_I^J \frac{\partial x^J}{\partial \xi^k} = \frac{\partial \xi^i}{\partial x^I} \frac{\partial x^I}{\partial \xi^k} = \delta_k^i, \quad (1.42)$$

in other words if we represent the components  $g_{ij}$  as a matrix, we can find the components  $g^{ij}$  by taking the inverse of that matrix, i.e.

$$g^{ij} = \frac{D^{ij}}{g}, \quad (1.43)$$

where  $D^{ij}$  is the matrix of cofactors of the matrix of coefficients  $g_{ij}$  and  $g$  is the determinant of the matrix  $g_{ij}$

$$g \equiv |g_{ij}| = \left| \frac{\partial x^I}{\partial \xi^J} \right|^2 \quad \text{and} \quad |g^{ij}| = \left| \frac{\partial \xi^i}{\partial x^j} \right|^2 = \frac{1}{g}, \quad (1.44)$$

which demonstrates that the determinants are positive.

## 1.4.1 Properties of the metric tensor

### Line elements

We have already seen that an infinitesimal line element corresponding to a change  $d\xi^i$  in the  $\xi^i$  coordinate is given by

$$\mathbf{t}_i = \mathbf{d}\mathbf{s}_i = \frac{\partial \mathbf{r}}{\partial \xi^i} d\xi^i = \mathbf{g}_i d\xi^i \quad (\text{not summed}).$$

Hence, a general infinitesimal line element corresponding to changes in all coordinates is given by

$$\mathbf{d}\mathbf{r} = \mathbf{g}_i d\xi^i, \quad (\text{summed over } i),$$

and its length squared is

$$ds^2 = \mathbf{d}\mathbf{r} \cdot \mathbf{d}\mathbf{r} = \mathbf{g}_i d\xi^i \cdot \mathbf{g}_j d\xi^j = g_{ij} d\xi^i d\xi^j.$$

Thus, the infinitesimal length change due to increments  $d\xi^i$  in general coordinates is given by  $\sqrt{g_{ij} d\xi^i d\xi^j}$ , which explains the use of the term metric to describe the tensor  $g_{ij}$ .

### Surface elements

Any surface  $\xi^1$  is constant is spanned by the covariant base vectors  $\mathbf{g}_2$  and  $\mathbf{g}_3$  and so an infinitesimal area within that surface is given by

$$dS_{(1)} = |\mathbf{d}\mathbf{s}_2 \times \mathbf{d}\mathbf{s}_3| = |\mathbf{g}_2 \times \mathbf{g}_3| d\xi^2 d\xi^3,$$

and

$$|\mathbf{g}_2 \times \mathbf{g}_3| = |\epsilon_{231} \mathbf{g}^1| = |\epsilon_{231}| |\mathbf{g}^1| = \left| \frac{\partial x^I}{\partial \xi^2} \frac{\partial x^J}{\partial \xi^3} \frac{\partial x^K}{\partial \xi^1} e_{IJK} \right| \sqrt{\mathbf{g}^1 \cdot \mathbf{g}^1}. \quad (1.45)$$

The term between the modulus signs is the determinant<sup>12</sup> of the matrix with components  $\partial x^I / \partial \xi^j$ , so from the definition of the determinant of the metric tensor (1.44),

$$dS_{(1)} = \sqrt{g g^{11}} d\xi^2 d\xi^3.$$

Hence, an element of area in the surface  $\xi^1$  is constant is

$$dS_{(i)} = \sqrt{g g^{ii}} d\xi^j d\xi^k, \quad (i \neq j \neq k \text{ and } i \text{ not summed}).$$

---

<sup>12</sup>It is the scalar triple product of the three rows of the matrix.

## Volume elements

An infinitesimal parallelepiped spanned by the vectors  $d\mathbf{s}_i$  has volume given by

$$dV = |\mathbf{ds}_1 \cdot (\mathbf{ds}_2 \times \mathbf{ds}_3)| = |\mathbf{g}_1 \cdot (\mathbf{g}_2 \times \mathbf{g}_3)| d\xi^1 d\xi^2 d\xi^3 = |\mathbf{g}_1 \cdot \epsilon_{231} \mathbf{g}^1| d\xi^1 d\xi^2 d\xi^3.$$

Now

$$|\mathbf{g}_1 \cdot \epsilon_{231} \mathbf{g}^1| = |\epsilon_{231}| = \sqrt{g},$$

using the same argument as in equation (1.45). Thus, the volume element is given by

$$dV = \sqrt{g} d\xi^1 d\xi^2 d\xi^3, \quad (1.46)$$

which is, of course, equivalent to the standard expression for change of coordinates in multiple integrals

$$dV = \left| \frac{\partial(x^1, x^2, x^3)}{\partial(\xi^1, \xi^2, \xi^3)} \right| d\xi^1 d\xi^2 d\xi^3.$$

## Index raising and lowering

We conclude this brief section by noting one of the most useful properties of the metric tensor. For a vector  $\mathbf{a} = a_j \mathbf{g}^j$ , we can find the components in the covariant basis by taking the dot product

$$a^i = \mathbf{a} \cdot \mathbf{g}^i = a_j \mathbf{g}^j \cdot \mathbf{g}^i = a_j g^{ji}.$$

Thus the contravariant metric tensor can be used to raise indices and similarly the covariant metric tensor can be used to lower indices

$$a_i = a^j g_{ji}.$$

# 1.5 Vector and tensor calculus

## 1.5.1 Covariant differentiation

We have already established that the derivative of the position vector  $\mathbf{r}$  with respect to the coordinate  $\xi^i$  gives a tangent vector  $\mathbf{g}_i$ . If we wish to express the rate of change of a vector field in the direction of a coordinate then we also need to be able to calculate

$$\frac{\partial \mathbf{a}}{\partial \xi^i} \equiv \mathbf{a}_{,i},$$

where we shall use a comma to indicate differentiation with respect to the  $i$ -th general coordinate.

Under a change of coordinates from  $\xi^i$  to  $\chi^{\bar{i}}$ , the derivative becomes

$$\frac{\partial \mathbf{a}}{\partial \chi^{\bar{i}}} = \frac{\partial \xi^j}{\partial \chi^{\bar{i}}} \frac{\partial \mathbf{a}}{\partial \xi^j},$$

so it transforms covariantly. If we decompose the vector into the covariant basis we have that

$$\mathbf{a}_{,j} = (a^i \mathbf{g}_i)_{,j} = a^i_{,j} \mathbf{g}_i + a^i \mathbf{g}_{i,j}, \quad (1.47)$$

by the product rule.

Vector equations are often written by just writing out the components (in an assumed basis)

$$\mathbf{v} = \mathbf{u} \quad \Rightarrow \quad v_I \mathbf{e}_I = u_I \mathbf{e}_I \quad \Rightarrow \quad v_I = u_I,$$

where formally the last equation is obtained via dot product with a suitable base vector. In Cartesian components we can simply take derivatives of the component equations with respect to the coordinates because the base vectors do not depend on the coordinates

$$u_I = v_I \quad \text{leads to} \quad u_{I,J} = v_{I,J} \quad \text{because} \quad (u_I \mathbf{e}_I)_{,J} = u_{I,J} \mathbf{e}_I.$$

In other words, the components of the differentiated vector equation are simply the derivatives of the components of the original equation.

In a general coordinate system, the base vectors are **not** independent of the coordinates and so the second term in equation (1.47) cannot be neglected. The vector equation

$$\mathbf{a} = \mathbf{b} \quad \Rightarrow \quad a^i \mathbf{g}_i = b^i \mathbf{g}_i \quad \Rightarrow \quad a^i = b^i,$$

where the last equation is now obtained by taking the dot product with the contravariant base vectors. However, in general coordinates,

$$a^i = b^i \quad \text{does not (directly) lead to} \quad a^i_{,j} = b^i_{,j}.$$

Although the final equation is (obviously) true from direct differentiation of each component the statement is not coordinate-independent because it does not obey the correct transformation properties for a mixed second-order tensor.

In fact, the derivatives of the base vectors are given by

$$\mathbf{g}_{i,j} = \frac{\partial^2 \mathbf{r}}{\partial \xi^j \partial \xi^i} = \mathbf{r}_{,ij} = \mathbf{r}_{,ji},$$

assuming symmetry of partial differentiation. If we decompose the position vector into the fixed Cartesian coordinates, we have

$$\mathbf{g}_{i,j} = \frac{\partial^2 x^K}{\partial \xi^j \partial \xi^i} \mathbf{e}_K,$$

which can be written in the contravariant basis as

$$\mathbf{g}_{i,j} = \Gamma_{ijk} \mathbf{g}^k,$$

where

$$\Gamma_{ijk} = \mathbf{g}_{i,j} \cdot \mathbf{g}_k = \frac{\partial^2 \mathbf{r}}{\partial \xi^j \partial \xi^i} \cdot \mathbf{g}_k = \frac{\partial^2 x^K}{\partial \xi^i \partial \xi^j} \frac{\partial x^K}{\partial \xi^k}. \quad (1.48)$$

The symbol  $\Gamma_{ijk}$  is called the Christoffel symbol of the first kind. Thus, we can write

$$\mathbf{a}_{,j} = a^i_{,j} \mathbf{g}_i + a^i \Gamma_{ijk} \mathbf{g}^k = a^i_{,j} \mathbf{g}_i + a^i \Gamma_{ijk} g^{kl} \mathbf{g}_l = (a^l_{,j} + \Gamma^l_{ij} a^i) \mathbf{g}_l,$$

where

$$\Gamma^l_{ij} = g^{kl} \Gamma_{ijk} = g^{kl} \mathbf{g}_k \cdot \mathbf{g}_{i,j} = \mathbf{g}^l \cdot \mathbf{g}_{i,j}, \quad (1.49)$$

and  $\Gamma^l_{ij}$  is called the Christoffel symbol of the second kind.

Finally, we obtain

$$\mathbf{a}_{,j} = a^k |_{,j} \mathbf{g}_k, \quad \text{where} \quad a^k |_{,j} = a^k_{,j} + \Gamma^k_{ij} a^i,$$

and the expression  $a^k|_j$  is the covariant derivative of the component  $a^k$ . In many books the covariant derivative is represented by a subscript semicolon  $a^k_{;j}$  and in some it is denoted by a comma, just to confuse things. The fact that  $\mathbf{a}_{,j}$  is a covariant vector and that  $\mathbf{g}_k$  is a covariant vector means that the covariant derivative  $a^k|_j$  is a (mixed) second-order tensor<sup>13</sup>. Thus when differentiating equations expressed in components in a general coordinate system we should write

$$a^i = b^i \quad \Rightarrow \quad a^i|_j = b^i|_j.$$

If we had decomposed the vector  $\mathbf{a}$  into the contravariant basis then

$$\mathbf{a}_{,j} = a_{i,j} \mathbf{g}^i + a_i \mathbf{g}^i_{,j}.$$

and from equation (1.12),

$$\mathbf{g}^k = \frac{\partial \xi^k}{\partial x^K} \mathbf{e}^K \quad \Rightarrow \quad \mathbf{g}^k \frac{\partial x^I}{\partial \xi^k} = \frac{\partial \xi^k}{\partial x^K} \frac{\partial x^I}{\partial \xi^k} \mathbf{e}^K = \delta^K_I \mathbf{e}^K = \mathbf{e}^I. \quad (1.50)$$

We differentiate equation (1.50) with respect to  $\xi^j$ , remembering that  $\mathbf{e}^I$  is a constant base vector, to obtain

$$\begin{aligned} \mathbf{g}^k_{,j} \frac{\partial x^I}{\partial \xi^k} + \mathbf{g}^k \frac{\partial^2 x^I}{\partial \xi^j \partial \xi^k} &= 0, \\ \Rightarrow \quad \mathbf{g}^k_{,j} \frac{\partial x^I}{\partial \xi^k} \frac{\partial \xi^i}{\partial x^I} &= -\mathbf{g}^k \frac{\partial \xi^i}{\partial x^I} \frac{\partial^2 x^I}{\partial \xi^j \partial \xi^k} \quad \Rightarrow \quad \mathbf{g}^i_{,j} = -\mathbf{g}_{k,j} \cdot \mathbf{g}^i \mathbf{g}^k = -\Gamma^i_{kj} \mathbf{g}^k. \end{aligned}$$

Thus the derivative of the vector  $\mathbf{a}$  when decomposed into the contravariant basis is

$$\mathbf{a}_{,j} = a_{i,j} \mathbf{g}^i - \Gamma^i_{jk} a_i \mathbf{g}^k = (a_{k,j} - \Gamma^i_{jk} a_i) \mathbf{g}^k;$$

and finally, we obtain

$$\mathbf{a}_{,j} = a_i|_j \mathbf{g}^i, \quad \text{where} \quad a_i|_j = a_{i,j} - \Gamma^k_{ji} a_k,$$

which gives the covariant derivative of the covariant components of a vector.

In Cartesian coordinate systems, the base vectors are independent of the coordinates, so  $\Gamma_{IJK} \equiv 0$  for all  $I, J, K$  and the covariant derivative reduces to the partial derivative:

$$a_I|_K = a^I|_K = a_{I,K}.$$

Therefore, in most cases, when generalising equations derived in Cartesian coordinates to other coordinate systems we simply need to replace the partial derivatives with covariant derivatives.

The partial derivative of a scalar already exhibits covariant transformation properties because

$$\frac{\partial \phi}{\partial \chi^i} = \frac{\partial \xi^j}{\partial \chi^i} \frac{\partial \phi}{\partial \xi^j},$$

and the partial derivative of a scalar coincides with the covariant derivative

$$\phi_{,i} = \phi|_i.$$

The covariant derivative of higher-order tensors can be constructed by considering the covariant derivative of invariants and insisting that the covariant derivative obeys the product rule.

---

<sup>13</sup>Thus, the covariant derivative exhibits tensor transformation properties, which the partial derivative does not because

$$\frac{\partial a^{\bar{i}}}{\partial \chi^{\bar{j}}} = \frac{\partial \xi^k}{\partial \chi^{\bar{j}}} \frac{\partial}{\partial \xi^k} \left( \frac{\partial \chi^{\bar{i}}}{\partial \xi^l} a^l \right) = \frac{\partial \xi^k}{\partial \chi^{\bar{j}}} \frac{\partial \chi^{\bar{i}}}{\partial \xi^l} \frac{\partial a^l}{\partial \xi^k} + \frac{\partial \xi^k}{\partial \chi^{\bar{j}}} \frac{\partial^2 \chi^{\bar{i}}}{\partial \xi^k \partial \xi^l} a^l,$$

and the presence of the second term violates the tensor transformation rule.

## 1.5.2 Vector differential operators

### Gradient

The gradient of a scalar field  $f(\mathbf{x})$  is denoted by  $\nabla f$ , or  $\text{grad}f$ , and is defined to be a vector field such that

$$f(\mathbf{x} + d\mathbf{x}) = f(\mathbf{x}) + \nabla f \cdot d\mathbf{x} + o(|d\mathbf{x}|); \quad (1.51)$$

*i.e.* the gradient is the vector that expresses the change in the scalar field with position.

Letting  $d\mathbf{x} = t\mathbf{h}$ , dividing by  $t$  and taking the limit as  $t \rightarrow 0$  gives the alternative definition

$$\nabla f \cdot \mathbf{h} = \lim_{t \rightarrow 0} \frac{f(\mathbf{x} + t\mathbf{h}) - f(\mathbf{x})}{t} = \frac{d}{dt} f(\mathbf{x} + t\mathbf{h})|_{t=0}. \quad (1.52)$$

Here, the derivative is the directional derivative in the direction  $\mathbf{h}$ .

If we decompose the vectors into components in the global Cartesian basis equation (1.52) becomes

$$[\nabla f]_K h_K = \frac{\partial f}{\partial x_K} \frac{d}{dt} (x_K + th_K)|_{t=0} = \frac{\partial f}{\partial x_K} h_K,$$

and so because  $\mathbf{h}$  is arbitrary, we can define

$$\nabla f = [\nabla f]_K \mathbf{e}_K = \frac{\partial f}{\partial x_K} \mathbf{e}_K, \quad (1.53)$$

which should be a familiar expression for the gradient. Relative to the general coordinate system  $\xi^i$  equation (1.53) becomes

$$\nabla f = \frac{\partial f}{\partial \xi^i} \frac{\partial \xi^i}{\partial x_K} \mathbf{e}_K = \frac{\partial f}{\partial \xi^i} \mathbf{g}^i = f_{,i} \mathbf{g}^i, \quad (1.54)$$

so we can write the vector differential operator  $\nabla = \mathbf{g}^i \frac{\partial}{\partial \xi^i}$ . Note that because the derivative transforms covariantly and the base vector transform contravariantly the gradient is invariant under coordinate transform.

#### Example 1.4. Gradient in standard coordinate systems

Find the gradient in a plane polar coordinate system.

**Solution 1.4.** The gradient of a scalar field  $f$  is simply

$$\nabla f = \mathbf{g}^i \frac{\partial f}{\partial \xi^i}.$$

In plane polar coordinates,

$$\mathbf{g}^1 = \cos \xi^2 \mathbf{e}_1 + \sin \xi^2 \mathbf{e}_2, \quad \mathbf{g}^2 = -\frac{\sin \xi^2}{\xi^1} \mathbf{e}_1 + \frac{\cos \xi^2}{\xi^1} \mathbf{e}_2,$$

so

$$\nabla f = (\cos \xi^2 \mathbf{e}_1 + \sin \xi^2 \mathbf{e}_2) \frac{\partial f}{\partial \xi^1} + \left( -\frac{\sin \xi^2}{\xi^1} \mathbf{e}_1 + \frac{\cos \xi^2}{\xi^1} \mathbf{e}_2 \right) \frac{\partial f}{\partial \xi^2},$$

which can be written in the (hopefully) familiar form

$$\nabla f = \frac{\partial f}{\partial r} \mathbf{e}_r + \frac{1}{r} \frac{\partial f}{\partial \theta} \mathbf{e}_\theta.$$

## Gradient of a Vector Field

The gradient of a vector field  $\mathbf{F}(\mathbf{x})$  is a second-order tensor  $\nabla\mathbf{F}(\mathbf{x})$  also often written as  $\nabla \otimes \mathbf{F}$  that arises when the vector differential operator is applied to a vector

$$\nabla \otimes \mathbf{F} = \mathbf{g}^i \otimes \mathbf{F}_{,i} = F^k|_i \mathbf{g}^i \otimes \mathbf{g}_k = F_k|_i \mathbf{g}^i \otimes \mathbf{g}^k.$$

Note that we have used the covariant derivative because we are taking the derivative of a vector decomposed into the covariant or contravariant basis.

## Divergence

The divergence of a vector field is the trace or contraction of the gradient of the vector field. It is formed by taking the dot product of the vector differential operator  $\nabla$  with the vector field:

$$\operatorname{div}\mathbf{F} = \nabla \cdot \mathbf{F} = \mathbf{g}^i \cdot \mathbf{F}_{,i} = \mathbf{g}^i \cdot \frac{\partial}{\partial \xi^i} (F^j \mathbf{g}_j) = \mathbf{g}^i \cdot F^j|_i \mathbf{g}_j = \delta_j^i F^j|_i = F^i|_i,$$

so that

$$\operatorname{div}\mathbf{F} = F^i|_i = F^i_{,i} + \Gamma_{ji}^i F^j = F^i_{,i} + \Gamma_{ij}^i F^j. \quad (1.55)$$

## Curl

The curl of a vector field is the cross product of our vector differential operator  $\nabla$  with the vector field

$$\operatorname{curl}\mathbf{F} = \nabla \times \mathbf{F} = \mathbf{g}^i \times \frac{\partial \mathbf{F}}{\partial \xi^i} = \mathbf{g}^i \times \frac{\partial (F_j \mathbf{g}^j)}{\partial \xi^i} = \mathbf{g}^i \times F_j|_i \mathbf{g}^j = \epsilon^{ijk} F_j|_i \mathbf{g}_k.$$

## Laplacian

The Laplacian of a scalar field  $f$  is defined by

$$\nabla^2 f = \Delta f = \nabla \cdot \nabla f.$$

Thus in general coordinates

$$\nabla^2 f = \mathbf{g}^j \cdot \frac{\partial}{\partial \xi^j} \left( \mathbf{g}^i \frac{\partial f}{\partial \xi^i} \right) = \mathbf{g}^j \cdot f_{,i}|_j \mathbf{g}^i = g^{ij} f_{,i}|_j = g^{ij} f|_{ij},$$

because the partial and covariant derivatives are the same for scalar quantities. Now  $g^{ij}|_j$  is a tensor (covariant first order), which means that we know how it transforms under change in coordinates, but in Cartesian coordinates  $g^{ij}|_j$  becomes  $\delta^{IJ}|_J = 0$ , so  $g^{ij}|_j = 0$  in all coordinate systems<sup>14</sup>. Thus,

$$\nabla^2 f = (g^{ij} f_{,i})|_j = \frac{1}{\sqrt{g}} \frac{\partial}{\partial \xi^j} \left( \sqrt{g} g^{ij} \frac{\partial f}{\partial \xi^i} \right).$$

The last equality is not obvious and is proved in the next section, equation (1.58)

---

<sup>14</sup>This is a special case of Ricci's Lemma.

## 1.6 Divergence theorem

The divergence theorem is a generalisation of the fundamental theorem of calculus into higher dimensions. Consider the volume given by  $a \leq \xi^1 \leq b$ ,  $c \leq \xi^2 \leq d$  and  $e \leq \xi^3 \leq f$ . If we want to calculate the outward flux of a vector field  $\mathbf{F}(\mathbf{x})$  from this volume then we must calculate

$$\iint \mathbf{F} \cdot \mathbf{n} \, dS,$$

where  $\mathbf{n}$  is the outer unit normal to the surface and  $dS$  is an infinitesimal element of the surface area.

From section 1.4.1, we know that on the faces  $\xi^1 = a, b$ ,

$$dS = \sqrt{gg^{11}} d\xi^2 d\xi^3,$$

and a normal to the surface is given by the contravariant base vector  $\mathbf{g}^1$ , with length  $|\mathbf{g}^1| = \sqrt{\mathbf{g}^1 \cdot \mathbf{g}^1} = \sqrt{g^{11}}$ . On the face  $\xi^1 = a$  the vector  $\mathbf{g}^1$  is directed into the volume (it is oriented along the line of increasing coordinate) and on  $\xi^1 = b$ ,  $\mathbf{g}^1$  is directed out of the volume, so the net outward flux through these faces is

$$\begin{aligned} & \int_c^d \int_e^f \mathbf{F}(a, \xi^2, \xi^3) \cdot (-\mathbf{g}^1 / \sqrt{g^{11}}) \sqrt{gg^{11}} d\xi^2 d\xi^3 + \int_c^d \int_e^f \mathbf{F}(b, \xi^2, \xi^3) \cdot (+\mathbf{g}^1 / \sqrt{g^{11}}) \sqrt{gg^{11}} d\xi^2 d\xi^3 \\ &= \int_c^d \int_e^f \left[ F^1(b, \xi^2, \xi^3) \sqrt{g(b, \xi^2, \xi^3)} - F^1(a, \xi^2, \xi^3) \sqrt{g(a, \xi^2, \xi^3)} \right] d\xi^2 d\xi^3 \end{aligned}$$

From the fundamental theorem of calculus, the integral becomes

$$\int_c^d \int_e^f \int_a^b (F^1 \sqrt{g})_{,1} d\xi^1 d\xi^2 d\xi^3,$$

and using similar arguments, the total outward flux from the volume is given by

$$\iint \mathbf{F} \cdot \mathbf{n} \, dS = \int_a^b \int_c^d \int_e^f (F^i \sqrt{g})_{,i} d\xi^1 d\xi^2 d\xi^3.$$

From section 1.4.1 we also know that an infinitesimal volume element is given by  $dV = \sqrt{g} d\xi^1 d\xi^2 d\xi^3$ , so

$$\iint \mathbf{F} \cdot \mathbf{n} \, dS = \iiint \frac{1}{\sqrt{g}} (F^i \sqrt{g})_{,i} \sqrt{g} d\xi^1 d\xi^2 d\xi^3 = \iiint \frac{1}{\sqrt{g}} (F^i \sqrt{g})_{,i} dV.$$

The volume integrand is

$$\frac{1}{\sqrt{g}} (F^i \sqrt{g})_{,i} = F^i_{,i} + \frac{1}{\sqrt{g}} F^i (\sqrt{g})_{,i},$$

but

$$\frac{1}{\sqrt{g}} (\sqrt{g})_{,i} = \frac{1}{2g} \frac{\partial g}{\partial \xi^i} = \frac{1}{2g} \frac{\partial g}{\partial g_{jk}} \frac{\partial g_{jk}}{\partial \xi^i},$$

by the chain rule. From equations (1.42) and (1.44)

$$g_{jk} g^{kl} = g_{jk} \frac{D^{kl}}{g} = \delta_j^l \quad \Rightarrow \quad g_{jk} D^{kl} = g \delta_j^l \quad \Rightarrow \quad \frac{\partial g}{\partial g_{jk}} \delta_j^l = \frac{\partial g}{\partial g_{lk}} = D^{kl} = g g^{kl} = g g^{lk}, \quad (1.56)$$



which means that

$$\frac{1}{\sqrt{g}}(\sqrt{g})_{,i} = \frac{1}{2g}gg^{jk}\frac{\partial g_{jk}}{\partial \xi^i} = \frac{1}{2}g^{jk}(\mathbf{g}_{j,i}\cdot\mathbf{g}_k + \mathbf{g}_j\cdot\mathbf{g}_{k,i}) = \frac{1}{2}(\mathbf{g}_{j,i}\cdot\mathbf{g}^j + \mathbf{g}_{k,i}\cdot\mathbf{g}^k) = \frac{1}{2}(\Gamma_{ji}^j + \Gamma_{ki}^k) = \Gamma_{ji}^j, \quad (1.57)$$

from the definition of the Christoffel symbol (1.49). Thus the volume integrand becomes

$$\frac{1}{\sqrt{g}}(F^i\sqrt{g})_{,i} = F^i_{,i} + F^i\Gamma_{ji}^j = F^i|_i, \quad (1.58)$$

and so,

$$\iint \mathbf{F} \cdot \mathbf{n} \, dS = \iiint (F^i_{,i} + F^i\Gamma_{ji}^j) \, dV = \iiint F^i|_i \, dV = \iiint \operatorname{div} \mathbf{F} \, dV,$$

using the definition of the divergence (1.55). In fact, one should argue that the definition of the divergence is chosen so that it is the quantity that appears within the volume integral in the divergence theorem. The form of the theorem that we shall use most often is

$$\iint F^i n_i \, dS = \iint F_i n^i \, dS = \iiint F^i|_i \, dV = \iiint \frac{1}{\sqrt{g}}(F^i\sqrt{g})_{,i} \, dV. \quad (1.59)$$

You might argue that the proof above is only valid for simple parallelepipeds in our general coordinate system, but we can extend the result to more general volumes by breaking the volume up into regions, each of which can be described by a parallelepiped in a general coordinate systems.

## 1.7 The Stokes Theorem

For completeness, we also state, but do not prove the Stokes theorem, which states that for any vector  $\mathbf{v}$ , the integral of its tangential component around a closed path is equal to the flux of the curl of the vector through any surface bounded by the path:

$$\oint_C \mathbf{v} \cdot d\mathbf{r} = \iint_A (\operatorname{curl} \mathbf{v}) \cdot d\mathbf{S}. \quad (1.60)$$

In general tensor form, equation (1.60) becomes

$$\oint_C \mathbf{v} \cdot \mathbf{g}_i \, d\xi^i = \oint v_i \, d\xi^i = \iint_A \epsilon^{ijk} v_j|_i \mathbf{g}_k \cdot d\mathbf{S} = \iint_A \epsilon^{ijk} v_j|_i n_k \, dS, \quad (1.61)$$

where the vector area is  $d\mathbf{S} = \mathbf{n} \, dS$ , and  $\mathbf{n}$  is a unit vector normal to the surface. Again we can make the argument that the definition of the curl is chosen so that this relationship is satisfied.

You may be interested to know that an entire theory (differential forms) can be constructed in which the Stokes theorem is a fundamental result

$$\int_{\partial C} \omega = \int_C d\omega,$$

where  $\omega$  is an  $n$ -form and  $d\omega$  is its differential. The theory encompasses all known integral theorems because the appropriate  $n$ -dimensional differential forms coincide with the divergence and curl and the theory provides a rational extension to higher dimensions.