Chapter 1

Linear vector spaces

1.1 Vectors, bases, components

A general vector[1] in three-dimensional space, may be written as $\mathsf{v} = \sum_i v_i \mathsf{e}_i$ where $\mathsf{e}_1, \mathsf{e}_2, \mathsf{e}_3$ form a *basis* and v_1, v_2, v_3 are *components* of the vector relative to the basis. With a matrix notation

$$\mathsf{v} = \begin{pmatrix} \mathsf{e}_1 & \mathsf{e}_2 & \mathsf{e}_3 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \tag{1.1}$$

where basis vectors are collected in a *row* matrix and vector components in a *column*. With an alternative choice of basis, the same vector might be written $\mathsf{v} = \sum_i \bar{v}_i \bar{\mathsf{e}}_i$, where $\bar{\mathsf{e}}_i = \mathsf{e}_1 T_{1i} + \mathsf{e}_2 T_{2i} + \mathsf{e}_3 T_{3i}$ or in matrix form

$$\begin{pmatrix} \bar{\mathsf{e}}_1 & \bar{\mathsf{e}}_2 & \bar{\mathsf{e}}_3 \end{pmatrix} = \begin{pmatrix} \mathsf{e}_1 & \mathsf{e}_2 & \mathsf{e}_3 \end{pmatrix} \begin{pmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{pmatrix} \tag{1.2}$$

The last two equations may be written more briefly as

$$\mathsf{v} = \mathbf{e}\mathbf{v} = \bar{\mathbf{e}}\bar{\mathbf{v}} \quad (a); \qquad \bar{\mathbf{e}} = \mathbf{e}\mathbf{T} \qquad (b). \tag{1.3}$$

The columns $\mathbf{v}, \bar{\mathbf{v}}$ *represent* the vector v with respect to the two alternative bases, while the square matrix $\mathbf{T}$ describes the *transformation* from the one basis to the other. Transformations are simply *changes of description*: the vector v is not affected by the change, being considered an *invariant*. To express v relative to the second basis we shall have to write $\mathsf{v} = \bar{\mathbf{e}}\bar{\mathbf{v}} = \mathbf{e}\mathbf{T}\bar{\mathbf{v}}$ and comparison with (1.3a) shows that

$$\mathbf{v} = \mathbf{T}\bar{\mathbf{v}}, \qquad \bar{\mathbf{v}} = \mathbf{T}^{-1}\mathbf{v}. \tag{1.4}$$

[1] Some knowledge of elementary vector algebra and of the use of matrix notation will be assumed. This may be found elsewhere (e.g. in Book 11 of the series Basic Books in Science, also published on this website). Remember that 'ordinary' three-dimensional vector space is *linear* (two vectors may be combined by 'addition' to give a third, and may be multiplied by arbitrary numbers) and that the term 'three-dimensional' means that three independent vectors are sufficient to determine all others.

Here $\mathbf{T}^{-1}$ is the *inverse* of $\mathbf{T}$, with the property $\mathbf{T}\mathbf{T}^{-1} = \mathbf{T}^{-1}\mathbf{T} = \mathbf{1}$, where $\mathbf{1}$ is the *unit matrix*, with elements $(\mathbf{1})_{ij} = \delta_{ij}$, where δ_{ij} is the 'Kronecker symbol' (=1, for $i = j$; =0, otherwise).

1.2 Transformations. The tensor concept

The last few equations allow us to distinguish two types of transformation law. In changing the basis according to (1.2), the basis vectors follow the rule

$$\mathsf{e}_i \to \bar{\mathsf{e}}_i = \sum_j \mathsf{e}_j T_{ji}, \tag{1.5}$$

(summation limits, 1 and 3, will not usually be shown) while, from (1.4),

$$v_i \to \bar{v}_i = \sum_j (\mathbf{T}^{-1})_{ij} v_j. \tag{1.6}$$

In order to compare these two transformation laws it is often convenient to *transpose*[2] equation (1.3b)and to write instead of (1.5)

$$\mathsf{e}_i \to \bar{\mathsf{e}}_i = \sum_j \tilde{T}_{ij} \mathsf{e}_j \tag{1.7}$$

so that the vectors and transformation matrices occur in the same order in both equations; and also to introduce

$$\mathbf{U} = \mathbf{T}^{-1}. \tag{1.8}$$

The 'standard' transformation laws then look nicer:

$$\bar{v}_i = \sum_j U_{ij} v_j, \qquad \bar{\mathsf{e}}_i = \sum_j \tilde{T}_{ij} \mathsf{e}_j. \tag{1.9}$$

Each matrix in (1.9), $\mathbf{U}$ and $\tilde{\mathbf{T}}$, may be obtained by taking the inverse of the other and transposing; and the alternative transformations in (1.9) are said to be *contragredient*. A special symbol is used to denote the double operation of inversion and transposition, writing for any matrix $\breve{\mathbf{A}} = \tilde{\mathbf{A}}^{-1}$; and since $\mathbf{T} = \mathbf{U}^{-1}$ it follows that $\tilde{\mathbf{T}} = \breve{\mathbf{U}}$. Both matrices in (1.9) may thus be written in terms of the matrix $\mathbf{U}(= \mathbf{T}^{-1})$, the second equation becoming $\bar{\mathsf{e}}_i = \sum_j \breve{U}_{ij} \mathsf{e}_j$.
The systematic study of transformations is the subject of the *tensor calculus*, to be developed later in more detail; but even at this point it is useful to anticipate a simple convention to distinguish quantities which follow the two transformation laws in (1.9): the indices that label quantities behaving like the basis vectors e_i will be left in the subscript position, but those that label quantities transforming like the vector components

[2]The transpose of a matrix $\mathbf{A}$ is indicated by $\tilde{\mathbf{A}}$ and is obtained by interchanging corresponding rows and columns. Note that in transposing a matrix product the factors will appear in reverse order.

v_i will be written with *superscripts*. The two types of transformation are also given the special names 'covariant' and 'contravariant', respectively, and assume the forms

$$
\begin{aligned}
\text{Covariant :} \qquad \mathsf{e}_i \to \bar{\mathsf{e}}_i &= \sum_j \breve{U}_{ij}\mathsf{e}_j \quad (a) \\
\text{Contravariant :} \qquad v^i \to \bar{v}^i &= \sum_j U_{ij}v^j \quad (b)
\end{aligned}
\tag{1.10}
$$

The fact that the vector v is, by definition, *invariant* against a change of coordinate system is now expressed in the equation

$$
\begin{aligned}
\mathsf{v} &= \sum_i \bar{\mathsf{e}}_i \bar{v}^i = \sum_i \sum_j \breve{U}_{ij}\mathsf{e}_j \sum_k U_{ik}v^k \\
&= \sum_{j,k} \left(\sum_i {U^{-1}}_{ji} U_{ik} \right) \mathsf{e}_j v^k = \sum_{j,k} \delta_{jk}\mathsf{e}_j v^k = \sum_j \mathsf{e}_j v^j
\end{aligned}
$$

and this invariance property is quite independent of the precise nature of the quantities with upper and lower subscripts.

Sets[3] of quantities $\{A_i\}$ and $\{B^i\}$ that transform, when the basis is changed, according to (1.10a) and (1.10b) respectively, are called "rank-1 tensors". Thus,

$$
\begin{aligned}
\text{Covariant :} \qquad A_i \to \bar{A}_i &= \sum_j \breve{U}_{ij}A_j \qquad (a) \\
\text{Contravariant :} \qquad B^i \to \bar{B}^i &= \sum_j U_{ij}B^j \qquad (b)
\end{aligned}
\tag{1.11}
$$

where the covariant quantities transform *co*grediently to the basis vectors and the contravariant quantities transform *contra*grediently. From one covariant set and one contravariant set we can always form an invariant

$$
\sum_i A_i B^i = \text{invariant}, \tag{1.12}
$$

which is a tensor of rank zero. In later Sections we meet tensors of higher rank.

1.3 Scalar products and the metric

In elementary vector analysis the three basis vectors $\mathsf{e}_1, \mathsf{e}_2, \mathsf{e}_3$ define the axes of a rectilinear Cartesian coordinate system are thus chosen to be orthogonal and of unit length: this means there exists a *scalar product*, $\mathsf{e}_i \cdot \mathsf{e}_j = |\mathsf{e}_i||\mathsf{e}_j| \cos\theta_{ij}$, $|\mathsf{e}_i|$ and $|\mathsf{e}_j|$ being the lengths

[3] A whole set is indicated by putting a typical member in braces, $\{...\}$.

of the vectors and θ_{ij} the angle between them, and that $\mathsf{e}_i \cdot \mathsf{e}_j = \delta_{ij}$. Such a space is said to possess a *metric*, which makes it possible to define the length of *any* vector and the angle between *any pair* of vectors. In general, the so-called 'metric matrix' is simply the array of all basis-vector scalar products:

$$\mathbf{g} = \begin{pmatrix} \mathsf{e}_1 \cdot \mathsf{e}_1 & \mathsf{e}_1 \cdot \mathsf{e}_2 & \mathsf{e}_1 \cdot \mathsf{e}_3 \\ \mathsf{e}_2 \cdot \mathsf{e}_1 & \mathsf{e}_2 \cdot \mathsf{e}_2 & \mathsf{e}_2 \cdot \mathsf{e}_3 \\ \mathsf{e}_3 \cdot \mathsf{e}_1 & \mathsf{e}_3 \cdot \mathsf{e}_2 & \mathsf{e}_3 \cdot \mathsf{e}_3 \end{pmatrix} \tag{1.13}$$

and when the basis vectors are orthogonal and of unit length this means $\mathbf{g} = \mathbf{1}$, with 1s on the diagonal and 0s elsewhere. It is, however, often useful to use an *oblique* basis in which the vectors are neither orthogonal nor of unit length: in such cases $\mathbf{g}$ is non-diagonal, but (by definition) always symmetric.

To express a scalar product $\mathsf{x} \cdot \mathsf{y}$ in terms of the vector *components* we then use

$$\mathsf{x} \cdot \mathsf{y} = (\sum_i x^i \mathsf{e}_i) \cdot (\sum_j y^j \mathsf{e}_j) = \sum_{i,j} x^i y^j (\mathsf{e}_i \cdot \mathsf{e}_j) = \sum_{i,j} x^i g_{ij} y^j,$$

or, in matrix notation,

$$\mathsf{x} \cdot \mathsf{y} = \tilde{\mathbf{x}}\mathbf{g}\mathbf{y} \tag{1.14}$$

where $\tilde{\mathbf{x}}$, the *transpose* of $\mathbf{x}$, is the *row* matrix $\tilde{\mathbf{x}} = (x^1 \ x^2 \ x^3)$. We note that when the basis is rectangular Cartesian, with $\mathbf{g} = \mathbf{1}$,

$$\mathsf{x} \cdot \mathsf{y} = x^1 y^1 + x^2 y^2 + x^3 y^3, \qquad |\mathsf{x}|^2 = \mathsf{x} \cdot \mathsf{x} = (x^1)^2 + (x^2)^2 + (x^3)^2. \tag{1.15}$$

When the vectors are *unit* vectors, $|\mathsf{x}| = |\mathsf{y}| = 1$ and the scalar product defines the angle between the two vectors:

$$\mathsf{x} \cdot \mathsf{y} = \cos\theta = x^1 y^1 + x^2 y^2 + x^3 y^3 \tag{1.16}$$

– expressions familiar from elementary Cartesian geometry. It is also to be noted that the metric matrix may be written as a column-row product:

$$\mathbf{g} = \begin{pmatrix} \mathsf{e}_1 \\ \mathsf{e}_2 \\ \mathsf{e}_3 \end{pmatrix} \cdot \begin{pmatrix} \mathsf{e}_1 & \mathsf{e}_2 & \mathsf{e}_3 \end{pmatrix} = \tilde{\mathbf{e}} \cdot \mathbf{e}, \tag{1.17}$$

where the dot applies to all pairs of elements from the column and the row.

1.4 The reciprocal basis

In the non-trivial case where $\mathbf{g}$ is not the unit matrix and the basis is consequently oblique, it is convenient to introduce, alongside the 'direct' basis $\mathbf{e}$, a *'reciprocal'* basis $\mathbf{f}$ defined by

$$\mathbf{f} = \mathbf{e}\mathbf{g}^{-1}, \quad (\mathsf{f}_i = \sum_j \mathsf{e}_j (\mathbf{g}^{-1})_{ji}) \tag{1.18}$$

in which the existence of the inverse matrix $\mathbf{g}^{-1}$ is guaranteed by the assumed linear independence of the two basis sets. According to (1.17), the scalar products of any vector e_i of the direct basis and any f_j of the reciprocal basis may be displayed in the matrix

$$\tilde{\mathbf{e}} \cdot \mathbf{f} = \tilde{\mathbf{e}} \cdot \mathbf{e}\mathbf{g}^{-1} = \mathbf{g}\mathbf{g}^{-1} = \mathbf{1}. \tag{1.19}$$

In other words,

$$\mathsf{e}_i \cdot \mathsf{f}_j = \mathsf{f}_j \cdot \mathsf{e}_i = \delta_{ij} \tag{1.20}$$

– which means that every reciprocal vector f_i is perpendicular to the other two *direct* vectors $\mathsf{e}_j, \mathsf{e}_k$ $(j, k \neq i)$ and has a length reciprocal to that of the direct vector e_i. Although it is commonly said that $\mathbf{e} = (\mathsf{e}_1\ \mathsf{e}_2\ \mathsf{e}_3)$ defines the 'direct space', while $\mathbf{f} = (\mathsf{f}_1\ \mathsf{f}_2\ \mathsf{f}_3)$ defines the 'reciprocal space', it should be clearly understood that there is only *one space*, in which $\mathbf{e}$ and $\mathbf{f}$ provide alternative *bases*, linked by the transformation (1.18). It is therefore possible to express any given vector v equally well in terms of either the e's or the f's. It is convenient to use e^i as a new name for the reciprocal vector f_i, noting that this vector is *not* the same as e_i, thus obtaining from equation (1.18) the definition

$$\mathsf{e}^i = \sum_j \mathsf{e}_j (\mathbf{g}^{-1})_{ji}. \tag{1.21}$$

Let us put $\mathsf{v} = \sum_i \mathsf{f}_i v_i$, where v_i will be the i-component of v with respect to the *reciprocal* basis: then, as a result of (1.20), it follows that $v_i = \mathsf{f}_i \cdot \mathsf{v}$. Consequently, we can write the vector v in two forms:

$$\begin{aligned} &\text{Direct space}: \qquad \mathsf{v} = \sum_i \mathsf{e}_i v^i \qquad v^i = \mathsf{e}^i \cdot \mathsf{v} \qquad (a) \\ &\text{Reciprocal space}: \qquad \mathsf{v} = \sum_i \mathsf{e}^i v_i \qquad v_i = \mathsf{e}_i \cdot \mathsf{v}. \qquad (b) \end{aligned} \tag{1.22}$$

The scalar-product expressions for the components both follow as a result of (1.20), which we now rewrite as

$$\mathsf{e}_i \cdot \mathsf{e}^j = \mathsf{e}^j \cdot \mathsf{e}_i = \delta_{ij}. \tag{1.23}$$

To summarize, then, we have obtained two equivalent forms of any vector:

$$\mathsf{v} = \sum_i \mathsf{e}_i v^i = \sum_i \mathsf{e}^i v_i. \tag{1.24}$$

The first sum in (1.24) is known to be invariant against *any* basis change (see (1.10) *et seq*). It must later be confirmed that the superscript/subscript notation for *reciprocal* vectors and corresponding components is appropriate i.e. that they *do* indicate contravariant and covariant character, respectively.

Chapter 2

Elements of tensor algebra

2.1 Tensors of higher rank

Let us return to the two fundamental transformation laws (1.10a,b), for basis vectors and vector components in a linear vector space, namely

$$\begin{aligned} \text{Covariant}: \quad & \mathsf{e}_i \to \bar{\mathsf{e}}_i = \sum_j \breve{U}_{ij}\mathsf{e}_j \qquad (a) \\ \text{Contravariant}: \quad & v^i \to \bar{v}^i = \sum_j U_{ij} v^j \qquad (b) \end{aligned} \tag{2.1}$$

The idea of classifying various entities according to their transformation properties is easily generalized: if, for example, $\{\phi^i\}$ and $\{\psi^j\}$ are two contravariant sets of rank 1, then the set of all products $\{\phi^i\psi^j\}$ forms a contravariant set of rank 2, with the transformation law

$$\phi^i\psi^j \to \bar{\phi}^i\bar{\psi}^j = \sum_{k,l} U_{ik}U_{jl}\phi^k\psi^l, \tag{2.2}$$

the individual products being the contravariant components of a tensor of rank 2. More generally, sets of quantities $\{A^{ij}\}$, $\{A^i_{\cdot j}\}, \{A_{ij}\}$ are, respectively, contravariant, 'mixed' and covariant tensor components when they transform in such a way that[1]

$$\begin{aligned} \bar{A}^{ij} &= \sum_{k,l} U_{ik}U_{jl}A^{kl}\,, \qquad (a) \\ \bar{A}^i_{\cdot j} &= \sum_{k,l} U_{ik}\breve{U}_{jl}A^k_{\cdot l}\,, \qquad (b) \\ \bar{A}_{ij} &= \sum_{k,l} \breve{U}_{ik}\breve{U}_{jl}A_{kl}\,, \qquad (c) \end{aligned} \tag{2.3}$$

and when there are m factors of U-type and n of $\breve{U}$-type then we speak of a tensor of rank $(m+n)$, with m degrees of contravariance and n of covariance. The *order* of the upper and

[1] A dot is used to indicate the position of a 'missing' index, to avoid possible ambiguity (see text).

lower indices is of course important; any missing indices in either sequence are commonly indicated using dots. As a simple example, it may be noted that the components of the metric, $g_{ij} = \mathsf{e}_i \cdot \mathsf{e}_j$, form a rank-2 covariant set because in the basis change (2.1a) they are replaced by

$$\bar{g}_{ij} = \sum_{k,l} \breve{U}_{ik} \breve{U}_{jl} g_{kl}, \tag{2.4}$$

as in the last equation of (2.3).

2.2 Metric of the reciprocal basis

We are now in a position to verify that the reciprocal vectors $\{\mathsf{e}^i\}$ in (1.24) do indeed follow the *contra*variant transformation law (1.11b). Thus, the analogue of (1.21), after a basis change, will be

$$\bar{\mathsf{e}}^i = \sum_j \bar{\mathsf{e}}_j (\bar{\mathbf{g}}^{-1})^{ji}, \tag{2.5}$$

where $\bar{\mathbf{g}}$ is defined like $\mathbf{g}$ but in terms of the new basis vectors $\{\bar{\mathsf{e}}_i\}$. Here we have anticipated the contravariant nature of the elements of $\mathbf{g}^{-1}$, whose transformation properties must now be confirmed. To do so we express $\mathbf{g}$ in the form (1.17): thus $\mathbf{g} = \tilde{\mathbf{e}} \cdot \mathbf{e}$ and on transformation (writing (1.10a) as $\bar{\mathbf{e}} = \mathbf{e}\mathbf{U}^{-1}$)

$$\bar{\mathbf{g}} = \tilde{\bar{\mathbf{e}}} \cdot \bar{\mathbf{e}} = \tilde{\mathbf{U}}^{-1} \tilde{\mathbf{e}} \cdot \mathbf{e}\mathbf{U}^{-1} = \tilde{\mathbf{U}}^{-1} \mathbf{g} \mathbf{U}^{-1}.$$

On remembering that the inverse of a product is the product of the inverses *in reverse order*, we may write $\bar{\mathbf{g}}^{-1} = \mathbf{U}\mathbf{g}^{-1}\tilde{\mathbf{U}}$ and hence the elements of the inverse metric transform according to

$$(\mathbf{g}^{-1})^{ij} \to (\bar{\mathbf{g}}^{-1})^{ij} = \sum_{k,l} U_{ik} U_{jl} (\mathbf{g}^{-1})^{kl} \tag{2.6}$$

The indices on the matrix $\mathbf{g}^{-1}$ are thus correctly shown in the upper position, its elements evidently forming a rank-2 *contra*variant set[2].

On using (2.6) in (2.5) it follows at once that

$$\begin{aligned} \bar{\mathsf{e}}^i &= \sum_j \bar{\mathsf{e}}_j (\bar{\mathbf{g}}^{-1})^{ji} \\ &= \sum_j \sum_m \mathsf{e}_m \breve{U}_{jm} \sum_{k,l} U_{jk} (\mathbf{g}^{-1})^{kl} U_{il} \\ &= \sum_k \mathsf{e}_k \sum_l (\mathbf{g}^{-1})^{kl} U_{il} = \sum_l U_{il} \mathsf{e}^l. \end{aligned}$$

[2]This also follows from the fact that $\mathbf{g}^{-1}$ is the metrical matrix for the *reciprocal* basis, according to (1.21), with elements $\mathsf{e}^i \cdot \mathsf{e}^j$ – justifying the contravariant position of the indices.

In other words the result of a basis change is

$$\mathsf{e}^i \rightarrow \bar{\mathsf{e}}^i = \sum_j U_{ij}\mathsf{e}^j \tag{2.7}$$

– confirming again that the reciprocal vectors $\{\mathsf{e}^i\}$ form a rank-1 contravariant set.

2.3 The metric tensors: raising and lowering of indices

At this point we notice that the practice of writing row/column indices on the transformation matrix $\mathbf{U}$ always in the subscript position (with no tensorial significance) introduces an anomaly: when summing over a repeated index in, for example, equations (2.2) and (2.6), the index appears once in the upper position and once in the lower; but the positions of the remaining indices do not correctly indicate the tensor character of the result. This anomaly may be removed by a simple change of notation: the row index in U_{ij} will be raised to the upper position, and similarly for the column index in $\breve{U}_{ij}$. From now on, then,

$$U_{ij} \rightarrow U^i_j, \qquad \breve{U}_{ij} \rightarrow \breve{U}^j_i. \tag{2.8}$$

With this new notation, (2.1a) and (2.7) may be rewritten as

$$\mathsf{e}_i \rightarrow \bar{\mathsf{e}}_i = \sum_j \breve{U}^j_i \mathsf{e}_j \quad (a), \qquad \mathsf{e}^i \rightarrow \bar{\mathsf{e}}^i = \sum_j U^i_j \mathsf{e}^j \quad (b), \tag{2.9}$$

where the tensor character of the result is in each case correctly indicated by the position of the un-summed index. It is easily verified that, with corresponding changes, all the tensor equations encountered so far become notationally consistent. Thus for example the equations (2.3) take the form

$$\bar{A}^{ij} = \sum_{k,l} U^i_k U^j_l A^{kl} \quad (a); \quad \bar{A}^i_{.j} = \sum_{k,l} U^i_k \breve{U}^l_j A^k_{.l} \quad (b); \quad \bar{A}_{ij} = \sum_{k,l} \breve{U}^k_i \breve{U}^l_j A_{kl} \quad (c). \tag{2.10}$$

The un-summed indices on the right then correctly show, in each case, the tensor character of the results on the left.

Finally, two important properties of the metric will be noted. These become evident when the reciprocal, vectors defined in (1.21), are written in subscript/superscript notation as

$$\mathsf{e}^i = \sum_j g^{ij}\mathsf{e}_j, \tag{2.11}$$

where g^{ji} is the ji-element of the reciprocal matrix $\mathbf{g}^{-1}$. This means that summation over a common index (j), one upper and one lower, has effectively removed the covariance of e_i and replaced it by contravariance. In the same way it is easy to verify that the contravariant vectors e_i are replaced by their covariant counterparts on multiplying by the doubly covariant quantities g_{ij} and summing over the common index j:

$$\mathsf{e}_i = \sum_j g_{ij}\mathsf{e}^j. \tag{2.12}$$

The metric tensors, g^{ij} and g_{ij}, thus provide, respectively, the means of 'raising or lowering a tensor index' i.e. changing one degree of tensor character from covariance to contravariance or vice versa. Of course, this operation does not change the tensor quantities themselves in any way: it changes only the mode in which they are described. We note in passing that the metric tensors in the direct and reciprocal bases are related by

$$g_{ij}g^{jk} = \delta_i^k \tag{2.13}$$

– this expressing (in tensor notation) the fact that the metric *matrices* are related by $\mathbf{g}\mathbf{g}^{-1} = \mathbf{1}$.

Chapter 3

The tensor calculus

3.1 General coordinate systems

So far, it was assumed always that the coordinate system was *rectilinear*: there was a unique set of basis vectors, defining the directions of the axes, and the position of every point in space was specified by giving the components of its *position vector* relative to those same axes. The only transformations considered were those in which the axes were simply rotated about the origin, through angles determined by the given elements of the matrix $\mathbf{T}$. The transformations were *linear*. It is often useful, however, to employ more general coordinates, q_1, q_2, ..., such that every point in space is identified as a point of intersection of the surfaces on which $q_1 =$ constant, $q_2 =$ constant, etc.. Thus, q_1, q_2, q_3 might be the spherical polar coordinates r, θ, ϕ (Fig.1) and the surfaces considered would then be a sphere (radius r); a cone (half-angle θ) around the z axis; and a plane (containing the z axis and inclined at an angle ϕ to the x axis). The same point, P say, could be specified in any other coordinate system by giving another set of numbers $\bar{q}_1, \bar{q}_2, \bar{q}_3$, possibly with a quite different significance e.g. they could be the Cartesian coordinates x, y, z. Every coordinate must be a *function of position* of P, expressible in terms of whatever coordinates we care to use; and consequently

$$\bar{q}_i = \bar{q}_i(q_1, q_2, ...), \quad q_i = q_i(\bar{q}_1, \bar{q}_2, ...). \tag{3.1}$$

It is clear, however, that the position vector of P can no longer be written in the form $\mathsf{r} = q_1\mathsf{e}_1 + q_2\mathsf{e}_2 + q_3\mathsf{e}_3$ because there are no basis vectors e_i corresponding to a universal set of rectilinear axes. On the other hand, it is possible to define at any point P a *local* set of basis vectors $\mathsf{e}_1, \mathsf{e}_2, \mathsf{e}_3$ which are simply unit normals to the three surfaces $q_1 =$ constant, $q_2 =$ constant, $q_3 =$ constant, respectively, at point P. Thus, in spherical polar coordinates, e_1 will be a vector normal to the sphere on which q_1 has a constant value, as in Fig.2, while e_2 will be normal to the surface of a cone, and e_3 to a plane containing the z axis. In terms of these vectors, the infinitesimal $d\mathsf{r}$ which leads to point P' with $\mathsf{r}' = \mathsf{r} + d\mathsf{r}$ will be expressed as

$$d\mathsf{r} = \mathsf{e}_1 dq^1 + \mathsf{e}_2 dq^2 + \mathsf{e}_3 dq^3, \tag{3.2}$$

where we anticipate the transformation properties of the components by using superscripts on the differentials, emphasizing also the fact that the vector $d\mathsf{r}$ is an invariant. In fact, if new coordinates $\bar{q}_1, \bar{q}_2, \bar{q}_3$ are introduced, so that point P is alternatively identified as the intersection of three *new* coordinate surfaces, then it will be possible to write also

$$d\mathsf{r} = \bar{\mathsf{e}}_1 d\bar{q}^1 + \bar{\mathsf{e}}_2 d\bar{q}^2 + \bar{\mathsf{e}}_3 d\bar{q}^3. \tag{3.3}$$

The change $\{\mathsf{e}_i\} \to \{\bar{\mathsf{e}}_i\}$ corresponds to a *local* rotation of basis; and if $d\mathsf{r}$ is to remain invariant then the change $\{dq^i\} \to \{d\bar{q}^i\}$ must follow a contragredient transformation law.

All the theory developed in Chapter 1 can now be taken over in its entireity, with matrices $\mathbf{U}, \breve{\mathbf{U}}$ describing, respectively, the changes of components and of basis vectors at any point P. The only new features will be

- the elements of the matrices vary continuously, from point to point in space, and will thus be *continuous functions of the coordinates*;
- the lengths and inclinations of the basis vectors $\{\mathsf{e}_i\}$ will also vary continuously with the coordinates of the point to which they refer
- the basic contravariant transformation law will refer to the change of components of an *infinitesimal vector* $d\mathsf{r}$, at the point P considered, when new coordinates are intoduced.

Such transformations are linear only because we are considering the *infinitesimal neighbourhood* of P: the relationship between the two sets of coordinates themselves (not their differentials) is clearly *nonlinear*.

To identify the elements of the matrices $\mathbf{U}, \breve{\mathbf{U}}$ we start from the functional relationships (3.1) and consider two infinitely near points, with coordinates $\{q_i\}$ and $\{q_i + dq_i\}$, in the one frame, and $\{\bar{q}_i\}$ and $\{\bar{q}_i + d\bar{q}_i\}$, in the other. By simple partial differential calculus,

$$d\bar{q}^i = \sum_j \left(\frac{\partial \bar{q}_i}{\partial q_j}\right) dq^j = \sum_j U_{ij} dq^j$$

where the indices on the differentials are put in the superscript position, as in (3.2), to anticipate their contravariant character[1].

It is also convenient to adopt the conventions introduced at the end of Chapter 1, writing the row and column indices of the matrix $\mathbf{U}$ in the upper and lower positions, respectively, so that the last equation takes the form

$$d\bar{q}^i = \sum_j U^i_j dq^j, \tag{3.4}$$

where

$$U^i_j = \left(\frac{\partial \bar{q}_i}{\partial q_j}\right). \tag{3.5}$$

[1]Note that the coordinates themselves, $\{q_i\}$ are *not* components of a tensor; and their indices are left in the subscript position.

Invariance of the infinitesimal vector $d\mathbf{r}$, against change of coordinate system, then implies that in the same change

$$\mathsf{e}_i \to \bar{\mathsf{e}}_i = \sum_j \breve{U}_i^j \mathsf{e}_j \tag{3.6}$$

where $\breve{U}_i^j = (\mathbf{U}^{-1})_{ji}$. The summations in (3.4) and (3.6) both involve one upper index and one lower, the remaining 'free' index showing the tensor character of the result.

To express the elements of $\breve{\mathbf{U}}$ directly as partial derivatives, without inverting a matrix, we may start from the second relationship in (3.1), expressing the differential dq_i first in terms of the $\bar{q}_j$ and then again in terms of the q_k, to obtain

$$dq_i = \sum_j \left(\frac{\partial q_i}{\partial \bar{q}_j}\right) d\bar{q}_j = \sum_{jk} \left(\frac{\partial q_i}{\partial \bar{q}_j}\right)\left(\frac{\partial \bar{q}_j}{\partial q_k}\right) dq_k.$$

This equation expresses dq_i in the form $\sum_k A_{ik} dq_k$ and is satisfied only when $A_{ik} = \delta_{ik}$. Thus

$$\sum_j \left(\frac{\partial q_i}{\partial \bar{q}_j}\right)\left(\frac{\partial \bar{q}_j}{\partial q_k}\right) = \delta_{ik}. \tag{3.7}$$

The second partial derivative is the jk-element of $\mathbf{U}$: the first must therefore be the ij-element of $\mathbf{U}^{-1}$ (i.e. the ji-element of $\breve{\mathbf{U}}$), so that the 'chain rule' for a matrix product will give the ik-element of the unit matrix. With the same notation as in (2.8) $(\breve{\mathbf{U}})_{ji} = \breve{U}_j^i$ and (3.7) becomes

$$\sum_j \breve{U}_j^i U_k^j = \delta_k^i, \tag{3.8}$$

where (cf.(3.3))

$$\breve{U}_j^i = \left(\frac{\partial q_i}{\partial \bar{q}_j}\right). \tag{3.9}$$

In the standard textbooks on tensor calculus it is customary to write all matrix elements explicitly as partial derivatives. The standard transformation laws for rank-1 contravariant and covariant tensors thus become, respectively,

$$A^i \to \bar{A}^i = \sum_j \left(\frac{\partial \bar{q}_i}{\partial q_j}\right) A^j, \tag{3.10}$$

$$B_i \to \bar{B}_i = \sum_j \left(\frac{\partial q_j}{\partial \bar{q}_i}\right) B_j, \tag{3.11}$$

and tensors of higher rank, with m degrees of contravariance and n of covariance, may be defined as in (2.3). Thus, for example,

$$C_k^{ij} \to \bar{C}_k^{ij} = \sum_r \sum_s \sum_t \left(\frac{\partial \bar{q}_i}{\partial q_r}\right)\left(\frac{\partial \bar{q}_j}{\partial q_s}\right)\left(\frac{\partial q_t}{\partial \bar{q}_k}\right) C_t^{rs} \tag{3.12}$$

is the transformation law for a rank-3 tensor with two degrees of contravariance and one of covariance.

The summation convention

It is already clear that the equations of the tensor calculus, although expressing only a simple generalization of the equations of Section 1.1 for *linear* vector spaces, are becoming quite cumbersome; and that their interpretation requires a fair amount of mental gymnastics. Such difficulties are greatly reduced by using a *summation convention*, due to Einstein – with which explicit summation symbols are completely eliminated. It may be stated as follows:

> Whenever an index is *repeated*, in the expression on one side of a tensor equation, then summation over all values of that index will be *understood*.

The resultant simplification may be carried one step further by using the contravariant and covariant matrix elements, $U^i_j, \breve{U}^i_j$ defined in (3.7) and (3.9), in place of the partial derivatives they represent. The economy of this notation is evident when we rewrite the last three equations to obtain

$$\begin{aligned} \bar{A}^i &= U^i_r\, A^r, \\ \bar{B}_i &= \breve{U}^r_i\, B_r, \\ \bar{C}^{ij}_k &= U^i_r U^j_s \breve{U}^t_k\, C^{rs}_t. \end{aligned} \qquad (3.13)$$

The text assumes a more peaceful appearance and the significance of the indices, as contravariant or covariant, becomes obvious: in particular, every summation involves one upper index and one lower; and the positions of the indices relate to the upper and lower parts of each partial derivative in (3.5) and (3.9). The 'summation convention' will be used in this and all later Chapters.

3.2 Volume elements, tensor densities, and volume integrals

In three dimensions, using rectangular Cartesian coordinates, an infinitesimal element of volume is easily defined as $dV = dxdydz$ – the product of the lengths of the sides of a rectangular box, normally taken positive. More generally, using an arbitrary coordinate system and tensor notation, dV will have to be related to three infinitesimal displacements d_1q^i, d_2q^j, d_3q^k – where it will be convenient to define each in terms of its contravariant components. But clearly the expression[2]

$$dV^{ijk} = d_1q^i,\ d_2q^j,\ d_3q^k \qquad (3.14)$$

will not serve the purpose, being simply one component of a rank-3 contravariant tensor and taking no account of the metric of the space.

With rectangular axes and displacements along the three axial directions, each vector would have only one non-zero component and $dV^{123} = d_1q^1 d_2q^2 d_3q^3$ would apparently

[2]It is understood that dV^{ijk} depends on the three displacements chosen, but to avoid confusion the labels 1,2,3 are suppressed.

be satisfactory; but even then it would lack uniqueness – for the arbitrary reversal of any axis would introduce a $-$ sign, while enumeration of the three displacements in a different order would make no difference to the product. Evidently we must seek a recipe for obtaining a single invariant number from the rank-3 tensor components, which will reduce to the elementary definition on specializing to a rectangular Cartesian system.

To this end let us consider the replacement of (3.14) by the determinant

$$dV^{ijk} = \begin{vmatrix} d_1q^i & d_1q^j & d_1q^k \\ d_2q^i & d_2q^j & d_2q^k \\ d_3q^i & d_3q^j & d_3q^k \end{vmatrix}. \tag{3.15}$$

This is an *antisymmetric* rank-3 tensor whose expansion contains only non-zero products of the form $\pm dV^{123}$; and it thus depends on the single number $dV = dV^{123}$ which is the product of the diagonal elements and, when the axes are rectangular Cartesian, coincides with the volume element determined by elementary geometry. To confirm that (3.15) provides a satisfactory definition, a considerable digression is necessary; and this will introduce the new concept of a *tensor density*[3]

We start by introducing a convenient method of manipulating determinants. This depends on the introduction of the *Levi-Civita symbol* (not in itself a tensor) defined by

$$\begin{aligned} \epsilon_{ijk} &= 0, \quad \textit{no index repeated}, \\ &= +1, \quad ijk\ \textit{an even permutation of}\ 123, \\ &= -1, \quad ijk\ \textit{an odd permutation of}\ 123. \end{aligned} \tag{3.16}$$

Any determinant

$$D = \begin{vmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{vmatrix}. \tag{3.17}$$

can then be written in the form

$$D = (3!)^{-1}\epsilon_{ijk}\epsilon_{lmn}d_{il}d_{jm}d_{kn} \tag{3.18}$$

– for the implied summations include all possible permutations of the *same* product of d-factors and the result must therefore be divided by 3!.

Two determinants of particular importance are (i) that of the metric matrix $|\mathbf{g}|$; and (ii) that of the matrix $|\mathbf{U}|$ in the basic contravariant transformation law. From the last result these can be written

$$\begin{aligned} g = |\mathbf{g}| &= (3!)^{-1}\epsilon_{ijk}\epsilon_{lmn}g_{il}g_{jm}g_{kn}, \quad (a) \\ U = |\mathbf{U}| &= (3!)^{-1}\epsilon_{ijk}\epsilon_{lmn}U^i_lU^j_mU^k_n. \quad (b) \end{aligned} \tag{3.19}$$

From (3.19a) it follows, on putting $\mathbf{g} = \mathbf{1}$, that

$$\epsilon_{ijk}\epsilon_{ijk} = 3!. \tag{3.20}$$

[3] Also referred to (e.g. by Fock) as a 'pseudo-tensor'; and sometimes as a 'relative tensor'.

Next let us define a rank-3 contravariant tensor, antisymmetric in all indices, by

$$E^{ijk} = E\epsilon_{ijk}, \tag{3.21}$$

so that $E = E^{123}$ is the single component that determines all others. In the transformation described by the matrix $\mathbf{U}$,

$$E^{ijk} \to \bar{E}^{ijk} = E^{lmn} U^i_l U^j_m U^k_n$$

and, on multiplying by ϵ_{ijk} and using (3.20), this gives

$$\bar{E}3! = E\epsilon_{lmn}\epsilon_{ijk} U^i_l U^j_m U^k_n = E3!U.$$

Consequently

$$\bar{E} = EU, \quad (U = \det|U^i_j|) \tag{3.22}$$

and ϵ_{ijk} in (3.16) is not a tensor until the scalar factor E is attached, as in (3.21). However, from (3.19a), $E^{ijk}E_{lmn}g_{il}g_{jm}g_{kn}$ is certainly an invariant (upper and lower indices being fully contracted) and use of (3.18) then reduces the expression to $3!E^2g$. Thus,

$$E^2 g = \text{invariant}, \qquad g = \det|g_{ij}|. \tag{3.23}$$

It follows that, in the transformation with matrix $\mathbf{U}$,

$$E^2 g = \bar{E}^2 \bar{g} = (EU)^2 \bar{g}$$

– where (3.22) has been used – and hence $g = U^2\bar{g}$ and $\sqrt{\bar{g}} = U^{-1}\sqrt{g}$. We know, however, that $U^{-1} = \breve{U}$, the determinant of $\breve{\mathbf{U}}$; and the transformation law for $\sqrt{g}$ therefore becomes

$$\sqrt{g} \to \sqrt{\bar{g}} = J\sqrt{g}, \tag{3.24}$$

where J is the *Jacobian* of the coordinate transformation, defined by (using (refc9) for the elements of $\breve{\mathbf{U}}$

$$J = \det\left|\frac{\partial q_i}{\partial \bar{q}_j}\right|. \tag{3.25}$$

Volume integrals and tensor densities

After this digression let us return to the problem of defining an invariant measure of the volume enclosed within some specified region of space (e.g. that bounded by surfaces on which $\bar{q}_1, \bar{q}_2, \bar{q}_3$ have given constant values). The whole volume may be divided into elements $dV(q_1, q_2, q_3)$ and we shall write

$$V = \int dV(q_1, q_2, q_3) \tag{3.26}$$

where dV is the 123-component of the tensor in (3.15) and, in rectangular Cartesian coordinates, will reduce to the elementary definition $dV = dq_1dq_2dq_3$ (a product of displacements along the axial directions, each with a single non-zero component). When a more general coordinate system is employed, the corresponding volume element will be expressed, using (3.26) and (3.24), as

$$J\sqrt{g}\,dV(\bar{q}_1, \bar{q}_2, \bar{q}_3).$$

For rectangular Cartesian coordinates, $g = 1$ and $J = 1$ and this reduces to the elementary definition; but it is now clear that the correct definition of the volume element is not $dV = dq_1 dq_2 dq_3$ but rather

$$dV = \sqrt{g}\,dq_1 dq_2 dq_3 \tag{3.27}$$

and that in a change of coordinate system

$$dV \to J\sqrt{g}\,d\bar{q}_1, d\bar{q}_2, d\bar{q}_3. \tag{3.28}$$

It is also clear that in evaluating a *quantity* as the integral of a scalar *density* $T = T(q_1, q_2, q_3)$ (i.e. "so much per unit volume"), the transformation law for the appropriate integral will be

$$\int T(q_1, q_2, q_3)\sqrt{g}\,dq_1 dq_2 dq_3 \to \int JT(\bar{q}_1 \bar{q}_2 \bar{q}_3)\sqrt{g} d\bar{q}_1, d\bar{q}_2, d\bar{q}_3. \tag{3.29}$$

In other words, given a scalar function T (a rank-0 tensor), then multiplication by $\sqrt{g}$ yields a rank-0 *tensor density*, usually denoted by $\mathcal{T}$,

$$\mathcal{T} = \sqrt{g}\,T,$$

whose transformation law under change of variables in an integral is simply $\mathcal{T} \to \bar{\mathcal{T}} = J\mathcal{T}$. The same argument applies to a tensor of any rank: thus, for a tensor T_k^{ij} with two degrees of contravariance and one of covariance, a related tensor *density* may be defined as

$$\mathcal{T}_k^{ij} = \sqrt{g}\,T_k^{ij}. \tag{3.30}$$

Since, according to (3.24), $\sqrt{g}$ is merely multiplied by the Jacobian J in a change of variables, the transformation law for the tensor density (3.30) will be

$$\bar{\mathcal{T}}_k^{ij} = JU_l^i U_m^j \breve{U}_k^n T_n^{lm}. \tag{3.31}$$

It is to be noted that, since $J = \pm 1$ for any transformation that leaves the length of a vector invariant ($+1$ for a proper rotation of axes, -1 for an improper rotation), a tensor *density* behaves in exactly the same way as the tensor itself except that the components suffer a sign change when the transformation is improper (e.g. changing from a right-handed to a left-handed system of coordinates).

An important example of a tensor density, usually referred to as a pseudo-vector, is provided by the tensor formed from two rank-1 covariant tensors, B_i, C_j. The rank-2 tensor T_{ij}, comprising (in three dimensions) all nine components $T_{ij} = B_i C_j$, has no particular symmetry under interchange of indices; but it may be written as the sum of symmetric and anti-symmetric parts $\frac{1}{2}(T_{ij} + T_{ji})$ and $\frac{1}{2}(T_{ij} - T_{ji})$. From the anti-symmetric tensor,

$$A_{ij} = (T_{ij} - T_{ji}),$$

a rank-1 tensor density may be constructed, by using the Levi-Civita symbol (3.16), in the form

$$\mathcal{A}_i = \tfrac{1}{2}\epsilon_{ijk}A_{jk}$$

and has components

$$\begin{array}{lll} \mathcal{A}_1 & = \tfrac{1}{2}(A_{23} - A_{32}) & = A_{23} \\ \mathcal{A}_2 & = \tfrac{1}{2}(A_{31} - A_{13}) & = A_{31} \\ \mathcal{A}_3 & = \tfrac{1}{2}(A_{12} - A_{21}) & = A_{12}. \end{array}$$

In this case, where the tensor A_{ij} is formed from two vectors B_i, C_j, it is evident that these components, namely

$$(B_2C_3 - B_3C_2), \quad (B_3C_1 - B_1C_3), \quad (B_1C_2 - B_2C_1),$$

are those of the so-called vector product $\mathsf{B} \times \mathsf{C}$. In fact the vector product is not a vector but rather a rank-2 tensor density or, more briefly, a *pseudo-vector*. Its components, more correctly indicated by the double indices, transform like those of a vector under pure rotations of axes but show an additional change of sign under a reflection or inversion.

3.3 Differential operators: the covariant derivative

The laws of physics are commonly expressed in the form of partial differential equations, which describe the behaviour of *fields*. The fields vary from point to point in space and may refer to scalar quantities, such as a temperature or an electric potential; or to a vector quantity, such as an electric field, with three components E_x, E_y, E_z; or to more general many-component quantities such as the 9-component stress tensor in an elastic medium. In all cases, the components are functions of position; and if the coordinate system is changed the components will change according to the tensor character of the quantities they describe. An important application of the tensor calculus is to the formulation of physical laws in a form which is invariant against changes of coordinate system. To this end it is necessary to study the *differential operators* which determine how a tensor quantity changes in moving from one point in space to a neighbouring point.

A differential operator familiar from elementary vector analysis is the *gradient* ∇, which describes the rate of change of a scalar field along the directions of three unit vectors. If we use a general coordinate system the gradient operator may be written (noting the summation convention)

$$\nabla = \mathsf{e}^j \mathsf{D}_j = g^{ij}\mathsf{e}_i\mathsf{D}_j \tag{3.32}$$

where e^j and e_i are, respectively, contravariant basis vectors and covariant reciprocal vectors at the point considered and $\mathsf{D}_j = \partial/\partial q_j$ is the partial derivative with respect to the generalized coordinate q_j. Both expressions for ∇ in (3.32) display the fact that the result is invariant against change of coordinate system, summation over the repeated indices (one upper, the other lower) eliminating any degrees of covariance or contravariance – to leave an *invariant*. In particular, the first expression for ∇ confirms that D_j is a member of a *co*variant set.

It is clearly a simple matter, once the metric tensor is known, to derive the form of the gradient operator in any given coordinate system.

To obtain the corresponding derivatives of vector and tensor fields, however, is less easy and requires the introduction of the 'covariant derivative'. The root of the difficulty is that a vector or tensor quantity implicitly refers to a set of axes in space i.e. to the vectors of a basis; and, as already noted, the basis vectors themselves vary from point to point in space. The *covariant derivative* is introduced simply to take account of this variation.

The covariant derivative

Let us consider first a *vector* field in which, at any point, $\mathsf{v} = \mathsf{e}_i v^i = \mathsf{e}^i v_i$, in terms of covariant or contravariant basis vectors. The change in any vector v, on moving from point P to a neighbouring point P', will be

$$d\mathsf{v} = dv_j \mathsf{e}^j + v_j d\mathsf{e}^j = dv^j \mathsf{e}_j + v^j d\mathsf{e}_j \tag{3.33}$$

and it will thus be necessary to know how e_j or e^j changes in an infinitesimal change of coordinates. The rate of change of e_j with respect to q_i will be a vector $\mathsf{D}_i\mathsf{e}_j$, expressible as a linear combination of the vectors e_k, and may thus be written

$$\mathsf{D}_i\mathsf{e}_j = \Gamma^k_{ij}\mathsf{e}_k. \tag{3.34}$$

The quantities Γ^k_{ij} are called the "coefficients of an affine connection", being definable even when the space does not possess a metric. But for a metric space they may be defined also as scalar products: thus, taking the scalar product of (3.34) with a particular reciprocal vector e^k, will leave only one term on the right-hand side and this will be

$$\Gamma^k_{ij} = \mathsf{e}^k \cdot \mathsf{D}_i\mathsf{e}_j. \tag{3.35}$$

In this context Γ^k_{ij} is usually referred to as a "Christoffel symbol of the second kind" and denoted by $\{ij, k\}$.

The Christoffel symbols may also be related to the elements of the metric. Thus,

$$\mathsf{D}_k g_{ij} = \mathsf{D}_k\mathsf{e}_i \cdot \mathsf{e}_j + \mathsf{e}_i \cdot \mathsf{D}_k\mathsf{e}_j = \Gamma^l_{ki} g_{lj} + \Gamma^l_{kj} g_{il},$$

where summation over l is implied after the last equality. Since multiplication by g_{il}, followed by the summation, has the effect of lowering a contravariant index l, it is natural to express the last result as

$$\mathsf{D}_k g_{ij} = \Gamma_{ki,j} + \Gamma_{kj,i}.$$

On making cyclic permutations of the indices in this expression ($kij \to ijk, jki$) and adding the results, it follows easily that

$$\Gamma_{ij,k} = \tfrac{1}{2}[\mathsf{D}_i g_{jk} + \mathsf{D}_j g_{ki} - \mathsf{D}_k g_{ij}]. \tag{3.36}$$

The quantity $\Gamma_{ij,k}$ is a "Christoffel symbol of the first kind" and is usually denoted by $[ij, k]$. It is clearly sufficient to know the latter since the index k may easily be raised:

$$\Gamma^k_{ij} = g^{kn}\Gamma_{ij,n}, \tag{3.37}$$

or, with Christoffel notation[4], $\{ij,k\} = g^{kn}[ij,n]$.

Now that the coefficients in (3.34) may be regarded as known, we can return to the discussion of how a vector changes on moving from P to a neighbouring point. Let us take the vector with covariant components, $\mathsf{v} = \mathsf{e}^j v_j$, and consider the differential $d\mathsf{v} = v_j d\mathsf{e}^j + dv_j \mathsf{e}^j$. On dividing by dq_l and passing to the limit, it follows that

$$\mathsf{D}_l\mathsf{v} = (\mathsf{D}_l\mathsf{e}^j)v_j + \mathsf{e}^j\mathsf{D}_l v_j, \tag{3.38}$$

where the first term, arising from variation of the basis vectors, would be absent in a Cartesian coordinate system. The derivative vector $\mathsf{D}_l\mathsf{v}$ may be expressed in terms of either direct or reciprocal vectors: choosing the latter, we define (note the semicolon in the subscript!)

$$\mathsf{D}_l\mathsf{v} = v_{i;l}\mathsf{e}^i,$$

where $v_{i;l}$ stands for the ith covariant component of the derivative with respect to coordinate q_l. As usual, this component can be picked out by taking a scalar product with the corresponding contragredient vector e_i. Thus, making use of (3.38),

$$v_{i;l} = \mathsf{e}_i \cdot (\mathsf{D}_l\mathsf{v}) = \mathsf{e}_i \cdot (\mathsf{D}_l\mathsf{e}^j)v_j + \delta_i^j\mathsf{D}_l v_j.$$

The first term on the right-hand side of this expression contains essentially a Christoffel symbol: for if we take the derivative of the scalar product $\mathsf{e}^k \cdot \mathsf{e}_i$, which is of course an invariant, we obtain

$$\mathsf{D}_j(\mathsf{e}^k \cdot \mathsf{e}_i) = \mathsf{e}_i\mathsf{D}_j\mathsf{e}^k + \mathsf{e}^k\mathsf{D}_j\mathsf{e}_i,$$

in which the second term in the sum is the Christoffel symbol $\{ij,k\}$ defined in (3.35) *et seq*. Evidently, then,

$$v_{i;l} = \frac{\partial v_i}{\partial q_l} - \{il,j\}v_j. \tag{3.39}$$

This equation defines the **covariant derivative** of the vector whose covariant components are v_i, with respect to the coordinate q_l; the first term (often denoted by $v_{i,l}$) is simply the derivative of the component v_i itself, while the second term arises from the change in the basis vector e^i to which it is referred. In a Cartesian system, the second term is always absent, all the Christoffel symbols being zero.

Absolute differentials

It is now important to establish that the quantities $v_{i;l}$ do indeed form a set with tensor properties, and to show how one can construct an *absolute differential* $d\mathsf{v}$ - an infinitesimal vector that will be invariant against change of coordinate system. To do this, we must evaluate the same quantities in a new coordinate system $\{\bar{q}_i\}$ and establish their transformation law. Instead of v_i, D_l, and e^j we shall then have corresponding quantities

$$\bar{v}_i = \breve{U}_i^r v_r, \quad \bar{\mathsf{D}}_l = \breve{U}_l^s\mathsf{D}_s, \quad \bar{\mathsf{e}}^j = U_t^j\mathsf{e}^t,$$

the first two being covariant, the third contravariant. These equations simply express the known tensor character of all three quantities. We now need to evaluate the analogue of

[4] Nowadays used less widely than in the classic texts.

equation (3.39) in the new coordinate system. Thus, the first term will become

$$\frac{\partial \bar{v}_i}{\partial \bar{q}_l} = \breve{U}_i^r \breve{U}_l^s \frac{\partial \bar{v}_r}{\partial \bar{q}_s}$$

and, remembering that $\{il, j\}$=$\mathsf{e}_i \cdot \mathsf{D}_l \mathsf{e}^j$, the new Christoffel symbol will be

$$\overline{\{il, j\}} = \breve{U}_i^s \mathsf{e}_s \cdot \breve{U}_l^t \mathsf{D}_t U_u^j \mathsf{e}^u = \breve{U}_i^s \breve{U}_l^t U_u^j \{st, u\}. \tag{3.40}$$

The second term on the right in (3.39) then becomes, using (3.40),

$$-\overline{\{il, j\}} = \breve{U}_j^k v_k \breve{U}_i^s \breve{U}_l^t U_u^j \{st, u\} = \breve{U}_i^s \breve{U}_l^t \{st, k\} v_k,$$

where it was noted that $\breve{U}_j^k U_u^j = \delta_u^k$. Finally then, substituting this result in (3.39),

$$\bar{v}_{i;l} = \breve{U}_i^s \breve{U}_l^t \left(\frac{\partial v_t}{\partial q_s} - \{st, k\} v_k \right) = \breve{U}_i^s \breve{U}_l^t v_{s,\,t} \tag{3.41}$$

and it follows that the set of covariant derivatives constitutes a rank-2 tensor, covariant in both indices i, l.

Exactly similar reasoning may be applied to a contravariant vector, with components v^i: in this case the covariant derivative is defined as

$$v^i_{;l} = \frac{\partial v^i}{\partial q_l} - \{jl, i\} v^j \tag{3.42}$$

and the transformation properties corresponding to (3.41) are found to be

$$\bar{v}^i_{;l} = U_s^i \breve{U}_l^t v^t_{,t} \;. \tag{3.43}$$

The covariant derivative of a contravariant vector, with components v^i, is thus again a rank-2 tensor, but is contravariant in the index i and covariant in the second index.

It is easily verified that

$$g_{ij} v^i_{;l} = v_{j;l} \tag{3.44}$$

– which confirms that $v^i_{;l}$ and $v_{i;l}$ are components of the same vector – namely the covariant derivative of the original v – and obey the usual rules for raising and lowering of the indices. It is also clear that if the components $v^i_{;l}$ or $v_{i;l}$ are multiplied by the infinitesimal coordinate change dq^l and summed the result will be

$$dv^i = v^i_{;\,l} dq^l \qquad dv_i = v_{i;\,l} dq^l. \tag{3.45}$$

These are, respectively, the contravariant and covariant components of an **absolute differential** $d\mathsf{v}$, which is invariant against change of coordinates and may therefore be associated with a real physical quantity.

3.4 Divergence, curl, and the Laplacian

Now that the covariant derivative has been defined, through (3.42) and (3.44), for either the contravariant or covariant components of any field vector, it is possible to pass to higher derivatives. An important second-order differential operator is the 'divergence', defined in Cartesian vector space as

$$\text{div } \mathsf{v} = \left(\frac{\partial v_x}{\partial x}\right) + \left(\frac{\partial v_y}{\partial y}\right) + \left(\frac{\partial v_z}{\partial z}\right) = \mathsf{D}_i v^i, \tag{3.46}$$

where summation is implied, as usual, in the last form. The natural generalization, on transforming to curvilinear coordinates, will then be

$$\text{div } \mathsf{v} = v^i_{;i} = \mathsf{D}_i v^i + v^j \Gamma^i_{ji}, \tag{3.47}$$

which involves the covariant derivatives given in (3.42) and reduces to (3.46) for the Cartesian case where all Christoffel symbols are zero. This quantity is a *scalar invariant* (tensor of rank zero) as a result of contraction over identical indices.

It remains only to determine the Christoffel symbols; and this may be done by relating them to the metric tensor, presented as a matrix in (3.13). The *determinant* of a matrix $\mathbf{g}$ is usually expressed in the form

$$g = |\mathbf{g}| = g_{ij} G^{ij}, \tag{3.48}$$

where G^{ij} is the *cofactor* of element g_{ij}. But it was shown in Section 1.6 that g_{ij} has a contravariant counterpart g^{ij}, whose elements are those of the inverse matrix $\mathbf{g}^{-1}$, and consequently $g_{ij} g^{ij} = 1$. Equation (3.48) is therefore a trivial identity, in which $G^{ij} = g\, g^{ij}$.

The Christoffel symbols appear when we differentiate the determinant with respect to the parameters q_i which define the coordinate system to be used. It is well known that the derivative with respect to any parameter, q_k say, is obtained by differentiating each element separately, multiplying by its cofactor, and summing. Thus, from (3.48), with $G^{ij} = gG^{ij}$ and $g_{ij} = \mathsf{e}_i \cdot \mathsf{e}_j$,

$$\begin{aligned}
\mathsf{D}_k g &= (\mathsf{D}_k g_{ij}) G^{ij} + g g^{ij} \mathsf{D}_k(\mathsf{e}_i \cdot \mathsf{e}_j) \\
&= g g^{ij} [(\mathsf{D}_k \mathsf{e}_i) \cdot \mathsf{e}_j + \mathsf{e}_i \cdot (\mathsf{D}_k \mathsf{e}_j)] \\
&= g g^{ij} [(\Gamma^l_{ki} \mathsf{e}_l) \cdot \mathsf{e}_j + \mathsf{e}_i \cdot (\Gamma^l_{kj} \mathsf{e}_l) \cdot \mathsf{e}_j],
\end{aligned}$$

where each of the terms on the right may be reduced separately by using the basic properties (equation (2.9) *et seq*) of the metric tensor g^{ij}. On taking the first term,

$$[\Gamma^l_{ki} \mathsf{e}_l \cdot \mathsf{e}_j] = [\Gamma^l_{ki} g_{lj}] = [\Gamma_{jk,i}],$$

where index l was replaced by covariant j in the summation over l. When this result is multiplied by $g g^{ij}$ and summed the result is $g \Gamma^i_{ki}$. The second term reduces in a similar way, to give an identical result, and consequently

$$\mathsf{D}_k g = 2g \Gamma^i_{ki}, \tag{3.49}$$

which could be used in (3.47) to give

$$\text{div } \mathsf{v} = \mathsf{D}_i v^i + v^j \mathsf{D}_j g/(2g).$$

A neater result is obtained, however, on multiplying this last equation by $\sqrt{g}$ and noting that the right-hand side then becomes the derivative $\mathsf{D}_i(v^i\sqrt{g})$. In this way we find

$$\text{div } \mathsf{v} = \frac{1}{\sqrt{g}} \mathsf{D}_i(v^i\sqrt{g}) = \frac{1}{\sqrt{g}} \frac{\partial}{\partial q_i} \left(v^i\sqrt{g})\right). \tag{3.50}$$

This is a completely general expression for the divergence of v at any point in a vector field, using curvilinear coordinates of any kind.

Chapter 4

Applications in Relativity Theory

4.1 Elements of special relativity

The special theory of relativity provides the simplest possible application of general tensor methodology. There is a fundamental invariant, on which the whole of the theory is based: it is the square of the *interval* between two *events*, each specified by *four* coordinates (x, y, z, ct) – the three spatial coordinates of a point, together with the time (multiplied by c, the velocity of light in free space). The interval is in fact taken to be

$$ds : \quad (dx,\ dy,\ dz,\ cdt)$$

and the fundamental invariant is

$$ds^2 = -dx^2 - dy^2 - dz^2 + c^2dt^2 = c^2d\tau^2 = \text{invariant}. \tag{4.1}$$

By writing the invariant as $c^2d\tau^2$ we simply introduce a *proper time interval* between the two events; and (4.1) implies that events will be recorded in a *four-dimensional* space with metric

$$g_{11} = g_{22} = g_{33} = -1, \quad g_{44} = +1, \tag{4.2}$$

all other elements of the matrix $\mathbf{g}$ being zero. This type of vector space was first used by Minkowski (1908): it is a linear space, in which orthogonal basis vectors have been set up, and the metric is taken to be independent of position (i.e. values of the variables x, y, z, t). The differential form of the squared distance between two 'points' therefore applies even when the points are not close together; and we may write

$$s^2 = -x^2 - y^2 - z^2 + c^2t^2 = \text{invariant}, \tag{4.3}$$

where x, y, z, ct may be the coordinates of any event, relative to an (arbitrary) origin with $x, y, z, ct\ = 0$. An event is typically a point on a wavefront, originating at the origin at time $t = 0$ and propagating outwards with velocity c, or a particle in motion at the given point.

If the same event is recorded by a second observer, in uniform motion relative to the first, it will be given coordinates (again including the time!) $\bar{x}, \bar{y}, \bar{z}, \bar{t}$ and the squared distance to the point in question will be

$$\bar{s}^2 = -\bar{x}^2 - \bar{y}^2 - \bar{z}^2 + c^2\bar{t}^2. \tag{4.4}$$

The 'invariance of the interval' requires that s^2 as measured by the two observers will be the same *for a common value (c) of the velocity of light propagation*. And this principle, experimentally verified, implies that the two sets of coordinates are related in a particular way, namely by a *Lorentz transformation*. The simplest example of such a transformation is the one that relates the space-time coordinates for two reference frames: (i) that located at the origin (Observer 1), and (ii) one travelling with speed u along the x axis (Observer 2). The coordinates of an event, as recorded by Observer 1, will be denoted by x, y, z, ct, while those for Observer 2 will be $\bar{x}, \bar{y}, \bar{z}, c\bar{t}$.

The 'common sense' relationship between the two sets of coordinates is simply

$$\bar{x} = x - ut, \quad \bar{y} = y, \quad \bar{z} = z, \quad \bar{t} = t \tag{4.5}$$

but $\bar{s}^2$ obtained by substituting (4.5) in (4.4) does not agree with s^2 given by (4.3). To achieve the required invariance, the transformation law must be modified. Instead of (4.5), the correct coordinates to use in Frame 2 turn out to be

$$\bar{x} = \beta(x - ut), \quad \bar{y} = y, \quad \bar{z} = z, \quad \bar{t} = \beta(t - ux/c^2), \tag{4.6}$$

where β also depends on the relative velocity u and is

$$\beta = \beta_u = \frac{1}{\sqrt{(1 - u^2/c^2)}}. \tag{4.7}$$

The invariance of s^2 is confirmed by substituting (4.6) in (4.4) and finding $\bar{s}^2$.

The equations in (4.6) define the 'standard' Lorentz transformation – which is easily inverted to give

$$x = \beta(\bar{x} + u\bar{t}), \quad y = \bar{y}, \quad z = z, \quad t = \beta(\bar{t} + u\bar{x}/c^2), \tag{4.8}$$

It is unnecessary to consider the more general situation, in which Frame 2 moves in *any* given direction relative to Frame 1, because we are setting up *tensor* equations – which must remain valid under any corresponding linear transformation of coordinates.

Some important four-vectors

Let us now introduce the general tensor notation of earlier chapters. It is customary to replace the coordinates x, y, z, ct by x_1, x_2, x_3, x_4 when the tensor formalism is used[1]. And in this case we express the interval through its contravariant components as

$$ds^\mu : \quad (dx^1 = dx, \quad dx^2 = dy, \quad dx^3 = dz, \quad dx^4 = cdt). \tag{4.9}$$

[1]It is not strictly necessary to use the general formalism in the special theory, where the distinction between spatial and temporal components may be recognised by introducing factors of i and a 'pseudo-Cartesian' metric with $\mathbf{g} = \mathbf{1}$. Here, however, we use the full notation that applies also in Einstein's *general* (1916) theory.

This choice will set the pattern for all subsequent transformation equations, as in previous Chapters; and covariant components may be obtained from it by lowering the indices using the metric tensor (4.2), $dx_\mu = g_{\mu\nu}dx^\nu$. Thus, the interval will have corresponding covariant components

$$ds_\mu : \quad (dx_1 = -dx, \quad dx_2 = -dy, \quad dx_3 = -dz, \quad dx_4 = cdt). \tag{4.10}$$

It is important to note that the proper time interval defined in (4.1) et seq has a special significance: A clock at rest relative to the the moving Frame 2 will show time $\bar{t}$ and, since its relative coordinates are constant (it is moving with the frame), $d\bar{x} = d\bar{y} = d\bar{z} = 0$ and thus $c^2d\tau^2$ will coincide with $c^2d\bar{t}^2$. This is true for *any* frame whose clock is at rest; and allows us to define immediately a *velocity* four-vector, for Frame 1, whose components will be

$$v^\mu = \frac{dx^\mu}{d\tau} = \frac{dx^\mu}{dt}\frac{dt}{d\tau} = \beta\frac{dx^\mu}{dt} \tag{4.11}$$

– since dt(Frame 1, clock at rest)$= \beta d\tau$, from (4.8), and hence $dt/d\tau = \beta$. The velocity four-vector will thus be, using a dot to indicate a time derivative,

$$v^\mu : \quad (v^1 = \beta\dot{x}^1, \quad v^2 = \beta\dot{x}^2, \quad v^3 = \beta\dot{x}^3, \quad v^4 = \beta c). \tag{4.12}$$

Similarly, for a particle with a postulated 'rest mass' m_0 (assumed to be an invariant natural characteristic), one may define a momentum four-vector, by multiplying (4.12) by m_0. Thus,

$$p^\mu : \quad (p^1 = m_0\beta\dot{x}^1, \quad p^2 = m_0\beta\dot{x}^2, \quad p^3 = m_0\beta\dot{x}^3, \quad p^4 = m_0\beta c). \tag{4.13}$$

For a particle at rest in Frame 2 its momentum relative to Frame 1 will be p^1 with $\dot{x}^1 = u$: and evidently the quantity

$$m = m_0\beta = \frac{m_0}{\sqrt{(1 - u^2/c^2)}} \tag{4.14}$$

will be an 'apparent mass' for Observer 1. The mass of a particle, moving relative to the observer, thus increases with its velocity, becoming infinite as u approaches c. On the other hand, when $u \ll c$, it follows that

$$m = m_0\beta \approx m_0(1 + \tfrac{1}{2}u^2/c^2) = m_0 + (\tfrac{1}{2}m_0u^2)/c^2 \tag{4.15}$$

and consequently that the classical kinetic energy $\frac{1}{2}m_0u^2$ makes a contribution to the mass. If we introduce Einstein's famous mass-energy relation, $E = mc^2$, this last result suggests that

$$E = mc^2 = m_0c^2 + \tfrac{1}{2}m_0u^2 + \dots \quad . \tag{4.16}$$

The four-vector in (4.13) is thus an *energy-momentum four-vector*, whose fourth component is

$$p^4 = m_0\beta c = E/c. \tag{4.17}$$

Before introducing *electromagnetic* four-vectors, we recall that the differential operators $\mathsf{D}_\mu = \partial/\partial x_\mu$ follow the covariant transformation (see xxx), and accordingly may be displayed as

$$\mathsf{D}_\mu : \quad \left(\mathsf{D}_1 = +\frac{\partial}{\partial x_1}, \quad \mathsf{D}_2 = +\frac{\partial}{\partial x_2}, \quad \mathsf{D}_3 = +\frac{\partial}{\partial x_3}, \quad \mathsf{D}_4 = \frac{\partial}{\partial x_4}\right), \tag{4.18}$$

where, since $dx_4 = cdt$, it follows that

$$\mathsf{D}_4 = \mathsf{D}^4 = \frac{1}{c}\frac{\partial}{\partial t}. \tag{4.19}$$

It is also clear that a four-dimensional equivalent of the operator ∇^2 may be formed by contraction: it is $\bar{\Box}^2$, defined as $\bar{\Box}^2 = \mathsf{D}_\mu \mathsf{D}^\mu$, where the contravariant D^μ is obtained by raising the index in the usual way. Thus

$$\bar{\Box}^2 = -\frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial x^2} + \frac{1}{c^2}\frac{\partial^2}{\partial t^2} = -\nabla^2 + \frac{1}{c^2}\frac{\partial^2}{\partial t^2}. \tag{4.20}$$

4.2 Tensor form of Maxwell's equations

Let us start from the three-dimensional Cartesian forms of the equations for the magnetic field due to a system of moving charges. The aim will be simply to cast the equations in a very succinct form by introducing four-vector notation and then to demonstrate that they are already Lorentz invariant.

In free space, the electromagnetic field at any point may be specified by giving the components (E_x, E_y, E_z) of the *electric field strength* $\mathbf{E}$, together with those of the *magnetic flux density* $\mathbf{B}$ (namely, B_x, B_y, B_z). We shall employ SI units throughout and use, where appropriate, the notation of elementary vector analysis. The differential equations satisfied by the fields, relating them to the electric charges and currents which produce them, are as follows:

$$\operatorname{div}\mathbf{E} = \epsilon_0 \rho, \quad (a) \qquad \operatorname{curl}\mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad (b) \tag{4.21}$$

where ρ is the *electric charge density* (charge per unit volume) at the field point, and

$$\operatorname{div}\mathbf{B} = 0, \quad (a) \qquad \operatorname{curl}\mathbf{B} = \frac{1}{c^2}\frac{\partial \mathbf{E}}{\partial t} + \mu_0 \mathbf{J}, \quad (b) \tag{4.22}$$

where $\mathbf{J}$ is the *electric current density* (a three-component vector density). The constants ϵ_0 and μ_0 are, respectively, the *permittivity* and the *permeability* of free space. It is shown in all standard textbooks that the fields propagate with a velocity c given by

$$c^2 = (\epsilon_0 \mu_0)^{-1}. \tag{4.23}$$

It is also known that the fields may be derived from a scalar (electric) potential ϕ and a (magnetic) vector potential $\mathbf{A}$, according to

$$\mathbf{E} = -\operatorname{grad}\phi - \frac{\partial \mathbf{A}}{\partial t}, \quad (a) \qquad \mathbf{B} = \operatorname{curl}\mathbf{A}; \quad (b) \tag{4.24}$$

and the potentials may be assumed, with no loss of generality, to satisfy the equation

$$\text{div}\,\mathbf{A} + \frac{1}{c^2}\frac{\partial \phi}{\partial t} = 0. \tag{4.25}$$

This equation corresponds to use of a particular 'gauge' in defining the potentials (the *Lorentz gauge*), which will be assumed throughout.

In terms of the potentials, the key equations (4.21)a,b and (4.22)a,b, may be presented in the alternative forms

$$\nabla^2\phi - \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} = -\epsilon_0\rho, \tag{4.26}$$

$$\nabla^2\mathbf{A} - \frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2} = -\mu_0\mathbf{J}. \tag{4.27}$$

Let us now replace the vectors $\mathbf{J}$ and $\mathbf{A}$ by new *four*-vectors, with contravariant components

$$J^\mu : \quad (J^1 = J_x,\ J^2 = J_y,\ J^3 = J_z,\ J^4 = c\rho) \tag{4.28}$$

and

$$A^\mu : \quad (A^1 = A_x,\ A^2 = A_y,\ A^3 = A_z,\ A^4 = \phi/c). \tag{4.29}$$

The three-divergence $\text{div}\,\mathbf{A}$, which appears in (4.25), may clearly be written as a contracted tensor product $\text{div}\,\mathbf{A} = \mathsf{D}_\mu A^\mu \quad (\mu = 1, 2, 3)$, the contraction being restricted to the first three components in (4.28): thus

$$\text{div}\,\mathbf{A} = \mathsf{D}_\mu A^\mu \frac{\partial}{\partial x}A_x + \frac{\partial}{\partial y}A_y + \frac{\partial}{\partial z}A_z.$$

But now we may define a *four*-divergence, distinguished by an upper-case "D", and contract over all components in (4.18) and (4.28) to obtain

$$\text{Div}\,\mathbf{A} = \mathsf{D}_\mu A^\mu = \frac{\partial}{\partial x}A_x + \frac{\partial}{\partial y}A_y + \frac{\partial}{\partial z}A_z + \frac{1}{c}\frac{\partial}{\partial t}\left(\frac{\phi}{c}\right), \tag{4.30}$$

the D_4 component being defined in (4.19). Consequently,

$$\text{Div}\,\mathbf{A} = \mathsf{D}_\mu A^\mu = \text{div}\,\mathbf{A} + \frac{1}{c^2}\frac{\partial\phi}{\partial t}. \tag{4.31}$$

The choice of the fourth component in (4.30) is thus appropriate: on putting $\text{Div}\,\mathbf{A} = 0$ we retrieve equation (4.25), which relates the potentials, in the form $\mathsf{D}_u A^\mu = 0$. As this is a *tensor* equation its form will be left unchanged by any transformation that leaves the metric invariant.

Next we define the antisymmetric rank-2 four-tensor (first lowering the indices on A^μ)

$$F_{\mu\nu} = \mathsf{D}_\mu A_\nu - \mathsf{D}_\nu A_\mu \quad (\mu, \nu = 1, 2, 3, 4) \tag{4.32}$$

and interpret particular components by referring to the three-space equations. Evidently, since $\partial/\partial x_4 = (1/c)\partial/\partial t$,

$$F_{41} = \frac{1}{c}\frac{\partial}{\partial t}(-A_x) - \frac{\partial}{\partial x}\left(\frac{\phi}{c}\right) = \frac{1}{c}\left(-\frac{\partial A_x}{\partial t} - \frac{\partial \phi}{\partial x}\right) = \left(\frac{E_x}{c}\right)$$

by virtue of (4.24)a; and in the same way it appears that $F_{42} = E_y/c$, $F_{43} = E_z/c$. If the $F_{\mu\nu}$ are set out in a 4×4 matrix, the last row ($\mu = 4$) will contain elements E_x/c, E_y/c, E_z/c, 0; and similarly the last column will contain the same elements with a sign reversal (the matrix being antisymmetric).

Next consider the case $\mu, \nu < 4$: the components may then be identified with *magnetic* field components. Thus, for example,

$$F_{12} = \frac{\partial}{\partial x}(-A_y) - \frac{\partial}{\partial y}(-A_x) = -B_z,$$

as follows from (4.24)b and the three-space definition of curl $\mathbf{A}$. In the same way we find $F_{13} = B_y$ and $F_{23} = -B_x$.

The full matrix $\mathbf{F}$ thus has elements

$$F_{\mu\nu} = \begin{pmatrix} 0 & -B_z & +B_x & -E_x/c \\ +B_z & 0 & -B_x & -E_y/c \\ -B_x & +B_x & 0 & -E_z/c \\ E_x/c & E_y/c & E_z/c & 0 \end{pmatrix}.$$

The corresponding contravariant tensor will be $F^{\mu\nu} = g^{\mu\rho}g^{\nu\sigma}F_{\rho\sigma}$: its upper left-hand block is identical with that of $F_{\mu\nu}$ but the last row and column are changed in sign. From the electro-magnetic field tensors $F_{\mu\nu}$ and $F^{\mu\nu}$ we can easily express all the Maxwell equations in a succinct tensor form.

First we recall that the equations in (4.21) and (4.22) involve not only the fields $\mathbf{E}$ and $\mathbf{B}$ themselves but also their divergence and curl; and they also involve ρ and $\mathbf{J}$, which are collected in the charge-current four-vector (4.28).

Let us now define two new tensors, by combining the differential operators $\mathsf{D}_m u$ with the field tensors:

$$X_{\lambda\mu\nu} = \mathsf{D}_\lambda F_{\mu\nu} + \mathsf{D}_\mu F_{\nu\lambda} + \mathsf{D}_\nu F_{\lambda\mu} \quad Y^\mu = \mathsf{D}_\lambda, \qquad Y^\mu = \mathsf{D}_\lambda F^{\lambda\mu}. \tag{4.33}$$

The first of these, $X_{\lambda\mu\nu}$, is a rank-3 covariant tensor, sometimes called the "cyclic curl" (the indices in the three terms following the cyclic order $\lambda\mu\nu \to \mu\nu\lambda \to \nu\lambda\mu$). The second, after the contraction over λ, is evidently a rank-1 contravariant vector. Again, one can evaluate particular components and establish their meaning by comparison with the Maxwell equations.

For example, putting $\lambda\mu\nu = 123$,

$$X_{123} = \mathsf{D}_1 F_{23} + \mathsf{D}_2 F_{31} + \mathsf{D}_3 F_{12} = \frac{\partial}{\partial x}(-B_x) + \frac{\partial}{\partial y}(-B_y) + \frac{\partial}{\partial z}(-B_z) = -\text{div}\,\mathbf{B},$$

while for $\lambda = 4$ (the time component), putting $\mu\nu = 23, 31, 12$ in turn, we find the x-,y- and z-components of

$$-\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} - \frac{1}{c}\mathrm{curl}\,\mathbf{E}.$$

On setting the X-tensor to zero (all components) we therefore obtain the equivalent statement

$$\mathrm{div}\,\mathbf{B} = 0, \qquad \mathrm{curl}\,\mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{4.34}$$

– coinciding with the Maxwell equations in (4.22)a and (4.21)b, respectively.

The two remaining equations, (4.21)a and (4.22)b, involve the charge and current densities ρ and $\mathbf{J}$, respectively, which are contained in the contravariant four-vector J^μ. By evaluating Y^μ, which has the same tensor character as J^μ, we find that the components for $\mu = 1, 2, 3$ coincide with the three components of

$$\left(\frac{1}{c^2}\frac{\partial \mathbf{E}}{\partial t} - \mathrm{curl}\,\mathbf{B}\right),$$

while the $\mu = 4$ component yields

$$\left(\frac{-\mathrm{div}\,\mathbf{E}}{c}\right).$$

Comparison of the first expression with (4.22)b identifies it as the three-vector part of $-\mu_0\mathbf{J}$; while the second expression coincides with the fourth component, namely $-\mu_0 c\rho$.

In summary, we have shown that the tensor equations

$$X_{\lambda\mu\nu} = 0, \quad (a) \qquad Y^\mu = -\mu_0 J^\mu \quad (b) \tag{4.35}$$

embody in themselves the whole system of three-vector equations due to Maxwell: they completely determine the electric and magnetic field vectors, at all points in space, arising from any given distribution of electric charges and currents. Furthermore, by introducing the 'quad-squared' operator defined in (4.20), the equations (4.26) and (4.27), which relate the scalar and vector potentials to the electric charge and current densities, may be condensed into the single four-vector equation

$$\bar{\Box}^2\,\mathbf{A} = -\mu_0\mathbf{J} \tag{4.36}$$

– as may be confirmed by using (4.20), (4.28), (4.29) and remembering that $\mu_0\epsilon_0 = c^{-2}$.

This is not the place to discuss the striking physical implications of the conclusions just presented (see, for example, W. Rindler *Essential Relativity*, 2nd edn, Springer-Verlag 1977, for a beautiful exposition) and we return at once to the tensor formalism.