13
Symmetry groups

13.1 Groups of transformations

SPACES that are symmetrical have a fundamental importance in modern physics. Why is this? It might be thought that completely exact symmetry is something that could arise only exceptionally, or perhaps just as some convenient approximation. Although a symmetrical object, such as a square or a sphere, has a precise existence as an idealized ('Platonic'; see §1.3) mathematical structure, any *physical* realization of such a thing would ordinarily be regarded as merely some kind of approximate representation of this Platonic ideal, therefore possessing no actual symmetry that can be regarded as exact. Yet, remarkably, according to the highly successful physical theories of the 20th century, all physical interactions (including gravity) act in accordance with an idea which, strictly speaking, depends crucially upon certain physical structures possessing a symmetry that, at a fundamental level of description, is indeed necessarily exact!

What is this idea? It is a concept that has come to be known as a 'gauge connection'. That name, as it stands, conveys little. But the idea is an important one, enabling us to find a subtle ('twisted') notion of differentiation that applies to general entities on a manifold (entities that are indeed more general than just those—the p-forms—which are subject to exterior differentiation, as described in Chapter 12). These matters will be the subject of the two chapters following this one; but as a prerequisite, we must first explore the basic notion of a *symmetry group*. This notion also has many other important areas of application in physics, chemistry, and crystallography, and also within many different areas of mathematics itself.

Let us take a simple example. What are the symmetries of a *square*? The question has two different answers depending upon whether or not we allow symmetries which reverse the orientation of the square (i.e. for which the square is turned over). Let us first consider the case in which these orientation-reversing symmetries are not allowed. Then the square's symmetries are generated from a single rotation through a right angle in the square's plane, repeated various numbers of times. For convenience, we can represent these motions in terms of complex numbers, as we did in

Chapter 5. We may, if we choose, think of the vertices of the square as occupying the points 1, i, -1, $-$ i in the complex plane (Fig. 13.1a), and our basic rotation represented by multiplication by i (i.e. by 'i×'). The various *powers* of i represent all our rotations, there being four distinct ones in all:

$$\mathrm{i}^0 = 1, \quad \mathrm{i}^1 = \mathrm{i}, \quad \mathrm{i}^2 = -1, \quad \mathrm{i}^3 = -\mathrm{i}$$

(Fig. 13.1b). The fourth power $\mathrm{i}^4 = 1$ gets us back to the beginning, so we have no more elements. The product of any two of these four elements is again one of them.

These four elements provide us with a simple example of a *group*. This consists of a set of elements and a law of 'multiplication' defined between pairs of them (denoted by juxtaposition of symbols) for which the *associative* multiplication law holds

$$a(bc) = (ab)c,$$

where there is an *identity* element 1 satisfying

$$1a = a1 = a,$$

and where each element a has an *inverse* a^{-1}, such that[13.1]

$$a^{-1}a = aa^{-1} = 1.$$

The symmetry operations which take an object (not necessarily a square) into itself always satisfy these laws, called the *group axioms*.

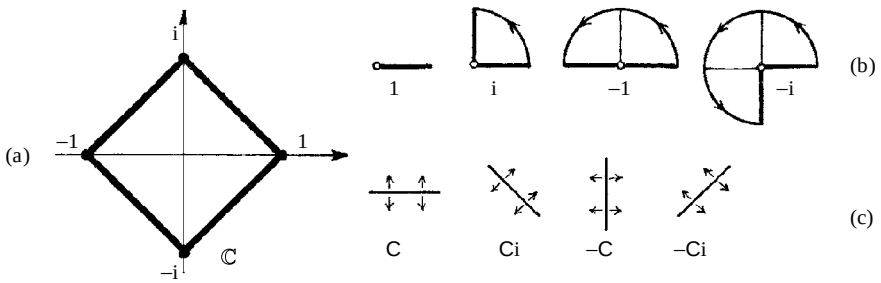

Fig. 13.1 Symmetry of a square. (a) We may represent the square's vertices by the points 1, i, $-$ 1, $-$ i in the complex plane $\mathbb{C}$. (b) The group of non-reflective symmetries are represented, in $\mathbb{C}$, as multiplication by $1 = \mathrm{i}^0$, $\mathrm{i} = \mathrm{i}^1$, $-1 = \mathrm{i}^2$, $-\mathrm{i} = \mathrm{i}^3$, respectively. (c) The reflective symmetries are given, in C, by C (complex conjugation), Ci, $-$ C, and $-$ Ci.

[13.1] Show that if we just assume $1a = a$ and $a^{-1}a = 1$ for all a, together with associativity $a(bc) = (ab)c$, then $a1 = a$ and $aa^{-1} = 1$ can be *deduced*. (*Hint*: Of course a is not the only element asserted to have an inverse.) Show why, on the other hand, $a1 = a$, $a^{-1}a = 1$, and $a(bc) = (ab)c$ are insufficient.

Recall the conventions recommended in Chapter 11, where we think of b acting first and a afterwards, in the product ab. We can regard these as operations as being performed upon some object appearing to the right. Thus, we could consider the motion, b, expressing a symmetry of an object $\boldsymbol{\Phi}$, as $\boldsymbol{\Phi} \mapsto b(\boldsymbol{\Phi})$, which we follow up by another such motion a, giving $b(\boldsymbol{\Phi}) \mapsto a(b(\boldsymbol{\Phi}))$. This results in the combined action $\boldsymbol{\Phi} \mapsto a(b(\boldsymbol{\Phi}))$, which we simply write $\boldsymbol{\Phi} \mapsto ab(\boldsymbol{\Phi})$, corresponding to the motion ab. The identity operation leaves the object alone (clearly always a symmetry) and the inverse is just the reverse operation of a given symmetry, moving the object back to where it came from.

In our particular example of non-reflective rotations of the square, we have the additional *commutative* property

$$ab = ba.$$

Groups that are commutative in this sense are called *Abelian*, after the tragically short-lived Norwegian mathematician Niels Henrik Abel.[1] Clearly any group that can be represented simply by the multiplication of complex numbers must be Abelian (since the multiplication of individual complex numbers always commutes). We saw other examples of this at the end of Chapter 5 when we considered the general case of a finite *cyclic* group $\mathbb{Z}_n$, generated by a single nth root of unity.[13.2]

Now let us allow the orientation-reversing *reflections* of our square. We can still use the above representation of the square in terms of complex numbers, but we shall need a new operation, which I denote by C, namely *complex conjugation*. (This flips the square over, about a horizontal line; see §10.1, Fig. 10.1.) We now find (see Fig. 13.1c) the 'multiplication laws'[13.3]

$$\mathsf{C}\mathrm{i} = (-\mathrm{i})\mathsf{C}, \quad \mathsf{C}(-1) = (-1)\mathsf{C}, \quad \mathsf{C}(-\mathrm{i}) = \mathrm{i}\mathsf{C}, \quad \mathsf{C}\mathsf{C} = 1$$

(where[2] I shall henceforth write $(-\mathrm{i})\mathsf{C}$ as $-\mathrm{i}\mathsf{C}$, etc.). In fact, we can obtain the multiplication laws for the entire group just from the basic relations[13.4]

$$\mathrm{i}^4 = 1, \quad \mathsf{C}^2 = 1, \quad \mathsf{C}\mathrm{i} = \mathrm{i}^3\mathsf{C},$$

the group being non-Abelian, as is manifested in the last equation. The total number of of distinct elements in a group is called its *order*. The order of this particular group is 8.

Now let us consider another simple example, namely the group of rotational symmetries of an ordinary sphere. As before, we can first consider the

[13.2] Explain why any vector space is an Abelian group—called an *additive* Abelian group—where the group 'multiplication' operation is the 'addition' operation of the vector space.

[13.3] Verify these relations (bearing in mind that $\mathsf{C}\mathrm{i}$ stands for 'the operation $\mathrm{i}\times$, followed by the operation C, etc.). (*Hint*: You can check the relations by just confirming their effects on 1 and i. Why?)

[13.4] Show this.

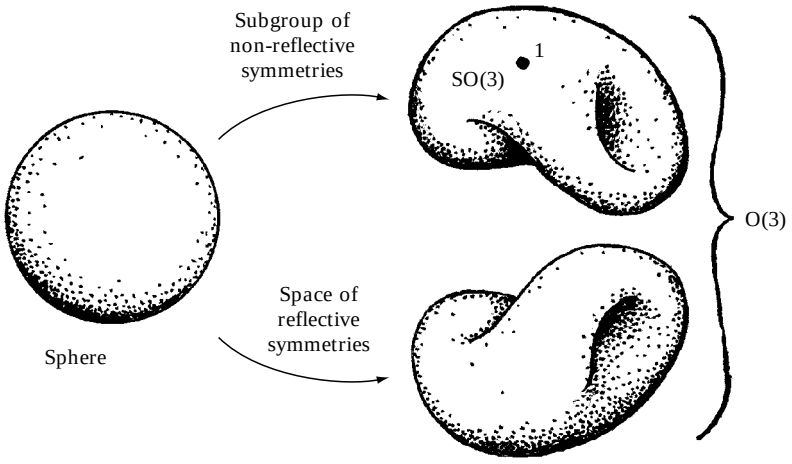

Fig. 13.2 Rotational symmetry of a sphere. The entire symmetry group, O(3), is a disconnected 3-manifold, consisting of two pieces. The component containing the identity element 1 is the (normal) subgroup SO(3) of non-reflective symmetries of the sphere. The remaining component is the 3-manifold of reflective symmetries.

case where reflections are excluded. This time, our symmetry group will have an infinite number of elements, because we can rotate through any angle about any axis direction in 3-space. The symmetry group actually constitutes a 3-dimensional space, namely the 3-manifold denoted by $\mathcal{R}$ in Chapter 12. Let me now give this group (3-manifold) its official name. It is called[3] SO(3), the non-reflective orthogonal group in 3 dimensions. If we now include the reflections, then we get a whole new set of symmetries—another 3-manifold's worth—which are disconnected from the first, namely those which involve a reversal of the orientation of the sphere. The entire family of group elements again constitutes a 3-manifold, but now it is a disconnected 3-manifold, consisting of two separate connected pieces (see Fig. 13.2). This entire group space is called O(3).

These two examples illustrate two of the most important categories of groups, the finite groups and the continuous groups (or *Lie* groups; see §13.6).[4] Although there is a great difference between these two types of group, there are many of the important properties of groups that are common to both.

13.2 Subgroups and simple groups

Of particular significance is the notion of a *subgroup* of a group. To exhibit a subgroup, we select some collection of elements within the group which themselves form a group, using the same multiplication and inversion

operations as in the whole group. Subgroups are important in many modern theories of particle physics. It tends to be assumed that there is some fundamental symmetry of Nature that relates different kinds of particles to one another and also relates different particle interactions to one another. Yet one may not see this full group acting as a symmetry in any manifest way, finding, instead, that this symmetry is 'broken' down to some subgroup of the original group where the *subgroup* plays a manifest role as a symmetry. Thus, it is important to know what the possible subgroups of a putative 'fundamental' symmetry group actually are, in order that those symmetries that are indeed manifest in Nature might be able to be thought about as subgroups of this putative group. I shall be addressing questions of this kind in §§25.5–8, §26.11, and §28.1.

Let us examine some particular cases of subgroups, for the examples that we have been considering. The *non-reflective* symmetries of the square constitute a 4-element subgroup $\{1, \mathrm{i}, -1, -\mathrm{i}\}$ of the entire 8-element group of symmetries of the square. Likewise, the non-reflective rotation group SO(3) constitutes a subgroup of the entire group O(3). Another subgroup of the symmetries of the square consists of the four elements $\{1, -1, \mathsf{C}, -\mathsf{C}\}$; yet another has just the two elements $\{1, -1\}$.[13.5] Moreover there is always the 'trivial' subgroup consisting of the identity alone $\{1\}$ (and the whole group itself is, equally trivially, always a subgroup).

All the various subgroups that I have just described have a special property of particular importance. They are examples of what are called *normal* subgroups. The significance of a normal subgroup is that, in an appropriate sense, the action of any element of the whole group leaves a normal subgroup alone or, more technically, we say that each element of the whole group *commutes* with the normal subgroup. Let me be more explicit. Call the whole group $\mathcal{G}$ and the subgroup $\mathcal{S}$. If I select any particular element g of the group $\mathcal{G}$, then I can denote by $\mathcal{S}g$ the set consisting of all elements of $\mathcal{S}$ each individually multiplied by g on the right (what is called *postmultiplied* by g). Thus, in the case of the particular subgroup $\mathcal{S} = \{1, -1, \mathsf{C}, -\mathsf{C}\}$, of the symmetry group of the square, if we choose $g = \mathrm{i}$, then we obtain $\mathcal{S}\mathrm{i} = \{\mathrm{i}, -\mathrm{i}, \mathsf{C}\mathrm{i}, -\mathsf{C}\mathrm{i}\}$. Likewise, the notation $g\mathcal{S}$ will denote the set consisting of all elements of $\mathcal{S}$, each individually multiplied by g on the left (*premultiplied* by g). Thus, in our example, we now have $\mathrm{i}\mathcal{S} = \{\mathrm{i}, -\mathrm{i}, \mathrm{i}\mathsf{C}, -\mathrm{i}\mathsf{C}\}$. The condition for $\mathcal{S}$ to be a normal subgroup of $\mathcal{G}$ is that these two sets are the same, i.e.

$$\mathcal{S}g = g\mathcal{S}, \quad \text{for all } g \text{ in } \mathcal{S}.$$

In our particular example, we see that this is indeed the case (since $\mathsf{C}\mathrm{i} = -\mathrm{i}\mathsf{C}$ and $-\mathsf{C}\mathrm{i} = \mathrm{i}\mathsf{C}$), where we must bear in mind that the collection

[13.5] Verify that all these in this paragraph are subgroups (and bear in mind Note 13.4).

of things inside the curly brackets is to be taken as an *unordered* set (so that it does not matter that the elements $-\mathsf{iC}$ and iC appear in reverse order in the collection of elements, when $\mathcal{S}\mathsf{i}$ and $\mathsf{i}\mathcal{S}$ are written out explicitly).

We can exhibit a *non*-normal subgroup of the group of symmetries of the square, as the subgroup of two elements $\{1, \mathsf{C}\}$. It is non-normal because $\{1, \mathsf{C}\}\mathsf{i} = \{\mathsf{i}, \mathsf{Ci}\}$ whereas $\mathsf{i}\{1, \mathsf{C}\} = \{\mathsf{i}, -\mathsf{Ci}\}$. Note that this subgroup arises as the new (reduced) symmetry group if we mark our square with a horizontal arrow pointing off to the right (see Fig. 13.3a). We can obtain another non-normal subgroup, namely $\{1, \mathsf{Ci}\}$ if we mark it, instead, with an arrow pointing diagonally down to the right (Fig. 13.3b).[13.6] In the case of O(3), there happens to be only one non-trivial normal subgroup,[13.7] namely SO(3), but there are many non-normal subgroups. Non-normal examples are obtained if we select some appropriate finite set of points on the sphere, and ask for the symmetries of the sphere with these points marked. If we mark just a single point, then the subgroup consists of rotations of the sphere about the axis joining the origin to this point (Fig. 13.3c). Alternatively, we could, for example, mark points that are the vertices of a regular polyhedron. Then the subgroup is finite, and consists of the symmetry group of that particular polyhedron (Fig. 13.3d).

One reason that normal subgroups are important is that, if a group $\mathcal{G}$ possesses a non-trivial normal subgroup, then we can break $\mathcal{G}$ down, in a sense, into smaller groups. Suppose that $\mathcal{S}$ is a normal subgroup of $\mathcal{G}$. Then the distinct sets $\mathcal{S}g$, where g runs through all the elements of $\mathcal{G}$, turn

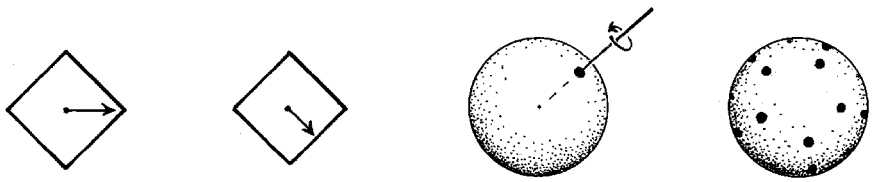

Fig. 13.3 (a) Marking the square of Fig. 13.1 with an arrow pointing to the right, reduces its symmetry group to a non-normal subgroup $\{1,\mathsf{C}\}$. (b) Marking it with an arrow pointing diagonally down to the right yields a different non-normal subgroup $\{1,\mathsf{Ci}\}$. (c) Marking the sphere of Fig. 13.2 with a single point reduces its symmetry to a (non-normal) O(2) subgroup of O(3): rotations about the axis joining the origin to this point. (d) If the sphere is marked with the vertices of a regular polyhedron (here a dodecahedron), its group of symmetries is a finite (non-normal) subgroup of O(3).

[13.6] Check these assertions, and find two more non-normal subgroups, showing that there are no further ones.

[13.7] Show this. (*Hint*: which *sets* of rotations can be rotation-invariant?)

out themselves to form a group. Note that for a given set $\mathcal{S}g$, the choice of g is generally not unique; we can have $\mathcal{S}g_1 = \mathcal{S}g_2$, for different elements g_1, g_2 of $\mathcal{G}$. The sets of the form $\mathcal{S}g$, for *any* subgroup $\mathcal{S}$, are called *cosets* of $\mathcal{G}$; but when $\mathcal{G}$ is normal, the cosets form a group. The reason for this is that if we have two such cosets $\mathcal{S}g$ and $\mathcal{S}h$ (g and h being elements of $\mathcal{G}$) then we can define the 'product' of $\mathcal{S}g$ with $\mathcal{S}h$ to be

$$(\mathcal{S}g)\ (\mathcal{S}h) = \mathcal{S}(gh),$$

and we find that all the group axioms are satisfied, provided that $\mathcal{S}$ is normal, essentially because the right-hand side is well defined, independently of which g and h were chosen in the representation of the cosets on the left-hand side of this equation.[13.8] The resulting group defined in this way is called the *factor group* of $\mathcal{G}$ by its normal subgroup $\mathcal{S}$. The factor group of $\mathcal{G}$ by $\mathcal{S}$ is written $\mathcal{G}/\mathcal{S}$. We can still write $\mathcal{G}/\mathcal{S}$ for the factor *space* (not a group) of distinct cosets $\mathcal{S}g$ even when $\mathcal{S}$ is not normal.[13.9]

Groups that possess no non-trivial normal subgroups at all are called *simple groups*. The group SO(3) is an example of a simple group. Simple groups are, in a clear sense, the basic building blocks of group theory. It is thus an important achievement of the 19th and 20th centuries in mathematics that all the finite simple groups and all the continuous simple groups are now known. In the continuous case (i.e. for Lie groups), this was a mathematical landmark, started by the highly influential German mathematician Wilhelm Killing (1847–1923), whose basic papers appeared in 1888–1890, and was essentially completed, in 1894, in one of the most important of mathematical papers ever written,[5] by the superb geometer and algebraist Élie Cartan (whom we have already encountered in Chapter 12, and whom we shall meet again in Chapter 17). This classification has continued to play a fundamental role in many areas of mathematics and physics, to the present day. It turns out that there are four families, known as A_m, B_m, C_m, D_m(for $m = 1, 2, 3, \ldots$), of respective dimension $m(m+2), m(2m+1), m(2m+1), m(2m-1)$, called the *classical groups* (see end of §13.10) and five *exceptional groups* known as E_6, E_7, E_8, F_4, G_2, of respective dimension 78, 133, 248, 52, 14.

The classification of the finite simple groups is a more recent (and even more difficult) achievement, carried out over a great many years during the 20th century by a considerable number of mathematicians (with the aid of computers in more recent cases), being completed only in 1982.[6] Again there are some systematic families and a finite collection of *exceptional*

[13.8] Verify this and show that the axioms fail if $\mathcal{S}$ is not normal.

[13.9] Explain why the number of elements in $\mathcal{G}/\mathcal{S}$, for any finite subgroup $\mathcal{S}$ of $\mathcal{G}$, is the order of $\mathcal{G}$ divided by the order of $\mathcal{S}$.

finite simple groups. The largest of these exceptional groups is referred to as the *monster*, which is of order

$$= 808017424794512875886459904961710757005754368000000000.$$
$$= 2^{46} \times 3^{20} \times 5^9 \times 7^6 \times 11^2 \times 13^3 \times 17 \times 19 \times 23 \times 29 \times 31 \times 41 \times 47 \times 59 \times 71.$$

Exceptional groups appear to have a particular appeal for many modern theoretical physicists. The group E_8 features importantly in string theory (§31.12), while various people have expressed a hope that the huge but finite monster may feature in some future theory.[7]

The classification of the simple groups may be regarded as a major step towards the classification of groups generally since, as indicated above, general groups may be regarded as being built up out of simple groups (together with Abelian ones). In fact, this is not really the whole story because there is further information in how one simple group can build upon another. I do not propose to enter into the details of this matter here, but it is worth just mentioning the simplest way that this can happen. If $\mathcal{G}$ and $\mathcal{H}$ are any two groups, then they can be combined together to form what is called the *product group* $\mathcal{G} \times \mathcal{H}$, whose elements are simply *pairs* (g, h), where g belongs to $\mathcal{G}$ and h belongs to $\mathcal{H}$, the rule of group multiplication between elements (g_1, h_1) and (g_2, h_2), of $\mathcal{G} \times \mathcal{H}$, being defined as

$$(g_1, h_1)\,(g_2, h_2) = (g_1 g_2, h_1 h_2),$$

and it is very easy to verify that the group axioms are satisfied. Many of the groups that feature in particle physics are in fact product groups of simple groups (or elementary modifications of such).[13.10]

13.3 Linear transformations and matrices

In the general study of groups, there is a particular class of symmetry groups that have been found to play a central role. These are the groups of symmetries of vector spaces. The symmetries of a vector space are expressed by the *linear transformations* preserving the vector-space structure.

Recall from §11.1 and §12.3 that, in a vector space V, we have, defining its structure, a notion of addition of vectors and multiplication of vectors by numbers. We may take note of the fact that the geometrical picture of addition is obtained by use of the parallelogram law, while multiplication by a number is visualized as scaling the vector up (or down) by that number (Fig. 13.4). Here we are picturing it as a *real* number, but complex vector spaces are also allowed (and are particularly important in many

[13.10] Verify that $\mathcal{G} \times \mathcal{H}$ is a group, for any two groups $\mathcal{G}$ and $\mathcal{H}$, and that we can identify the factor group $(\mathcal{G} \times \mathcal{H})/\mathcal{G}$ with $\mathcal{H}$.

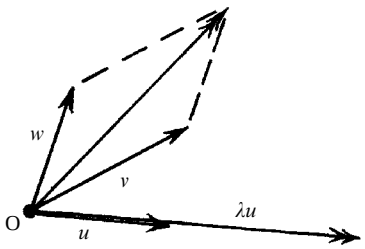

Fig. 13.4 A linear transformation preserves the vector-space structure of the space on which it acts. This structure is defined by the operations of addition (illustrated by the parallelogram law) and multiplication by a scalar λ (which could be a real number or, in the case of a complex vector space, a complex number). Such a transformation preserves the 'straightness' of lines and the notion of 'parallel', keeping the origin O fixed.

contexts, because of complex magic!), though hard to portray in a diagram. A linear transformation of V is a transformation that takes V to itself, preserving its structure, as defined by these basic vector-space notions. More generally, we can also consider linear transformations that take one vector space to another.

A linear transformation can be explicitly described using an array of numbers called a *matrix*. Matrices are important in many mathematical contexts. We shall examine these extremely useful entities with their elegant algebraic rules in this section (and in §§13.4,5). In fact, §§13.3–7 may be regarded as a rapid tutorial in matrix theory and its application to the theory of continuous groups. The notions described here are vital to a proper understanding of quantum theory, but readers already familiar with this material—or else who prefer a less detailed comprehension of quantum theory when we come to that—may prefer to skip these sections, at least for the time being.

To see what a linear transformation looks like, let us first consider the case of a 3-dimensional vector space and see its relevance to the rotation group O(3) (or SO(3)), discussed in §13.1, giving the symmetries of the sphere. We can think of this sphere as embedded in Euclidean 3-space $\mathbb{E}^3$ (this space being regarded as a vector space with respect to the origin O at the sphere's centre[8]) as the locus

$$x^2 + y^2 + z^2 = 1$$

in terms of ordinary Cartesian coordinates (x, y, z).[13.11] Rotations of the sphere are now expressed in terms of linear transformation of $\mathbb{E}^3$, but of a very particular type known as *orthogonal* which we shall be coming to in §§13.1,8 (see also §13.1).

General linear transformations, however, would squash or stretch the sphere into an *ellipsoid*, as illustrated in Fig. 13.5. Geometrically,

[13.11] Show how this equation, giving the points of unit distance from O, follows from the Pythagorean theorem of §2.1.

a linear transformation is one that preserves the 'straightness' of lines and the notion of 'parallel' lines, keeping the origin O fixed. But it need not preserve right angles or other angles, so shapes can be squashed or stretched, in a uniform but anisotropic way.

How do we express linear transformations in terms of the coordinates x, y, z? The answer is that each new coordinate is expressed as a (homogeneous) *linear combination* of the original ones, i.e. by a separate expression like $\alpha x + \beta y + \gamma z$, where α, β, and γ are constant numbers.[13.12] We have 3 such expressions, one for each of the new coordinates. To write all this in a compact form, it will be useful to make contact with the *index notation* of Chapter 12. For this, we re-label the coordinates as (x^1, x^2, x^3), where

$$x^1 = x, \quad x^2 = y, \quad x^3 = z$$

(bearing in mind, again, that these upper indices do *not* denote powers see §12.2). A general point in our Euclidean 3-space has coordinates x^a, where $a = 1, 2, 3$. An advantage of using the index notation is that the discussion applies in any number of dimensions, so we can consider that a (and all our other index letters) run over $1, 2, \ldots, n$, where n is some fixed positive integer. In the case just considered, $n = 3$.

In the index notation, with Einstein's summation convention (§12.7), the general linear transformation now takes the form[9,[13.13]]

$$x^a \mapsto T^a{}_b \, x^b.$$

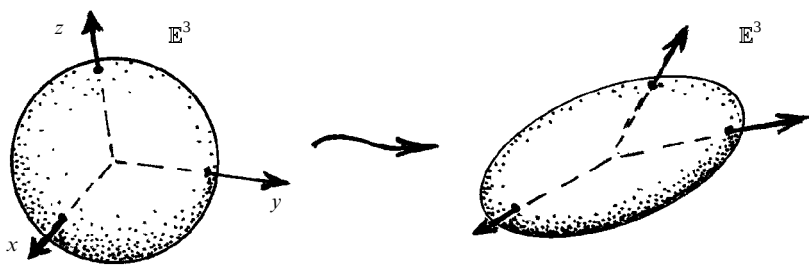

Fig. 13.5 A linear transformation acting on $\mathbb{E}^3$ (expressed in terms of Cartesian x, y, z coordinates) would generally squash or stretch the unit sphere $x^2 + y^2 + z^2 = 1$ into an ellipsoid. The orthogonal group O(3) consists of the linear transformations of $\mathbb{E}^3$ which preserve the unit sphere.

[13.12] Can you explain why? Just do this in the 2-dimensional case, for simplicity.

[13.13] Show this explicitly in the 3-dimensional case.

Calling this linear transformation $\boldsymbol{T}$, we see that $\boldsymbol{T}$ is determined by this set of *components* $T^a{}_b$. Such a set of *components* is referred to as an $n \times n$ *matrix*, usually set out as a *square*—or, in other contexts (see below) $m \times n$-rectangular—array of numbers. The above displayed equation, in the 3-dimensional case is then written

$$\begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} \mapsto \begin{pmatrix} T^1{}_1 & T^1{}_2 & T^1{}_3 \\ T^2{}_1 & T^2{}_2 & T^2{}_3 \\ T^3{}_1 & T^3{}_2 & T^3{}_3 \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix},$$

this standing for three separate relations, starting with $x^1 \mapsto T^1{}_1 x^1 + T^1{}_2 x^2 + T^1{}_3 x^3$.[13.14]

We can also write this without indices or explicit coordinates, as $\boldsymbol{x} \mapsto \boldsymbol{Tx}$. If we prefer, we can adopt the *abstract–index* notation (§12.8) whereby '$x^a \mapsto T^a{}_b x^b$' is *not* a component expression, but actually represents this abstract transformation $\boldsymbol{x} \mapsto \boldsymbol{Tx}$. (When it is important whether an indexed expression is to be read abstractly or as components, this will be made clear by the wording.) Alternatively, we can use the *diagrammatic* notation, as depicted in Fig. 13.6a. In my descriptions, the matrix of numbers $(T^a{}_b)$ or the abstract linear transformation $\boldsymbol{T}$ will be used interchangeably when I am not concerned with the technical distinctions between these two concepts (the former depending upon a specific coordinate description of our vector space V, the latter not).

Let us consider a second linear transformation $\boldsymbol{S}$, applied following the application of $\boldsymbol{T}$. The product $\boldsymbol{R}$ of the two, written $\boldsymbol{R} = \boldsymbol{ST}$, would have a component (or abstract–index) description

$$R^a{}_c = S^a{}_b \, T^b{}_c$$

(summation convention for components!).[13.15] The diagrammatic form of the product $\boldsymbol{ST}$ is given in Fig. 13.6b. Note that, in the diagrammatic notation, to form a successive product of linear transformations, we string

[13.14] Write this all out in full, explaining how this expresses $x^a \mapsto T^a{}_b x^b$.

[13.15] What is this relation between $\boldsymbol{R}$, $\boldsymbol{S}$, and $\boldsymbol{T}$, written out explicitly in terms of the elements of 3×3 square arrays of components. You may recognize this, the normal law for 'multiplication of matrices', if this is familiar to you.

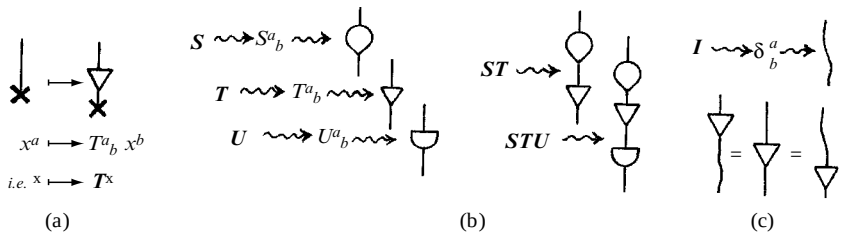

Fig. 13.6 (a) The linear transformation $x^a \mapsto T^a{}_b x^b$, or written without indices as $\boldsymbol{x} \mapsto \boldsymbol{Tx}$ (or read with the indices as abstract, as in §12.8), in diagrammatic form. (b) Diagrams for linear transformations $\boldsymbol{S}$, $\boldsymbol{T}$, $\boldsymbol{U}$, and their products $\boldsymbol{ST}$ and $\boldsymbol{STU}$. In a successive product, we string them in a line downwards. (c) The Kronecker delta δ^a_b, or identity transformation $\boldsymbol{I}$, is depicted as a 'disembodied' line, so relations $T^a{}_b\delta^b_c = T^a{}_c = \delta^a_b T^b{}_c$ become automatic in the notation (see also Fig. 12.17).

them in a line downwards. This happens to work out conveniently in the notation, but one could perfectly well adopt a different convention in which the connecting 'index lines' are drawn horizontally. (Then there would be a closer correspondence between algebraic and diagrammatic notations.)

The *identity* linear transformation $\boldsymbol{I}$ has components that are normally written δ^a_b (the *Kronecker delta*—the standard convention being that these indices are not normally staggered), for which

$$\delta^a_b = \begin{cases} 1 & \text{if } a = b, \\ 0 & \text{if } a \neq b, \end{cases}$$

and we have[13.16]

$$T^a{}_b\delta^b_c = T^a{}_c = \delta^a_b T^b{}_c$$

giving the algebraic relations $\boldsymbol{TI} = \boldsymbol{T} = \boldsymbol{IT}$. The square matrix of components δ^a_b has 1s down what is called the *main diagonal*, which extends from the top-left corner to bottom-right. In the case $n = 3$, this is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

In the diagrammatic notation, we simply represent the Kronecker delta by a 'disembodied' line, and the above algebraic relations become automatic in the notation; see Fig. 13.6c.

[13.16] Verify.

Those linear transformations which map the entire vector space down to a region (subspace) of smaller dimension within that space are called *singular*.[10] An equivalent condition for $\boldsymbol{T}$ to be singular is the existence of a non-zero vector $\boldsymbol{v}$ such that[13.17]

$$\boldsymbol{T}\boldsymbol{v} = 0.$$

Provided that the transformation is non-singular, then it will have an inverse,[13.18] where the inverse of $\boldsymbol{T}$ is written $\boldsymbol{T}^{-1}$, so that

$$\boldsymbol{T}\boldsymbol{T}^{-1} = \boldsymbol{I} = \boldsymbol{T}^{-1}\boldsymbol{T},$$

as is required of an inverse. We can give the explicit expression for this inverse conveniently in the diagrammatic notation; see Fig. 13.7, where I have introduced the useful diagrams for the antisymmetrical (*Levi-Civita*) quantities $\varepsilon_{a\ldots c}$ and $\in^{a\ldots c}$ (with normalization $\varepsilon_{a\cdots c}\ \in^{a\cdots c} = n!$) that were introduced in §12.7 and Fig. 12.18.[13.19]

The algebra of matrices (initiated by the highly prolific English mathematician and lawyer Arthur Cayley in 1858)[11] finds a very broad range of application (e.g. statistics, engineering, crystallography, psychology, computing—not to mention quantum mechanics). This generalizes the algebra of quaternions and the Clifford and Grassmann algebras studied in §§11.3,5,6. I use bold-face *upright* letters (**A**, **B**, **C**, ...) for the arrays of components that constitute actual matrices (rather than abstract linear transformations, for which bold-face *italic* letters are being used).

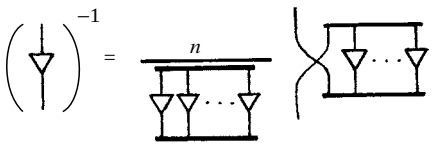

Fig. 13.7 The inverse $\boldsymbol{T}^{-1}$ of a non-singular $(n \times n)$ matrix $\boldsymbol{T}$ given here explicitly in diagrammatic form, using the diagrammatic form of the Levi-Civita antisymmetric quantities $\varepsilon_{a\ldots c}$ and $\in^{a\ldots c}$ (normalized by $\varepsilon_{a\ldots c}\ \in^{a\ldots c} = n!$) introduced in §12.7 and depicted in Fig. 12.18.

[13.17] Why? Show that this would happen, in particular, if the array of components has an entire column of 0s or two identical columns. Why does this also hold if there are two identical rows? *Hint*: For this last part, consider the determinant condition below.

[13.18] Show why, not using explicit expressions.

[13.19] Prove directly, using the diagrammatic relations given in Fig. 12.18, that this definition gives $\boldsymbol{T}\boldsymbol{T}^{-1} = \boldsymbol{I} = \boldsymbol{T}^{-1}\boldsymbol{T}$.

Restricting attention to $n \times n$ matrices for fixed n, we have a system in which notions of addition and multiplication are defined, where the standard algebraic laws

$$\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A}, \qquad \mathbf{A}+(\mathbf{B}+\mathbf{C})=(\mathbf{A}+\mathbf{B})+\mathbf{C}, \qquad \mathbf{A}(\mathbf{BC})=(\mathbf{AB})\mathbf{C},$$
$$\mathbf{A}(\mathbf{B}+\mathbf{C})=\mathbf{AB}+\mathbf{AC}, \qquad (\mathbf{A}+\mathbf{B})\mathbf{C}=\mathbf{AC}+\mathbf{BC}$$

hold. (Each element of $\mathbf{A}+\mathbf{B}$ is simply the sum of the corresponding elements of $\mathbf{A}$ and $\mathbf{B}$.) However, we do not usually have the commutative law of multiplication, so that generally $\mathbf{AB} \neq \mathbf{BA}$. Moreover, as we have seen above, non-zero $n \times n$ matrices do not always have inverses.

It should be remarked that the algebra also extends to the *rectangular* cases of $m \times n$ matrices, where m need not be equal to n. However, addition is defined between an $m \times n$ matrix and a $p \times q$ matrix only when $m = p$ and $n = q$; multiplication is defined between them only when $n = p$, the result being an $m \times q$ matrix. This extended algebra subsumes products like the $\mathbf{Tx}$ considered above, where the 'column vector' $\mathbf{x}$ is thought of as being an $n \times 1$ matrix.[13.20]

The *general linear group* $\mathrm{GL}(n)$ is the group of symmetries of an n-dimensional vector space, and it is realized explicitly as the multiplicative group of $n \times n$ non-singular matrices. If we wish to emphasize that our vector space is *real*, and that the numbers appearing in our matrices are correspondingly real numbers, then we refer to this full linear group as $\mathrm{GL}(n, \mathbb{R})$. We can also consider the complex case, and obtain the *complex* full linear group $\mathrm{GL}(n, \mathbb{C})$. Each of these groups has a normal subgroup, written respectively $\mathrm{SL}(n, \mathbb{R})$ and $\mathrm{SL}(n, \mathbb{C})$—or, more briefly when the underlying field (see §16.1) $\mathbb{R}$ or $\mathbb{C}$ is understood, $\mathrm{SL}(n)$—called the *special* linear group. These are obtained by restricting the matrices to have their *determinants* equal to 1. The notion of a determinant will be explained next.

13.4 Determinants and traces

What is the determinant of an $n \times n$ matrix? It is a *single number* calculated from the elements of the matrix, which vanishes if and only if the matrix is singular. The diagrammatic notation conveniently describes the determinant explicitly; see Fig. 13.8a. The index-notation form of this is

$$\frac{1}{n!}\epsilon^{ab\ldots d}T^{e}{}_{a}T^{f}{}_{b}\ldots T^{h}{}_{d}\mathcal{E}_{ef\ldots h}$$

[13.20] Explain this, and give the full algebraic rules for rectangular matrices.

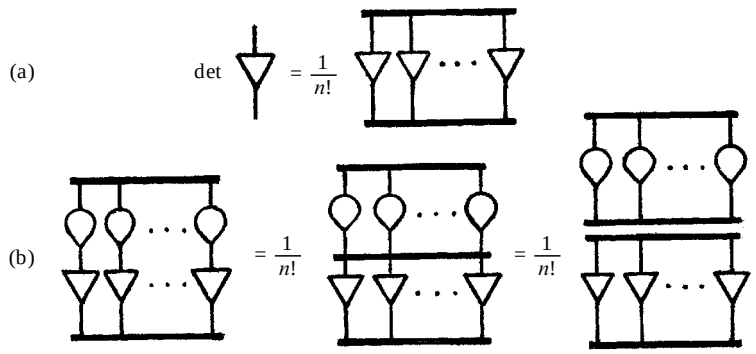

Fig. 13.8 (a) Diagrammatic notation for det $(T^a{}_b) = \det \boldsymbol{T} = |\boldsymbol{T}|$. (b) Diagrammatic proof that $\det(\boldsymbol{ST}) = \det \boldsymbol{S} \det \boldsymbol{T}$. The antisymmetrizing bar can be inserted in the middle term because there is already antisymmetry in the index lines that it crosses. See Figs. 12.17, 12.18.

where the quantities $\epsilon^{a\ldots d}$ and $\varepsilon_{e\ldots h}$ are antisymmetric (Levi-Civita) tensors, normalized accoring to

$$\epsilon^{a\ldots d}\varepsilon_{a\ldots d} = n!$$

for an n-dimensional space (and recall that $n! = 1 \times 2 \times 3 \times \cdots \times n$), where the indices $a, \ldots, d$ and $e, \ldots, h$ are each n in number.

We can refer to this determinant as $\det(T^a{}_b)$ or $\det \boldsymbol{T}$ (or sometimes $|\boldsymbol{T}|$ or as the array constituting the matrix but with vertical bars replacing the parentheses). In the particular cases of a 2×2 and a 3×3 matrix, the determinant is given by[13.21]

$$\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc,$$

$$\det\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & j \end{pmatrix} = aej - afh + bfg - bdj + cdh - ceg.$$

The determinant satisfies the important and rather remarkable relation

$$\det \mathbf{AB} = \det \mathbf{A} \det \mathbf{B},$$

which can be seen to be true quite neatly in the diagrammatic notation (Fig. 13.8b). The key ingredients are the formulae illustrated in Fig. 12.18[13.22] which, when written in the index notation, look like

[13.21] Derive these from the expression of Fig. 13.8a.

[13.22] Show why these hold.

$$\epsilon^{a\ldots c}\,\varepsilon_{f\ldots h} = n!\;\delta^{[a}_{f}\cdots\delta^{c]}_{h}$$

(see §11.6 for the bracket/index notation) and

$$\epsilon^{ab\ldots c}\varepsilon_{fb\ldots c} = (n-1)!\;\delta^{a}_{f}.$$

We also have the notion of the *trace* of a matrix (or linear transformation)

$$\text{trace}\;\mathbf{T} = T^{a}{}_{a} = T^{1}{}_{1} + T^{2}{}_{2} + \cdots + T^{n}{}_{n}$$

(i.e. the sum of the elements along the main diagonal—see §13.3), this being illustrated diagrammatically in Fig. 13.9. Unlike the case of a determinant, there is no particular relation between the trace of the product **AB** of two matrices and the traces of **A** and **B** individually. Instead, we have the relation[13.23]

$$\text{trace}\,(\mathbf{A}+\mathbf{B}) = \text{trace}\,\mathbf{A} + \text{trace}\,\mathbf{B}.$$

There is an important connection between the determinant and the trace which has to do with the determinant of an 'infinitesimal' linear transformation, given by an $n \times n$ matrix $\mathbf{I} + \varepsilon\mathbf{A}$ for which the number ε is considered to be 'infinitesimally small' so that we can ignore its square ε^2 (and also higher powers ε^3, ε^4, etc.). Then we find[13.24]

$$\det\,(\mathbf{I}+\varepsilon\mathbf{A}) = 1 + \varepsilon\,\text{trace}\,\mathbf{A}$$

(ignoring ε^2, etc.). In particular, infinitesimal elements of SL(n), i.e. elements of SL(n) representing infinitesimal rotations, being of *unit* determinant (as opposed to those of GL(n)), are characterized by the **A** in $\mathbf{I} + \varepsilon\mathbf{A}$ having zero trace. We shall be seeing the significance of this in §13.10. In fact the above formula can be extended to *finite* (that is, non-infinitesimal) linear transformations through the expression[13.25]

$$\det\,\mathrm{e}^{\mathbf{A}} = \mathrm{e}^{\text{trace}\,\mathbf{A}},$$

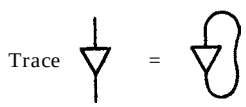

Fig. 13.9 Diagrammatic notation for trace $\mathbf{T}(= T^{a}{}_{a})$.

[13.23] Show this.

[13.24] Show this.

[13.25] Establish the expression for this. *Hint*: Use the 'canonical form' for a matrix in terms of its eigenvalues—as described in §13.5—assuming first that these eigenvalues are unequal (and see Exercise [13.27]). Then use a general argument to show that the equality of some eigenvalues cannot invalidate identities of this kind.

where '$e^{\mathbf{A}}$' for matrices has just the same definition as it has for ordinary numbers (see §5.3), i.e.

$$e^{\mathbf{A}} = \mathbf{I} + \mathbf{A} + 1/2\mathbf{A}^2 + 1/6\mathbf{A}^3 + 1/24\mathbf{A}^4 + \cdots.$$

We shall return to these issues in §13.6 and §14.6.

13.5 Eigenvalues and eigenvectors

Among the most important notions associated with linear transformations are what are called 'eigenvalues' and 'eigenvectors'. These are vital to quantum mechanics, as we shall be seeing in §21.5 and §§22.1,5, and to many other areas of mathematics and applications. An *eigenvector* of a linear transformation $\boldsymbol{T}$ is a non-zero complex vector $\boldsymbol{v}$ which $\boldsymbol{T}$ sends to a multiple of itself. That is to say, there is a complex number λ, the corresponding *eigenvalue*, for which

$$\boldsymbol{T}\boldsymbol{v} = \lambda\boldsymbol{v}, \text{ i.e. } \boldsymbol{T}^a{}_b v^b = \lambda v^a.$$

We can also write this equation as $(\boldsymbol{T} - \lambda\boldsymbol{I})\boldsymbol{v} = 0$, so that, if λ is to be an eigenvalue of $\boldsymbol{T}$, the quantity $\boldsymbol{T} - \lambda\boldsymbol{I}$ must be *singular*. Conversely, if $\boldsymbol{T} - \lambda\boldsymbol{I}$ is singular, then λ is an eigenvalue of $\boldsymbol{T}$. Note that if $\boldsymbol{v}$ is an eigenvector, then so also is any non-zero complex multiple of $\boldsymbol{v}$. The complex 1-dimensional space of these multiples is unchanged by the transformation $\boldsymbol{T}$, a property which characterizes $\boldsymbol{v}$ as an eigenvector (Fig. 13.10).

From the above, we see that this condition for λ to be an eigenvalue of $\boldsymbol{T}$ is

$$\det(\boldsymbol{T} - \lambda\boldsymbol{I}) = 0.$$

Writing this out, we obtain a polynomial equation[13.26] of degree n in λ. By the 'fundamental theorem of algebra', §4.2, we can factorize the λ-polynomial $\det(\boldsymbol{T} - \lambda\boldsymbol{I})$ into linear factors. This reduces the above equation to

$$(\lambda_1 - \lambda)(\lambda_2 - \lambda)(\lambda_3 - \lambda)\ldots(\lambda_n - \lambda) = 0$$

where the complex numbers $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_n$ are the various eigenvalues of $\boldsymbol{T}$. In particular cases, some of these factors may coincide, in which case we have a *multiple* eigenvalue. The multiplicity m of an eigenvalue λ_r is the number of times that the factor $\lambda_r - \lambda$ appears

[13.26] See if you can express the coefficients of this polynomial in diagrammatic form. Work them out for $n = 1$ and $n = 2$.

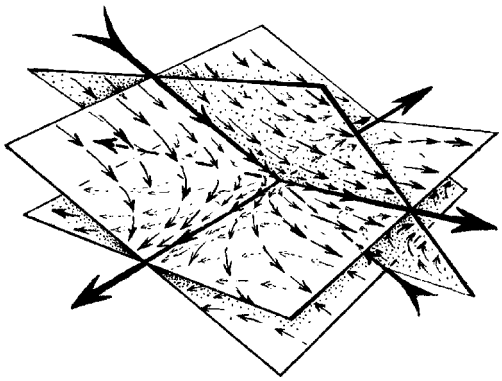

Fig. 13.10 The action of a linear transformation $\boldsymbol{T}$. Its eigenvectors always constitute linear spaces through the origin (here three lines). These spaces are unaltered by $\boldsymbol{T}$. (In this example, there are two (unequal) positive eigenvalues (outward pointing arrows) and one negative one (inward arrows).

in the above product. The total number of eigenvalues of $\boldsymbol{T}$, counted appropriately with multiplicities, is always equal to n, for an $n \times n$ matrix.[13.27]

For a particular eigenvalue λ of multiplicity r, the space of corresponding eigenvectors constitutes a linear space, of dimensionality d, where $1 \leq d \leq r$. For certain types of matrix, including the unitary, Hermitian, and normal matrices of most interest in quantum mechanics (see §13.9, §§22.4,6), we always have the maximum dimensionality $d = r$ (despite the fact that $d = 1$ is the most 'general' case, for given r). This is fortunate, because the (more general) cases for which $d < r$ are more difficult to handle. In quantum mechanics, eigenvalue multiplicities are referred to as *degeneracies* (cf. §§22.6,7).

A *basis* for an n-dimensional vector space $\mathbf{V}$ is an ordered set $\boldsymbol{e} = (\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$ of n vectors $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ which are *linearly independent*, which means that there is no relation of the form $\alpha_1\boldsymbol{e}_1 + \cdots + \alpha_n\boldsymbol{e}_n = 0$ with $\alpha_1, \ldots, \alpha_n$ not all zero. Every element of $\mathbf{V}$ is then uniquely a linear combination of these basis elements.[13.28] In fact, this property is what characterizes a basis in the more general case when $\mathbf{V}$ can be infinite-dimensional, when the linear independence by itself is not sufficient.

Thus, given a basis $\boldsymbol{e} = (\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$, *any* element $\boldsymbol{x}$ of $\mathbf{V}$ can be uniquely written

$$\begin{aligned}\boldsymbol{x} &= x^1\boldsymbol{e}_1 + x^2\boldsymbol{e}_2 + \cdots + x^n\boldsymbol{e}_n \\ &= x^j\boldsymbol{e}_j\,,\end{aligned}$$

[13.27] Show that $\det \boldsymbol{T} = \lambda_1\lambda_2\cdots\lambda_n$, $\operatorname{trace} \boldsymbol{T} = \lambda_1 + \lambda_2 + \cdots + \lambda_n$.

[13.28] Show this.

(the indices *j not* being abstract here) where $(x^1, x^2, \ldots, x^n)$ is the ordered set of *components* of $\boldsymbol{x}$ with respect to $\boldsymbol{e}$ (compare §12.3). A non-singular linear transformation $\boldsymbol{T}$ always sends a basis to another basis; moreover, if $\boldsymbol{e}$ and $\boldsymbol{f}$ are any two given bases, then there is a unique $\boldsymbol{T}$ sending each $\boldsymbol{e}_a$ to its corresponding $\boldsymbol{f}_j$:

$$\boldsymbol{T}\boldsymbol{e}_j = \boldsymbol{f}_j.$$

In terms of components taken with respect to $\boldsymbol{e}$, the components of the basis elements $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ themselves are, respectively, $(1, 0, 0, \ldots, 0)$, $(0, 1, 0, \ldots, 0), \ldots, (0, 0, \ldots, 0, 1)$. In other words, the components of $\mathbf{e}_j$ are $(\delta^1_j, \delta^2_j, \delta^3_j, \ldots, \delta^n_j)$.[13.29] When all components are taken with respect to the $\boldsymbol{e}$ basis, we find that $\boldsymbol{T}$ is represented as the matrix $(T^i{}_j)$, where the components of $\boldsymbol{f}_j$ in the $\boldsymbol{e}$ basis would be[13.30]

$$(T^1{}_j, T^2{}_j, T^3{}_j, \ldots, T^n{}_j).$$

It should be recalled that the conceptual difference between a linear transformation and a matrix is that the latter refers to some basis-dependent presentation, whereas the former is abstract, not depending upon a basis.

Now, provided that each multiple eigenvalue of $\boldsymbol{T}$ (if there are any) satisfies $d = r$, i.e. its eigenspace dimensionality equals its multiplicity, it is possible to find a basis $(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$ for $\mathbf{V}$, each of which is an eigenvector of $\boldsymbol{T}$.[13.31] Let the corresponding eigenvalues be $\lambda_1, \lambda_2, \ldots, \lambda_n$:

$$\boldsymbol{T}\boldsymbol{e}_1 = \lambda_1\boldsymbol{e}_1, \quad \boldsymbol{T}\boldsymbol{e}_2 = \lambda_2\boldsymbol{e}_2, \ldots, \boldsymbol{T}\boldsymbol{e}_n = \lambda_n\boldsymbol{e}_n.$$

If, as above, $\boldsymbol{T}$ takes the $\boldsymbol{e}$ basis to the $\boldsymbol{f}$ basis, then the $\boldsymbol{f}$ basis elements are as above, so we have $\boldsymbol{f}_1 = \lambda_1\boldsymbol{e}_1, \boldsymbol{f}_2 = \lambda_2\boldsymbol{e}_2, \ldots, \boldsymbol{f}_n = \lambda_n\boldsymbol{e}_n$. It follows that $\boldsymbol{T}$, referred to the $\boldsymbol{e}$ basis, takes the *diagonal* matrix form

$$\begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \vdots\vdots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix},$$

that is $T^1_1 = \lambda_1, T^2_2 = \lambda_2, \ldots, T^n_n = \lambda_n$, the remaining components being zero. This *canonical form* for a linear transformation is very useful both conceptually and calculationally.[12]

[13.29] Explain this notation.

[13.30] Why? What are the components of $\boldsymbol{e}_i$ in the $\boldsymbol{f}$ basis?

[13.31] See if you can prove this. *Hint*: For each eigenvalue of multiplicity r, choose r linearly independent eigenvectors. Show that a linear relation between vectors of this entire collection leads to a contradiction when this relation is pre-multiplied by $\boldsymbol{T}$, successively.

13.6 Representation theory and Lie algebras

There is an important body of ideas (particularly significant for quantum theory) called the *representation theory* of groups. We saw a very simple example of a group representation in the discussion in §13.1, when we observed that the non-reflective symmetries of a square can be represented by complex numbers, the group multiplication being faithfully represented as actual multiplication of the complex numbers. However, nothing quite so simple can apply to non-Abelian groups, since the multiplication of complex numbers is commutative. On the other hand, linear transformations (or matrices) usually do not commute, so we may regard it as a reasonable prospect to represent non-Abelian groups in terms of them. Indeed, we already encountered this kind of thing at the beginning of §13.3, where we represented the rotation group O(3) in terms of linear transformations in three dimensions.

As we shall be seeing in Chapter 22, quantum mechanics is all to do with linear transformations. Moreover, various symmetry groups have crucial importance in modern particle physics, such as the rotation group O(3), the symmetry groups of relativity theory (Chapter 18), and the symmetries underlying particle interactions (Chapter 25). It is not surprising, therefore, that representations of these groups in particular, in terms of linear transformations, have fundamental roles to play in quantum theory.

It turns out that, quantum theory (particularly the quantum field theory of Chapter 26) is frequently concerned with linear transformations of *infinite*-dimensional spaces. For simplicity, however, I shall phrase things here just for representations by linear transformations in the finite-dimensional case. Most of the ideas that we shall encounter apply also in the case of infinite-dimensional representations, although there are differences that can be important in some circumstances.

What is a group representation? Consider a group $\mathcal{G}$. Representation theory is concerned with finding a subgroup of GL(n) (i.e. a multiplicative group of $n \times n$ matrices) with the property that, for any element g in $\mathcal{G}$, there is a corresponding linear transformation $\boldsymbol{T}(g)$ (belonging to GL(n)) such that the multiplication law in $\mathcal{G}$ is preserved by the operations of GL(n), i.e. for any two elements g, h of $\mathcal{G}$, we have

$$\boldsymbol{T}(g)\boldsymbol{T}(h) = \boldsymbol{T}(gh).$$

The representation is called *faithful* if $\boldsymbol{T}(g)$ is different from $\boldsymbol{T}(h)$ whenever g is different from h. In this case we have an *identical copy* of the group $\mathcal{G}$, as a *subgroup* of GL(n).

In fact, every finite group has a faithful representation in GL(n, $\mathbb{R}$), where n is the order of $\mathcal{G}$,[13.32] and there are frequently many non-faithful representations. On the other hand, it is not quite true that every (finite-dimensional) continuous group has a faithful representation in some GL(n). However, if we are not worried about the global aspects of the group, then a representation *is* always (locally) possible.[13]

There is a beautiful theory, due to the profoundly original Norwegian mathematician Sophus Lie (1842–1899), which leads to a full treatment of the local theory of continuous groups. (Indeed, continuous groups are commonly called 'Lie groups'; see §13.1.) This theory depends upon a study of *infinitesimal* group elements.[14] These infinitesimal elements define a kind of algebra—referred to as a *Lie algebra*—which provides us with complete information as to the local structure of the group. Although the Lie algebra may not provide us with the full *global* structure of the group, this is normally considered to be a matter of lesser importance.

What is a Lie algebra? Suppose that we have a matrix (or linear transformation) $\boldsymbol{I} + \varepsilon \boldsymbol{A}$ to represent an 'infinitesimal' element a of some continuous group $\mathcal{G}$, where ε is taken as 'small' (compare end of §13.4). When we form the matrix product of $\boldsymbol{I} + \varepsilon \boldsymbol{A}$ and $\boldsymbol{I} + \varepsilon \boldsymbol{B}$ to represent the product ab of two such elements a and b, we obtain

$$\begin{aligned}(\boldsymbol{I} + \varepsilon \boldsymbol{A})\ (\boldsymbol{I} + \varepsilon \boldsymbol{B}) &= \boldsymbol{I} + \varepsilon(\boldsymbol{A} + \boldsymbol{B}) + \varepsilon^2 \boldsymbol{A}\boldsymbol{B} \\ &= \boldsymbol{I} + \varepsilon(\boldsymbol{A} + \boldsymbol{B})\end{aligned}$$

if we are allowed to ignore the quantity ε^2, as being 'too small to count'. In accordance with this, the matrix *sum* $\boldsymbol{A} + \boldsymbol{B}$ represents the *group product* ab of two infinitesimal elements a and b.

Indeed, the sum operation is part of the Lie algebra of the quantities $\boldsymbol{A}, \boldsymbol{B}, \ldots$. But the sum is commutative, whereas the group $\mathcal{G}$ could well be non-Abelian, so we do not capture much of the structure of the group if we consider only sums (in fact, only the dimension of $\mathcal{G}$). The non-Abelian nature of $\mathcal{G}$ is expressed in the *group commutators* which are the expressions[13.33]

$$a\ b\ a^{-1}\ b^{-1}.$$

[13.32] Show this. *Hint*: Label each column of the representing matrix by a separate element of the finite group $\mathcal{G}$, and also label each row by the corresponding group element. Place a 1 in any position in the matrix for which a certain relation holds (find it!) between the element of $\mathcal{G}$ labelling the row, that labelling the column, and the element of $\mathcal{G}$ that this particular matrix is representing. Place a 0 whenever this relation does not hold.

[13.33] Why is this expression just the *identity* group element when a and b commute?

Let us write this out in terms of $\boldsymbol{I}+\varepsilon\boldsymbol{A}$, etc., taking note of the power series expression $(\boldsymbol{I}+\varepsilon\boldsymbol{A})^{-1}=\boldsymbol{I}-\varepsilon\boldsymbol{A}+\varepsilon^2\boldsymbol{A}^2-\varepsilon^3\boldsymbol{A}^3+\cdots$ (this series being easily checked by multiplying both sides by $\boldsymbol{I}+\varepsilon\boldsymbol{A}$). Now it is ε^3 that we ignore as being 'too small to count', but we keep ε^2, whence[13.34]

$$(\boldsymbol{I}+\varepsilon\boldsymbol{A})\,(\boldsymbol{I}+\varepsilon\boldsymbol{B})\,(\boldsymbol{I}+\varepsilon\boldsymbol{A})^{-1}\,(\boldsymbol{I}+\varepsilon\boldsymbol{B})^{-1}$$

$$=(\boldsymbol{I}+\varepsilon\boldsymbol{A})\,(\boldsymbol{I}+\varepsilon\boldsymbol{B})\,(\boldsymbol{I}-\varepsilon\boldsymbol{A}+\varepsilon^2\boldsymbol{A}^2)\,(\boldsymbol{I}-\varepsilon\boldsymbol{B}+\varepsilon^2\boldsymbol{B}^2)$$

$$=\boldsymbol{I}+\varepsilon^2(\boldsymbol{AB}-\boldsymbol{BA})$$

This tells us that if we are to keep track of the precise way in which the group $\mathcal{G}$ is non-Abelian, we must take note of the 'commutators', or *Lie brackets*

$$[\boldsymbol{A},\,\boldsymbol{B}]=\boldsymbol{AB}-\boldsymbol{BA}.$$

The Lie algebra is now constructed by means of repeated application of the operations $+$, its inverse $-$, and the bracket operation $[\ ,\ ]$, where it is customary also to allow the multiplication by ordinary numbers (which might be real or complex). The 'additive' aspect of the algebra has the usual vector-space structure (as with quaternions, in §11.1). In addition, Lie bracket satisfies distributivity, etc., namely

$$[\boldsymbol{A}+\boldsymbol{B},\,\boldsymbol{C}]=[\boldsymbol{A},\,\boldsymbol{C}]+[\boldsymbol{B},\,\boldsymbol{C}],\ [\lambda\boldsymbol{A},\,\boldsymbol{B}]=\lambda[\boldsymbol{A},\,\boldsymbol{B}],$$

the antisymmetry property

$$[\boldsymbol{A},\,\boldsymbol{B}]=-[\boldsymbol{B},\,\boldsymbol{A}],$$

(whence also $[\boldsymbol{A},\,\boldsymbol{C}+\boldsymbol{D}]=[\boldsymbol{A},\,\boldsymbol{C}]+[\boldsymbol{A},\boldsymbol{D}]$, $[\boldsymbol{A},\,\lambda\boldsymbol{B}]=\lambda[\boldsymbol{A},\,\boldsymbol{B}]$), and an elegant relation known as the *Jacobi identity*[13.35]

$$[\boldsymbol{A},\,[\boldsymbol{B},\,\boldsymbol{C}]]+[\boldsymbol{B},[\boldsymbol{C},\boldsymbol{A}]]+[\boldsymbol{C},[\boldsymbol{A},\boldsymbol{B}]]=0$$

(a more general form of which will be encountered in §14.6).

We can choose a basis $(\boldsymbol{E}_1,\,\boldsymbol{E}_2\,,\ldots,\,\boldsymbol{E}_N)$ for the vector space of our matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, ... (where N is the dimension of the group $\mathcal{G}$, if the representation is faithful). Forming their various commutators $[\boldsymbol{E}_\alpha,\,\boldsymbol{E}_\beta]$, we express these in terms of the basis elements, to obtain relations (using the summation convention)

$$[\boldsymbol{E}_\alpha,\,\boldsymbol{E}_\beta]=\gamma_{\alpha\beta}{}^{\chi}\boldsymbol{E}_\chi.$$

[13.34] Spell out this 'order ε^2' calculation.

[13.35] Show all this.

The N^3 component quantities $\gamma_{\alpha\beta}{}^{\chi}$ are called *structure constants* for $\mathcal{G}$. They are not all independent because they satisfy (see §11.6 for bracket notation)

$$\gamma_{\alpha\beta}{}^{\chi} = -\gamma_{\beta\alpha}{}^{\chi}, \quad \gamma_{[\alpha\beta}{}^{\xi}\gamma_{\chi]\xi}{}^{\zeta} = 0,$$

by virtue of the above antisymmetry and Jacobi identity.[13.36] These relations are given in diagrammatic form in Fig. 13.11.

It is a remarkable fact that the structure of the Lie algebra for a faithful representation (basically, the knowledge of the structure constants $\gamma_{\alpha\beta}{}^{\chi}$) is sufficient to determine the precise local nature of the group $\mathcal{G}$. Here, 'local' means in a (sufficiently small) N-dimensional open region $\mathcal{N}$ surrounding the identity element $\boldsymbol{I}$ in the 'group manifold' $\widetilde{\mathcal{G}}$ whose points represent the different elements of $\mathcal{G}$ (see Fig. 13.12). In fact, starting from a Lie group element $\boldsymbol{A}$, we can construct a corresponding actual *finite* (i.e. non-infinitesimal) group element by means of the 'exponentiation' operation $e^{\boldsymbol{A}}$ defined at the end of §13.4. (This will be considered a little more fully in §14.6.) Thus, the theory of representations of continuous groups by linear transformations (or by matrices) may be largely transferred to the study of representations of Lie algebras by such transformations—which, indeed, is the normal practice in physics.

This is particularly important in quantum mechanics, where the Lie algebra elements themselves, in a remarkable way, frequently have direct interpretations as physical quantities (such as angular momentum, when the group $\mathcal{G}$ is the rotation group, as we shall be seeing later in §22.8).

The Lie algebra matrices tend to be considerably simpler in structure than the corresponding Lie group matrices, being subject to linear rather

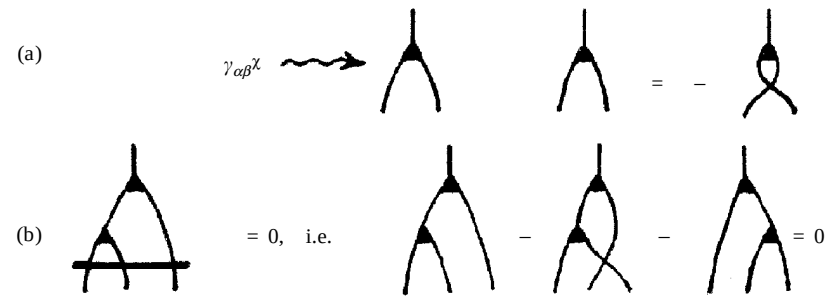

Fig. 13.11 (a) Structure constants $\gamma_{\alpha\beta}{}^{\chi}$ in diagrammatic form, depicting antisymmetry in α, β and (b) the Jacobi identity.

[13.36] Show this.

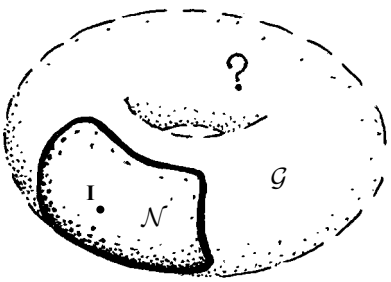

Fig. 13.12 The Lie algebra for a (faithful) representation of a Lie group $\mathcal{G}$ (basically, knowledge of the structure constants $\gamma_{\alpha\beta}{}^{\chi}$) determines the local structure of $\mathcal{G}$, i.e. it fixes the structure of $\mathcal{G}$ within some (sufficiently small) open region $\mathcal{N}$ surrounding the identity element **I**, but it does not tell us about the global nature of $\mathcal{G}$.

than nonlinear restrictions (see §13.10 for the case of the classical groups). This procedure is beloved of quantum physicists!

13.7 Tensor representation spaces; reducibility

There are ways of building up more elaborate representations of a group $\mathcal{G}$, starting from some particular one. How are we to do that? Suppose that $\mathcal{G}$ is represented by some family $\mathcal{T}$ of linear transformations, acting on an n-dimensional vector space $\mathbf{V}$. Such a $\mathbf{V}$ is called a *representation space* for $\mathcal{G}$. Any element t of $\mathcal{G}$ is now represented by a corresponding linear transformation $\boldsymbol{T}$ in $\mathcal{T}$, where $\boldsymbol{T}$ effects $\boldsymbol{x} \mapsto \boldsymbol{T}\boldsymbol{x}$ for each $\boldsymbol{x}$ belonging to $\mathbf{V}$. In the (abstract) index notation (§12.7) we write this $x^a \mapsto T^a{}_b x^b$, as in §13.3, or in diagrammatic form, as in Fig. 13.6a. Let us see how we can find other representation spaces for $\mathcal{G}$, starting from the given one $\mathbf{V}$.

As a first example, recall, from §12.3, the definition of the *dual space* $\mathbf{V}^*$ of $\mathbf{V}$. The elements of $\mathbf{V}^*$ are defined as linear maps from $\mathbf{V}$ to the scalars. We can write the action of $\boldsymbol{y}$ (in $\mathbf{V}^*$) on an element $\boldsymbol{x}$ in $\mathbf{V}$ as $y_a x^a$, in the index notation (§12.7). The notation $\boldsymbol{y} \bullet \boldsymbol{x}$ would have been used earlier (§12.3) for this ($\boldsymbol{y} \bullet \boldsymbol{x} = y_a x^a$), but now we can also use the *matrix* notation

$$\mathbf{y}\mathbf{x} = y_a x^a,$$

where we take $\mathbf{y}$ to be a *row* vector (i.e. a $1 \times n$ matrix) and $\mathbf{x}$ a *column* vector (an $n \times 1$ matrix). In accordance with our transformation $\mathbf{x} \mapsto \boldsymbol{T}\mathbf{x}$, now thought of as a *matrix* transformation, the dual space $\mathbf{V}^*$ undergoes the linear transformation

$$\mathbf{y} \mapsto \mathbf{y}\mathbf{S}, \quad \text{i.e.} \quad y_a \mapsto y_b S^b{}_a,$$

where $\mathbf{S}$ is the *inverse* of $\mathbf{T}$:

$$\boldsymbol{S} = \boldsymbol{T}^{-1}, \quad \text{so} \quad S^a{}_b T^b{}_c = \delta^a_c,$$

since, if $\mathbf{x} \mapsto \mathbf{T}\mathbf{x}$, we need $\mathbf{y} \mapsto \mathbf{y}\mathbf{T}^{-1}$ to ensure that $\mathbf{y}\mathbf{x}$ is preserved by $\mapsto$.

The use of a row vector $\mathbf{y}$, in the above, gives us a non-standard multiplication ordering. It is more usual to write things the other way around, by employing the notation of the *transpose* $\mathbf{A}^{\mathsf{T}}$ of a matrix $\mathbf{A}$. The elements of the matrix $\mathbf{A}^{\mathsf{T}}$ are the same as those of $\mathbf{A}$, but with rows and columns interchanged. If $\mathbf{A}$ is square ($n \times n$), then so is $\mathbf{A}^{\mathsf{T}}$, its elements being those of $\mathbf{A}$ reflected in its main diagonal (see §13.3). If $\boldsymbol{A}$ is rectangular ($m \times n$), then $\mathbf{A}^{\mathsf{T}}$ is $n \times m$, correspondingly reflected. Thus $\mathbf{y}^{\mathsf{T}}$ is a standard column vector, and we can write the above $\mathbf{y} \mapsto \mathbf{y}\mathbf{S}$ as

$$\mathbf{y}^{\mathsf{T}} \mapsto \mathbf{S}^{\mathsf{T}}\mathbf{y}^{\mathsf{T}},$$

since the transpose operation $^{\mathsf{T}}$ reverses the order of multiplication: $(\mathbf{A}\mathbf{B})^{\mathsf{T}} = \mathbf{B}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}$. We thus see that the dual space V^*, of any representation space V is itself a representation space of $\mathcal{G}$. Note that the inverse operation $^{-1}$ also reverses multiplication order, $(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$,[13.37] so the multiplication ordering needed for a representation is restored.

The same kinds of consideration apply to the various vector spaces of tensors constructed from V; see §12.8. We recall that a tensor $\boldsymbol{Q}$ of valence $[^p_q]$ (over the vector space V) has an index description as a quantity

$$Q^{f\ldots h}_{a\ldots c},$$

with q lower and p upper indices. We can add tensors to other tensors of the same valence and we can multiply them by scalars; tensors of fixed valence $[^p_q]$ form a vector space of dimension n^{p+q} (the total number of components).[13.38] Abstractly, we think of $\boldsymbol{Q}$ as belonging to a vector space that we refer to as the *tensor product*

$$\mathsf{V}^* \otimes \mathsf{V}^* \otimes \ldots \otimes \mathsf{V}^* \otimes \mathsf{V} \otimes \mathsf{V} \otimes \ldots \otimes \mathsf{V}$$

of q copies of the dual space V^* and p copies of V ($p, q \geq 0$). (We shall come to this notion of 'tensor product' a little more fully in §23.3.) Recall the abstract definition of a tensor, given in §12.8, as a multilinear function.

[13.37] Why?

[13.38] Why this number?

This will suffice for our purposes here (although there are certain subtleties in the case of an infinite-dimensional $\mathbf{V}$, of relevance to the applications to many-particle quantum states, needed in §23.8).[15]

Whenever a linear transformation $x^a \mapsto T^a{}_b x^b$ is applied to $\mathbf{V}$, this induces a corresponding linear transformation on the above tensor product space, given explicitly by[13.39]

$$Q^{f\ldots h}_{a\ldots c} \mapsto S^{a'}{}_a \ldots S^{c'}{}_c T^f{}_{f'} \ldots T^h{}_{h'} Q^{f'\ldots h'}_{a'\ldots c'}.$$

All these indices require good eyesight and careful scrutiny, in order to make sure of what is summed with what; so I recommend the diagrammatic notation, which is clearer, as illustrated in Fig. 13.13. We see that each lower index of $Q_{\cdots}^{\cdots}$ transforms by the inverse matrix $\mathbf{S} = \mathbf{T}^{-1}$ (or, rather, by $\mathbf{S}^{\mathsf{T}}$), as with y_a and each upper index by $\mathbf{T}$, as with x^a. Accordingly, the space of $\left[{}^p_q\right]$-valent tensors over $\mathbf{V}$ is also a representation space for $\mathcal{G}$, of dimension n^{p+q}.

These representation spaces are, however, likely to be what is called *reducible*. To illustrate this situation, consider the case of a $\left[{}^2_0\right]$-valent tensor Q^{ab}. Any such tensor can be split into its *symmetric* part $Q^{(ab)}$ and its *antisymmetric* part $Q^{[ab]}$ (§12.7 and §11.6):

$$Q^{ab} = Q^{(ab)} + Q^{[ab]},$$

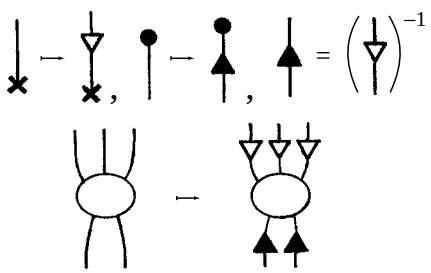

Fig. 13.13 The linear transformation $x^a \mapsto T^a{}_b x^b$, applied to $\mathbf{x}$ in the vector space V (with $\boldsymbol{T}$ depicted as a white triangle), extends to the dual space $\mathbf{V}^*$ by use of the inverse $\boldsymbol{S} = \boldsymbol{T}^{-1}$ (depicted as a black triangle) and thence to the spaces $\mathbf{V}^* \bigotimes \ldots \bigotimes \mathbf{V}^* \bigotimes \mathbf{V} \bigotimes \ldots \bigotimes \mathbf{V}$ of $\left[{}^p_q\right]$-valent tensors $\boldsymbol{Q}$. The case $p = 3$, $q = 2$ is illustrated, with $\boldsymbol{Q}$ shown as an oval with three arms and two legs undergoing $Q_{ab}{}^{cde} \propto S^{a'}{}_a S^{b'}{}_b T^c{}_{c'} T^d{}_{d'} T^e{}_{e'} Q_{a'b'}{}^{c'd'e'}$.

[13.39] Show this.

where

$$Q^{(ab)} = \tfrac{1}{2}(Q^{ab} + Q^{ba}), \quad Q^{[ab]} = \tfrac{1}{2}(Q^{ab} - Q^{ba}).$$

The dimension of the *symmetric* space $\mathbf{V}_+$ is $\frac{1}{2}n(n+1)$, and that of the *antisymmetric* space $\mathbf{V}_-$ is $\frac{1}{2}n(n-1)$.[13.40] It is not hard to see that, under the transformation $x^a \mapsto T^a{}_b x^b$, so that $Q^{ab} \mapsto T^a{}_c T^b{}_d Q^{cd}$, the symmetric and antisymmetric parts transform to tensors which are again, respectively, symmetric, and antisymmetric.[13.41] Accordingly, the spaces $\mathbf{V}_+$ and $\mathbf{V}_-$ are, separately, representation spaces for $\mathcal{G}$. By choosing a basis for $\mathbf{V}$ where the first $\frac{1}{2}n(n+1)$ basis elements are in $\mathbf{V}_+$ and the remaining $\frac{1}{2}n(n-1)$ are in $\mathbf{V}_-$, we obtain our representation with all matrices being of the $n^2 \times n^2$ 'block-diagonal' form

$$\begin{pmatrix} \mathbf{A} & \mathbf{O} \\ \mathbf{O} & \mathbf{B} \end{pmatrix},$$

where $\mathbf{A}$ stands for a $\frac{1}{2}n(n+1) \times \frac{1}{2}n(n+1)$ matrix and $\mathbf{B}$ for a $\frac{1}{2}n(n-1) \times \frac{1}{2}n(n-1)$ matrix, the two $\mathbf{O}$s standing for the appropriate rectangular blocks of zeros.

A representation of this form is referred to as the *direct sum* of the representation given by the $\mathbf{A}$ matrices and that given by the $\mathbf{B}$ matrices. The representation in terms of $\left[{}^2_0\right]$-valent tensors is therefore *reducible*, in this sense.[13.42] The notion of 'direct sum' also extends to any number (perhaps infinite) of smaller representations.

In fact there is a more general meaning for the term 'reducible representation', namely one for which there is a choice of basis for which all the matrices of the representation can be put in the somewhat more complicated form

$$\begin{pmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{O} & \mathbf{B} \end{pmatrix},$$

where $\mathbf{A}$ is $p \times p$, $\mathbf{B}$ is $q \times q$, and $\mathbf{C}$ is $p \times q$, with $p, q \geq 1$ (for fixed p and q). Note that, if the representing matrices all have this form, then the $\mathbf{A}$ matrices and the $\mathbf{B}$ matrices each individually constitute a (smaller) representation of $\mathcal{G}$.[13.43] If the $\mathbf{C}$ matrices are all zero, we get the earlier case where the representation is the direct sum of these two smaller representations. A representation is called *irreducible* if it is not reducible (with C present or

[13.40] Show this.

[13.41] Explain this.

[13.42] Show that the representation space of $\left[{}^1_1\right]$-valent tensors is also reducible. *Hint*: Split any such tensor into a 'trace-free' part and a 'trace' part.

[13.43] Confirm this.

not). A representation is called *completely reducible* if we never get the above situation (with non-zero C), so that it is a direct sum of irreducible representations.

There is an important class of continuous groups, known as *semi-simple* groups. This extensively studied class includes the simple groups referred to in §13.2. Compact semi-simple groups have the pleasing property that *all* their representations are completely reducible. (See §12.6, Fig. 12.13 for the definition of 'compact'.) It is sufficient to study *irreducible* representations of such a group, every representation being just a direct sum of these irreducible ones. In fact, every irreducible representation of such a group is finite-dimensional (which is not the case if we allow a semi-simple group to be non-compact, when representations that are *not* completely reducible can also occur).

What is a semi-simple group? Recall the 'structure constants' $\gamma_{\alpha\beta}^{\chi}$ of §13.6, which specify the Lie brackets and define the local structure of the group $\mathcal{G}$. There is a quantity of considerable importance known[16] as the '*Killing form*' κ that can be constructed from $\gamma_{\alpha\beta}{}^{\chi}$:[13.44]

$$\kappa_{\alpha\beta} = \gamma_{\alpha\zeta}{}^{\xi}\ \gamma_{\beta\xi}{}^{\zeta} = \kappa_{\beta\alpha}.$$

The diagrammatic form of this expression is given in Fig. 13.14. The condition for $\mathcal{G}$ to be semi-simple is that the matrix $\kappa_{\alpha\beta}$ be non-singular.

Some remarks are appropriate concerning the condition of compactness of a semi-simple group. For a given set of structure constants $\gamma_{\alpha\beta}{}^{\chi}$, assuming that we can take them to be real numbers, we could consider either the real or the complex Lie algebra obtained from them. In the complex case, we do not get a compact group $\mathcal{G}$, but we might do so in the real case. In fact, compactness occurs in the real case when $-\kappa_{\beta\alpha}$ is what is called *positive definite* (the meaning of which term we shall come to in §13.8). For fixed $\gamma_{\alpha\beta}{}^{\chi}$, in the case of a real group $\mathcal{G}$, we can always construct the *complexification* $\mathbb{C}\mathcal{G}$ (at least locally) of $\mathcal{G}$ which comes about merely by using the same $\gamma_{\alpha\beta}{}^{\chi}$, but with complex coefficients in the Lie algebra. However, different real groups $\mathcal{G}$ might sometimes give rise to the same[17] $\mathbb{C}\mathcal{G}$. These different real groups are called different *real forms* of the complex group. We shall be seeing important

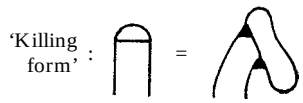

Fig. 13.14 The 'Killing form' $\kappa_{\alpha\beta}$ defined from the structure constants $\gamma_{\alpha\zeta}{}^{\xi}$ by $\kappa_{\alpha\beta} = \gamma_{\alpha\zeta}{}^{\xi}\gamma_{\beta\xi}{}^{\zeta}$.

[13.44] Why does $\kappa_{\alpha\beta} = \kappa_{\beta\alpha}$?

instances of this in later chapters, especially in §18.2, where the Euclidean motions in 4 dimensions and the Lorentz/Poincaré symmetries of special relativity are compared. It is a remarkable property of any complex semi-simple Lie group that it has exactly one real form $\mathcal{G}$ which is compact.

13.8 Orthogonal groups

Now let us return to the orthogonal group. We already saw at the beginning of §13.3 how to represent O(3) or SO(3) faithfully as linear transformations of a 3-dimensional real vector space, with ordinary Cartesian coordinates (x,y,z), where the sphere

$$x^2 + y^2 + z^2 = 1$$

is to be left invariant (the upper index 2 meaning the usual 'squared'). Let us write this equation in terms of the index notation (§12.7), so that we can generalize to n dimensions. The equation of our sphere can now be written

$$g_{ab}x^a x^b = 1,$$

which stands for $(x^1)^2 + \cdots + (x^n)^2 = 1$, the components g_{ab} being given by

$$g_{ab} = \begin{cases} 1 & \text{if } a = b, \\ 0 & \text{if } a \neq b. \end{cases}$$

In the diagrammatic notation, I recommend simply using a 'hoop' for g_{ab}, as indicated in Fig. 13.15a. I shall also use the notation g^{ab} (with the *same* explicit components as g_{ab}) for the *inverse* quantity ('inverted hoop' in Fig. 13.15a):

$$g_{ab}\, g^{bc} = \delta_a^c = g^{cb}g_{ba}.$$

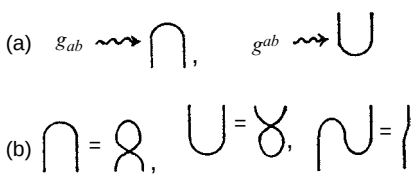

Fig. 13.15 (a) The metric g_{ab} and its inverse g^{ab} in the 'hoop' diagrammatic notation. (b) The relations $g_{ab} = g_{ba}$ (i.e. $\boldsymbol{g}^{\mathrm{T}} = \boldsymbol{g}$), $g_{ab} = g^{ba}$, and $g_{ab}g^{bc} = \delta_a^c$ in diagrammatic notation.

The puzzled reader might very reasonably ask why I have introduced two new notations, namely g_{ab} and g^{ab} for precisely the same matrix components that I denoted by δ^a_b in §13.3! The reason has to do with the consistency of the notation and with what happens when a linear transformation is applied to the *coordinates*, according to some replacement

$$x^a \mapsto t^a{}_b x^b,$$

$t^a{}_b$ being non-singular, so that it has an inverse $s^a{}_b$:

$$t^a{}_b s^b{}_c = \delta^a_c = s^a{}_b t^b{}_c.$$

This is formally the same as the type of linear transformation that we considered in §§13.3,7, but we are now thinking of it in a quite different way. In those sections, our linear transformation was thought of as *active*, so that the vector space **V** was viewed as being actually *moved* (over itself). Here we are thinking of the transformation as *passive* in that the objects under consideration—and, indeed, the vector space **V** itself—remain pointwise fixed, but the representations in terms of coordinates are changed. Another way of putting this is that the basis $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$ that we had previously been using (for the representation of vector/tensor quantities in terms of components[18]) is to be replaced by some other basis. See Fig. 13.16.

In direct correspondence with what we saw in §13.7 for the active transformation of a tensor, we find that the corresponding passive change in the components $Q^{a\ldots c}_{p\ldots r}$ of a tensor $\boldsymbol{Q}$ is given by[13.45]

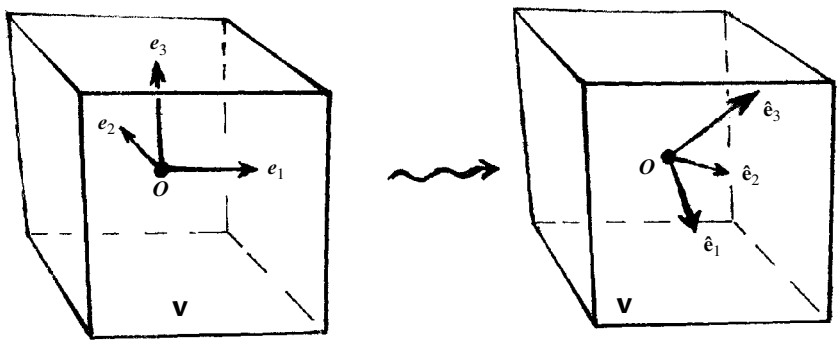

Fig. 13.16 A passive transformation in a vector space V leaves V pointwise fixed, but changes its coordinate description, i.e. the basis $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ is replaced by some other basis (case $n = 3$ illustrated).

[13.45] Use Note 13.18 to establish this.

$$Q^{a\ldots c}_{p\ldots r} \mapsto t^a{}_d \cdots t^c{}_f \; Q^{d\ldots f}_{j\ldots l} \; s^j{}_p \ldots s^l{}_r.$$

Applying this to δ^a_b, we find that its components are completely unaltered,[13.46] whereas this is *not* the case for g_{ab}. Moreover, after a general such coordinate change, the components g^{ab} will be quite different from g_{ab} (inverse matrices). Thus, the reason for the additional symbols g^{ab} and g_{ab} is simply that they can only represent the same matrix of components as does δ^a_b in special types of coordinate system ('Cartesian' ones) and, in general, the components are just *different*. This has a particular importance for general relativity, where the coordinate system cannot normally be arranged to have this special (Cartesian) form.

A general coordinate change can make the matrix of components g_{ab} a more complicated although not completely general matrix. It retains the property of symmetry between a and b giving a *symmetric* matrix. The term 'symmetric' tells us that the square array of components is symmetrical about its main diagonal, i.e. $\mathbf{g}^{\mathsf{T}} = \mathbf{g}$ (using the 'transpose' notation of §13.3). In index-notation terms, this symmetry is expressed as either of the two equivalent[13.47] forms

$$g_{ab} = g_{ba},\; g^{ab} = g^{ba},$$

and see Fig. 13.15b for the diagrammatic form of these relations.

What about going in the opposite direction? Can any non-singular $n \times n$ real symmetric matrix be reduced to the component form of a Kronecker delta? Not quite—not by a real linear transformation of coordinates. What it can be reduced to by such means is this same form except that there are some terms 1 and some terms -1 along the main diagonal. The number, p, of these 1 terms and the number, q, of -1 terms is an *invariant*, which is to say we cannot get a different number by trying some other real linear transformation. This invariant (p, q) is called the *signature* of $\mathbf{g}$. (Sometimes it is $p - q$ that is called the signature; sometimes one just writes $+\ldots+-\ldots-$ with the appropriate number of each sign.) In fact, this works also for a *singular* $\boldsymbol{g}$, but then we need some 0s along the main diagonal also and the number of 0s becomes part of the signature as well as the number of 1s and the number of -1s. If we only have 1s, so that $\mathbf{g}$ is non-singular and also $q = 0$, then we say that $\boldsymbol{g}$ is *positive-definite*. A non-singular $\mathbf{g}$ for which $p = 1$ and $q \neq 0$ (or $q = 1$ and $p \neq 0$) is called *Lorentzian*, in honour of the Dutch physicist H.A. Lorentz (1853–1928), whose important work in this connection provided one of the foundation stones of relativity theory; see §§17.6–9 and §§18.1–3.

[13.46] Why?

[13.47] Why equivalent?

An alternative characterization of a positive-definite matrix **A**, of considerable importance in certain other contexts (see §20.3, §24.3, §29.3) is that the real symmetric matrix **A** satisfy

$$\mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x} > 0$$

for all $\boldsymbol{x} \neq 0$. In index notation, this is: '$A_{ab}x^a x^b > 0$ unless the vector x^a vanishes'.[13.48] We say that A is *non-negative-definite* (or *positive-semi-definite*) if this holds but with $\geq$ in place of $>$ (so we now allow $\mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x} = 0$ for some non-zero **x**).

Under appropriate circumstances, a symmetric non-singular $\left[{}^{0}_{2}\right]$-tensor g_{ab}, is called a *metric*—or sometimes a *pseudometric* when $\boldsymbol{g}$ is not positive definite. This terminology applies if we are to use the quantity ds, defined by its square $\mathrm{d}s^2 = g_{ab}\mathrm{d}x^a\mathrm{d}x^b$, as providing us with some notion of 'distance' along curves. We shall be seeing in §14.7 how this notion applies to curved manifolds (see §10.2, §§12.1,2), and in §17.8 how, in the Lorentzian case, it provides us with a 'distance' measure which is actually the *time* of relativity theory. We sometimes refer to the quantity

$$|\boldsymbol{v}| = (g_{ab}v^a v^b)^{\frac{1}{2}}$$

as the *length* of the vector $\boldsymbol{v}$, with index form v^a.

Let us return to the definition of the orthogonal group O(n). This is simply the group of linear transformations in n dimensions—called *orthogonal* transformations—that preserve a given positive-definite $\boldsymbol{g}$. 'Preserving' $\boldsymbol{g}$ means that an orthogonal transformation $\boldsymbol{T}$ has to satisfy

$$g_{ab}T^a{}_c T^b{}_d = g_{cd}.$$

This is an example of the (active) tensor transformation rule described in §13.7, as applied to g_{ab} (and see Fig. 13.17 for the diagrammatic form of this equation). Another way of saying this is that the metric form $\mathrm{d}s^2$ of the previous paragraph is unchanged by orthogonal transformations. We can, if we please, insist that the components g_{ab} be actually the Kronecker delta—this, in effect, providing the definition of O(3) given in §§13.1,3—but the group comes out the same[19] whatever positive-definite $n \times n$ array of g_{ab} we choose.[13.49]

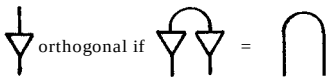

Fig. 13.17 $\boldsymbol{T}$ is an orthogonal transformation if $g_{ab}T^a{}_c T^b{}_d = g_{cd}$.

[13.48] Can you confirm this characterization?

[13.49] Explain why.

With the particular component realization of g_{ab} as the Kronecker delta, the matrices describing our orthogonal transformations are those satisfying[13.50]

$$\boldsymbol{T}^{-1} = \boldsymbol{T}^{\mathsf{T}},$$

called *orthogonal matrices*. The real orthogonal $n \times n$ matrices provide a concrete realization of the group $\mathrm{O}(n)$. To specialize to the non-reflective group $\mathrm{SO}(n)$, we require that the determinant be equal to unity:[13.51]

$$\det \boldsymbol{T} = 1.$$

We can also consider the corresponding *pseudo-orthogonal* groups $\mathrm{O}(p, q)$ and $\mathrm{SO}(p, q)$ that are obtained when $\boldsymbol{g}$, though non-singular, is not necessarily positive definite, having the more general signature (p, q). The case when $p = 1$ and $q = 3$ (or equivalently $p = 3$ and $q = 1$), called the *Lorentz* group, plays a fundamental role in relativity theory, as indicated above. We shall also be finding (if we ignore time-reflections) that the Lorentz group is the same as the group of symmetries of the hyperbolic 3-space that was described in §2.7, and also (if we ignore space reflections) of the group of symmetries of the Riemann sphere, as achieved by the bilinear (Möbius) transformations as studied in §8.2. It will be better to delay the explanations of these remarkable facts until our investigation of the Minkowski spacetime geometry of special relativity theory (§§18.4,5). We shall also be seeing in §33.2 that these facts have a seminal significance for twistor theory.

How 'different' are the various groups $\mathrm{O}(p, q)$, for $p + q = n$, for fixed n? (The positive-definite and Lorentzian cases are contrasted, for $n = 2$ and $n = 3$, in Fig. 13.18.) They are closely related, all having the same dimension $\frac{1}{2}n(n-1)$; they are what are called *real forms* of one and the same complex group $\mathrm{O}(n, \mathbb{C})$, the *complexification* of $\mathrm{O}(n)$. This complex group is defined in the same way as $\mathrm{O}(n)$ $(= \mathrm{O}(n, \mathbb{R}))$, but where the linear transformations are allowed to be *complex*. Indeed, although I have phrased my considerations in this chapter in terms of real linear transformations, there is a parallel discussion where 'complex' replaces 'real' throughout. (Thus the coordinates x^a become complex and so do the components of our matrices.) The only essential difference, in what has been said above, arises with the concept of *signature*. There are complex linear coordinate transformations that can convert a -1 in a diagonal realization of g_{ab} into a $+1$ and *vice versa*,[13.52] so we do not now have a

[13.50] Explain this. What is $\boldsymbol{T}^{-1}$ in the pseudo-orthogonal cases (defined in the next paragraph)?

[13.51] Explain why this is equivalent to preserving the volume form $\varepsilon_{a\ldots c}$, i.e. $\varepsilon_{a\ldots c}T^a_p \ldots T^c_r = \varepsilon_{p\ldots r}$? Moreover, why is the preservation of its sign sufficient?

[13.52] Why?

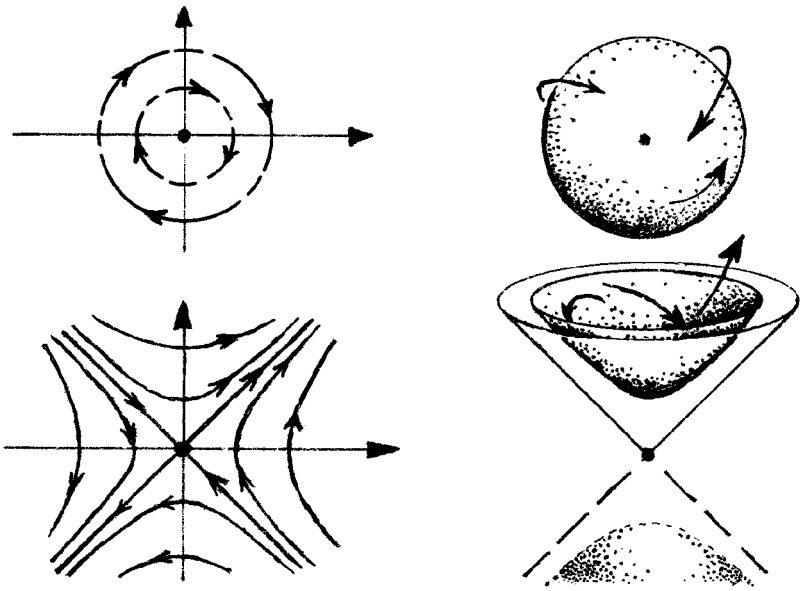

Fig. 13.18 (a) O(2,0) and O(1,1) are contrasted. (b) O(3,0) and O(1,2) are similarly contrasted, the 'unit sphere' being illustrated in each case. For O(1,2) (see §§2.4,5, §18.4), this 'sphere' is a hyperbolic plane (or two copies of such).

meaningful notion of signature. The only invariant[20] of $\boldsymbol{g}$, in the complex case, is what is called its *rank*, which is the number of non-zero terms in its diagonal realization. For a non-singular $\boldsymbol{g}$, the rank has to be maximal, i.e. n.

When is the difference between these various real forms important and when is it not? This can be a delicate question, but physicists are often rather cavalier about the distinctions, even though these can be important. The positive-definite case has the virtue that the group is compact, and much of the mathematics is easier for such situations (see §13.7). Sometimes people blithely carry over results from the compact case to the non-compact cases ($p \neq 0 \neq q$), but this is often not justified. (For example, in the compact case, one need only be concerned with representations that are finite-dimensional, but in the non-compact case additional infinite-dimensional representations arise.) On the other hand, there are other situations in which considerable insights can be obtained by ignoring the distinctions. (We may compare this with Lambert's discovery of the formula, in terms of angles, of the area of a hyperbolic triangle, given in §2.4. He obtained his formula by allowing his sphere to have an imaginary radius. This is similar to a signature change, which amounts to allowing some coordinates to have imaginary values. In §18.4, Fig. 18.9, I shall try

to make the case that Lambert's approach to non-Euclidean geometry is perfectly justifiable.)

The different possible real forms of $\mathrm{O}(n, \mathbb{C})$ are distinguished by certain set of *inequalities* on the matrix elements (such as $\det \boldsymbol{T} > 0$). A feature of *quantum theory* is that such inequalities are often *violated* in physical processes. For example, *imaginary* quantities can, in a sense, have a physically *real* significance in quantum mechanics, so the distinction between different signatures can become blurred. On the other hand, it is my impression that physicists are often somewhat less careful about these matters than they should be. Indeed, this question will have considerable relevance for us in our examination of a number of modern theories (§28.9, §31.11, §32.3). But more of this later. This is the 'can of worms' that I hinted at in §11.2!

13.9 Unitary groups

The group $\mathrm{O}(n, \mathbb{C})$ provides us with *one* way in which the notion of a 'rotation group' can be generalized from the real numbers to the complex. But there is another way which, in certain contexts, has an even greater significance. This is the notion of a *unitary* group.

What does 'unitary' mean? The orthogonal group is concerned with the preservation of a *quadratic form*, which we can write equivalently as $g_{ab}x^a x^b$ or $\mathbf{x}^{\mathsf{T}}\mathbf{g}\mathbf{x}$. For a unitary group, we use *complex* linear transformations which preserve instead what is called a *Hermitian* form (after the important 19th century French mathematician Charles Hermite 1822–1901).

What is a Hermitian form? Let us first return to the orthogonal case. Rather than a quadratic form (in $\boldsymbol{x}$), we could equally have used the symmetric *bilinear* form (in $\boldsymbol{x}$ and $\boldsymbol{y}$)

$$\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) = g_{ab}x^a y^b = \mathbf{x}^{\mathsf{T}}\mathbf{g}\mathbf{y}.$$

This arises as a particular instance of the 'multilinear function' definition of a tensor given in §12.8, as applied to the $\left[\begin{smallmatrix}2\\0\end{smallmatrix}\right]$ tensor $\boldsymbol{g}$ (and putting $\boldsymbol{y} = \boldsymbol{x}$, we retrieve the quadratic form above). The symmetry of $\boldsymbol{g}$ would then be expressed as

$$\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{g}(\boldsymbol{y}, \boldsymbol{x}),$$

and linearity in the second variable $\mathbf{y}$ as

$$\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y} + \boldsymbol{w}) = \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) + \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{w}), \quad \boldsymbol{g}(\boldsymbol{x}, \lambda\boldsymbol{y}) = \lambda\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}).$$

For *bilinearity*, we also require linearity in the *first* variable $\boldsymbol{x}$, but this now follows from the symmetry.

A *Hermitian form* $\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{y})$ satisfies, instead, Hermitian symmetry

$$\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{y}) = \overline{\boldsymbol{h}(\boldsymbol{y}, \boldsymbol{x})},$$

together with linearity in the second variable $\boldsymbol{y}$:

$$\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{y} + \boldsymbol{w}) = \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{y}) + \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{w}), \quad \boldsymbol{h}(\boldsymbol{x}, \lambda\boldsymbol{y}) = \lambda\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{y}).$$

The Hermitian symmetry now implies what is called *antilinearity* in the first variable:

$$\boldsymbol{h}(\boldsymbol{x} + \boldsymbol{w}, \boldsymbol{y}) = \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{y}) + \boldsymbol{h}(\boldsymbol{w}, \boldsymbol{y}), \quad \boldsymbol{h}(\lambda\boldsymbol{x}, \boldsymbol{y}) = \overline{\lambda}\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{y}).$$

Whereas an orthogonal group preserves a (non-singular) symmetric bilinear form, the complex linear transformations preserving a non-singular Hermitian form give us a unitary group.

What do such forms do for us? A (not necessarily symmetric) non-singular bilinear form $\boldsymbol{g}$ provides us with a means of identifying the vector space $\mathbf{V}$, to which $\boldsymbol{x}$ and $\boldsymbol{y}$ belong, with the dual space $\mathbf{V}^*$. Thus, if $\boldsymbol{v}$ belongs to $\mathbf{V}$, then $\boldsymbol{g}(\boldsymbol{v}, \ )$ provides us with a linear map on $\mathbf{V}$, mapping the element $\boldsymbol{x}$ of $\mathbf{V}$ to the number $\boldsymbol{g}(\boldsymbol{v}, \boldsymbol{x})$. In other words, $\boldsymbol{g}(\boldsymbol{v}, \ )$ is an element of $\mathbf{V}^*$ (see §12.3). In index form, this element of $\mathbf{V}^*$ is the covector $v^a g_{ab}$, which is customarily written with the same kernel letter $\boldsymbol{v}$, but with the index lowered (see also §14.7) by g_{ab}, according to

$$v_b = v^a g_{ab}.$$

The inverse of this operation is achieved by the raising of the index of v_a by use of the inverse metric $\left[{}^2_0\right]$-tensor g^{ab}:

$$v^a = g^{ab} v_b.$$

We shall need the analogue of this in the Hermitian case. As before, each choice of element $\boldsymbol{v}$ from the vector space $\mathbf{V}$ provides us with an element $\boldsymbol{h}(\boldsymbol{v}, \ )$ of the dual space $\mathbf{V}^*$. However, the difference is that now $\boldsymbol{h}(\boldsymbol{v}, \ )$ depends antilinearly on $\boldsymbol{v}$ rather than linearly; thus $\boldsymbol{h}(\lambda\boldsymbol{v}, \ ) = \bar{\lambda}\boldsymbol{h}(\boldsymbol{v}, \ )$.

An equivalent way of saying this is that $\boldsymbol{h}(\boldsymbol{v}, \ )$ is *linear* in $\bar{\boldsymbol{v}}$, this vector quantity $\bar{\boldsymbol{v}}$ being the 'complex conjugate' of $\boldsymbol{v}$. We consider these complex-conjugate vectors to constitute a separate vector space $\bar{\boldsymbol{v}}$. This viewpoint is particularly useful for the (abstract) index notation, where a separate 'alphabet' of indices is used, say a', b', c', ..., for these complex-conjugate elements, where contractions (summations) are not permitted between primed and unprimed indices. The operation of complex conjugation interchanges the primed with the unprimed indices. In the index notation, our Hermitian form is represented as an array of quantities $h_{a'b}$ with one (lower) index of each type, so

$$\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{y}) = h_{a'b}\bar{x}^{a'}y^{b}$$

(with $\bar{x}^{a'}$ being the complex conjugate of the element x^{a}), where 'Hermiticity' is expressed as

$$h_{a'b} = \overline{h_{b'a}}$$

The array of quantities $h_{a'b}$ allows us to lower or raise an index, but it now changes primed indices to unprimed ones, and vice versa, so it refers us to the dual of the complex-conjugate space:

$$\bar{v}_{a} = \bar{v}^{a'}h_{a'b}, \quad v_{a'} = h_{a'b}v^{b}.$$

For the inverses of these operations—where the Hermitian form is assumed non-singular (i.e. the matrix of components $h^{ab'}$ is non-singular)—we need the inverse $h^{ab'}$ of $h_{a'b}$

$$h^{ab'}h_{b'c} = \delta^{a}_{c}, \quad h_{a'b}h^{bc'} = \delta^{c'}_{a'},$$

whence[13.53]

$$\bar{v}^{a'} = \bar{v}_{b}h^{ba'}, \quad v^{a} = h^{ab'}v_{b'}.$$

Note that all primed indices can be eliminated using $h_{a'b}$ (and the corresponding inverse $h^{ab'}$) by virtue of the above relations, which can be applied index-by-index to any tensor quantity. The complex-conjugate space is thereby 'identified' with the dual space, instead of having to be a quite separate space.

The operation of 'complex conjugation'—usually called *Hermitian conjugation*—which incorporates this identification with the dual into the notion of complex conjugation (though not commonly written in the index notation) is of central importance to quantum mechanics, as well as to many other areas of mathematics and physics (such as twistor theory, see §33.5). In the quantum-mechanical literature this is often denoted by a *dagger* '†', but sometimes by an *asterisk* '$*$'.

I prefer the asterisk, which is more usual in the mathematical literature, so I shall use this here—in bold type. The asterisk is appropriate here because it interchanges the roles of the vector space $\mathbf{V}$ and its dual $\mathbf{V}^{*}$. A complex tensor of valence $\left[\begin{smallmatrix}p\\q\end{smallmatrix}\right]$ (all primed indices having been eliminated, as above) is mapped by $*$ to a tensor of valence $\left[\begin{smallmatrix}q\\p\end{smallmatrix}\right]$. Thus, upper indices become lower and lower indices become upper under the action of $*$. As applied to scalars, $*$ is simply the ordinary operation of complex conjugation. The operation $*$ is an equivalent notion to the Hermitian form $\boldsymbol{h}$ itself.

The most familiar Hermitian conjugation operation (which occurs when the components $h_{a'b}$ are taken to be the Kronecker delta) simply

[13.53] Verify these relations, explaining the notational consistency of $h^{ab'}$.

takes the complex conjugate of each component, reorganizing the components so as to read upper indices as lower ones and lower indices as upper ones. Accordingly, the matrix of components of a linear transformation is taken to the transpose of its complex conjugate (sometimes called the *conjugate transpose* of the matrix), so in the 2×2 case we have

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^* = \begin{pmatrix} \bar{a} & \bar{c} \\ \bar{b} & \bar{d} \end{pmatrix}.$$

A *Hermitian matrix* is a matrix that is equal to its Hermitian conjugate in this sense. This concept, and the more general abstract *Hermitian operator*, are of great importance in quantum theory.

We note that $*$ is *antilinear* in the sense

$$(\boldsymbol{T} + \boldsymbol{U})^* = \boldsymbol{T}^* + \boldsymbol{U}^*,$$
$$(z\boldsymbol{T})^* = \bar{z}\boldsymbol{T}^*,$$

applied to tensors $\boldsymbol{T}$ and $\boldsymbol{U}$, both of the same valence, and for any complex number z. The action of $*$ must also preserve products of tensors but, because of the reversal of the index positions, it reverses the order of contractions; in particular, when $*$ is applied to linear transformations (regarded as tensors with one upper and one lower index), the order of multiplication is reversed:

$$(\boldsymbol{LM})^* = \boldsymbol{M}^*\boldsymbol{L}^*.$$

It is very handy, in the diagrammatic notation, to depict such a conjugation operation as reflection in a horizontal plane. This interchanges upper and lower indices, as required; see Fig. 13.19.

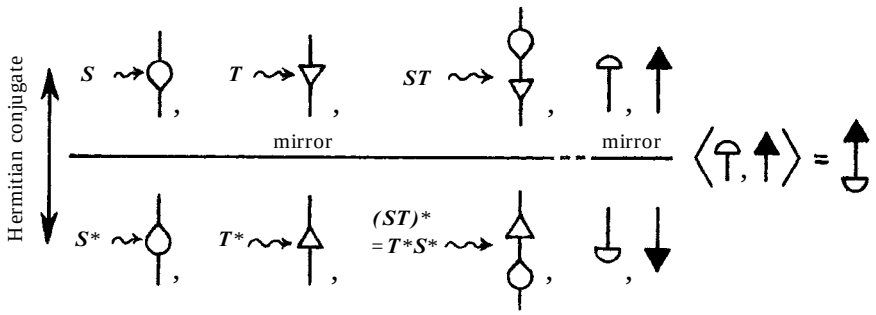

Fig. 13.19 The operation of Hermitian conjugation (*) conveniently depicted as reflection in a horizontal plane. This interchanges 'arms' with 'legs' and reverses the order of multiplication: $(\boldsymbol{ST})^* = \boldsymbol{T}^*\boldsymbol{S}^*$. The diagrammatic expression for the Hermitian scalar product $\langle \boldsymbol{v}|\boldsymbol{w}\rangle = \boldsymbol{v}^*\boldsymbol{w}$ is given (so that taking its complex conjugate would reflect the diagram on the far right upside-down).

The operation $*$ enables us to define a *Hermitian scalar product* between two elements $\boldsymbol{v}$ and $\boldsymbol{w}$, of $\mathbf{V}$, namely the scalar product of the covector $\boldsymbol{v}^*$ with the vector $\boldsymbol{w}$ (the different notations being useful in different contexts):

$$\langle \boldsymbol{v} \mid \boldsymbol{w} \rangle = \boldsymbol{v}^* \bullet \boldsymbol{w} = \boldsymbol{h}(\boldsymbol{v}, \boldsymbol{w})$$

(and see Fig. 13.19), and we have

$$\langle \boldsymbol{v} \mid \boldsymbol{w} \rangle = \overline{\langle \boldsymbol{w} \mid \boldsymbol{v} \rangle}.$$

In the particular case $\boldsymbol{w} = \boldsymbol{v}$, we get the *norm* of $\boldsymbol{v}$, with respect to $*$:

$$\| \boldsymbol{v} \| = \langle \boldsymbol{v} \mid \boldsymbol{v} \rangle.$$

We can choose a *basis* $(\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n)$ for $\mathbf{V}$, and then the components $h_{a'b}$ in this basis are simply the n^2 complex numbers

$$h_{a'b} = \boldsymbol{h}(\boldsymbol{e}_a, \boldsymbol{e}_b) = \langle \boldsymbol{e}_a \mid \boldsymbol{e}_b \rangle,$$

constituting the elements of a Hermitian matrix. The basis $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$ is called *pseudo-orthonormal*, with respect to $*$, if

$$\langle \boldsymbol{e}_i \mid \boldsymbol{e}_j \rangle = \begin{cases} \pm 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases};$$

in the case when all the $\pm$ signs are $+$, i.e. when each ± 1 is just 1, the basis is *orthonormal*.

A pseudo-orthonormal basis can always be found, but there are many choices. With respect to any such basis, the matrix $h_{a'b}$ is diagonal, with just 1s and -1s down the diagonal. The total number of 1s, p, always comes out the same, for a given $*$, independently of any particular choice of basis, and so also does the total number of -1s, q. This enables us to define the invariant notion of *signature* (p, q) for the operation $*$.

If $q = 0$, we say that $*$ is *positive-definite*. In this case,[21] the norm of any non-zero vector is always positive:[13.54]

$$\boldsymbol{v} \neq 0 \quad \textit{implies} \quad \| \boldsymbol{v} \| > 0.$$

Note that this notion of 'positive-definite' generalizes that of §13.8 to the complex case.

A linear transformation $\boldsymbol{T}$ whose inverse is $\boldsymbol{T}^*$, so that

$$\boldsymbol{T}^{-1} = \boldsymbol{T}^*, \text{ i.e. } \boldsymbol{T}\ \boldsymbol{T}^* = \boldsymbol{I} = \boldsymbol{T}^*\boldsymbol{T},$$

[13.54] Show this.

is called *unitary* in the case when $*$ is positive-definite, and *pseudo-unitary* in the other cases.[13.55] The term 'unitary matrix' refers to a matrix $\boldsymbol{T}$ satisfying the above relation when $*$ stands for the usual conjugate transpose operation, so that $\boldsymbol{T}^{-1} = \overline{\boldsymbol{T}}$.

The group of unitary transformations in n dimensions, or of $(n \times n)$ unitary matrices, is called the *unitary group* $\mathrm{U}(n)$. More generally, we get the pseudo-unitary group $\mathrm{U}(p, q)$ when $*$ has signature (p, q).[22] If the transformations have unit determinant, then we correspondingly obtain $\mathrm{SU}(n)$ and $\mathrm{SU}(p, q)$. Unitary transformations play an essential role in quantum mechanics (and they have great value also in many pure-mathematical contexts).

13.10 Symplectic groups

In the previous two sections, we encountered the orthogonal and unitary groups. These are examples of what are called *classical groups*, namely the simple Lie groups other than the exceptional ones; see §13.2. The list of classical groups is completed by the family of *symplectic* groups. Symplectic groups have great importance in *classical* physics, as we shall be seeing particularly in §20.4—and also in quantum physics, particularly in the infinite-dimensional case (§26.3).

What is a symplectic group? Let us return again to the notion of a bilinear form, but where instead of the symmetry $(\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{g}(\boldsymbol{y}, \boldsymbol{x}))$ required for defining the orthogonal group, we impose *antisymmetry*

$$\boldsymbol{s}(\boldsymbol{x}, \boldsymbol{y}) = -\boldsymbol{s}(\boldsymbol{y}, \boldsymbol{x}),$$

together with linearity

$$\boldsymbol{s}(\boldsymbol{x}, \boldsymbol{y} + \boldsymbol{w}) = \boldsymbol{s}(\boldsymbol{x}, \boldsymbol{y}) + \boldsymbol{s}(\boldsymbol{x}, \boldsymbol{w}), \quad \boldsymbol{s}(\boldsymbol{x}, \lambda\boldsymbol{y}) = \lambda\boldsymbol{s}(\boldsymbol{x}, \boldsymbol{y}),$$

where linearity in the first variable $\boldsymbol{x}$ now follows from the *anti*symmetry. We can write our antisymmetric form variously as

$$\boldsymbol{s}(\boldsymbol{x}, \boldsymbol{y}) = x^a s_{ab} y^b = \mathbf{x}^{\mathsf{T}}\mathbf{S}\mathbf{y},$$

just as in the symmetric case, but where s_{ab} is *antisymmetric*:

$$s_{ba} = -s_{ab} \quad \text{i.e.} \quad \mathbf{S}^{\mathsf{T}} = -\mathbf{S},$$

$\mathbf{S}$ being the matrix of components of s_{ab}. We require $\boldsymbol{S}$ to be non-singular. Then s_{ab} has an inverse s^{ab}, satisfying[23]

[13.55] Show that these transformations are precisely those which preserve the Hermitian correspondence between vectors $\boldsymbol{v}$ and covectors $\boldsymbol{v}^*$, and that they are those which preserve $h_{ab'}$.

$$s_{ab}s^{bc} = \delta_a^c = s^{cb}s_{ba},$$

where $s^{ab} = -s^{ba}$.

We note that, by analogy with a symmetric matrix, an antisymmetric matrix $\mathbf{S}$ equals *minus* its transpose. It is important to observe that an $n \times n$ antisymmetric matrix $\mathbf{S}$ can be non-singular only if n is even.[13.56] Here n is the dimension of the space $\mathbf{V}$ to which $\mathbf{x}$ and $\mathbf{y}$ belong, and we indeed take n to be even.

The elements $\mathbf{T}$ of GL(n) that preserve such a non-singular antisymmetric s_{ab} (or, equivalently, the bilinear form $\mathbf{s}$), in the sense that

$$s_{ab}T^a_c T^b_d = s_{cd}, \quad \text{i.e. } \mathbf{T}^\mathsf{T}\mathbf{S}\,\mathbf{T} = \mathbf{S},$$

are called *symplectic*, and the group of these elements is called a *symplectic group* (a group of very considerable importance in classical mechanics, as we shall be seeing in §20.4). However, there is some confusion in the literature concerning this terminology. It is mathematically more accurate to define a (real) symplectic group as a real form of the *complex* symplectic group $\mathrm{Sp}(\frac{1}{2}n, \mathbb{C})$, which is the group of *complex* $T^a{}_b$ (or $\mathbf{T}$) satisfying the above relation. The particular real form just defined is non-compact; but in accordance with the remarks at the end of §13.7—$\mathrm{Sp}(\frac{1}{2}n, \mathbb{C})$ being semi-simple—there is another real form of this complex group which *is* compact, and it is this that is normally referred to as the (real) symplectic group $\mathrm{Sp}(\frac{1}{2}n)$.

How do we find these different real forms? In fact, as with the orthogonal groups, there is a notion of *signature* which is not so well known as in the cases of the orthogonal and unitary groups. The symplectic group of real transformations preserving s_{ab} would be the 'split-signature' case of signature $(\frac{1}{2}n, \frac{1}{2}n)$. In the compact case, the symplectic group has signature $(n, 0)$ or $(0, n)$.

How is this signature defined? For each pair of natural numbers p and q such that $p + q = n$, we can define a corresponding 'real form' of the complex group $\mathrm{Sp}(\frac{1}{2}n, \mathbb{C})$ by taking only those elements which are also pseudo-unitary for signature (p, q)—i.e. which belong to $\mathrm{U}(p, q)$ (see §13.9). This gives[24] us the (pseudo-)symplectic group $\mathrm{Sp}(p, q)$. (Another way of saying this is to say that $\mathrm{Sp}(p, q)$ is the intersection of $\mathrm{Sp}(\frac{1}{2}n, \mathbb{C})$ with $\mathrm{U}(p, q)$.) In terms of the index notation, we can define $\mathrm{Sp}(p, q)$ to be the group of complex linear transformations T^a_b that preserve both the antisymmetric s_{ab}, as above, and also a Hermitian matrix $\mathbf{H}$ of components $h_{a'b}$, in the sense that

$$\bar{T}^{a'}_{b'} T^a_b h_{a'a} = h_{b'b},$$

[13.56] Prove this.

where **H** has signature (p, q) (so we can find a pseudo-orthonormal basis for which **H** is diagonal with p entries 1 and q entries -1; see §13.9).[25] The compact *classical* symplectic group $\mathrm{Sp}(\frac{1}{2}n)$ is my $\mathrm{Sp}(n, 0)$ (or $\mathrm{Sp}(0, n)$), but the form of most importance in classical physics is $\mathrm{Sp}(\frac{1}{2}n, \frac{1}{2}n)$.[13.57]

As with the orthogonal and unitary groups, we can find choices of basis for which the components s_{ab} have a particularly simple form. We cannot now take this form to be diagonal, however, because the only antisymmetric diagonal matrix is zero! Instead, we can take the matrix of s_{ab} to consist of 2×2 blocks down the main diagonal, of the form

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

In the familiar split-signature case $\mathrm{Sp}(\frac{1}{2}n, \frac{1}{2}n)$, we can take the *real* linear transformations preserving this form. The general case $\mathrm{Sp}(p, q)$ is exhibited by taking, rather than real transformations, pseudo-unitary ones of signature (p, q).[13.58]

For various (small) values of p and q, some of the orthogonal, unitary, and symplectic groups are the same ('isomorphic') or at least locally the same ('locally isomorphic'), in the sense of having the same Lie algebras (cf. §13.6).[26] The most elementary example is the group SO(2), which describes the group of non-reflective symmetries of a circle, being the same as the unitary group U(1), the multiplicative group of unit-modulus complex numbers $e^{i\theta}$ (θ real).[13.59] Of a particular importance for physics is the fact that SU(2) and Sp(1) are the same, and are locally the same as SO(3) (being the twofold cover of this last group, in accordance with the twofold nature of the quaternionic representation of rotations in 3-space, as described in §11.3). This has great importance for the quantum physics of *spin* (§22.8). Of significance in relativity theory is the fact that $\mathrm{SL}(2, \mathbb{C})$, being the same as $\mathrm{Sp}(1, \mathbb{C})$, is locally the same as the non-reflective part of the Lorentz group O(1, 3) (again a twofold cover of it). We also find that SU(1, 1), Sp(1, 1), and SO(2, 1) are the same, and there are several other examples. Particularly noteworthy for twistor theory is the local identity between SU(2, 2) and the non-reflective part of the group O(2, 4) (see §33.3).

The Lie algebra of a symplectic group is obtained by looking for solutions **X** of the matrix equation

$$\mathbf{X}^{\mathsf{T}}\mathbf{S} + \mathbf{S}\,\mathbf{X} = 0, \quad \text{i.e. } \mathbf{S}\,\mathbf{X} = (\mathbf{S}\,\mathbf{X})^{\mathsf{T}},$$

[13.57] Find explicit descriptions of Sp(1) and Sp(1, 1) using this prescription. Can you see why the groups $\mathrm{Sp}(n, 0)$ are compact?

[13.58] Show why these two different descriptions for the case $p = q = \frac{1}{2}n$ are equivalent.

[13.59] Why are they the same?

so the infinitesimal transformation (Lie algebra element) $\mathbf{X}$ is simply $\mathbf{S}^{-1}$ times a symmetric $n \times n$ matrix. This enables the dimensionality $\frac{1}{2}n(n+1)$ of the symplectic group to be directly seen. Note that $\mathbf{X}$ has to be trace-free (i.e. trace $\mathbf{X} = 0$—see §13.4).[13.60] The Lie algebras for orthogonal and unitary groups are also readily obtained, in terms, respectively, of anti-symmetric matrices and pure-imaginary multiples of Hermitian matrices, the respective dimensions being $n(n-1)/2$ and n^2.[13.61]

We note from §13.4 that, for the transformations to have unit determinant, the trace of the infinitesimal element $\boldsymbol{X}$ must vanish. This is automatic in the symplectic case (noted above), and in the orthogonal case the infinitesimal elements all have unit determinant.[13.62] In the unitary case, restriction to SU(n) is one further condition (trace $\boldsymbol{X} = 0$), so the dimension of the group is reduced to $n^2 - 1$.

The *classical groups* referred to in §13.2, sometimes labelled A_m, B_m, C_m, D_m (for $m = 1, 2, 3, \ldots$), are simply the respective groups SU($m+1$), SO($2m+1$), Sp(m), and SO($2m$), that we have been examining in §§13.8–10, and we see from the above that they indeed have respective dimensionalities $m(m+2)$, $(2m+1)$, $m(2m+1)$, and $m(2m-1)$, as asserted in §13.2. Thus, the reader has now had the opportunity to catch a significant glimpse of all the classical simple groups. As we have seen, such groups, and some of the various other 'real forms' (of their complexifications) play important roles in physics. We shall be gaining a little acquaintance with this in the next chapter. As mentioned at the beginning of this chapter, according to modern physics, all physical interactions are governed by 'gauge connections' which, technically, depend crucially on spaces having exact symmetries. However, we still need to know what a 'gauge theory' actually is. This will be revealed in Chapter 15.

Notes

Section 13.1

13.1. Abel was born in 1802 and died of consumption (tuberculosis) in 1829, aged 26. The more general non-Abelian ($ab \neq ba$) group theory was introduced by the even more tragically short-lived French mathematician Evariste Galois (1811–1832), who was killed in a duel before he reached 21, having been up the entire previous night feverishly writing down his revolutionary ideas involving the use of these groups to investigate the solubility of algebraic equations, now called *Galois theory*.

[13.60] Explain where the equation $\boldsymbol{X}^{\mathsf{T}}\boldsymbol{S} + \boldsymbol{S}\boldsymbol{X} = 0$ comes from and why $\boldsymbol{S}\boldsymbol{X} = (\boldsymbol{S}\boldsymbol{X})^{\mathsf{T}}$. Why does trace $\boldsymbol{X}$ vanish? Give the Lie algebra explicitly. Why is it of this dimension?

[13.61] Describe these Lie algebras and obtain these dimensions.

[13.62] Why, and what does this mean geometrically?

13.2. We should also take note that '–C' means 'take the complex conjugate, then multiply by -1', i.e. $-\mathsf{C} = (-1)\mathsf{C}$.

13.3. The S stands for 'special' (meaning 'of unit determinant') which, in the present context just tells us that orientation-reversing motions are excluded. The O stands for 'orthogonal' which has to do with the fact that the motions that it represents preserves the 'orthogonality' (i.e. the right-angled nature) of coordinate axes. The 3 stands for the fact that we are considering rotations in three dimensions.

13.4. There is a remarkable theorem that tells us that not only is every continuous group also *smooth* (i.e. C^0 implies C^1, in the notation of §§6.3,6, and even C^0 implies C^∞), but it is also *analytic* (i.e. C^0 implies C^ω). This famous result, which represented the solution of what had become known as 'Hilbert's 5th problem', was obtained by Andrew Mattei Gleason, Deane Montgomery, Leo Zippin, and Hidehiko Yamabe in 1953; see Montgomery and Zippin (1955). This justifies the use of power series in §13.6.

Section 13.2

13.5. See van der Waerden (1985), pp. 166–74.

13.6. See Devlin (1988).

13.7. See Conway and Norton (1972); Dolan (1996).

Section 13.3

13.8. We shall be seeing in §14.1 that a Euclidean space is an example of an *affine* space. If we select a particular point (origin) O, it becomes a vector space.

13.9. In many places in this book it will be convenient—and sometimes essential—to stagger the indices on a tensor-type symbol. In the case of a linear transformation, we need this to express the order of matrix multiplication.

13.10. This region is a vector space of dimension r (where $r < n$). We call r the *rank* of the matrix or linear transformation $\boldsymbol{T}$. A *non*-singular $n \times n$ matrix has rank n. (The concept of 'rank' applies also to rectangular matrices.) Compare Note 12.18.

13.11. For a history of the theory of matrices, see MacDuffee (1933).

Section 13.5

13.12. In those degenerate situations where the eigenvectors do not span the whole space (i.e., some d is less than the corresponding r), we can still find a canonical form, but we now allow 1s to appear just above the main diagonal, these residing just within square blocks whose diagonal terms are *equal* eigenvalues (*Jordan normal form*); see Anton and Busby (2003). Apparently Weierstrauss had (effectively) found this normal form in 1868, two years before Jordan; See Hawkins (1977).

Section 13.6

13.13. To illustrate this point, consider $\mathrm{SL}(n, \mathbb{R})$ (i.e. the unit-determinant elements of $\mathrm{GL}(n, \mathbb{R})$ itself). This group has a 'double cover' $\widetilde{\mathrm{SL}}(n, \mathbb{R})$ (provided that $n \geq 3$) which is obtained from $\mathrm{SL}(n, \mathbb{R})$ in basically the same way whereby we effectively found the double cover $\widetilde{\mathrm{SO}}(3)$ of $\mathrm{SO}(3)$ when we considered the rotations of a book, with belt attachment, in §11.3. Thus, $\widetilde{\mathrm{SO}}(3)$ is the group of (non-reflective) rotations of a *spinorial object* in ordinary 3-space. In the same way, we can consider 'spinorial objects' that are subject to the more general linear transformations that allow 'squashing' or 'stretching', as discussed in §13.3. In this way, we arrive at the group $\widetilde{\mathrm{SL}}(n, \mathbb{R})$, which is *locally* the same as $\mathrm{SL}(n, \mathbb{R})$, but which cannot, in fact, be faithfully represented in any $\mathrm{GL}(m)$. See Note 15.9.

13.14. This notion is well defined; cf. Note 13.4.

Section 13.7

13.15. See Thirring (1983).

13.16. Here, again, we have an instance of the capriciousness of the naming of mathematical concepts. Whereas many notions of great importance in this subject, to which Cartan's name is conventionally attached (e.g. 'Cartan subalgebra, Cartan integer') were originally due to Killing (see §13.2), what we refer to as the 'Killing form' is actually due to Cartan (and Hermann Weyl); see Hawkins (2000), §6.2. However, the 'Killing vector' that we shall encounter in §30.6 *is* actually due to Killing (Hawkins 2000, note 20 on p. 128).

13.17. I am (deliberately) being mathematically a little sloppy in my use of the phrase 'the same' in this kind of context. The strict mathematical term is 'isomorphic'.

Section 13.8

13.18. I have not been very explicit about this procedure up to this point. A basis $\boldsymbol{e} = (\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$ for $\mathbf{V}$ is associated with a *dual basis*—which is a basis $\boldsymbol{e}^* = (\boldsymbol{e}^1, \ldots, \boldsymbol{e}^n)$ for $\mathbf{V}^*$—with the property that $\boldsymbol{e}^i \bullet \boldsymbol{e}_j = \delta^i_j$. The components of a $\left[\begin{smallmatrix}p\\q\end{smallmatrix}\right]$-valent tensor $\boldsymbol{Q}$ are obtained by applying the multilinear function of §12.8 to the various collections of p dual basis elements and q basis elements: $Q^{f\ldots h}_{a\ldots c} = \boldsymbol{Q}(\boldsymbol{e}^f, \ldots, \boldsymbol{e}^h; \boldsymbol{e}_a, \ldots, \boldsymbol{e}_c)$.

13.19. See Note 13.3.

13.20. See Note 13.10. The reader may be puzzled about why the $T^a{}_b$ of §13.5 can have lots of invariants, namely all its *eigenvalues* $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_n$, whereas g_{ab} does not. The answer lies simply in the difference in transformation behaviour implicit in the different index positioning.

Section 13.9

13.21. Note that, in the positive-definite case, $(\boldsymbol{e}^*_1, \boldsymbol{e}^*_2, \ldots, \boldsymbol{e}^*_n)$ is a *dual basis* to $(\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n)$, in the sense of Note 13.18.

13.22. The groups $\mathrm{U}(p, q)$, for fixed $p + q = n$, as well as $\mathrm{GL}(n, \mathbb{R})$, all have the same complexification, namely $\mathrm{GL}(n, \mathbb{C})$, and these can all be regarded as different real forms of this complex group.

Section 13.10

13.23. We can then use s_{ab} and s^{ab} to raise and lower indices of tensors, just as with g_{ab} and g^{ab}, so $v_a = s_{ab}v^b$ $v^a = s^{ab}v_b$ (see §13.8); but, because of the antisymmetry, we must be a little careful to make the ordering of the indices consistent. Those readers who are familiar with the 2-spinor calculus (see Penrose and Rindler 1984, vol.1) may notice a slight notational discrepancy between our s_{ab} and the ε_{AB} of that calculus.

13.24. I am not aware of a standard terminology or notation for these various real forms, so the notation $\mathrm{Sp}(p, q)$ has been concocted for the present purposes.

13.25. In fact, every element of $\mathrm{Sp}(\frac{1}{2}n, \mathbb{C})$ has unit determinant, so we do not need an '$\mathrm{SSp}(\frac{1}{2}n)$' by analogy with $\mathrm{SO}(n)$ and $\mathrm{SU}(n)$. The reason is that there is an expression (the 'Pfaffian') for Levi-Civita's $\varepsilon\ldots$ in terms of the s_{ab}, which must be preserved whenever the s_{ab} are.

13.26. See Note 13.17.