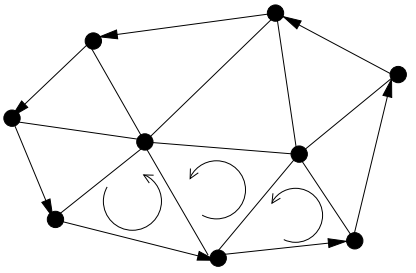

Figure 6.2 *A triangulated surface.* The surface is represented by a series of points, and each set of three adjacent points defines a triangle, or simplex. As more points are added the simplices become a closer fit to the true surface. Each simplex is given the same orientation by ensuring that for adjacent simplices, the common edge in traversed in opposite directions.

the directed measure dX is still defined everywhere. A directed surface integral then takes the form

$$\int F\,dX = \int Fe_1 \wedge e_2\,dx^1\,dx^2, \tag{6.101}$$

or a sum of such terms if more than one coordinate patch is required. Again, we form the geometric product between the integrand and the measure. As in the case of a line integral, this is not the most general surface integral that can be considered, as the integrand can multiply the measure from the left or the right, giving rise to different integrals.

As an example of a surface integral, consider a closed surface in three dimensions, with unit outward normal $\boldsymbol{n}$. We let F be given by the bivector-valued function $\phi \boldsymbol{n} I^{-1}$, where ϕ is a scalar field. The surface integral is then

$$\oint \phi \boldsymbol{n} I^{-1}\,dX = \oint \phi |dS|. \tag{6.102}$$

Here $|dS| = I^{-1}\boldsymbol{n}\,dX$ is the scalar-valued measure over the surface. The directed measure is usually chosen so that $\boldsymbol{n}\,dX$ has the same orientation as I. As a second example, suppose that $F = 1$. In this case we can show that

$$\oint dX = 0, \tag{6.103}$$

which holds for any closed surface (see later). If the surface is open, the result of the directed surface integral depends entirely on the boundary, since all the internal simplices cancel out. This result is sometimes called the vector area, though in geometric algebra the result is a bivector.

In order to construct proofs of some of the more important results it is necessary to express the surface integral (6.101) in terms of a limit of a sum. This involves the idea of a triangulated surface (figure 6.2). A set of points are chosen

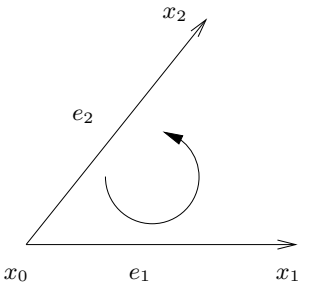

Figure 6.3 *A planar simplex.* The points x_0, x_1, x_2 define a triangle. The order specifies how the boundary is traversed, which defines an orientation for the simplex.

on the surface, and adjacent sets of three points define a series of planar triangles, or *simplices*. As more points are added these triangles become smaller and are an ever better model for the surface. (In computer graphics programs this is precisely how 'smooth' surfaces are represented internally.) Each simplex has an orientation attached such that, for a pair of adjacent simplices, the common edge is traversed in opposite directions. In this way an initial simplex builds up to define an orientation for the entire surface. For some surfaces, such as the Mobius strip, it is not possible to define a consistent orientation over the entire surface. For these it is not possible to define a directed integral, so our presentation is restricted to *orientable* surfaces.

Suppose now that the three points x_0, x_1, x_2 define the corners of a simplex, with orientation specified by traversing the edges in the order $x_0 \mapsto x_1 \mapsto x_2$ (see figure 6.3). We define the vectors

$$e_1 = x_1 - x_0, \qquad e_2 = x_2 - x_0. \tag{6.104}$$

The surface measure is then defined by

$$\Delta X = \tfrac{1}{2} e_1 \wedge e_2 = \tfrac{1}{2}(x_1 \wedge x_2 + x_2 \wedge x_0 + x_0 \wedge x_1). \tag{6.105}$$

ΔX has the orientation defined by the boundary, and an area equal to that of the simplex. The final expression makes it clear that ΔX is invariant under even permutations of the vertices. With this definition of ΔX we can express the surface integral (6.101) as the limit:

$$\int F \, dX = \lim_{n \mapsto \infty} \sum_{k=1}^{n} \bar{F}^k \Delta X^k. \tag{6.106}$$

The sum here runs over all simplices making up the surface, and for each simplex $\bar{F}$ is the average value of F over the simplex. For well-behaved integrals the value in the limit is independent of the precise nature of the limiting process.

6.4.3 n-dimensional surfaces

The simplex structure introduced in the previous section provides a means of defining a directed integral for any dimension of surface. We discretise the surface by considering a series of points, and adjacent sets of points are combined to define a simplex. Suppose that we have an n-dimensional surface, and that one simplex for the discretised surface has vertices $x_0, \ldots, x_n$, with the order specifying the desired orientation. For this simplex we define vectors

$$e_i = x_i - x_0, \quad i = 1, \ldots, n, \tag{6.107}$$

and the directed volume element is

$$\Delta X = \frac{1}{n!} e_1 \wedge \cdots \wedge e_n. \tag{6.108}$$

A point in the simplex can be described in terms of coordinates $\lambda^1, \ldots, \lambda^n$ by writing

$$x = x_0 + \sum_{i=1}^{n} \lambda^i e_i. \tag{6.109}$$

Each coordinate lies in the range $0 \leq \lambda^i \leq 1$, and the coordinates also satisfy

$$\sum_{i=1}^{n} \lambda^i \leq 1. \tag{6.110}$$

Now suppose we have a multivector field $F(x)$ defined over the surface. We denote the value at each vertex by $F_i = F(x_i)$. A new function $f(x)$ is then introduced which linearly interpolates the F_i over the simplex. This can be written

$$f(x) = F_0 + \sum_{i=1}^{n} \lambda^i (F_i - F_0). \tag{6.111}$$

As the number of points increases and the simplices grow smaller, $f(x)$ becomes an ever better approximation to $F(x)$, and the triangulated surface approaches the true surface.

The directed integral of F over the surface is now approximated by the integral of f over each simplex in the surface. To evaluate the integral over each simplex we use the λ^i as coordinates, so that

$$dX = e_1 \wedge \cdots \wedge e_n \, d\lambda^1 \, \cdots d\lambda^n. \tag{6.112}$$

It is then a straightforward exercise in integration to establish that

$$\int dX = \Delta X \tag{6.113}$$

and

$$\int \lambda^i \, dX = \frac{1}{n+1} \Delta X, \quad \forall \lambda^i. \tag{6.114}$$

Combining these two results we find that the integral of $f(x)$ over a single simplex evaluates to

$$\int f \, dX = \frac{1}{n+1} \left(\sum_{i=0}^{n} F_i \right) \Delta X. \tag{6.115}$$

The function is therefore replaced by its average value over the simplex. We write this as $\bar{F}$. Summing over all the simplices making up the surface we can now define

$$\int F \, dX = \lim_{n \mapsto \infty} \sum_{k=1}^{n} \bar{F}^k \, \Delta X^k, \tag{6.116}$$

where k runs over all of the simplices in the surface. More generally, suppose that $\mathsf{L}(A_n)$ is a position-dependent linear function of a grade-n multivector A_n. We can then write

$$\int \mathsf{L}(dX) = \lim_{n \mapsto \infty} \sum_{k=1}^{n} \bar{\mathsf{L}}^k(\Delta X^k), \tag{6.117}$$

with $\bar{\mathsf{L}}^k(\Delta X^k)$ the average value of $\mathsf{L}(\Delta X^k)$ over the vertices of each simplex.

6.4.4 The fundamental theorem of geometric calculus

Most physicists are familiar with a number of integral theorems, including the divergence and Stokes' theorems, and the Cauchy integral formula of complex analysis. We will now show that these are all special cases of a more general theorem in geometric calculus. In this section we will sketch of proof of this important theorem. Readers who are not interested in the details of the proof may want to jump straight to the following section, where some applications are discussed. The proof given here uses simplices and triangulated surfaces, which means that it is relevant to methods of discretising integrals for numerical computation.

We start by introducing a notation for simplices which helps clarify the nature of the boundary operator. We let $(x_0, x_1, \ldots, x_k)$ denote the k-simplex defined by the $k+1$ points $x_0, \ldots, x_k$. This is abbreviated to

$$(x)_{(k)} = (x_0, x_1, \ldots, x_k). \tag{6.118}$$

The order of points is important, as it specifies the orientation of the simplex. If any two adjacent points are swapped then the simplex changes sign. The

boundary operator for a simplex is denoted by ∂ and is defined by

$$\partial(x)_{(k)} = \sum_{i=0}^{k} (-1)^i (x_0, \ldots, \check{x}_i, \ldots, x_k)_{(k-1)}, \tag{6.119}$$

where the check denotes that the term is missing from the product. So, for example,

$$\partial(x_0, x_1) = (x_1) - (x_0), \tag{6.120}$$

which returns the two points at the end of a line segment. The boundary of a boundary vanishes,

$$\partial\partial(x)_{(k)} = 0. \tag{6.121}$$

Proofs of this can be found in most differential geometry textbooks.

So far we have dealt only with ordered lists of points, not geometric sums or products. To add some geometry we introduce the operator Δ which returns the directed content of a simplex,

$$\Delta(x)_{(k)} = \frac{1}{k!}(x_1 - x_0) \wedge (x_2 - x_0) \wedge \cdots \wedge (x_k - x_0). \tag{6.122}$$

This is the result of integrating the directed measure over a simplex

$$\int_{(x)_{(k)}} dX = \Delta(x)_{(k)} = \Delta X. \tag{6.123}$$

The directed content of a boundary vanishes,

$$\Delta(\partial(x)_{(k)}) = 0. \tag{6.124}$$

As an example, consider a planar simplex consisting of three points. We have

$$\partial(x_0, x_1, x_2) = (x_1, x_2) - (x_0, x_2) + (x_0, x_1). \tag{6.125}$$

So the directed content of the boundary is

$$\Delta(\partial(x_0, x_1, x_2)) = (x_2 - x_1) - (x_2 - x_0) + (x_1 - x_0) = 0. \tag{6.126}$$

The general result of equation (6.124) can be established by induction from the case of a triangle. These results are sufficient to establish that the directed integral over the surface of a simplex is zero:

$$\oint_{\partial(x)_{(k)}} dS = \sum_{i=0}^{k} (-1)^i \int_{(\check{x}_i)_{(k-1)}} dX = \Delta(\partial(x)_{(k)}) = 0. \tag{6.127}$$

A general volume is built up from a chain of simplices. Simplices in the chain are defined such that, at any common boundary, the directed areas of the bounding faces of two simplices are equal and opposite. It follows that the surface integrals over two simplices cancel out over their common face. The

surface integral over the boundary of the volume can therefore be replaced by the sum of the surface integrals over each simplex in the chain. If the boundary is closed we establish that

$$\oint dS = \lim_{n\mapsto\infty} \sum_{a=1}^{n} \oint dS^a = 0. \tag{6.128}$$

The sum runs over each simplex in the surface, with a labeling the simplex. It is implicit in this proof that the surface bounds a volume which can be filled by a connected set of simplices. So, as well as being oriented, the surface must be closed and simply connected.

Next, we return to equation (6.114) and introduce a constant vector b. If we define $b_i = b\cdot e_i$ we see that

$$\sum_{i=1}^{k} b_i \lambda^i = b\cdot(x - x_0), \tag{6.129}$$

which is valid for all vectors x in the simplex of interest. Multiplying equation (6.114) by b_i and summing over i we obtain

$$\int\limits_{(x)_{(k)}} b\cdot(x - x_0)\, dX = \frac{1}{k+1} \sum_{i=1}^{k} b\cdot e_i\, \Delta X, \tag{6.130}$$

where the integral runs over a simplex defined by $k+1$ vertices. A simple reordering yields

$$\begin{aligned} \int b\cdot x\, dX &= \frac{1}{k+1} \left(\sum_{i=1}^{k} b\cdot(x_i - x_0) + (k+1) b\cdot x_0 \right) \Delta X \\ &= b\cdot\bar{x}\, \Delta X, \end{aligned} \tag{6.131}$$

where $\bar{x}$ is the vector representing the (geometric) centre of the simplex,

$$\bar{x} = \frac{1}{k+1} \sum_{i=0}^{k} x_i. \tag{6.132}$$

Now suppose we have a k-simplex specified by the $k+1$ points $(x_0, \ldots, x_k)$ and we form the directed surface integral of $b\cdot x$. We obtain

$$\oint\limits_{\partial(x)_{(k)}} b\cdot x\, dS = \frac{1}{k+1} \sum_{i=0}^{k} (-1)^i b\cdot(x_0 + \cdots \check{x}_i \cdots + x_n) \Delta(\check{x}_i)_{(k-1)}. \tag{6.133}$$

To evaluate the final sum we need the result that

$$\sum_{i=0}^{k} (-1)^i b\cdot(x_0 + \cdots \check{x}_i \cdots + x_n) \Delta(\check{x}_i)_{(k-1)} = \frac{1}{k!} b\cdot(e_1 \wedge \cdots \wedge e_n). \tag{6.134}$$

The proof of this result is purely algebraic and is left as an exercise. We have now established the simple result that

$$\oint_{\partial(x)_{(k)}} b{\cdot}x\, dS = b{\cdot}(\Delta X), \tag{6.135}$$

where $\Delta X = \Delta((x)_{(k)})$. The order and orientations in this result are important. The simplex $(x)_{(k)}$ is oriented, and the order of points specifies how the boundary is traversed. With dS the oriented element over each boundary, and ΔX the volume element for the simplex, we find that the correct expression for the surface integral is $b{\cdot}(\Delta X)$.

We are now in a position to apply these results to the interpolated function $f(x)$ of equation (6.111). Suppose that we are working in a (flat) n-dimensional space and consider a simplex with points $(x_0, \ldots, x_n)$. The simplex is chosen such that its volume is non-zero, so the n vectors $e_i = x_i - x_0$ define a (non-orthonormal) frame. We therefore write

$$\mathsf{e}_i = x_i - x_0, \tag{6.136}$$

and introduce the reciprocal frame $\{\mathsf{e}^i\}$. These vectors satisfy

$$\mathsf{e}^i{\cdot}(x - x_0) = \lambda^i. \tag{6.137}$$

It follows that the surface integral of $f(x)$ over the simplex is given by

$$\begin{aligned}\oint_{\partial(x)_{(k)}} f(x)dS &= \sum_{i=1}^{n}(F_i - F_0) \oint \mathsf{e}^i{\cdot}(x - x_0)dS \\ &= \sum_{i=1}^{n}(F_i - F_0)\mathsf{e}^i{\cdot}(\Delta X).\end{aligned} \tag{6.138}$$

But if we consider the directional derivatives of $f(x)$ we find that

$$\frac{\partial f(x)}{\partial \lambda^i} = F_i - F_0. \tag{6.139}$$

The result of the surface integral can therefore be written

$$\begin{aligned}\oint_{\partial(x)_{(k)}} f(x)dS &= \sum_{i=1}^{n}(F_i - F_0)\mathsf{e}^i{\cdot}(\Delta X) \\ &= \sum_{i=1}^{n}\frac{\partial f}{\partial \lambda^i}\mathsf{e}^i{\cdot}(\Delta X) = \dot{f}\dot{\nabla}{\cdot}(\Delta X).\end{aligned} \tag{6.140}$$

Here we have used the result that $\nabla = \mathsf{e}^i\partial_i$, which follows from using the λ^i as a set of coordinates.

We now consider a chain of simplices, and add the result of equation (6.140)

over each simplex in the chain. The interpolated function $f(x)$ takes on the same value over the common boundary of two adjacent simplices, since $f(x)$ is only defined by the values at the common vertices. In forming a sum over a chain, all of the internal faces cancel and only the surface integral over the boundary remains. We therefore arrive at

$$\oint f(x)\,dS = \sum_a \dot{f}\dot{\nabla}\cdot(\Delta X^a), \tag{6.141}$$

with the sum running over all of the simplices in the chain. Taking the limit as more points are added and each simplex is shrunk in size we arrive at our first statement of the fundamental theorem,

$$\oint_{\partial V} F\,dS = \int_V \dot{F}\dot{\nabla}\,dX. \tag{6.142}$$

We have replaced the interpolated function f with F, which is obtained in the limit as more points are added. We have also used the fact that ∇ lies entirely within the space defined by the pseudoscalar measure dX to remove the contraction on the right-hand side and write a geometric product.

The above proof is easily adapted for the case where the function sits to the right of the measure, giving

$$\oint_{\partial V} dS\,G = \int_V \dot{\nabla}\,dX\,\dot{G}. \tag{6.143}$$

Since ∇ is a vector, the commutation properties with dX will depend on the dimension of the space. A yet more general statement of the fundamental theorem can be constructed by introducing a linear function $\mathsf{L}(A_{n-1}) = \mathsf{L}(A_{n-1}; x)$. This function takes a multivector A_{n-1} of grade $n-1$ as its linear argument, and returns a general multivector. L is also position-dependent, and its linear interpolation over a simplex is defined by

$$L(A) = \mathsf{L}(A; x_0) + \sum_{i=1}^{n} \lambda^i \big(\mathsf{L}(A; x_i) - \mathsf{L}(A, x_0)\big). \tag{6.144}$$

The linearity of $L(A)$ means that sums and integrals can be moved inside the argument, and we establish that

$$\begin{aligned}\oint L(dS) &= \mathsf{L}\left(\oint dS; x_0\right) + \sum_{i=1}^{n} \mathsf{L}\left(\oint \lambda^i dS; x_i\right) - \sum_{i=1}^{n} \mathsf{L}\left(\oint \lambda^i dS; x_0\right) \\ &= \sum_{i=1}^{n} \mathsf{L}(\mathsf{e}^i \Delta X; x_i) - \mathsf{L}(\mathsf{e}^i \Delta X; x_0) \\ &= \dot{L}(\dot{\nabla}\Delta X). \end{aligned} \tag{6.145}$$

There is no position dependence in the final term as the derivative is constant

over the simplex. Building up a chain of simplices and taking the limit we prove the general result

$$\oint_{\partial V} \mathsf{L}(dS) = \int_V \dot{\mathsf{L}}(\dot{\nabla} dX). \tag{6.146}$$

This holds for any linear function $\mathsf{L}(A_{n-1})$ integrated over a closed region of an n-dimensional flat space. This is still not the most general statement of the fundamental theorem, as we will later prove a version valid for surfaces embedded in a curved space, but equation (6.146) is sufficient to make contact with the main integral theorems of vector calculus.

6.4.5 The divergence and Green's theorems

To see the fundamental theorem of geometric calculus in practice, first consider the scalar-valued function

$$\mathsf{L}(A) = \langle JAI^{-1} \rangle. \tag{6.147}$$

Here J is a vector, and I is the (constant) unit pseudoscalar for the n-dimensional space. The argument A is a multivector of grade $n-1$. Equation (6.146) gives

$$\int_V \langle \dot{J}\dot{\nabla} dX I^{-1} \rangle = \int_V \nabla \cdot J \, |dX| = \oint_{\partial V} \langle J dS I^{-1} \rangle, \tag{6.148}$$

where $|dX| = I^{-1} dX$ is the scalar measure over the volume of interest. The normal to the surface, n is defined by

$$n|dS| = dS\, I^{-1}, \tag{6.149}$$

where $|dS|$ is the scalar-valued measure over the surface. This definition ensures that, in Euclidean spaces, $n\, dS$ has the orientation defined by I, and in turn that n points outwards. With this definition we arrive at

$$\int_V \nabla \cdot J \, |dX| = \oint_{\partial V} n \cdot J \, |dS|, \tag{6.150}$$

which is the familiar divergence theorem. This way of writing the theorem hides the fact that $n|dS|$ should be viewed as a single entity, which can be important in spaces of mixed signature.

Now return to the fundamental theorem in the form of equation (6.143), and let G equal the vector $\boldsymbol{J}$ in two-dimensional Euclidean space. We find that

$$\oint_{\partial V} dS\, \boldsymbol{J} = \int_V \dot{\boldsymbol{\nabla}}\, dX\, \dot{\boldsymbol{J}} = -\int_V \boldsymbol{\nabla} \boldsymbol{J}\, dX, \tag{6.151}$$

where we have used the fact that dX is a pseudoscalar, so it anticommutes with

vectors in two dimensions. Introducing Cartesian coordinates we have $dX = I dx\, dy$, so

$$\oint_{\partial V} dS\, \boldsymbol{J} = -\int_V \boldsymbol{\nabla} \boldsymbol{J} I\, dx\, dy. \tag{6.152}$$

If we let $\boldsymbol{J} = P\mathsf{e}_1 + Q\mathsf{e}_2$ and take the scalar part of both sides, we prove Green's theorem in the plane

$$\oint P dx + Q dy = \int \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx\, dy. \tag{6.153}$$

The line integral is taken around the perimeter of the area in a positive sense, as specified by $I = \mathsf{e}_1\mathsf{e}_2$.

6.4.6 Cauchy's integral formula

The fundamental theorem of geometric calculus enables us to view the Cauchy integral theorem of complex variable theory in a new light. We let ψ denote an even-grade multivector, which therefore commutes with dX, so we can write

$$\int \boldsymbol{\nabla} \psi\, dX = \oint d\boldsymbol{s}\, \psi = \oint \frac{\partial \boldsymbol{r}}{\partial \lambda} \psi\, d\lambda. \tag{6.154}$$

In the final expression λ is a parameter along the (closed) curve. Now recall from section 6.3.1 that we form the complex number z by $z = \mathsf{e}_1 \boldsymbol{r}$. We therefore have

$$\oint \psi dz = \int \mathsf{e}_1 \boldsymbol{\nabla} \psi\, dX, \tag{6.155}$$

where the term on the left is now a complex line integral. The condition that ψ is analytic can be written $\boldsymbol{\nabla}\psi = 0$ so we have immediately proved that the line integral of an analytic function around a closed curve always vanishes.

Cauchy's integral formula states that, for an analytic function,

$$f(a) = \frac{1}{2\pi i} \oint_C \frac{f(z)}{z - a} dz, \tag{6.156}$$

where the contour C encloses the point a and is traversed in a positive sense. The precise form of the contour is irrelevant, because the difference between two contour integrals enclosing a is a contour integral around a region not enclosing a (see figure 6.4). In such a region $f(z)/(z-a)$ is analytic so the difference has zero contribution.

To understand Cauchy's theorem in terms of geometric calculus we need to focus on the properties of the Cauchy kernel $1/(z-a)$. We first write

$$\frac{1}{z-a} = \frac{(z-a)^\dagger}{|(z-a)|^2} = \frac{\boldsymbol{r} - \boldsymbol{a}}{(\boldsymbol{r} - \boldsymbol{a})^2} \mathsf{e}_1, \tag{6.157}$$

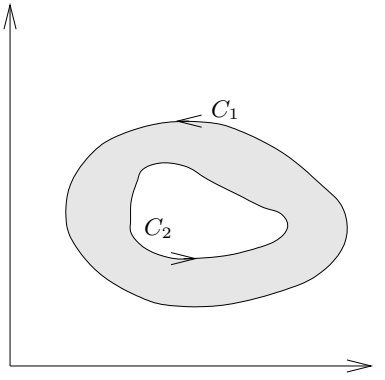

Figure 6.4 *Contour integrals in the complex plane.* The two contours C_1 and C_2 can be deformed into one another, provided the function to be integrated has no singularities in the intervening region. In this case the difference vanishes, by Cauchy's theorem.

where $\boldsymbol{a} = \mathsf{e}_1 a$ is the vector corresponding to the complex number a. The essential quantity here is the vector $(\boldsymbol{r}-\boldsymbol{a})/(\boldsymbol{r}-\boldsymbol{a})^2$, which we can write as

$$\frac{\boldsymbol{r}-\boldsymbol{a}}{(\boldsymbol{r}-\boldsymbol{a})^2} = \boldsymbol{\nabla}\ln|\boldsymbol{r}-\boldsymbol{a}|. \tag{6.158}$$

But $\ln|\boldsymbol{r}-\boldsymbol{a}|$ is the Green's function for the Laplacian operator in two dimensions,

$$\boldsymbol{\nabla}^2 \ln|\boldsymbol{r}-\boldsymbol{a}| = 2\pi\delta(\boldsymbol{r}-\boldsymbol{a}). \tag{6.159}$$

It follows that the vector part of the Cauchy kernel satisfies

$$\boldsymbol{\nabla}\frac{\boldsymbol{r}-\boldsymbol{a}}{(\boldsymbol{r}-\boldsymbol{a})^2} = 2\pi\delta(\boldsymbol{r}-\boldsymbol{a}). \tag{6.160}$$

The Cauchy kernel is the Green's function for the two-dimensional vector derivative! The existence of this Green's function proves that the vector derivative is invertible, which is not true of its separate divergence and curl components.

The Cauchy integral formula now follows from the fundamental theorem of geometric calculus in the form of equation (6.155),

$$\begin{aligned}\oint \frac{f(z)}{z-a}dz &= \mathsf{e}_1 \int \boldsymbol{\nabla}\left(\frac{\boldsymbol{r}-\boldsymbol{a}}{(\boldsymbol{r}-\boldsymbol{a})^2}\mathsf{e}_1 f(x)\right) dX \\ &= \mathsf{e}_1 \int \left(2\pi\delta(x-\boldsymbol{a})\mathsf{e}_1 f(z) + \boldsymbol{\nabla}f(z)\frac{\boldsymbol{r}-\boldsymbol{a}}{(\boldsymbol{r}-\boldsymbol{a})^2}\mathsf{e}_1\right) I|dX| \\ &= 2\pi I f(a),\end{aligned} \tag{6.161}$$

where we have assumed that f is analytic, $\boldsymbol{\nabla}f(z) = 0$. We can now understand precisely the roles of each term in the theorem:

(i) The dz encodes the tangent vector and forms a *geometric* product in the integrand.
(ii) The $(z-a)^{-1}$ is the Green's function for the vector derivative $\boldsymbol{\nabla}$ and ensures that the area integral only picks up the value at a.
(iii) The I (which replaces i) comes from the directed volume element $dX = I\,dx\,dy$.

Much of this is hidden in conventional accounts, but all of these insights are crucial to generalising the theorem. Indeed, we have already proved a more general theorem in two dimensions applying to non-analytic functions. For these we can now write, following section 6.3.1,

$$2\pi I f(a) = \oint \frac{f}{z-a} dz - 2\int \frac{\partial f}{\partial z^\dagger}\frac{1}{z-a} I|dX|. \tag{6.162}$$

A second key ingredient in complex analysis is the series expansion of a function. In particular, if $f(z)$ is analytic apart from a pole of order n at $z=a$, the function has a Laurent series of the form

$$f(z) = \frac{a_{-n}}{(z-a)^n} \cdots \frac{a_{-1}}{z-a} + \sum_{i=0}^{\infty} a_i (z-a)^i. \tag{6.163}$$

The powerful residue theorem states that for such a function

$$\oint_C f(z)\,dz = 2\pi i a_{-1}. \tag{6.164}$$

We now have a new interpretation for the residue term in a Laurent expansion — it is a weighted Green's function. The residue theorem just recovers the weight! Geometric calculus unifies the theory of poles and residues, supposedly unique to complex analysis, with that of Green's functions and δ-functions.

We now have an alternative picture of complex variable theory in terms of Green's functions and surface data. Suppose, for example, that we start with a function $f(x)$ on the real axis. We seek to propagate this function into the upper half-plane, subject to the boundary conditions that f falls to zero as $|z| \mapsto \infty$. The Cauchy formula tells us that we should propagate according to the formula

$$f(a) = \frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{f(x)}{x-a}\,dx. \tag{6.165}$$

But suppose now that we form the Fourier transform of the initial function $f(x)$,

$$f(x) = \int_{-\infty}^{\infty} \frac{dk}{2\pi} \bar{f}(k) \mathrm{e}^{ikx}. \tag{6.166}$$

We now have

$$f(a) = \frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{dk}{2\pi} \bar{f}(k) \int_{-\infty}^{\infty} \frac{\mathrm{e}^{ikx}}{x-a}\,dx. \tag{6.167}$$

Now we only close the x integral in the upper half-plane for positive k. For negative k there is no residue term, since a lies in the the upper half-plane. The Cauchy integral formula now returns

$$f(a) = \int_0^\infty \frac{dk}{2\pi} \bar{f}(k) \mathrm{e}^{ika}. \tag{6.168}$$

This shows that only the part of the function consistent with the desired boundary conditions is propagated in the positive y direction. The remaining part of the function propagates in the $-y$ direction, if similar boundary conditions are imposed in the lower half plane. In this way the boundary conditions and the Green's function between them specify precisely which parts of a function are propagated in the desired direction. No restrictions are placed on the boundary values $f(x)$, which need not be part of an analytic function.

A second example, which generalises nicely, is the unit circle. Suppose we have initial data $f(\theta)$ defined over the unit circle. We write $f(\theta)$ as

$$f(\theta) = \sum_{-\infty}^{\infty} f_n \mathrm{e}^{in\theta}. \tag{6.169}$$

The terms in $\exp(in\theta)$ are replaced by z^n over the unit circle, and we then choose whether to evaluate in interior or exterior closure of the Cauchy integral. The result is that only the negative powers are propagated outwards from the circle, resulting in the function

$$f(z) = \sum_{n=1}^{\infty} f_{-n} z^{-n}, \qquad |z| > 1. \tag{6.170}$$

(The constant component f_0 is technically propagated as well, but this can be removed trivially.) These observations are simple from the point of view of complex variable theory, but are considerably less obvious in propagator theory.

6.4.7 Green's functions in Euclidean spaces

The extension of complex variable theory to arbitrary Euclidean spaces is now straightforward. The analogue of an analytic function is a multivector ψ satisfying $\nabla\psi = 0$. We choose to work with even-grade multivectors to simplify matters. The fundamental theorem states that

$$\oint_{\partial V} dS\, \psi = \int \nabla\psi \, dX = 0. \tag{6.171}$$

where we have used the fact that ψ commutes with the pseudoscalar measure dX. For any *monogenic* function ψ, the directed integral of ψ over a closed surface must vanish.

The Green's function for the vector derivative in n dimensions is simply

$$G(x;y) = \frac{1}{S_n}\frac{x-y}{|x-y|^n}, \tag{6.172}$$

where x and y are vectors and S_n is the surface area of the unit ball in n-dimensional space. The Green's function satisfies

$$\nabla G(x;y) = \nabla\cdot G(x;y) = \delta(x-y). \tag{6.173}$$

In order to allow for the lack of commutativity between G and ψ we use the fundamental theorem in the form

$$\begin{aligned}\oint_{\partial V} G\, dS\, \psi &= \int_V (\dot{G}\dot{\nabla}\,\psi + G\,\nabla\psi)\, dX \\ &= \int_V \dot{G}\dot{\nabla}\psi\, dX,\end{aligned} \tag{6.174}$$

where we have used the fact that ψ is a monogenic function. Setting G equal to the Green's function of equation (6.172) we find that Cauchy's theorem in n dimensions can be written in the form

$$\psi(y) = \frac{1}{IS_n}\oint_{\partial V}\frac{x-y}{|x-y|^n} dS\,\psi(x). \tag{6.175}$$

This relates the value of a monogenic function at a point to the value of a surface integral over a region surrounding the point.

One consequence of equation (6.175) is that a generalisation of Liouville's theorem applies to monogenic functions in Euclidean spaces. We define the modulus function

$$|M| = \langle MM^\dagger\rangle^{1/2}, \tag{6.176}$$

which is a well-defined positive-definite function for all multivectors M in a Euclidean algebra. The modulus function is easily shown to satisfy Schwarz inequality in the form

$$|A+B| \le |A| + |B|. \tag{6.177}$$

If we let a denote a unit vector and let ∇_y denote the derivative with respect to the vector y we find that

$$a\cdot\nabla_y\psi(y) = -\frac{1}{IS_n}\oint_{\partial V}\frac{a(x-y)^2 + na\cdot(x-y)\,(x-y)}{|x-y|^{n+2}} dS\,\psi(x). \tag{6.178}$$

It follows that

$$|a\cdot\nabla_y\psi(y)| \le \frac{1}{S_n}\oint_{\partial V}\frac{n+1}{|x-y|^n}|dS|\,|\psi(x)|. \tag{6.179}$$

But if ψ is bounded, $|\psi(x)|$ never exceeds some given value. Taking the surface of integration out to large radius $r = |x|$, we find that the right-hand side falls off as $1/r$. This is sufficient to prove that the directional derivative of ψ must

vanish in all directions, and the only monogenic function that is bounded over all space is constant ψ.

Equation (6.175) enables us to propagate a function off an initial surface in Euclidean space, subject to suitable boundary conditions. Suppose, for example, that we wish to propagate ψ off the surface of the unit ball, subject to the condition that the function falls to zero at large distance. Much like the two-dimensional case, we can write

$$\psi = \sum_{l=-\infty}^{\infty} \alpha_l \psi_l, \tag{6.180}$$

where the ψ_l are angular monogenics, satisfying

$$x \wedge \nabla \psi = -l\psi. \tag{6.181}$$

Each angular monogenic is multiplied by r^l to yield a full monogenic function, and only the negative powers have their integral closed over the exterior region. The result is the function

$$\psi = \sum_{l=1}^{\infty} \alpha_{-l} r^{-l} \psi_{-l}, \quad r > 1. \tag{6.182}$$

Similarly, the positive powers are picked up if we solve the interior problem.

6.4.8 Spacetime propagators

Propagation in mixed signature spaces is somewhat different to the Euclidean case. There is no analogue of Liouville's theorem to call on, so one can easily construct bounded solutions to the monogenic equation which are non-singular over all space. Plane wave solutions to the massless Dirac equation are an example of such functions. Furthermore, the existence of characteristic surfaces has implications for the how boundary values are specified. To see this, consider a two-dimensional Lorentzian space with basis vectors $\{\gamma_0, \gamma_1\}$, $\gamma_0^2 = -\gamma_1^2 = 1$, and pseudoscalar $I = \gamma_1\gamma_0$. The monogenic equation is $\nabla\psi = 0$, where ψ is an even-grade multivector built from a scalar and pseudoscalar terms. We define the null vectors

$$n_\pm = \gamma_0 \pm \gamma_1. \tag{6.183}$$

Pre-multiplying the monogenic equation by n_+ we find that

$$n_+ \cdot \nabla\psi = -n_+ \wedge \nabla\psi = I\,(n_+ I) \cdot \nabla\psi = -I n_+ \cdot \nabla\psi. \tag{6.184}$$

where we have used the result that $I n_+ = n_+$. It follows that

$$(1 + I) n_+ \cdot \nabla\psi = 0, \tag{6.185}$$

and similarly,

$$(1 - I)n_- \cdot \nabla\psi = 0. \tag{6.186}$$

If we take ψ and decompose it into $\psi = \psi_+ + \psi_-$,

$$\psi_\pm = \tfrac{1}{2}(1 \pm I)\psi, \tag{6.187}$$

we see that the values of the separate $\psi_\pm$ components have vanishing derivatives along the respective null vectors $n_\pm$. Propagation of ψ from an initial surface is therefore quite straightforward. The function is split into $\psi_\pm$, and the values of these are transported along the respective null vectors. That is, ψ_+ has the same value along each vector in the n_+ direction, and the same for ψ_-. There is no need for a complicated contour integral.

The fact that the values of ψ are carried along the characteristics illustrates a key point. Any surface on which initial values are specified can cut a characteristic surface only once. Otherwise the initial values are unlikely to be consistent with the differential equation. For the monogenic equation, $\nabla\psi = 0$, suitable initial conditions consist of specifying ψ along the γ_1 axis, for example. But the fundamental theorem involves integrals around closed loops. The theorem is still valid in a Lorentzian space, so it is interesting to see what happens to the boundary data if we attempt to construct an interior solution with arbitrary surface data. The first step is to construct the Lorentzian Green's function. This can be found routinely via its Fourier transformation. With $x = x^0\gamma_0 + x^1\gamma_1$ we find

$$\begin{aligned} G(x) &= i\int \frac{d\omega}{2\pi}\frac{dk}{2\pi}\,\frac{\omega\gamma_0 + k\gamma_1}{\omega^2 - k^2}\mathrm{e}^{i(kx^1 - \omega x^0)} \\ &= \frac{i}{2}\int \frac{d\omega}{2\pi}\frac{dk}{2\pi}\left(\frac{\gamma_0 + \gamma_1}{\omega - k} + \frac{\gamma_0 - \gamma_1}{\omega + k}\right)\mathrm{e}^{i(kx^1 - \omega x^0)} \\ &= \frac{\epsilon(x^0)}{4}\big(\delta(x^1 - x^0)(\gamma_0 + \gamma_1) + \delta(x^1 + x^0)(\gamma_0 - \gamma_1)\big). \end{aligned} \tag{6.188}$$

The function $\epsilon(x^0)$ takes the value $+1$ or -1, depending on whether x^0 is positive or negative respectively.

To apply the fundamental theorem, suppose we take the contour of figure 6.5, which runs along the γ_1 axis for two different times $t_i < t_f$ and is closed at spatial infinity. We assume that the function we are propagating, ψ, falls off at large spatial distance, and write $\psi(x)$ as $\psi(x^0, x^1)$. The fundamental theorem

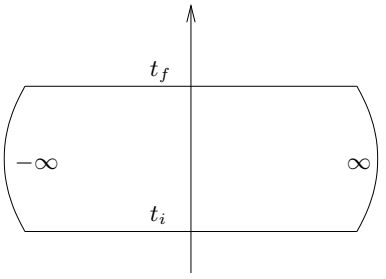

Figure 6.5 *A spacetime contour.* The contour is closed at spatial infinity.

then gives

$$\begin{aligned}\psi(y) =& I\int_{-\infty}^{\infty} d\lambda\, G(t_i\gamma_0 + \lambda\gamma_1 - y)\gamma_1\psi(t_i,\lambda)\\ &- I\int_{-\infty}^{\infty} d\lambda\, G(t_f\gamma_0 + \lambda\gamma_1 - y)\gamma_1\psi(t_f,\lambda)\\ =&\frac{1}{4}(1+I)\big(\psi(t_i, y^1 - y^0 + t_i) + \psi(t_f, y^1 - y^0 + t_f)\big)\\ &-\frac{1}{4}(1-I)\big(\psi(t_i, -y^1 + y^0 + t_i) + \psi(t_f, -y^1 + y^0 + t_f)\big).\end{aligned} \tag{6.189}$$

The construction of $\psi(y)$ in the interior region has a simple interpretation. For the function $\psi_+(y)$, for example, we form the null vector n_+ through y. The value at y is then the average value at the two intersections with the boundary. A similar construction holds for ψ_-. Much like the Euclidean case, only the part of the function on the boundary that is consistent with the monogenic equation is propagated to the interior.

These insights hold in other Lorentzian spaces, such as four-dimensional spacetime. The Green's functions become more complicated, and typically involve derivatives of δ-functions. These are more usefully handled via their Fourier transforms, and are discussed in more detail in section 8.5. In addition, the lack of a Liouville's theorem means that any monogenic function can be added to a Green's function to generate a new Green's function. This has no consequences if one rigorously applies surface integral formulae. In quantum theory, however, this is not usually the case. Rather than a rigorous application of the generalised Green's theorem, it is common instead to talk about propagators which transfer initial data from one timeslice to a later one. Used in this role, the Green's functions we have derived are referred to as *propagators*. As we are not specifying data over a closed surface, adding further terms to our Green's function can have an effect. These effects are related to the desired boundary conditions and are crucial to the formulation of a relativistic quantum field theory. There one is led

to employ the complex-valued Feynman propagator, which ensures that positive frequency modes are propagated forwards in time, and negative frequency modes are propagated backwards in time. We will meet this object in greater detail in section 8.5.

6.5 Embedded surfaces and vector manifolds

We now seek a generalisation of the preceding results where the volume integral is taken over a curved surface. We will do this in the setting of the *vector manifold* theory developed by Hestenes and Sobczyk (1984). The essential concept is to treat a manifold as a surface embedded in a larger, flat space. Points in the manifold are then treated as vectors, which simplifies a number of derivations. Furthermore, we can exploit the coordinate freedom of geometric algebra to derive a set of general results without ever needing to specify the dimension of the background space. The price we pay for this approach is that we are working with a more restrictive concept of a manifold than is usually the case in mathematics. For a start, the surface naturally inherits a metric from the embedding space, so we are already restricting to Riemannian manifolds. We will also insist that a pseudoscalar can be uniquely defined throughout the surface, making it orientable.

While this may all appear quite restrictive, in fact these criteria rule out hardly any structures of interest in physics. This approach enables us to quickly prove a number of key results in Riemannian geometry, and to unite these with results for the exterior geometry of the manifold, achieving a richer general theory. We are not prevented from discussing topological features of surfaces either. Rather than build up a theory of topology which makes no reference to the metric, we instead build up results that are unaffected if the embedding is (smoothly) transformed.

We define a vector manifold as a set of points labelled by vectors lying in a geometric algebra of arbitrary dimension and signature. If we consider a path in the surface $x(\lambda)$, the tangent vector is defined in the obvious way by

$$x' = \left.\frac{\partial x(\lambda)}{\partial \lambda}\right|_{\lambda_0} = \lim_{\epsilon \mapsto 0} \frac{x(\lambda_0 + \epsilon) - x(\lambda_0)}{\epsilon}. \tag{6.190}$$

An advantage of the embedding picture is that the meaning of the limit is well defined, since the numerator exists for all ϵ. This is true even if, for finite epsilon, the difference vector does not lie entirely in the tangent space and only becomes a tangent vector in the limit. Standard formulations of differential geometry avoid any mention of an embedding, however, so have to resort to a more abstract definition of a tangent vector.

An immediate consequence of this approach is that we can define the path

length as

$$s = \int_{\lambda_1}^{\lambda_2} |x' \cdot x'|^{1/2} \, d\lambda. \tag{6.191}$$

The embedded surface therefore inherits a metric from the 'ambient' background space. All finite-dimensional Riemannian manifolds can be studied in this way since, given a manifold, a natural embedding in a larger flat space can always be found. In applications such as general relativity one is usually not interested in the properties of the embedding, since they are physically unmeasurable. But in many other applications, particularly those involving constrained systems, the embedding arises naturally and useful information is contained in the extrinsic geometry of a manifold.

6.5.1 The pseudoscalar and projection

Suppose that we next introduce a set of paths in the surface all passing through the same point x. The paths define a set of tangent vectors $\{\mathsf{e}_1, \ldots, \mathsf{e}_n\}$. We assume that these are independent, so that they form a basis for the n-dimensional tangent space at the point x. The exterior product of the tangent vectors defines the pseudoscalar for the tangent space $I(x)$:

$$I(x) \equiv \mathsf{e}_1 \wedge \mathsf{e}_2 \wedge \cdots \wedge \mathsf{e}_n / |\mathsf{e}_1 \wedge \mathsf{e}_2 \wedge \cdots \wedge \mathsf{e}_n|. \tag{6.192}$$

The modulus in the denominator is taken as a positive number, so that I has the orientation specified by the tangent vectors. The pseudoscalar will satisfy

$$I^2 = \pm 1, \tag{6.193}$$

with the sign depending on dimension and signature. Clearly, to define I in this manner requires that the denominator in (6.192) is non-zero. This provides a restriction on the vector manifolds we consider here, and rules out certain structures in mixed signature spaces. The unit circle in the Lorentzian plane (figure 6.1), for example, falls outside the class of surfaces of studied here, as the tangent space has vanishing norm where the tangent vectors become null. Of course, there is no problem in referring to a closed spacetime curve as a vector manifold. The problem arises when attempting to generalise the integral theorems of the previous sections to such spaces.

The pseudoscalar $I(x)$ contains all of the geometric information about the surface and unites both its intrinsic and extrinsic properties. As well as assuming that $I(x)$ can be defined globally, we will also assume that $I(x)$ is continuous and differentiable over the entire surface, that it has the same grade everywhere, and that it is single-valued. The final assumption implies that the manifold is *orientable*, and rules out objects such as the Mobius strip, where the pseudoscalar is double-valued. Many of the restrictions on the pseudoscalar mentioned above

can be relaxed to construct a more general theory, but this is only achieved at some cost to the ease of presentation. We will follow the simpler route, as the results developed here are sufficiently general for our purposes in later chapters.

The pseudoscalar $I(x)$ defines an operator which projects from an arbitrary multivector onto the component that is intrinsic to the manifold. This operator is

$$\mathsf{P}(A_r(x),x)=\begin{cases}A_r(x)\cdot I(x)\,I^{-1}(x)=A_r\cdot I\,I^{-1}, & r\leq n\\ 0 & r>n\end{cases}. \tag{6.194}$$

which defines an operator at every point x on the manifold. It is straightforward to prove that P satisfies the essential requirement of a projection operator, that is,

$$\mathsf{P}^2(A)=\mathsf{P}\big(\mathsf{P}(A)\big)=\mathsf{P}(A). \tag{6.195}$$

The effect of P on a vector a is to project onto the component of a that lies entirely in the tangent space at the point x. Such vectors are said to be *intrinsic* to the manifold. The complement,

$$\mathsf{P}_\perp(a)=a-\mathsf{P}(a), \tag{6.196}$$

lies entirely outside the tangent space, and is said to be extrinsic to the manifold.

Suppose now that $A(x)$ is a multivector field defined over some region of the manifold. We do not assume that A is intrinsic to the manifold. Given a vector a in the tangent space, the directional derivative along a is defined in the obvious manner:

$$a\cdot\nabla A(x)=\lim_{\epsilon\mapsto 0}\frac{A(x+\epsilon a)-A(x)}{\epsilon}. \tag{6.197}$$

Again, the presence of the embedding enables us to write this limit without ambiguity. The derivative operator $a\cdot\nabla$ is therefore simply the vector derivative in the ambient space contracted with a vector in the tangent space. Given a set of linearly independent tangent vectors $\{\mathsf{e}_i\}$, we can now define a vector derivative ∂ intrinsic to the manifold by

$$\partial=\mathsf{e}^i\,\mathsf{e}_i\cdot\nabla=\mathsf{P}(\nabla). \tag{6.198}$$

This is simply the ambient space vector derivative projected onto the tangent space. The use of the ∂ symbol should not cause confusion with the boundary operator introduced in section 6.4.4. The definition of ∂ requires the existence of the reciprocal frame $\{\mathsf{e}^i\}$, which is why we restricted to manifolds over which I is globally defined. The projection of the vector operator ∂ satisfies

$$\mathsf{P}(\partial)=\partial. \tag{6.199}$$

The contraction of ∂ with a tangent vector a satisfies $a\cdot\partial=a\cdot\nabla$, which is simply the directional derivative in the a direction.

6.5.2 Directed integration for embedded surfaces

Now that we have defined the ∂ operator it is a straightforward task to write down a generalized version of the fundamental theorem of calculus appropriate for embedded surfaces. We can essentially follow through the derivation of section 6.4.4 with little modification. The volume to be integrated over is again triangulated into a chain of simplices. The only difference now is that the pseudoscalar for each simplex varies from one simplex to another. This changes very little. For example we still have

$$\oint dS = 0, \tag{6.200}$$

which holds for the directed integral over the closed boundary of any simply-connected vector manifold.

The linear interpolation results used in deriving equation (6.138) are all valid, because we can again fall back on the embedding picture. In addition, the assumption that the pseudoscalar $I(x)$ is globally defined means that the reciprocal frame required in equation (6.138) is well defined. The only change that has to be made is that the ambient derivative ∇ is replaced by its projection into the manifold, because we naturally assemble the inner product of ∇ with the pseudoscalar. The most general statement of the fundamental theorem can now be written as

$$\oint_{\partial V} \mathsf{L}(dS) = \int_V \dot{\mathsf{L}}(\dot{\partial} dX) = \int_V \dot{\mathsf{L}}(\dot{\nabla}\cdot dX). \tag{6.201}$$

The form of the volume integral involving ∂ is generally more useful as it forms a geometric product with the volume element. The function L can be any multivector-valued function in this equation — it is not restricted to lie in the tangent space. An important feature of this more general theorem is that if we write $dX = I|dX|$ we see that the directed element dX is position-dependent. But this position dependence is *not* differentiated in equation (6.201). It is only the integrand that is differentiated.

There are two main applications of the general theorem derived here. The first is a generalisation of the divergence theorem to curved spaces. We again write

$$\mathsf{L}(A) = \langle JAI^{-1}\rangle, \tag{6.202}$$

where J is a vector field in the tangent space, and I is the unit pseudoscalar for the n-dimensional curved space. Equation (6.201) now gives

$$\oint_{\partial V} n\cdot J|dS| = \int_V (\partial\cdot J + \langle J\dot{\partial}\dot{I}^{-1}I\rangle)|dX|, \tag{6.203}$$

where $|dX| = I^{-1}dX$ and $n|dS| = dS\, I^{-1}$. The final term in the integral vanishes, as can be shown by first writing $I^{-1} = \pm I$ and using

$$\langle J\dot{\partial}\dot{I}I\rangle = \tfrac{1}{2}\langle J\dot{\partial}(\dot{I}I + I\dot{I})\rangle = \tfrac{1}{2}\langle J\partial(I^2)\rangle = 0. \tag{6.204}$$

It follows that the divergence theorem in curved space is essentially unchanged from the flat-space version, so

$$\int_V \partial\cdot J\,|dX| = \oint_{\partial V} n\cdot J\,|dS|. \tag{6.205}$$

As a second application we derive Stokes' theorem in three dimensions. Suppose that σ denotes an open, connected surface in three dimensions, with boundary $\partial\sigma$. The linear function L takes a vector as its linear argument and we define

$$\mathsf{L}(\boldsymbol{a}) = \boldsymbol{J}\cdot\boldsymbol{a}. \tag{6.206}$$

Equation (6.201) now gives

$$\oint_{\partial\sigma} \boldsymbol{J}\cdot d\boldsymbol{l} = \int_\sigma \langle \dot{\boldsymbol{J}}\,\dot{\boldsymbol{\nabla}}\cdot dX\rangle = -\int_\sigma (\boldsymbol{\nabla}\wedge\boldsymbol{J})\cdot dX, \tag{6.207}$$

where the line integral is taken around the boundary of the surface, and since the embedding is specified we have chosen a form of the integral theorem involving the three-dimensional derivative $\boldsymbol{\nabla}$. We now define the normal vector to the surface by

$$dX = I\boldsymbol{n}|dX|, \tag{6.208}$$

where I is the three-dimensional (right-handed) pseudoscalar. This equation defines the vector $\boldsymbol{n}$ normal to the surface. The direction in which this points depends on the orientation of dX. Around the boundary, for example, we can denote the tangent vector at the boundary by $\boldsymbol{l}$, and the vector pointing into the surface as $\boldsymbol{m}$. Then dX has the orientation specified by $\boldsymbol{l}\wedge\boldsymbol{m}$, and from equation (6.208) we see that $\boldsymbol{l},\boldsymbol{m},\boldsymbol{n}$ must form a right-handed set. This extends inwards to define the normal vector $\boldsymbol{n}$ over the surface (see figure 6.6). We now have

$$\oint_{\partial\sigma} \boldsymbol{J}\cdot d\boldsymbol{l} = \int_\sigma -(I\boldsymbol{\nabla}\wedge\boldsymbol{J})\cdot\boldsymbol{n}\,|dX| = \int_\sigma (\operatorname{curl}\boldsymbol{J})\cdot\boldsymbol{n}\,|dX|, \tag{6.209}$$

which is the familiar Stokes' theorem in three dimensions. This is only the scalar part of a more general (and less familiar) theorem which holds in three dimensions. To form this result we remove the projection onto the scalar part, to obtain

$$\oint_{\partial\sigma} d\boldsymbol{l}\,\boldsymbol{J} = -I\int_\sigma \boldsymbol{n}\wedge\boldsymbol{\nabla}\,\boldsymbol{J}\,|dX|. \tag{6.210}$$

A version of this result holds for any open n-dimensional surface embedded in a flat space of dimension $n+1$.

6.5.3 Intrinsic and extrinsic geometry

Suppose now that the directional derivative $a\cdot\partial$ acts on a tangent vector field $b(x) = \mathsf{P}(b(x))$. There is no guarantee that the resulting vector also lies entirely

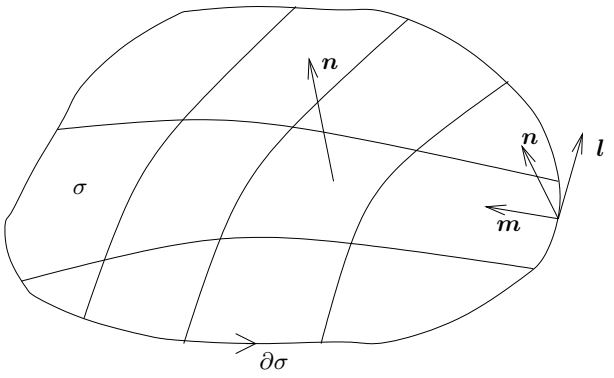

Figure 6.6 *Orientations for Stokes' theorem.* The bivector measure dX defines an orientation over the surface and at the boundary. With $\boldsymbol{l}$ and $\boldsymbol{m}$ the tangent and inward directions at the boundary, the normal $\boldsymbol{n}$ is defined so that $\boldsymbol{l}, \boldsymbol{m}, \boldsymbol{n}$ form a right-handed set.

in the tangent space, even if a does. For example, consider the simple case of a circle in the plane. The derivative of the tangent vector around the circle is a radial vector, which is entirely extrinsic to the manifold. In order to restrict to quantities intrinsic to the manifold we define a new derivative — the covariant derivative D — as follows:

$$a\cdot DA(x) = \mathsf{P}(a\cdot\partial A(x)). \tag{6.211}$$

The operator $a\cdot D$ acts on multivectors in the tangent space, returning a new multivector field in the tangent space. Since the $a\cdot\partial$ operator satisfies Leibniz's rule, the covariant derivative $a\cdot D$ must as well,

$$a\cdot D(AB) = \mathsf{P}\big(a\cdot\partial(AB)\big) = (a\cdot DA)B + A\,a\cdot DB. \tag{6.212}$$

The vector operator D is then defined in the obvious way from the covariant directional derivatives,

$$D = \mathsf{e}^i\,\mathsf{e}_i\cdot D. \tag{6.213}$$

So, for example, we can write

$$DA_r = \mathsf{e}^i(\mathsf{e}_i\cdot DA_r) = \mathsf{P}(\partial A_r). \tag{6.214}$$

The result decomposes into grade-raising and grade-lowering terms, so we write

$$\begin{aligned} D\cdot A_r &= \langle DA_r\rangle_{r-1}, \\ D\wedge A_r &= \langle DA_r\rangle_{r+1}. \end{aligned} \tag{6.215}$$

So, like ∂, D has the algebraic properties of a vector in the tangent space. Acting on a scalar function $\alpha(x)$ defined over the manifold the two derivatives coincide,

so

$$\partial\alpha(x) = D\alpha(x). \tag{6.216}$$

Suppose now that a is a tangent vector to the manifold, and we look at how the pseudoscalar changes along the a direction. It should be obvious, from considering a 2-sphere for example, that the resulting quantity must lie at least partly outside the manifold. We let $\{\mathsf{e}_i\}$ denote an orthonormal frame, so

$$I = \mathsf{e}_1\mathsf{e}_2\cdots\mathsf{e}_n. \tag{6.217}$$

It follows that

$$\begin{aligned} a\cdot\partial\, I\, I^{-1} &= \sum_{i=1}^{n} \mathsf{e}_1\cdots\big(a\cdot D\mathsf{e}_i + \mathsf{P}_\perp(a\cdot\partial\,\mathsf{e}_i)\big)\cdots\mathsf{e}_n\, I^{-1} \\ &= a\cdot D\, I\, I^{-1} + \mathsf{P}_\perp(a\cdot\partial\,\mathsf{e}_i)\wedge\mathsf{e}^i. \end{aligned} \tag{6.218}$$

The final term is easily shown to be independent of the choice of frame. But $a\cdot DI$ must remain in the tangent space, so it can only be a multiple of the pseudoscalar I. It follows that

$$(a\cdot D\, I)I = \langle(a\cdot D\, I)I\rangle = \tfrac{1}{2}\langle a\cdot D(I^2)\rangle = 0, \tag{6.219}$$

so

$$a\cdot D\, I = 0. \tag{6.220}$$

That is, the (unit) pseudoscalar is a covariant constant over the manifold. Equation (6.218) now simplifies to give

$$a\cdot\partial\, I = \mathsf{P}_\perp(a\cdot\partial\,\mathsf{e}_i)\wedge\mathsf{e}^i\, I = -S(a)I, \tag{6.221}$$

which defines the *shape tensor* $S(a)$. This is a bivector-valued, linear function of its vector argument a, where a is a tangent vector. Since the result of $a\cdot\partial\, I$ has the same grade as I, we can write

$$a\cdot\partial I = I\times S(a) \tag{6.222}$$

with

$$S(a)\cdot I = S(a)\wedge I = 0. \tag{6.223}$$

The fact that $S(a)\cdot I = 0$ confirms that $S(a)$ lies partly outside the manifold, so that $\mathsf{P}(S(a)) = 0$.

The shape tensor $S(a)$ unites the intrinsic and extrinsic geometry of the manifold in a single quantity. It can be thought of as the 'angular momentum' of $I(x)$ as it slides over the manifold. The shape tensor provides a compact relation between directional and covariant derivatives. We first form

$$b\cdot S(a) = b^i\mathsf{P}_\perp(a\cdot\partial\,\mathsf{e}_i) = \mathsf{P}_\perp(a\cdot\partial\, b), \tag{6.224}$$

where a and b are tangent vectors. It follows that

$$a\cdot\partial\, b = \mathsf{P}(a\cdot\partial\, b) + \mathsf{P}_\perp(a\cdot\partial\, b) = a\cdot D\, b + b\cdot S(a), \tag{6.225}$$

which we can rearrange to give the neat result

$$a\cdot D\, b = a\cdot\partial\, b + S(a)\cdot b. \tag{6.226}$$

Applying this result to the geometric product bc we find that

$$\begin{aligned} a\cdot D(bc) &= (a\cdot\partial\, b)c + S(a)\cdot b\, c + b(a\cdot\partial\, c) + b\, S(a)\cdot c \\ &= a\cdot\partial(bc) + S(a)\times(bc), \end{aligned} \tag{6.227}$$

where $\times$ is the commutator product, $A\times B = (AB - BA)/2$. It follows that for any multivector field A taking its values in the tangent space we have

$$a\cdot DA = a\cdot\partial A + S(a)\times A. \tag{6.228}$$

The fact that $S(a)$ is bivector-valued ensures that $S(a)\times A$ does not alter the grade of A. As a check, setting $A = I$ recovers equation (6.222). If we now write

$$a\cdot\partial\, b = a\cdot\partial\, \mathsf{P}(b) = a\cdot\dot\partial\, \dot{\mathsf{P}}(b) + \mathsf{P}(a\cdot\partial b) = a\cdot\dot\partial\, \dot{\mathsf{P}}(b) + a\cdot Db \tag{6.229}$$

we establish the further relation

$$a\cdot\dot\partial\, \dot{\mathsf{P}}(b) = b\cdot S(a). \tag{6.230}$$

This holds for any pair of tangent vectors a and b.

6.5.4 Coordinates and derivatives

A number of important results can be derived most simply by introducing a coordinate frame. In a region of the manifold we introduce local coordinates x^i and define the frame vectors

$$\mathsf{e}_i = \frac{\partial x}{\partial x^i}. \tag{6.231}$$

From the definition of ∂ it follows that $\mathsf{e}^i = \partial x^i$. The $\{\mathsf{e}_i\}$ are usually referred to as *tangent* vectors and the reciprocal frame $\{\mathsf{e}^i\}$ as *cotangent* vectors (or 1-forms). The fact that the space is curved implies that it may not be possible to construct a global coordinate system. The 2-sphere is the simplest example of this. In this case we simply patch together a series of local coordinate systems. The covariant derivative along a coordinate vector, $\mathsf{e}_i\cdot D$, satisfies

$$\mathsf{e}_i\cdot DA = D_iA = \mathsf{e}_i\cdot\partial A + S(\mathsf{e}_i)\times A = \partial_iA + S_i\times A, \tag{6.232}$$

which defines the D_i and S_i symbols.

The tangent frame vectors satisfy

$$\partial_i\mathsf{e}_j - \partial_j\mathsf{e}_i = (\partial_i\partial_j - \partial_j\partial_i)x = 0. \tag{6.233}$$

Projecting this result into the manifold establishes that

$$D_i \mathsf{e}_j - D_j \mathsf{e}_i = 0. \tag{6.234}$$

Projecting out of the manifold we similarly establish the result

$$\mathsf{e}_i \cdot S_j = \mathsf{e}_j \cdot S_i. \tag{6.235}$$

In terms of arbitrary tangent vectors a and b this can be written as

$$a \cdot S(b) = b \cdot S(a). \tag{6.236}$$

The shape tensor can be written in terms of the coordinate vectors as

$$S(a) = \mathsf{e}^k \wedge \mathsf{P}_\perp(a \cdot \partial \mathsf{e}_k). \tag{6.237}$$

It follows that

$$S_i = \mathsf{e}^k \wedge \mathsf{P}_\perp(\partial_i \mathsf{e}_k) = \mathsf{e}^k \wedge \mathsf{P}_\perp(\partial_k \mathsf{e}_i). \tag{6.238}$$

The tangent vectors therefore satisfy

$$\partial \wedge \mathsf{e}_i = \mathsf{e}^k \wedge \big(\mathsf{P}(\partial_k \mathsf{e}_i) + \mathsf{P}_\perp(\partial_k \mathsf{e}_i)\big) = D \wedge \mathsf{e}_i + S_i. \tag{6.239}$$

If we decompose a vector in the tangent space as $a = a^i \mathsf{e}_i$ we establish the general result that

$$\partial \wedge a = D \wedge a + S(a). \tag{6.240}$$

This gives a further interpretation to the shape tensor. It is the object which picks up the component of the curl of a tangent vector which lies outside the tangent space. As we can write

$$\partial \wedge a = \partial \wedge \big(\mathsf{P}(a)\big) = \dot{\partial} \wedge \dot{\mathsf{P}}(a) + \mathsf{P}(\partial \wedge a) = D \wedge a + \dot{\partial} \wedge \dot{\mathsf{P}}(a), \tag{6.241}$$

we establish the further result

$$\dot{\partial} \wedge \dot{\mathsf{P}}(a) = S(a). \tag{6.242}$$

This is easily seen to be consistent with the definition of the shape tensor in terms of the derivative of pseudoscalar.

If we now apply the preceding to the case of the curl of a gradient of a scalar, we find that

$$\partial \wedge \partial \phi = \mathsf{P}(\nabla) \wedge \mathsf{P}(\nabla \phi) = \mathsf{P}(\nabla \wedge \nabla \phi) + \dot{\partial} \wedge \dot{\mathsf{P}}(\nabla \phi). \tag{6.243}$$

But the ambient derivative satisfies the integrability condition $\nabla \wedge \nabla = 0$. It follows that we have

$$\partial \wedge \partial \phi = S(\nabla \phi), \tag{6.244}$$

which lies outside the manifold. The covariant derivative therefore satisfies

$$D \wedge (D\phi) = 0. \tag{6.245}$$

An important application of this result is to the coordinate scalars themselves. We find that

$$D\wedge(Dx^i) = D\wedge\mathsf{e}^i = 0, \tag{6.246}$$

which can also be proved directly from equation (6.234). Applying this result to an arbitrary vector $a = a_i\mathsf{e}^i$ we find that

$$D\wedge a = D\wedge(a_j\mathsf{e}^j) = \mathsf{e}^i\wedge\mathsf{e}^j(\partial_i a_j) = \tfrac{1}{2}\mathsf{e}^i\wedge\mathsf{e}^j(\partial_i a_j - \partial_j a_i). \tag{6.247}$$

This demonstrates that the $D\wedge$ operator is precisely the *exterior derivative* of differential geometry.

6.5.5 Riemannian geometry

To understand further how the shape tensor can specify the intrinsic geometry of a surface, we now make contact with Riemannian geometry. In Riemannian geometry one focuses entirely on the intrinsic properties of a manifold. It is customary to formulate the subject using the metric tensor as the starting point. In terms of the $\{\mathsf{e}_i\}$ coordinate frame the metric tensor is defined in the expected manner:

$$g_{ij} = \mathsf{e}_i\cdot\mathsf{e}_j. \tag{6.248}$$

In what follows we will not place any restriction on the signature of the tangent space. Some texts prefer to use the adjective 'Riemannian' to refer to extensions of Euclidean geometry to curved spaces (as Riemann originally intended). But in the physics literature it is quite standard now to refer to general relativity as a theory of Riemannian geometry, despite the Lorentzian signature.

After the metric, the next main object in Riemannian geometry is the Christoffel connection. The directional covariant derivative, D_i, restricts the result of its action to the tangent space. The result of its action on one of the $\{\mathsf{e}_i\}$ vectors can therefore be decomposed uniquely in the $\{\mathsf{e}_i\}$ frame. The coefficients of this define the Christoffel connection by

$$\Gamma^i_{jk} = (D_j\mathsf{e}_k)\cdot\mathsf{e}^i. \tag{6.249}$$

The components of the connection are clearly dependent on the choice of coordinate system, as well as the underlying geometry. It follows that a connection is necessary even when working in a curvilinear coordinate system in a flat space. A connection on its own does not imply that a space is curved. A typical use of the Christoffel connection is in finding the components in the $\{\mathsf{e}^i\}$ frame of a covariant derivative $a\cdot D\,b$, for example. We form

$$(a\cdot D\,b)\cdot\mathsf{e}^i = a^j\big(D_j(b^k\mathsf{e}_k)\big)\cdot\mathsf{e}^i = a^j(\partial_j b^i + \Gamma^i_{jk}b^k), \tag{6.250}$$

which shows how the connection accounts for the position dependence in the coordinate frame.

The components of the Christoffel connection can be found directly from the metric without referring to the frame vectors themselves. To achieve this we first establish a pair of results. The first is that the connection Γ^i_{jk} is symmetric on the jk indices. This follows from

$$\Gamma^i_{jk} - \Gamma^i_{kj} = (D_j \mathsf{e}_k - D_k \mathsf{e}_j)\cdot \mathsf{e}^i = 0, \tag{6.251}$$

where we have used equation (6.234). The second result is for the curl of a frame vector,

$$D\wedge \mathsf{e}_i = D\wedge(g_{ij}\mathsf{e}^j) = (Dg_{ij})\wedge \mathsf{e}^j. \tag{6.252}$$

We can now write

$$\begin{aligned}
\Gamma^i_{jk} &= \tfrac{1}{2}\mathsf{e}^i\cdot(D_j\mathsf{e}_k + D_k\mathsf{e}_j) \\
&= \tfrac{1}{2}\mathsf{e}^i\cdot\big(\mathsf{e}_j\cdot(Dg_{kl}\wedge\mathsf{e}^l) + \mathsf{e}_k\cdot(Dg_{jl}\wedge\mathsf{e}^l) + Dg_{jk}\big) \\
&= \tfrac{1}{2}\mathsf{e}^i\cdot(\partial_j g_{kl}\mathsf{e}^l + \partial_k g_{jl}\mathsf{e}^l - Dg_{jk}) \\
&= \tfrac{1}{2}g^{il}(\partial_j g_{kl} + \partial_k g_{jl} - \partial_l g_{jk}),
\end{aligned} \tag{6.253}$$

which recovers the familiar definition of the Christoffel connection.

We now seek a method of encoding the intrinsic curvature of a Riemannian manifold. Suppose we form the commutator of two covariant derivatives

$$\begin{aligned}
[D_i, D_j]A = {} & \partial_i(\partial_j A + S_j\times A) + S_i\times(\partial_j A + S_j\times A) \\
& -\partial_j(\partial_i A + S_i\times A) - S_j\times(\partial_i A + S_i\times A) \\
= {} & (\partial_i S_j - \partial_j S_i)\times A + (S_i\times S_j)\times A,
\end{aligned} \tag{6.254}$$

where we have used the Jacobi identity of section 4.1.3. Remarkably, all derivatives of the multivector A have cancelled out and what remains is a commutator with a bivector. To simplify this we form

$$\begin{aligned}
\partial_i S_j - \partial_j S_i &= -\partial_i(\partial_j I\, I^{-1}) + \partial_j(\partial_i I\, I^{-1}) \\
&= -S_j I S_i I^{-1} + S_i I S_j I^{-1} \\
&= -2S_i\times S_j,
\end{aligned} \tag{6.255}$$

where we have used the fact that $S(a)$ anticommutes with I. On substituting this result in equation (6.254) we obtain the simple result

$$[D_i, D_j]A = -(S_i\times S_j)\times A. \tag{6.256}$$

The commutator of covariant derivatives defines the *Riemann tensor*. We denote this by $\mathsf{R}(a\wedge b)$, where

$$\mathsf{R}(\mathsf{e}_i\wedge\mathsf{e}_j)\times A = [D_i, D_j]A. \tag{6.257}$$

$\mathsf{R}(a\wedge b)$ is a bivector-valued linear function of its bivector argument. In terms of the shape tensor we have

$$\mathsf{R}(a\wedge b) = \mathsf{P}\big(S(b)\wedge S(a)\big). \tag{6.258}$$

The projection is required here because the Riemann tensor is defined to be entirely intrinsic to the manifold. The Riemann tensor (and its derivatives) fully encodes all of the local intrinsic geometry of a manifold. Since it can be derived easily from the shape tensor, it follows that the shape tensor also captures all of the intrinsic geometry. In addition to this, the shape tensor tells us about the extrinsic geometry — how the manifold is embedded in the larger ambient space.

The Riemann tensor can also be expressed entirely in terms of intrinsic quantities. To achieve this we first write

$$\mathsf{R}(\mathsf{e}_i \wedge \mathsf{e}_j)\cdot\mathsf{e}_k = [D_i, D_j]\mathsf{e}_k = D_i(\Gamma^a_{jk}\mathsf{e}_a) - D_j(\Gamma^a_{ik}\mathsf{e}_a). \tag{6.259}$$

It follows that

$$\begin{aligned}\mathsf{R}_{ijk}{}^{l} &= \mathsf{R}(\mathsf{e}_i \wedge \mathsf{e}_j)\cdot(\mathsf{e}_k \wedge \mathsf{e}^l) \\ &= \partial_i\Gamma^l_{jk} - \partial_j\Gamma^l_{ik} + \Gamma^a_{jk}\Gamma^l_{ia} - \Gamma^a_{ik}\Gamma^l_{ja},\end{aligned} \tag{6.260}$$

recovering the standard definition of Riemannian geometry. An immediate advantage of the geometric algebra route is that many of the symmetry properties of $\mathsf{R}_{ijk}{}^{l}$ follow immediately from the fact that $\mathsf{R}(a\wedge b)$ is a bivector-valued linear function of a bivector. This immediately reduces the number of degrees of freedom to $n^2(n-1)^2/4$.

A further symmetry of the Riemann tensor can be found as follows:

$$\begin{aligned}\mathsf{R}(\mathsf{e}_i \wedge \mathsf{e}_j)\cdot\mathsf{e}_k &= D_iD_j\mathsf{e}_k - D_jD_i\mathsf{e}_k \\ &= D_iD_k\mathsf{e}_j - D_jD_k\mathsf{e}_i \\ &= [D_i, D_k]\mathsf{e}_j - [D_j, D_k]\mathsf{e}_i + D_k(D_i\mathsf{e}_j - D_j\mathsf{e}_i) \\ &= \mathsf{R}(\mathsf{e}_i \wedge \mathsf{e}_k)\cdot\mathsf{e}_j - \mathsf{R}(\mathsf{e}_j \wedge \mathsf{e}_k)\cdot\mathsf{e}_i.\end{aligned} \tag{6.261}$$

It follows that

$$a\cdot\mathsf{R}(b\wedge c) + c\cdot\mathsf{R}(a\wedge b) + b\cdot\mathsf{R}(c\wedge a) = 0, \tag{6.262}$$

for any three vectors a, b, c in the tangent space. This equation tells us that a vector quantity vanishes for all trivectors $a\wedge b\wedge c$, which provides a set of $n^2(n-1)(n-2)/6$ scalar equations. The number of independent degrees of freedom in the Riemann tensor is therefore reduced to

$$\frac{1}{4}n^2(n-1)^2 - \frac{1}{6}n^2(n-1)(n-2) = \frac{1}{12}n^2(n^2-1). \tag{6.263}$$

This gives the values 1, 6 and 20 for two, three and four dimensions respectively. Further properties of the Riemann tensor are covered in more detail in later chapters, where in particular we are interested in its relevance to gravitation.

The fact that Riemannian geometry is founded on the covariant derivative D, as opposed to the projected vector derivative ∂ limits the application of the integral theorem of equation (6.201). If one attempts to add multivectors from

different points in the surface, there is no guarantee that the result remains intrinsic. The only quantities that can be combined from different points on the surface are scalars, or functions taking their values in a different space (such as a Lie group). The most significant integral theorem that remains is a generalization of Stokes' theorem, applicable to a grade-r multivector A_r and an open surface σ of dimension $r+1$. For this case we have

$$\oint_{\partial\sigma} A_r \cdot dS = \int_\sigma (\dot{A}_r \wedge \dot{\partial}) \cdot dX = (-1)^r \int_\sigma (D \wedge A_r) \cdot dX, \tag{6.264}$$

which only features intrinsic quantities. A particular case of this is when $r = n-1$, which recovers the divergence theorem. This is important for constructing conservation theorems in curved spaces.

6.5.6 Transformations and maps

The study of maps between vector manifolds helps to clarify some of the relationships between the structures defined in this chapter and more standard formulations of differential geometry. Suppose that $f(x)$ defines a map from one vector manifold to another. We denote these $\mathcal{M}$ and $\mathcal{M}'$, so that

$$x' = f(x) \tag{6.265}$$

associates a point in the manifold $\mathcal{M}'$ with one in $\mathcal{M}$. We will only consider smooth, differentiable, invertible maps between manifolds. In the mathematics literature these are known as *diffeomorphisms*. These are a subset of the more general concept of a *homeomorphism*, which maps continuously between spaces without the restriction of smoothness. Somewhat surprisingly, these two concepts are not equivalent. It is possible for two manifolds to be homeomorphic, but not admit a diffeomorphism between them. This implies that it is possible for a single topological space to admit more than one differentiable structure. The first example of this to be discovered was the sphere S^7, which admits 28 distinct differentiable structures! In 1983 Donaldson proved the even more striking result that four-dimensional space R^4 admits an infinite number of differentiable structures.

A path in $\mathcal{M}$, $x(\lambda)$, maps directly to a path in $\mathcal{M}'$. The map accordingly induces a map between tangent vectors, as seen by forming

$$\frac{\partial x'(\lambda)}{\partial \lambda} = \frac{\partial f\big(x(\lambda)\big)}{\partial \lambda} = \mathsf{f}(v), \tag{6.266}$$

where v is the tangent vector in $\mathcal{M}$, $v = \partial_\lambda x(\lambda)$ and the linear function f is defined by

$$\mathsf{f}(a) = a \cdot \partial f(x) = \mathsf{f}(a; x). \tag{6.267}$$

The function $\mathsf{f}(a)$ takes a tangent vector in $\mathcal{M}$ as its linear argument, and returns

the image tangent vector in $\mathcal{M}'$. If we denote the latter by a', and write out the position dependence explicitly, we have

$$a'(x') = \mathsf{f}(a(x); x). \tag{6.268}$$

This map is appropriate for tangent vectors, so applies to the coordinate frame vectors $\{\mathsf{e}_i\}$. These map to an equivalent frame for the tangent space to $\mathcal{M}'$,

$$\mathsf{e}'_i = \mathsf{f}(\mathsf{e}_i). \tag{6.269}$$

The reciprocal frame in the transformed space is therefore given by

$$\mathsf{e}^{i'} = \bar{\mathsf{f}}^{-1}(\mathsf{e}^i). \tag{6.270}$$

The fact that the map $x \mapsto f(x)$ is assumed to be invertible ensures that the adjoint function $\bar{\mathsf{f}}(a)$ is also invertible.

Under transformations, therefore, vectors in one space can transform in two different ways. If they are tangent vectors they transform under the action of $\mathsf{f}(a)$. If they are cotangent vectors they transform under action of $\bar{\mathsf{f}}^{-1}(a)$. In differential geometry it is standard practice to maintain a clear distinction between these types of vectors, so one usually thinks of tangent and cotangent vectors as lying in separate linear spaces. The contraction relation $\mathsf{e}^i \cdot \mathsf{e}_j = \delta^i_j$ identifies the spaces as dual to each other. This relation is metric-independent and is preserved by arbitrary diffeomorphisms. These maps relate differentiable manifolds, and two diffeomorphic spaces are usually viewed as the same manifold.

A metric is regarded as an additional construct on a differentiable manifold, which maps between the tangent and cotangent spaces. In the vector manifold picture this map is achieved by constructing the reciprocal frame using equation (4.94). In using this relation we are implicitly employing a metric in the contraction with the pseudoscalar. For the theory of vector manifolds it is therefore useful to distinguish objects and operations that transform simply under diffeomorphisms. These will define the metric-independent features of a vector manifold. Metric-dependent quantities, like the Riemann tensor, invariably have more complicated transformation laws.

The exterior product of a pair of tangent vectors transforms as

$$\mathsf{e}_i \wedge \mathsf{e}_j \mapsto \mathsf{f}(\mathsf{e}_i) \wedge \mathsf{f}(\mathsf{e}_j) = \mathsf{f}(\mathsf{e}_i \wedge \mathsf{e}_j). \tag{6.271}$$

For example, if I' is the unit pseudoscalar for $\mathcal{M}'$ we have

$$\mathsf{f}(I) = \det(\mathsf{f}) I' \tag{6.272}$$

and for invertible maps we must have $\det(\mathsf{f}) \neq 0$. Similarly, for cotangent vectors we see that

$$\mathsf{e}^i \wedge \mathsf{e}^j \mapsto \bar{\mathsf{f}}^{-1}(\mathsf{e}^i) \wedge \bar{\mathsf{f}}^{-1}(\mathsf{e}^j) = \bar{\mathsf{f}}^{-1}(\mathsf{e}^i \wedge \mathsf{e}^j). \tag{6.273}$$

So exterior products of like vectors give rise to higher grade objects in a manner

that is unchanged by diffeomorphisms. Metric invariants are constructed from inner products between tangent and cotangent vectors. Since the derivative of a scalar field is

$$\partial\phi = \mathsf{e}^i \partial_i \phi, \tag{6.274}$$

we see that $\partial\phi$ is a cotangent vector, and we can write

$$\partial' = \bar{\mathsf{f}}^{-1}(\partial). \tag{6.275}$$

A similar result holds for the covariant derivative D. If a is a tangent vector the directional derivative of a scalar field $a\cdot\partial\phi$ is therefore an invariant,

$$a'\cdot\partial'\phi' = \mathsf{f}(a)\cdot\bar{\mathsf{f}}^{-1}(\partial)\phi = a\cdot\partial\phi, \tag{6.276}$$

where $\phi'(x') = \phi(x)$.

In constructing the covariant derivative in section 6.5.3, we made use of the projection operation $\mathsf{P}(a)$. This *is* a metric operation, as it relies on a contraction with I. Hence the covariant derivatives $D_i\mathsf{e}_j$ do depend on the metric (via the connection). To establish a metric-independent operation we let a and b represent tangent vectors and form

$$\begin{aligned} a\cdot\partial b - b\cdot\partial a &= a\cdot Db - b\cdot Da + a\cdot S(b) - b\cdot S(a) \\ &= a\cdot Db - b\cdot Da. \end{aligned} \tag{6.277}$$

The shape terms cancel, so the result is intrinsic to the manifold. Under a diffeomorphism the result transforms to

$$a\cdot\partial\mathsf{f}(b) - b\cdot\partial\mathsf{f}(a) = \mathsf{f}(a\cdot\partial b - b\cdot\partial a) + a\cdot\dot{\partial}\dot{\mathsf{f}}(b) - b\cdot\dot{\partial}\dot{\mathsf{f}}(a). \tag{6.278}$$

But $\mathsf{f}(a)$ is the differential of the map $f(x)$, so we have

$$(\partial_i\partial_j - \partial_j\partial_i)f(x) = \partial_i\mathsf{f}(\mathsf{e}_j) - \partial_j\mathsf{f}(\mathsf{e}_i) = \dot{\partial}_i\dot{\mathsf{f}}(\mathsf{e}_j) - \dot{\partial}_j\dot{\mathsf{f}}(\mathsf{e}_i) = 0. \tag{6.279}$$

It follows that, for tangent vectors a and b,

$$a\cdot\dot{\partial}\dot{\mathsf{f}}(b) - b\cdot\dot{\partial}\dot{\mathsf{f}}(a) = 0. \tag{6.280}$$

We therefore define the *Lie derivative* $\mathcal{L}_a b$ by

$$\mathcal{L}_a b = a\cdot\partial b - b\cdot\partial a. \tag{6.281}$$

This results in a new tangent vector, and transforms under diffeomorphisms as

$$\mathcal{L}_a b \mapsto \mathcal{L}'_{a'} b' = \mathsf{f}(\mathcal{L}_a b). \tag{6.282}$$

Relations between tangent vectors constructed from the Lie derivative will therefore be unchanged by diffeomorphisms.

A similar construction is possible for cotangent vectors. If we contract equation (6.279) with $\bar{\mathsf{f}}^{-1}(\mathsf{e}^k)$ we obtain

$$\mathsf{f}(\mathsf{e}_j)\cdot\big(\partial_j\bar{\mathsf{f}}^{-1}(\mathsf{e}^k)\big) - \mathsf{f}(\mathsf{e}_i)\cdot\big(\partial_i\bar{\mathsf{f}}^{-1}(\mathsf{e}^k)\big) = 0. \tag{6.283}$$

Now multiplying by $\bar{\mathsf{f}}^{-1}(\mathsf{e}^i \wedge \mathsf{e}^j)$ and summing we find that

$$\mathsf{P}'\big(\bar{\mathsf{f}}^{-1}(\partial) \wedge \bar{\mathsf{f}}^{-1}(\mathsf{e}^k)\big) = 0. \tag{6.284}$$

This result can be summarised simply as

$$D' \wedge \mathsf{e}^{k\prime} = D' \wedge \bar{\mathsf{f}}^{-1}(\mathsf{e}^k) = 0. \tag{6.285}$$

This is sufficient to establish that the exterior derivative of a cotangent vector results in a cotangent bivector (equivalent to a 2-form). The result transforms in the required manner:

$$D \wedge A \mapsto D' \wedge A' = \bar{\mathsf{f}}^{-1}(D \wedge A). \tag{6.286}$$

This is the result that makes the exterior algebra of cotangent vectors so powerful for studying the topological features of manifolds. This algebra is essentially that of differential forms, as is explained in section 6.5.7. For example, a form is said to be closed if its exterior derivative is zero, and to be exact if it can be written as the exterior derivative of a form of one degree lower. Both of these properties are unchanged by diffeomorphisms, so the size of the space of functions that are closed but not exact is a topological feature of a space. This is the basis of de Rham cohomology.

It is somewhat less common to see diffeomorphisms discussed when studying Riemannian geometry. More usually one focuses attention on the restricted class of *isometries*, which are diffeomorphisms that preserve the metric. These define symmetries of a Riemannian space. In the vector manifold setting, however, it is natural to study the effect of maps on metric-dependent quantities. The reason being that vector manifolds inherit their metric structure from the embedding, and if the embedding is changed by a diffeomorphism, the natural metric is changed as well. One does not have to inherit the metric from an embedding. One can easily impose a metric on a vector manifold by defining a linear transformation over the manifold. This takes us into the subject of induced geometries, which is closer to the spirit of the approach to gravity adopted in chapter 14. Similarly, when transforming a vector manifold, one need not insist that the transformed metric is that inherited by the new embedding. One can instead simply define a new metric on the transformed space directly from the original one.

The simplest example of a diffeomorphism inducing a new geometry is to consider a flat plane in three dimensions. If the plane is distorted in the third direction, and the new metric taken as that implied by the embedding, the surface clearly becomes curved. Formulae for the effects of such transformations are generally quite complex. Most can be derived from the transformation properties of the projection operation,

$$\mathsf{P}' = \mathsf{f}\mathsf{P}\mathsf{f}^{-1}. \tag{6.287}$$

This identity ensures that the projection and transformation formulae can be applied in either order. If we now form

$$\begin{aligned}\mathsf{e}'_i\cdot S'_j &= \mathsf{P}'_\perp\big(\partial_j \mathsf{f}(\mathsf{e}_i)\big)\\ &= \mathsf{f}(\mathsf{e}_i\cdot S_j) + \mathsf{P}'_\perp\big(\dot\partial_j \dot{\mathsf{f}}(\mathsf{e}_i)\big),\end{aligned} \tag{6.288}$$

we see that the shape tensor transforms according to

$$a'\cdot S'(b') = \mathsf{f}(a\cdot S(b)) + \mathsf{P}'_\perp\big(b\cdot\dot\partial\dot{\mathsf{f}}(a)\big). \tag{6.289}$$

Further results can be built up from this. For example, the new Riemann tensor is constructed from the commutator of the transformed shape tensor.

6.5.7 Differential geometry and forms

So far we have been deliberately loose in relating objects in vector manifold theory to those of modern differential geometry texts. In this section we clarify the relations and distinctions between the viewpoints. In the subject of differential geometry it is now common practice to identify directional derivatives as tangent vectors, so that the tangent vector a is the scalar operator

$$a = a^i\frac{\partial}{\partial x^i}. \tag{6.290}$$

Tangent vectors form a linear space, denoted $T_x\mathcal{M}$, where x labels a point in the manifold $\mathcal{M}$. This notion of a tangent vector is slightly different from that adopted in the vector manifold theory, where we explicitly let the directional derivative act on the vector x. As explained earlier, the limit implied in writing $\partial x/\partial x^i$ is only well defined if an embedding picture is assumed. The reason for the more abstract definition of a tangent vector in the differential geometry literature is to remove the need for an embedding, so that a topological space can be viewed as a single distinct entity. There are arguments in favour, and against, both viewpoints. For all practical purposes, however, the philosophies behind the two viewpoints are largely irrelevant, and calculations performed in either scheme will return the same results.

The dual space to $T_x\mathcal{M}$ is called the cotangent space and is denoted $T_x^*\mathcal{M}$. Elements of $T_x^*\mathcal{M}$ are called cotangent vectors, or 1-forms. The inner product between a tangent and cotangent vector can be written as $\langle\omega, a\rangle$. A basis for the dual space is defined by the coordinate differentials dx^i, so that

$$\langle dx^i, \partial/\partial x^j\rangle = \delta^i_j. \tag{6.291}$$

A 1-form therefore implicitly contains a directed measure on a manifold. So, if α is a 1-form we have

$$\alpha = \alpha_i dx^i = A\cdot(dx), \tag{6.292}$$

where A is a grade-1 multivector in the vector manifold sense. Similarly, if dX is a directed measure over a two-dimensional surface, we have

$$dX = e_i \wedge e_j \, dx^i \, dx^j, \tag{6.293}$$

so that

$$(e^j \wedge e^i) \cdot dX = dx^i \, dx^j - dx^j \, dx^i. \tag{6.294}$$

An arbitrary 2-form can be written as

$$\alpha_2 = \frac{1}{2!} \alpha_{ij} (dx^i \, dx^j - dx^j \, dx^i) = A_2^\dagger \cdot dX. \tag{6.295}$$

Here A_2 is the multivector

$$A_2 = \frac{1}{2!} \alpha_{ij} \, e^i \wedge e^j, \tag{6.296}$$

which has the same components as the differential form. More generally, an r-form α_r can be written as

$$\alpha_r = A_r^\dagger \cdot dX_r = A_r \cdot dX_r^\dagger. \tag{6.297}$$

Clearly there is little difference in working with the r-form α_r or the equivalent multivector A_r. So, for example, the outer product of two 1-forms results in the 2-form

$$\alpha_1 \wedge \beta_1 = \alpha_i \beta_i (e^i \wedge e^j) \cdot dX_2^\dagger = (A_1 \wedge B_1) \cdot dX_2^\dagger, \tag{6.298}$$

where dX_2 is a two-dimensional surface measure and A_1, B_1 are the grade-1 multivectors with components α_i and β_i respectively. Similarly, the exterior derivative of an r-form is given by

$$d\alpha_r = (D \wedge A_r) \cdot dX_{r+1}^\dagger. \tag{6.299}$$

The fact that forms come packaged with an implicit measure allows for a highly compact statement of Stokes' theorem, as given in equation (6.264). In ultra-compact notation this says that

$$\int_{\sigma_r} d\alpha = \oint_{\partial \sigma_r} \alpha, \tag{6.300}$$

where α is an $(r-1)$-form integrated over an open r-surface σ_r. This is entirely equivalent to equation (6.264), as can be seen by writing

$$\int_{\sigma_r} d\alpha = \int_{\sigma_r} (\dot{A}_{r-1}^\dagger \wedge \dot{D}) \cdot dX_r = \oint_{\partial \sigma_r} (A_{r-1}^\dagger) \cdot dS_{r-1} = \oint_{\partial \sigma_r} \alpha. \tag{6.301}$$

One can proceed in this manner to establish a direct translation scheme between the languages of differential forms and vector manifolds. Many of the expressions are so similar that there is frequently little point in maintaining a distinction.

If the language of differential forms is applied in a metric setting, an important

additional concept is that of a duality transformation, also known as the *Hodge* $*$ (star) operation. To define this we first introduce the volume form

$$\Omega = \sqrt{|g|}dx^1 \wedge dx^2 \wedge \cdots \wedge dx^n = \sqrt{|g|}(e^n \wedge e^{n-1} \wedge \cdots \wedge e^1) \cdot dX. \tag{6.302}$$

The pseudoscalar for a vector manifold, given a coordinate frame with the specified orientation, is given by

$$I = \frac{1}{\sqrt{|g|}}(e_1 \wedge e_2 \wedge \cdots \wedge e_n). \tag{6.303}$$

This definition was chosen earlier to ensure that $I^2 = \pm 1$ and that I keeps the orientation specified by the frame. It follows that

$$\Omega = I^{-1} \cdot dX, \tag{6.304}$$

so that the equivalent multivector is $I^{-1\dagger}$. This will equal $\pm I$, depending on signature. The Hodge $*$ of an r-form α_r is the $(n-r)$-form

$$*\alpha_r = \frac{\sqrt{|g|}}{r!(m-r)!}\omega_{i_1,\ldots,i_r}\epsilon^{i_1,\ldots,i_r}{}_{j_{r+1},\ldots,j_n}\, dx^{j_{r+1}} \wedge \cdots \wedge dx^{j_n}, \tag{6.305}$$

where $\epsilon_{i_1,\ldots,i_n}$ denotes the alternating tensor. If A_r is the multivector equivalent of α_r, the Hodge $*$ takes on the rather simpler expression

$$*A_r = (I^{-1}A_r)^\dagger = (I^{-1} \cdot A_r)^\dagger. \tag{6.306}$$

In effect, we are multiplying by the pseudoscalar, as one would expect for a duality relation. Applied twice we find that

$$**A_r = \left(I^{-1}(I^{-1} \cdot A_r)^\dagger\right)^\dagger = (-1)^{r(m-r)}A_r(I^\dagger I). \tag{6.307}$$

In spaces with Euclidean signature, $I^\dagger I = +1$. In spaces of mixed signature the sign depends on whether there are an even or odd number of basis vectors with negative norm. It is a straightforward exercise to prove the main results for the Hodge $*$ operation, given equation (6.307) and the fact that I is covariantly conserved.

6.6 Elasticity

As a more extended application of some of the ideas developed in this chapter, we discuss the foundations of the subject of elasticity. The behaviour of a solid object is modelled by treating the object as a continuum. Locally, the strains in the object will tend to be small, but these can build up to give large global displacements. As such, it is important to treat the full, non-linear theory of elasticity. Only then can one be sure about the validity of various approximation schemes, such as assuming small deflections.

Our discussion is based on a generalisation of the ideas employed in the treatment of a rigid body. We first introduce an undeformed, *reference* configuration, with points in this labelled with the vector x. This is sometimes referred to as the material configuration. Points in the *spatial* configuration, y, are obtained by a non-linear displacement f of the reference configuration, so that

$$y = y(x,t) = f(x,t). \tag{6.308}$$

We use non-bold vectors to label points in the body, and bold to label tangent vectors in either the reference or spatial body. We assume that the background space is flat, three-dimensional Euclidean space.

6.6.1 Body strains

To calculate the strains in the body, consider the image of the vector between two nearby points in the reference configuration,

$$(x+\epsilon\boldsymbol{a}) - x \mapsto y(x+\epsilon\boldsymbol{a}) - y(x) = \epsilon \mathsf{f}(\boldsymbol{a}) + O(\epsilon^2), \tag{6.309}$$

where f is the deformation gradient,

$$\mathsf{f}(\boldsymbol{a}) = \boldsymbol{a}\cdot\boldsymbol{\nabla} y = \boldsymbol{a}\cdot\boldsymbol{\nabla} f(x,t). \tag{6.310}$$

The function f maps a tangent vector in the reference configuration to the equivalent vector in the spatial configuration. That is, if $x(\lambda)$ is a curve in the reference configuration with tangent vector

$$\boldsymbol{x}' = \frac{\partial x(\lambda)}{\partial \lambda}, \tag{6.311}$$

then the spatial curve has tangent vector $\mathsf{f}(\boldsymbol{v})$. The length of the curve $\boldsymbol{x}(\lambda)$ in the reference configuration is

$$\int \left|\frac{\partial x}{\partial \lambda}\right| d\lambda = \int |\boldsymbol{x}'|\, d\lambda. \tag{6.312}$$

The length of the induced curve in the spatial configuration is therefore

$$\int d\lambda \big(\mathsf{f}(\boldsymbol{x}')^2\big)^{1/2} = \int d\lambda \big(\boldsymbol{x}'\cdot\bar{\mathsf{f}}\mathsf{f}(\boldsymbol{x}')\big)^{1/2}. \tag{6.313}$$

We define the (right) *Cauchy–Green tensor* C, by

$$\mathsf{C}(\boldsymbol{a}) = \bar{\mathsf{f}}\mathsf{f}(\boldsymbol{a}). \tag{6.314}$$

This tensor is a symmetric, positive-definite map between vectors in the reference configuration. It describes a set of positive dilations along the principal directions in the reference configuration. The eigenvalues of C can be written as $(\lambda_1^2, \lambda_2^2, \lambda_3^2)$, where the λ_i define the *principal stretches*. The deviations of these from unity measure the strains in the material.

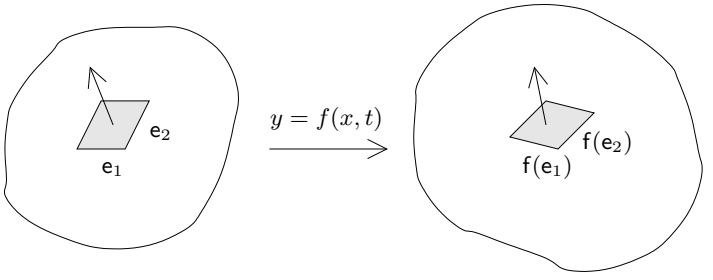

Figure 6.7 *An elastic body.* The function $f(x,t)$ maps points in the reference configuration to points in the spatial configuration. Coordinate curves e_1 and e_2 map to $\mathsf{f}(\mathsf{e}_1)$ and $\mathsf{f}(\mathsf{e}_2)$. The normal vector in the spatial configuration therefore lies in the $\bar{\mathsf{f}}^{-1}(\mathsf{e}^3)$ direction.

6.6.2 Body stresses

If we take a cut through the body then the contact force between the surfaces will be a function of the normal to the surface (and position in the body). Cauchy showed that, under reasonable continuity conditions, this force must be a linear function of the normal, which we write $\boldsymbol{\sigma}(\boldsymbol{n}) = \boldsymbol{\sigma}(\boldsymbol{n}; x)$. The tensor $\boldsymbol{\sigma}(\boldsymbol{n})$ maps a vector normal to a surface in the spatial configuration onto the force vector, also in the spatial configuration. We will verify shortly that $\boldsymbol{\sigma}$ is symmetric.

The total force on a volume segment in the body involves integrating $\boldsymbol{\sigma}(\boldsymbol{n})$ over the surface of the volume. But, as with the rigid body, it is simpler to perform all calculations back in the reference copy. To this end we let x^i denote a set of coordinates for position in the reference body. The associated coordinate frame is $\{\mathsf{e}_i\}$, with reciprocal frame $\{\mathsf{e}^i\}$. Suppose now that x^1 and x^2 are coordinates for a surface in the reference configuration. The equivalent normal in the spatial configuration is (see figure 6.7)

$$\boldsymbol{n} = \mathsf{f}(\mathsf{e}_1) \wedge \mathsf{f}(\mathsf{e}_2)\, I^{-1} = \det\,(\mathsf{f})\, \bar{\mathsf{f}}^{-1}(\mathsf{e}^3). \tag{6.315}$$

The force over this surface is found by integrating the quantity

$$\boldsymbol{\sigma}\big(\mathsf{f}(\boldsymbol{e}_1 \wedge \boldsymbol{e}_2) I^{-1}\big) dx^1\, dx^2 = \det\,(\mathsf{f}) \boldsymbol{\sigma}(\bar{\mathsf{f}}^{-1}(\boldsymbol{e}^3)) dx^1\, dx^2. \tag{6.316}$$

We therefore define the *first Piola–Kirchoff stress tensor* T by

$$\mathsf{T}(\boldsymbol{a}) = \det\,(\mathsf{f}) \boldsymbol{\sigma} \bar{\mathsf{f}}^{-1}(\boldsymbol{a}). \tag{6.317}$$

The stress tensor T takes as its argument a vector normal to a surface in the reference configuration, and returns the contact force in the spatial body. The

force balance equation tells us that, for any sub-body, we have

$$\frac{d}{dt}\int d^3x\,\rho\boldsymbol{v} = \oint \mathsf{T}(d\boldsymbol{s}) + \int d^3x\,\rho\boldsymbol{b}, \tag{6.318}$$

where ρ is the density in the reference configuration, $\boldsymbol{v} = \dot{y}$ is the spatial velocity, and $\boldsymbol{b}$ is the applied *body force*. The fundamental theorem immediately converts this to the local equation

$$\rho\dot{\boldsymbol{v}} = \check{\mathsf{T}}(\check{\boldsymbol{\nabla}}) + \rho\boldsymbol{b}. \tag{6.319}$$

The check symbol is used for the scope of the derivative, to avoid confusion with time derivatives (denoted with an overdot). This equation is sensible as $\boldsymbol{\nabla}$ is the vector derivative in the reference configuration, and $\check{\mathsf{T}}(\check{\boldsymbol{\nabla}})$ is a vector in the spation configuration.

The total torque on a volume element, centred on y_0, is (ignoring body forces)

$$M = \oint (y - y_0)\wedge\mathsf{T}(d\boldsymbol{s}). \tag{6.320}$$

This integral runs over the reference body, and returns a torque in the spatial configuration. This must be equated with the rate of change of angular momentum, which is

$$\begin{aligned}\frac{d}{dt}\int d^3x\,\rho(y - y_0)\wedge\dot{y} &= \int d^3x\,(y - y_0)\wedge\check{\mathsf{T}}(\check{\boldsymbol{\nabla}}) \\ &= \oint (y - y_0)\wedge\mathsf{T}(d\boldsymbol{s}) - \int d^3x\,\check{y}\wedge\mathsf{T}(\check{\boldsymbol{\nabla}}).\end{aligned} \tag{6.321}$$

Equating this with M we see that

$$\check{y}\wedge\mathsf{T}(\check{\boldsymbol{\nabla}}) = (\partial_i f(x))\wedge\mathsf{T}(\mathsf{e}^i) = \mathsf{f}(\mathsf{e}_i)\wedge\mathsf{T}(\mathsf{e}^i) = 0. \tag{6.322}$$

It follows that

$$\mathsf{f}(\mathsf{e}_i)\wedge\mathsf{T}(\mathsf{e}^i) = \det(\mathsf{f})\,\mathsf{f}(\mathsf{e}_i)\wedge\boldsymbol{\sigma}\bar{\mathsf{f}}^{-1}(\mathsf{e}^i) = 0, \tag{6.323}$$

and we see that $\boldsymbol{\sigma}$ must be a symmetric tensor in order for angular momentum to be conserved.

It is often convenient to work with a version of T that is symmetric and defined entirely in the material frame. We therefore define the *second Piola–Kirchoff stress tensor* $\mathcal{T}$ by

$$\mathcal{T}(\boldsymbol{a}) = \mathsf{f}^{-1}\mathsf{T}(\boldsymbol{a}). \tag{6.324}$$

It is meaningless to talk about symmetries of T, since it maps between different spaces, whereas $\mathcal{T}$ is defined entirely in the reference configuration and, by construction, is symmetric.

The equations of motion for an elastic material are completed by defining a constitutive relation. This relates the stresses to the strains in the body. These relations are most easily expressed in the reference copy as a relationship between

$\mathcal{T}$ and C. There is no universal definition of the strain tensor $\mathcal{E}$, though for certain applications a useful definition is

$$\mathcal{E}(\boldsymbol{a}) = \mathsf{C}^{1/2}(\boldsymbol{a}) - \boldsymbol{a}. \tag{6.325}$$

This tensor is zero if the material is undeformed. Linear materials have the property that $\mathcal{T}$ and $\mathcal{E}$ are linearly related by a rank-4 tensor. This can, in principle, have 36 independent degrees of freedom, all of which may need to be determined experimentally. If the material is homogeneous then the components of the rank-4 tensor are constants. If the material is also isotropic then the 36 degrees of freedom reduce to two. These are usually given in terms of the bulk modulus B and shear modulus G, with $\mathcal{T}$ and $\mathcal{E}$ related by an expression of the form

$$\mathcal{T}(\boldsymbol{a}) = 2G\mathcal{E}(\boldsymbol{a}) + (B - \tfrac{2}{3}G)\mathrm{tr}(\mathcal{E})\boldsymbol{a}. \tag{6.326}$$

In many respects this is the simplest material one can consider, though even in this case the non-linearity of the force law makes the full equations very hard to analyse. The analysis can be aided by the fact that these materials are described by an action principle, as discussed in section 12.4.1.

6.7 Notes

The treatment of vector manifolds presented here is a condensed version of the theory developed by Hestenes & Sobczyk in the book *Clifford Algebra to Geometric Calculus* (1984) and in a series of papers by Garret Sobczyk. There are a number of differences in our presentation, however. Most significant is our definition of the orientations in the fundamental theorem of integral calculus. Our definition of the boundary operator ensures that a boundary inherits its orientation from the directed volume measure. Hestenes & Sobczyk used the opposite specification for their boundary operator, which gives rise to a number of (fairly trivial) differences. A significant advantage of our conventions is that in two dimensions the pseudoscalar has the correct orientation implied by the imaginary in the Cauchy integral formula.

A further difference is that from the outset we have emphasised both the implied embedding of a vector manifold, and the fact that this gives rise to a metric. A vector manifold thus has greater structure than a differentiable manifold in the sense of differential geometry. For applications to finite-dimensional Riemannian geometry the different approaches are entirely equivalent, as any finite-dimensional Riemannian manifold can be embedded in a larger dimensional flat space in such a way that the metric is generated by the embedding. This result was proved by John Nash in 1956. His remarkable story is the subject of the book *A Beautiful Mind* by Sylvia Nasar (1998) and, more recently, a film of the same name. In other applications of differential geometry the full range of validity of the vector manifold approach has yet to be fully established. The

approach certainly does give streamlined proofs of a number of key results. But whether this comes with some loss of generality is an open question.

A final, small difference in our approach here to the original one of Hestenes & Sobczyk is our definition of the shape tensor. We have only considered the shape tensor $S(a)$ taking intrinsic vectors as its linear argument. This concept can be generalised to define a function that can act linearly on general vectors. One of the most interesting properties of this generalized version of the shape tensor is that it provides a natural square root of the Ricci tensor. This theory is developed in detail in chapter 5 of *Clifford Algebra to Geometric Calculus*, to which readers are referred for further information. There is no shortage of good textbooks on modern differential geometry. The books by Nakahara (1990), Schutz (1980) and Göckeler & Schucker (1987) are particularly strong on emphasising physical applications. Elasticity is described in the books by Marsden & Hughes (1994) and Antman (1995).

6.8 Exercises

6.1 Confirm that the vector derivative is independent of the choice of coordinate system.

6.2 If we denote the curl of a vector field $\boldsymbol{J}$ in three dimensions by $\boldsymbol{\nabla}\times\boldsymbol{J}$, show that

$$\boldsymbol{\nabla}\times\boldsymbol{J} = -I\boldsymbol{\nabla}\wedge J.$$

Hence prove that

$$\begin{aligned}\boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}\times\boldsymbol{J}) &= 0,\\ \boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\boldsymbol{J}) &= \boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\boldsymbol{J}) - \boldsymbol{\nabla}^2\boldsymbol{J}.\end{aligned}$$

6.3 An oblate spheroidal coordinate system can be defined by

$$\begin{aligned}a\cosh(u)\sin(v) &= \sqrt{(x^2+y^2)},\\ a\sinh(u)\cos(v) &= z,\\ \tan(\phi) &= y/x,\end{aligned}$$

where (x, y, z) denote standard Cartesian coordinates and a is a scalar. Prove that

$$\mathsf{e}_u^2 = \mathsf{e}_v^2 = a^2\big(\sinh^2(u)+\cos^2(v)\big) = \rho^2,$$

which defines the quantity ρ. Hence prove that the Laplacian becomes

$$\begin{aligned}\boldsymbol{\nabla}^2\psi = &\frac{1}{\rho^2\cosh(u)}\frac{\partial}{\partial u}\left(\cosh(u)\frac{\partial\psi}{\partial u}\right) + \frac{1}{\rho^2\sin(v)}\frac{\partial}{\partial v}\left(\sin(v)\frac{\partial\psi}{\partial v}\right)\\ &+ \frac{1}{a^2\cosh^2(u)\,\sin^2(v)}\frac{\partial^2\psi}{\partial\phi^2},\end{aligned}$$

and investigate the properties of separable solutions in oblate spheroidal coordinates.

6.4 Prove that over the surface of a tetrahedron the directed surface integral satisfies

$$\oint dS = 0.$$

By considering pairs of adjacent tetrahedra, prove that this integral vanishes for all orientable, connected closed surfaces.

6.5 For a circle in a plane confirm that the line integral around the perimeter satisfies

$$\oint b\cdot x\, dl = b\cdot A,$$

where A is the oriented area of the circle.

6.6 Prove that

$$\sum_{i=0}^{k}(-1)^i b\cdot(x_0+\cdots\check{x}_i\cdots+x_n)\Delta(\check{x}_i)_{(k-1)} = \frac{1}{k!}b\cdot(e_1\wedge\cdots\wedge e_n),$$

where the notation follows section 6.4.4.

6.7 Suppose that σ is an n-dimensional surface embedded in a flat space of dimensions $n+1$ with (constant) unit pseudoscalar I. Prove that

$$\oint_{\partial\sigma} dSJ = -I\int_\sigma l\wedge\nabla J\,|dX|,$$

where the normal l is defined by $dX = Il\,|dX|$.

6.8 The shape tensor is defined by

$$a\cdot\partial I = IS(a) = I\times S(a).$$

Prove that the shape tensor satisfies

$$a\cdot S(b) = b\cdot S(a)$$

and

$$\dot\partial\wedge\dot{\mathsf{P}}(a) = S(a),$$

where P projects into the tangent space, and a and b are tangent vectors.

6.9 An open two-dimensional surface in three-dimensional space is defined by

$$\boldsymbol{r}(x,y) = x\mathsf{e}_1 + y\mathsf{e}_2 + \alpha(r)\mathsf{e}_3,$$

where $r = (x^2+y^2)^{1/2}$ and the $\{\mathsf{e}_i\}$ are a standard Cartesian frame. Prove that the Riemann tensor can be written

$$\mathsf{R}(a\wedge b) = \frac{\alpha'\alpha''}{r(1+\alpha'^2)^2}\,a\wedge b,$$

where the primes denote differentiation with respect to r. The scalar factor κ in $\mathsf{R}(a\wedge b) = \kappa a\wedge b$ is called the *Gaussian curvature*.

6.10 A linear, isotropic, homogeneous material is described by a bulk modulus B and shear modulus G. By linearising the elasticity equations, show that the longitudinal and transverse sound speeds v_l and v_t are given by

$$v_l^2 = \frac{1}{3\rho}\big(3B + 4G\big), \qquad v_t^2 = \frac{G}{\rho}.$$

6.11 Consider an infinite linear, isotropic, homogeneous material containing a spherical hole into which air is pumped. Show that, in the linearised theory, the radial stress τ_r is related to the radius of the hole r by $\tau_r \propto r^{-3}$. Discuss how the full non-linear theory might modify this result.

7

Classical electrodynamics

Geometric algebra offers a number of new techniques for studying problems in electromagnetism and electrodynamics. These are described in this chapter. We will not attempt a thorough development of electrodynamics, which is a vast subject with numerous specialist areas. Instead we concentrate on a number of selected applications which highlight the advantages that geometric algebra can bring. There are two particularly significant new features that geometric algebra adds to traditional formulations of electrodynamics. The first is that, through employing the spacetime algebra, all equations can be studied in the appropriate spacetime setting. This is much more transparent than the more traditional approach based on a $3+1$ formulation involving retarded times. The spacetime algebra simplifies the study of how electromagnetic fields appear to different observers, and is particularly powerful for handling accelerated charges and radiation. These results build on the applications of spacetime algebra described in section 5.5.3.

The second major advantage of the geometric algebra treatment is a new, compact formulation of Maxwell's equations. The spacetime vector derivative and the geometric product enable us to unite all four of Maxwell's equations into a single equation. This is one of the most impressive results in geometric algebra. And, as we showed in chapter 6, this is more than merely a cosmetic exercise. The vector derivative is invertible directly, without having to pass via intermediate, second-order equations. This has many implications for scattering and propagator theory. Huygen's principle is encoded directly, and the first-order theory is preferable for numerical computation of diffraction effects. In addition, the first-order formulation of electromagnetism means that plane waves are easily handled, as are their polarisation states.

7.1 Maxwell's equations

Before writing down the Maxwell equations, we remind ourselves of the notation introduced in chapter 5. We denote an orthonormal spacetime frame by $\{\gamma_\mu\}$, with coordinates $x_\mu = \gamma_\mu \cdot x$. The spacetime vector derivative is

$$\nabla = \gamma^\mu \partial_\mu, \qquad \partial_\mu = \frac{\partial}{\partial x^\mu}. \tag{7.1}$$

The spacetime split of the vector derivative is

$$\nabla\gamma_0 = (\gamma^0\partial_t + \gamma^i\partial_i)\gamma_0 = \partial_t - \boldsymbol{\sigma}_i\partial_i = \partial_t - \boldsymbol{\nabla}, \tag{7.2}$$

where the $\boldsymbol{\sigma}_i = \gamma_i\gamma_0$ denote a right-handed orthonormal frame for the relative space defined by the timelike vector γ_0. The three-dimensional vector derivative operator is

$$\boldsymbol{\nabla} = \boldsymbol{\sigma}_i \frac{\partial}{\partial x^i} = \boldsymbol{\sigma}_i \partial_i, \tag{7.3}$$

and all relative vectors are written in bold.

The four Maxwell equations, in SI units, are

$$\begin{aligned} \boldsymbol{\nabla}\cdot\boldsymbol{D} &= \rho, & \boldsymbol{\nabla}\cdot\boldsymbol{B} &= 0, \\ -\boldsymbol{\nabla}\times\boldsymbol{E} &= \frac{\partial}{\partial t}\boldsymbol{B}, & \boldsymbol{\nabla}\times\boldsymbol{H} &= \frac{\partial}{\partial t}\boldsymbol{D} + \boldsymbol{J}, \end{aligned} \tag{7.4}$$

where

$$\begin{aligned} \boldsymbol{D} &= \epsilon_0\boldsymbol{E} + \boldsymbol{P}, \\ \boldsymbol{H} &= \frac{1}{\mu_0}\boldsymbol{B} - \boldsymbol{M}, \end{aligned} \tag{7.5}$$

and the $\times$ symbol denotes the vector cross product. The cross product is ubiquitous in electromagnetic theory, and it will be encountered at various points in this chapter. To avoid any confusion, the commutator product (denoted by $\times$) will not be employed in this chapter.

The first step in simplifying the Maxwell equations is to assume that we are working in a vacuum region outside isolated sources and currents. We can then remove the polarisation and magnetisation fields $\boldsymbol{P}$ and $\boldsymbol{M}$. We also replace the cross product with the exterior product, and revert to natural units ($c = \epsilon_0 = \mu_0 = 1$), so that the equations now read

$$\begin{aligned} \boldsymbol{\nabla}\cdot\boldsymbol{E} &= \rho, & \boldsymbol{\nabla}\cdot\boldsymbol{B} &= 0, \\ \boldsymbol{\nabla}\wedge\boldsymbol{E} &= -\partial_t(I\boldsymbol{B}), & \boldsymbol{\nabla}\wedge\boldsymbol{B} &= I(\boldsymbol{J} + \partial_t\boldsymbol{E}). \end{aligned} \tag{7.6}$$

We naturally assemble equations for the separate divergence and curl parts of the vector derivative. We know that there are many advantages in uniting these

into a single equation involving the vector derivative. First we take the two equations for $\boldsymbol{E}$ and combine them into the single equation

$$\boldsymbol{\nabla}\boldsymbol{E} = \rho - \partial_t(I\boldsymbol{B}). \tag{7.7}$$

A similar manipulation combines the $\boldsymbol{B}$-field equations into

$$\boldsymbol{\nabla}(I\boldsymbol{B}) = -\boldsymbol{J} - \partial_t\boldsymbol{E}, \tag{7.8}$$

where we have multiplied through by I. This equation is a combination of (spatial) bivector and pseudoscalar terms, whereas equation (7.7) contains only scalar and vector parts. It follows that we can combine all of these equations into the single multivector equation

$$\boldsymbol{\nabla}(\boldsymbol{E} + I\boldsymbol{B}) + \partial_t(\boldsymbol{E} + I\boldsymbol{B}) = \rho - \boldsymbol{J}. \tag{7.9}$$

This is already a significant compactification of the original equations. We have not lost any information in writing this, since each of the separate Maxwell equations can be recovered by picking out terms of a given grade.

In section 5.5.3 we introduced the Faraday bivector F. This represents the *electromagnetic field strength* and is defined by

$$F = \boldsymbol{E} + I\boldsymbol{B}. \tag{7.10}$$

The combination of relative vectors and bivectors tells us that this quantity is a spacetime bivector. Many authors have noticed that the Maxwell equations can be simplified if expressed in terms of the complex quantity $\boldsymbol{E} + i\boldsymbol{B}$. The reason is that the spacetime pseudoscalar has negative square, so can be represented by the unit imaginary for certain applications. It is important, however, to work with I in the full spacetime setting, as I anticommutes with spacetime vectors.

In terms of the field strength the Maxwell equations reduce to

$$\boldsymbol{\nabla}F + \partial_t F = \rho - \boldsymbol{J}. \tag{7.11}$$

We now wish to convert this to manifestly Lorentz covariant form. We introduce the spacetime current J, which has

$$\rho = J\cdot\gamma_0, \qquad \boldsymbol{J} = J\wedge\gamma_0. \tag{7.12}$$

It follows that

$$\rho - \boldsymbol{J} = \gamma_0\cdot J + \gamma_0\wedge J = \gamma_0 J. \tag{7.13}$$

But we know that $\partial_t + \boldsymbol{\nabla} = \gamma_0\nabla$. We can therefore pre-multiply equation (7.11) by γ_0 to assemble the covariant equation

$$\nabla F = J. \tag{7.14}$$

This unites all four Maxwell equations into a single spacetime equation based on

the *geometric* product with the vector derivative. An immediate consequence is seen if we multiply through by ∇, giving

$$\nabla^2 F = \nabla J = \nabla\cdot J + \nabla\wedge J. \tag{7.15}$$

Since ∇^2 is a scalar operator, the left-hand side can only contain bivector terms. It follows that the current J must satisfy the conservation equation

$$\nabla\cdot J = \frac{\partial\rho}{\partial t} + \boldsymbol{\nabla}\cdot\boldsymbol{J} = 0. \tag{7.16}$$

This equation tells us that the total charge generating the fields must be conserved.

The equation $\nabla F = J$ separates into a pair of spacetime equations for the vector and trivector parts,

$$\nabla\cdot F = J, \qquad \nabla\wedge F = 0. \tag{7.17}$$

In tensor language, these correspond to the pair of spacetime equations

$$\partial_\mu F^{\mu\nu} = J^\nu, \qquad \epsilon^{\mu\nu\rho\sigma}\partial_\nu F_{\rho\sigma} = 0. \tag{7.18}$$

These two tensor equations are as compact a formulation of the Maxwell equations as tensor algebra can achieve, and the same is true of differential forms. Only geometric algebra enables us to combine the Maxwell equations (7.17) into the single equation $\nabla F = J$.

7.1.1 The vector potential

The fact that $\nabla\wedge F = 0$ tells us that we can introduce a vector field A such that

$$F = \nabla\wedge A. \tag{7.19}$$

The equation $\nabla\wedge F = \nabla\wedge\nabla\wedge A = 0$ then follows automatically. The field A is known as the *vector potential*. We shall see in later chapters that the vector potential is key to the quantum theory of how matter interacts with radiation. The vector potential is also the basis for the Lagrangian treatment of electromagnetism, described in chapter 12.

The remaining source equation tells us that the vector potential satisfies

$$\nabla\cdot(\nabla\wedge A) = \nabla^2 A - \nabla(\nabla\cdot A) = J. \tag{7.20}$$

There is some residual freedom in A beyond the restriction of equation (7.19). We can always add the gradient of a scalar field to A, since

$$\nabla\wedge(A + \nabla\lambda) = \nabla\wedge A + \nabla\wedge(\nabla\lambda) = F. \tag{7.21}$$

For historical reasons, this ability to alter A is referred to as a *gauge* freedom.

Before we can solve the equations for A, we must therefore specify a gauge. A natural way to absorb this freedom is to impose the *Lorentz condition*

$$\nabla \cdot A = 0. \tag{7.22}$$

This does not totally specify A, as the gradient of a solution of the wave equation can still be added, but this remaining freedom can be removed by imposing appropriate boundary conditions. The Lorentz gauge condition implies that $F = \nabla A$. We then recover a wave equation for the components of A, since

$$\nabla F = \nabla^2 A = J. \tag{7.23}$$

One route to solving the Maxwell equations is to solve the associated wave equation $\nabla^2 A = J$, with appropriate boundary conditions applied, and then compute F at the end. In this chapter we explore alternative, more direct routes.

The fact that a gauge freedom exists in the formulation in terms of A suggests that some *conjugate* quantity should be conserved. This is the origin of the current conservation law derived in equation (7.16). Conservation of charge is therefore intimately related to gauge invariance. A more detailed understanding of this will be provided by the Lagrangian framework.

7.1.2 The electromagnetic field strength

In uniting the Maxwell equations we introduced the *electromagnetic field strength* $F = \boldsymbol{E} + I\boldsymbol{B}$. This is a covariant spacetime bivector. Its components in the $\{\gamma^\mu\}$ frame give rise to the tensor

$$F^{\mu\nu} = \gamma^\nu \cdot (\gamma^\mu \cdot F) = (\gamma^\nu \wedge \gamma^\mu) \cdot F. \tag{7.24}$$

These are the components of a rank-2 antisymmetric tensor which, written out as a matrix, has entries

$$F^{\mu\nu} = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & -B_z & B_y \\ E_y & B_z & 0 & -B_x \\ E_z & -B_y & B_x & 0 \end{pmatrix}. \tag{7.25}$$

This matrix form of the field strength is often presented in textbooks on relativistic electrodynamics. It has a number of disadvantages. Amongst these are that Lorentz transformations cannot be handled elegantly and the natural complex structure is hidden.

Writing $F = \boldsymbol{E} + I\boldsymbol{B}$ decomposes F into the sum of a relative vector $\boldsymbol{E}$ and a relative bivector $I\boldsymbol{B}$. The separate $\boldsymbol{E}$ and $I\boldsymbol{B}$ fields are recovered from

$$\begin{aligned} \boldsymbol{E} &= \tfrac{1}{2}(F - \gamma_0 F \gamma_0), \\ I\boldsymbol{B} &= \tfrac{1}{2}(F + \gamma_0 F \gamma_0). \end{aligned} \tag{7.26}$$

This shows clearly how the split into $\boldsymbol{E}$ and $I\boldsymbol{B}$ fields depends on the observer velocity (γ_0 here). Observers in relative motion see different fields. For example, suppose that a second observer has velocity $v = R\gamma_0\tilde{R}$ and constructs the rest frame basis vectors

$$\gamma'_\mu = R\gamma_\mu\tilde{R}. \tag{7.27}$$

This observer measures components of an electric field to be

$$E'_i = (\gamma'_i\gamma'_0)\cdot F = (R\boldsymbol{\sigma}_i\tilde{R})\cdot F = \boldsymbol{\sigma}_i\cdot(\tilde{R}FR). \tag{7.28}$$

The effect of a Lorentz transformation can therefore be seen by taking F to $\tilde{R}FR$. The fact that bivectors are subject to the same rotor transformation law as vectors is extremely useful for computations.

Suppose now that two observers measure the F-field at a point. One has 4-velocity γ_0, and the other is moving at relative velocity $\boldsymbol{v}$ in the γ_0 frame. This observer has 4-velocity

$$v = R\gamma_0\tilde{R}, \qquad R = \exp(\alpha\hat{\boldsymbol{v}}/2), \tag{7.29}$$

where $\boldsymbol{v} = \tanh(\alpha)\hat{\boldsymbol{v}}$. The second observer measures the $\{\gamma_\mu\}$ components of $\tilde{R}FR$. To find these we decompose F into terms parallel and perpendicular to $\boldsymbol{v}$,

$$F = F_\parallel + F_\perp, \tag{7.30}$$

where

$$\boldsymbol{v}F_\parallel = F_\parallel\boldsymbol{v}, \qquad \boldsymbol{v}F_\perp = -F_\perp\boldsymbol{v}. \tag{7.31}$$

We quickly see that the parallel components are unchanged, but the perpendicular components transform to

$$\tilde{R}F_\perp R = \exp(-\alpha\hat{\boldsymbol{v}})F_\perp = \gamma(1-\boldsymbol{v})F_\perp, \tag{7.32}$$

where γ is the Lorentz factor $(1-\boldsymbol{v}^2)^{-1/2}$. This result is sufficient to immediately establish the transformation law

$$\begin{aligned} \boldsymbol{E}'_\perp &= \gamma(\boldsymbol{E} + \boldsymbol{v}\times\boldsymbol{B})_\perp, \\ \boldsymbol{B}'_\perp &= \gamma(\boldsymbol{B} - \boldsymbol{v}\times\boldsymbol{E})_\perp. \end{aligned} \tag{7.33}$$

Here the primed vectors are formed from $\boldsymbol{E}' = E'_i\boldsymbol{\sigma}_i$, for example. These have the components of F in the new frame, but combined with the original basis vectors.

Further useful information about the F field is contained in its square, which defines a pair of Lorentz-invariant terms. We form

$$F^2 = \langle FF\rangle + \langle FF\rangle_4 = a_0 + Ia_4, \tag{7.34}$$

which is easily seen to be Lorentz-invariant,

$$(\tilde{R}FR)(\tilde{R}FR) = \tilde{R}FFR = a_0 + Ia_4. \tag{7.35}$$

Both the scalar and pseudoscalar terms are independent of the frame in which they are measured. In the γ_0 frame these are

$$\alpha = \langle(\boldsymbol{E} + I\boldsymbol{B})(\boldsymbol{E} + I\boldsymbol{B})\rangle = \boldsymbol{E}^2 - \boldsymbol{B}^2 \tag{7.36}$$

and

$$\beta = -\langle I(\boldsymbol{E} + I\boldsymbol{B})(\boldsymbol{E} + I\boldsymbol{B})\rangle = 2\boldsymbol{E}\cdot\boldsymbol{B}. \tag{7.37}$$

The former yields the Lagrangian density for the electromagnetic field. The latter is seen less often. It is perhaps surprising that $\boldsymbol{E}\cdot\boldsymbol{B}$ is a full Lorentz invariant, rather than just being invariant under rotations.

7.1.3 Dielectric and magnetic media

The Maxwell equations inside a medium, with polarisation and magnetisation fields $\boldsymbol{P}$ and $\boldsymbol{M}$, were given in equation (7.4). These separate into a pair of spacetime equations. We introduce the spacetime bivector field G by

$$G = \boldsymbol{D} + I\boldsymbol{H}. \tag{7.38}$$

Maxwell's equations are now given by the pair of equations

$$\begin{aligned} \nabla\wedge F &= 0, \\ \nabla\cdot G &= J. \end{aligned} \tag{7.39}$$

The first tells us that F has vanishing curl, so can still be obtained from a vector potential, $F = \nabla\wedge A$. The second equation tells us how the $\boldsymbol{D}$ and $\boldsymbol{H}$ fields respond to the presence of free sources. These equations on their own are insufficient to fully describe the behaviour of electromagnetic fields in matter. They must be augmented by constitutive relations which relate F and G. The simplest examples of these are for linear, isotropic, homogeneous materials, in which case the constitutive relations amount to specifying a relative permittivity ϵ_r and permeability μ_r. The fields are then related by

$$\boldsymbol{D} = \epsilon_r\boldsymbol{E}, \qquad \boldsymbol{B} = \mu_r\boldsymbol{H}. \tag{7.40}$$

More complicated models for matter can involve considering responses to different frequencies, and the presence of preferred directions on the material. The subject of suitable constitutive relations is one of heuristic model building. We are, in effect, seeking models which account for the quantum properties of matter in bulk, without facing the full multiparticle quantum equations.

7.2 Integral and conservation theorems

A number of important integral theorems exist in electromagnetism. Indeed, the subject of integral calculus was largely shaped by considering applications to electromagnetism. Here the results are all derived as examples of the fundamental theorem of integral calculus, derived in chapter 6.

7.2.1 Static fields

We start by deriving a number of results for static field configurations. When the fields are static the Maxwell equations reduce to the pair

$$\boldsymbol{\nabla} \boldsymbol{E} = \frac{\rho}{\epsilon_0}, \qquad \boldsymbol{\nabla} \boldsymbol{B} = \mu_0 I \boldsymbol{J}, \tag{7.41}$$

where (for this section) we have reinserted the constants ϵ_0 and μ_0. A current $\boldsymbol{J}$ is static if the charge flows at a constant rate. The fact that $\boldsymbol{\nabla} \wedge \boldsymbol{E} = 0$ implies that around any closed path

$$\oint_{\partial\sigma} \boldsymbol{E} \cdot d\boldsymbol{l} = 0, \tag{7.42}$$

which applies for all static configurations. We can therefore introduce a potential ϕ such that

$$\boldsymbol{E} = -\boldsymbol{\nabla}\phi. \tag{7.43}$$

The potential ϕ is the timelike component of the vector potential A, $\phi = \gamma_0 \cdot A$. One can formulate many of the main results of electrostatics directly in terms of ϕ. Here we adopt a different approach and work directly with the $\boldsymbol{E}$ and $\boldsymbol{B}$ fields.

An extremely important integral theorem is a straightforward application of Gauss' law (indeed this *is* Gauss' original law)

$$\oint_{\partial V} \boldsymbol{E} \cdot \boldsymbol{n} \, |dA| = \frac{1}{\epsilon_0} \int_V \rho \, |dX| = \frac{Q}{\epsilon_0}, \tag{7.44}$$

where Q is the enclosed charge. In this formula $\boldsymbol{n}$ is the outward pointing normal, formed from $d\boldsymbol{A} = I\boldsymbol{n}|dA|$, where $d\boldsymbol{A}$ is the directed measure over the surface, and the scalar measure $|dX|$ is simply

$$|dX| = dx \, dy \, dz. \tag{7.45}$$

For the next application, recall from section 6.4.7 the form of the Green's function for the vector derivative in three dimensions,

$$\boldsymbol{G}(\boldsymbol{r}; \boldsymbol{r}') = \frac{1}{4\pi} \frac{\boldsymbol{r} - \boldsymbol{r}'}{|\boldsymbol{r} - \boldsymbol{r}'|^3}. \tag{7.46}$$

An application of the fundamental theorem tells us that

$$\int_V (\dot{\boldsymbol{G}}\dot{\boldsymbol{\nabla}}\boldsymbol{E} + \boldsymbol{G}\,\boldsymbol{\nabla}\boldsymbol{E})|dX| = -I\oint_{\partial V}\boldsymbol{G}\,d\boldsymbol{A}\,\boldsymbol{E}. \tag{7.47}$$

If we assume that the sources are localised, so that $\boldsymbol{E}$ falls off at large distance, we can take the integral over all space and the right-hand side will vanish. Replacing $\boldsymbol{G}$ by the Green's function above we find that the field from a static charge distribution is given by

$$\boldsymbol{E}(\boldsymbol{r}) = \frac{1}{4\pi\epsilon_0}\int \frac{\rho(\boldsymbol{r}')(\boldsymbol{r}-\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|^3}\,|dX'|. \tag{7.48}$$

If ρ is a single δ-function source, $\rho = Q\delta(\boldsymbol{r}'-\boldsymbol{r}_0)$, we immediately recover the Coulomb field

$$\boldsymbol{E}(\boldsymbol{r}) = \frac{Q}{4\pi\epsilon_0}\frac{(\boldsymbol{r}-\boldsymbol{r}_0)}{|\boldsymbol{r}-\boldsymbol{r}_0|^3}. \tag{7.49}$$

Unsurprisingly, this is simply a weighted Green's function.

For the magnetic field $\boldsymbol{B}$, the absence of magnetic monopoles is encoded in the integral equation

$$\oint \boldsymbol{B}\cdot d\boldsymbol{A} = 0. \tag{7.50}$$

This tells us that the integral curves of $\boldsymbol{B}$ always form closed loops. This is true both inside and outside matter, and holds in the time-dependent case as well. Next we apply the integral theorem of equation (7.47) with $\boldsymbol{E}$ replaced by $\boldsymbol{B}$. If we again assume that the fields are produced by localised charges and fall off at large distances, we derive

$$I\boldsymbol{B}(\boldsymbol{r}) = -\frac{\mu_0}{4\pi}\int \frac{(\boldsymbol{r}-\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|^3}\boldsymbol{J}(\boldsymbol{r}')\,|dX'|. \tag{7.51}$$

The scalar term in the integrand vanishes as a consequence of the static conservation law $\boldsymbol{\nabla}\cdot\boldsymbol{J} = 0$. The bivector term gives the magnetic field bivector $I\boldsymbol{B}$. Now suppose that the current is carried entirely in an 'ideal' wire. This is taken as an infinitely thin wire carrying a current J,

$$\boldsymbol{J} = J\int d\lambda\,\frac{d\boldsymbol{y}(\lambda)}{d\lambda}\delta\big(\boldsymbol{r}-\boldsymbol{y}(\lambda)\big) = J\int d\boldsymbol{l}\,\delta\big(\boldsymbol{r}-\boldsymbol{y}(\lambda)\big). \tag{7.52}$$

We have little option but to use J for the current as the more standard symbol I is already taken for the pseudoscalar. The result is that the $\boldsymbol{B}$-field is determined by a line integral along the wire. This is the Biot–Savart law, which can be written

$$\boldsymbol{B}(\boldsymbol{r}) = \frac{\mu_0 J}{4\pi}\int \frac{d\boldsymbol{l}'\times(\boldsymbol{r}-\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|^3}, \tag{7.53}$$

where $\boldsymbol{r}'$ is the position vector to the line element $d\boldsymbol{l}'$.

A further integral theorem for magnetic fields is found if we consider the integral around a loop enclosing a surface σ. We have

$$\oint_{\partial\sigma} \boldsymbol{B}\cdot d\boldsymbol{l} = \int_\sigma (\dot{\boldsymbol{B}}\wedge\dot{\boldsymbol{\nabla}})\cdot d\boldsymbol{A} = \mu_0 \int_\sigma \boldsymbol{J}\cdot(-I\, d\boldsymbol{A}). \tag{7.54}$$

Again, we write $d\boldsymbol{A} = I\boldsymbol{n}|dA|$, where $\boldsymbol{n}$ is the unit right-handed normal. That is, if we grip the surface in our right hands in the manner specified by the line integral, our thumbs point in the normal direction. The result is that we integrate $\boldsymbol{J}\cdot\boldsymbol{n}$ over the surface. This returns the total current through the loop, J, recovering Ampère's law,

$$\oint_{\partial\sigma} \boldsymbol{B}\cdot d\boldsymbol{l} = \mu_0 J. \tag{7.55}$$

This is routinely used for finding the magnetic fields surrounding electrical circuits.

7.2.2 Time-varying fields

If the fields vary in time, some of the preceding formulae remain valid, and others only require simple modifications. The two applications of Gauss' law, equations (7.44) and (7.50), remain unchanged. The two applications of Stokes' theorem acquire an additional term. For the $\boldsymbol{E}$-field we have

$$\oint_{\partial\sigma} \boldsymbol{E}\cdot d\boldsymbol{l} = \frac{d}{dt}\int_\sigma (I\boldsymbol{B})\cdot d\boldsymbol{A} = -\frac{d\Phi}{dt}, \tag{7.56}$$

where Φ is the linked magnetic flux. The flux is the integral of $\boldsymbol{B}\cdot\boldsymbol{n}$ over the area enclosed by the loop, with $\boldsymbol{n}$ the unit normal. Magnetic flux is an important concept for understanding inductance in circuits.

For the magnetic field we can derive a similar formula,

$$\oint_{\partial\sigma} \boldsymbol{B}\cdot d\boldsymbol{l} = \mu_0 J + \epsilon_0\mu_0 \frac{d}{dt}\int_\sigma \boldsymbol{E}\cdot\boldsymbol{n}\,|dA|. \tag{7.57}$$

This is useful when studying boundary conditions at surfaces of media carrying time-varying currents. The equations involving the Euclidean Green's function are no longer valid when the sources vary with time. In section 7.5 we discuss an alternative Green's function suitable for the important case of electromagnetic radiation.

7.2.3 The energy-momentum tensor

The energy density contained in a vacuum electromagnetic field, measured in the γ_0 frame, is

$$\varepsilon = \tfrac{1}{2}(\boldsymbol{E}^2 + \boldsymbol{B}^2), \tag{7.58}$$

where we have reverted to natural units. In section 7.1.2 we saw that the quantity $\boldsymbol{E}^2 - \boldsymbol{B}^2$ is Lorentz-invariant. This is not true of the energy density, which should clearly depend on the observer performing the measurement. The total energy in a volume V is found by integrating ε over the volume. If we look at how this varies in time, assuming no sources are present, we find that

$$\begin{aligned}\frac{d}{dt}\int_V |dX|\,\tfrac{1}{2}(\boldsymbol{E}^2+\boldsymbol{B}^2) &= \int_V |dX|\,\langle -\boldsymbol{E}\boldsymbol{\nabla}(I\boldsymbol{B}) + I\boldsymbol{B}\boldsymbol{\nabla}\boldsymbol{E}\rangle \\ &= \oint_{\partial V} |dA|\,\boldsymbol{n}\cdot\big(\boldsymbol{E}\cdot(I\boldsymbol{B})\big). \end{aligned} \tag{7.59}$$

We therefore establish that the field momentum is described by the Poynting vector

$$\boldsymbol{P} = -\boldsymbol{E}\cdot(I\boldsymbol{B}) = \boldsymbol{E}\times\boldsymbol{B}. \tag{7.60}$$

The energy and momentum should be the components of a spacetime 4-vector P, so we form

$$\begin{aligned} P = (\varepsilon + \boldsymbol{P})\gamma_0 &= \tfrac{1}{2}(\boldsymbol{E}^2+\boldsymbol{B}^2)\gamma_0 + \tfrac{1}{2}(I\boldsymbol{B}\boldsymbol{E} - \boldsymbol{E}I\boldsymbol{B})\gamma_0 \\ &= \tfrac{1}{2}(\boldsymbol{E}+I\boldsymbol{B})(\boldsymbol{E}-I\boldsymbol{B})\gamma_0 \\ &= \tfrac{1}{2}F(-\gamma_0 F\gamma_0)\gamma_0 = -\tfrac{1}{2}F\gamma_0 F. \end{aligned} \tag{7.61}$$

This quantity is still observer-dependent as it contains a factor of γ_0. We have in fact constructed the *energy-momentum tensor* of the electromagnetic field. We write this as

$$\mathsf{T}(a) = -\tfrac{1}{2}FaF = \tfrac{1}{2}Fa\tilde{F}. \tag{7.62}$$

This is clearly a linear function of a and, since it is equal to its own reverse, the result is automatically a vector. It is instructive to contrast our neat form of the energy-momentum tensor with the tensor formula

$$\mathsf{T}^{\mu}{}_{\nu} = \tfrac{1}{4}\delta^{\mu}_{\nu}F^{\alpha\beta}F_{\alpha\beta} + F^{\mu\alpha}F_{\alpha\nu}. \tag{7.63}$$

The geometric algebra form of equation (7.62) does a far better job of capturing the geometric content of the electromagnetic energy-momentum tensor.

The energy-momentum tensor $\mathsf{T}(a)$ returns the flux of 4-momentum across the hypersurface perpendicular to a. This is the relativistic extension of the stress tensor, and it is as fundamental to field theory as momentum is to the mechanics of point particles. All relativistic fields, classical or quantum, have an associated energy-momentum tensor that contains information about the distribution of energy in the fields, and acts as a source of gravitation. The electromagnetic energy-momentum tensor demonstrates a number of properties that turn out to be quite general. The first is that the energy-momentum tensor is (usually) symmetric. For example, we have

$$a\cdot\mathsf{T}(b) = -\tfrac{1}{2}\langle aFbF\rangle = -\tfrac{1}{2}\langle FaFb\rangle = \mathsf{T}(a)\cdot b. \tag{7.64}$$

The reason for qualifying the above statement is that quantum spin gives rise to an antisymmetric contribution to the (matter) energy-momentum tensor. This will be discussed in more details when we look at Dirac theory.

A second property of the electromagnetic energy-momentum tensor is that the energy density $v\cdot\mathsf{T}(v)$ is positive for any timelike vector v. This is clear from the definition of ε in equation (7.58). The expression for ε is appropriate the γ_0 frame, but the sign of ε cannot be altered by transforming to a different frame. The reason is that

$$\langle vFvF\rangle = \langle R\gamma_0\tilde{R}FR\gamma_0\tilde{R}F\rangle = \langle\gamma_0F'\gamma_0F'\rangle, \tag{7.65}$$

where $F' = \tilde{R}FR$. Transforming to a different velocity is equivalent to back-transforming the fields in the γ_0 frame, so keeps the energy density positive. Matter which does not satisfy the inequality $v\cdot\mathsf{T}(v)\geq 0$ is said to be '*exotic*', and has curious properties when acting as a source of gravitational fields.

The third main property of energy-momentum tensors is that, in the absence of external sources, they give rise to a set of conserved vectors. This is because we have

$$\nabla\cdot\mathsf{T}(a) = 0 \quad \forall \text{ constant } a. \tag{7.66}$$

Equivalently, we can use the symmetry of $\mathsf{T}(a)$ to write

$$\dot{\mathsf{T}}(\dot{\nabla})\cdot a = 0, \quad \forall\, a, \tag{7.67}$$

which implies that

$$\dot{\mathsf{T}}(\dot{\nabla}) = 0. \tag{7.68}$$

For the case of electromagnetism, this result is straightforward to prove:

$$\dot{\mathsf{T}}(\dot{\nabla}) = -\tfrac{1}{2}[\dot{F}\dot{\nabla}F + F\nabla F] = 0, \tag{7.69}$$

which follows since $\nabla F = \dot{F}\dot{\nabla} = 0$ in the absence of sources.

Conservation of the energy-momentum tensor implies that the total flux of energy-momentum over a closed hypersurface is zero:

$$\int_{\partial V} |dA|\,\mathsf{T}(n) = 0, \tag{7.70}$$

where ∂V is a closed 3-surface with directed measure $dA = nI\,|dA|$. That the flux vanishes is a simple application of the fundamental theorem of integral calculus (in flat spacetime),

$$\int_{\partial V}\mathsf{T}(n\,|dA|) = \int_{\partial V}\mathsf{T}(dAI^{-1}) = \int_V \dot{\mathsf{T}}(\dot{\nabla})\,dX\,I^{-1} = 0. \tag{7.71}$$

Given that $\mathsf{T}(\gamma_0)$ is the energy-momentum density in the γ_0 frame, the total 4-momentum is

$$P_{tot} = \int |dX|\,\mathsf{T}(\gamma_0). \tag{7.72}$$

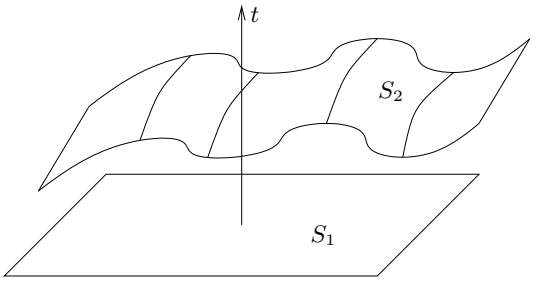

Figure 7.1 *Hypersurface integration.* The integral over a hypersurface of a (spacetime) conserved current is independent of the chosen hypersurface. The two surfaces S_1 and S_2 can be joined at spatial infinity (provided the fields vanish there). The difference is therefore the integral over a closed 3-surface, which vanishes by the divergence theorem.

The conservation equation (7.68) guarantees that, in the absence of charges, the total energy-momentum is conserved. We see that

$$\frac{d}{dt}P_{tot} = \int |dX|\,\partial_t \mathsf{T}(\gamma_0) = \int |dX|\,\dot{\mathsf{T}}(\dot{\boldsymbol{\nabla}}\gamma_0), \tag{7.73}$$

where we have used the fact that $\nabla = \gamma_0\partial_t - \boldsymbol{\nabla}\gamma_0$. The final integral here is a total derivative and so gives rise to a boundary term, which vanishes provided the fields fall off sufficiently fast at large distances. Similarly, we can also see that P_{tot} is independent of the chosen timelike axis. It is a covariant (non-local) property of the field configuration. The proof comes from considering the integral over two distinct spacelike hypersurfaces (figure 7.1). If the integrals are joined at infinity (which introduces zero contribution) we form a closed integral of $\mathsf{T}(n)$. This vanishes from the conservation equation, so the total energy-momentum is independent of the choice of hypersurface.

In the presence of additional sources the electromagnetic energy-momentum tensor is no longer conserved. The total energy-momentum tensor, including both the matter and electromagnetic content will be conserved, however. This is a general feature of field theory in a flat spacetime, though the picture is altered somewhat if gravitational fields are present. The extent to which the separate tensors for each field are not conserved contains useful information about the flow of energy-momentum. For example, suppose that an external current is present, so that

$$\dot{\mathsf{T}}(\dot{\nabla}) = -\tfrac{1}{2}(-JF + FJ) = J\cdot F. \tag{7.74}$$

An expression of the form $J\cdot F$ was derived in the Lorentz force law, discussed in section 5.5.3. In the γ_0 frame, $J\cdot F$ decomposes into

$$J\cdot F = \langle(\rho + \boldsymbol{J})\gamma_0(\boldsymbol{E} + I\boldsymbol{B})\rangle_1 = -\big(\boldsymbol{J}\cdot\boldsymbol{E} + \rho\boldsymbol{E} + \boldsymbol{J}\times\boldsymbol{B}\big)\gamma_0. \tag{7.75}$$

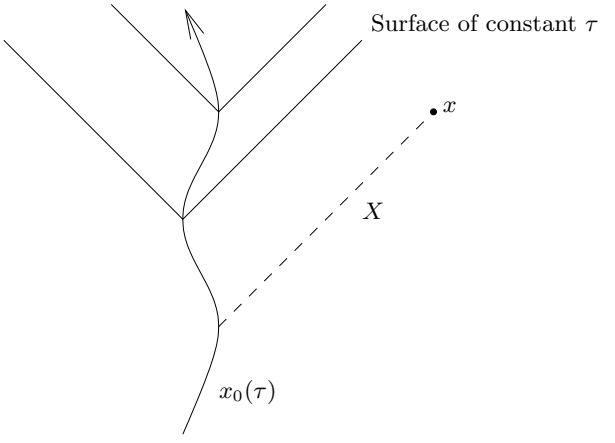

Figure 7.2 *Field from a moving point charge.* The charge follows the trajectory $x_0(\tau)$, and $X = x - x_0(\tau)$ is the retarded null vector connecting the point x to the worldline. The time τ can be viewed as a scalar field with each value of τ extended out over the forward null cone.

The timelike component, $\boldsymbol{J}\cdot\boldsymbol{E}$, is the work done — the rate of change of energy density. The relative vector term is the rate of change of field momentum, and so is closely related to the force on a point particle.

7.3 The electromagnetic field of a point charge

We now derive a formula for the electromagnetic fields generated by a radiating charge. This is one of the most important results in classical electromagnetic theory. Suppose that a charge q moves along a worldline $x_0(\tau)$, where τ is the proper time along the worldline (see figure 7.2). An observer at spacetime position x receives an electromagnetic influence from the point where the charge's worldline intersects the observer's past light-cone. The vector

$$X = x - x_0(\tau) \tag{7.76}$$

is the separation vector down the light-cone, joining the observer to this intersection point. Since this vector must be null, we can view the equation

$$X^2 = 0 \tag{7.77}$$

as defining a map from spacetime position x to a value of the particle's proper time τ. That is, for every spacetime position x there is a unique value of the (retarded) proper time along the charge's worldline for which the vector connecting x to the worldline is null. In this sense, we can write $\tau = \tau(x)$, and treat τ as a scalar field.

The Liénard–Wiechert potential for the retarded field from a point charge moving with an arbitrary velocity $v = \dot{x}_0$ is

$$A = \frac{q}{4\pi}\frac{v}{|X\cdot v|}. \tag{7.78}$$

This solution is obtained from the wave equation $\nabla^2 A = J$ using the appropriate retarded Green's function

$$G_{ret}(\boldsymbol{r}, t) = \frac{1}{4\pi|\boldsymbol{r}|}\delta(|\boldsymbol{r}| - t). \tag{7.79}$$

A similar solution exists if the advanced Green's function is used. The question of which is the correct one to use is determined experimentally by the fact that no convincing detection of an advanced (acausal) field has ever been reported. A deeper understanding of these issues is provided by the quantum treatment of radiation.

If the charge is at rest in the γ_0 frame, we have

$$x_0(\tau) = \tau\gamma_0 = (t - r)\gamma_0, \tag{7.80}$$

where r is the relative 3-space distance from the observer to the charge. The null vector X is therefore

$$X = r(\gamma_0 + e_r). \tag{7.81}$$

For this simple case the 4-potential A is a pure $1/r$ electrostatic field:

$$A = \frac{q}{4\pi}\frac{\gamma_0}{|X\cdot\gamma_0|} = \frac{q}{4\pi r}\gamma_0. \tag{7.82}$$

The same result is obtained if the advanced Green's function is used. The difference between the advanced and retarded solutions is only seen when the charge radiates. We know that radiation is not handled satisfactorily in the classical theory because it predicts that atoms are not stable and should radiate. Issues concerning the correct Green's function cannot be fully resolved without a quantum treatment.

7.3.1 The field strength

The aim now is to differentiate the potential of equation (7.78) to find the field strength. First, we differentiate the equation $X^2 = 0$ to obtain

$$\begin{aligned} 0 = \gamma^\mu(\partial_\mu X)\cdot X &= \dot{\nabla}\,\dot{x}\cdot X - \nabla\tau\,(\partial_\tau x_0)\cdot X \\ &= X - \nabla\tau\,(v\cdot X). \end{aligned} \tag{7.83}$$

It follows that

$$\nabla\tau = \frac{X}{X\cdot v}. \tag{7.84}$$

The gradient of τ points along X, which is the direction of constant τ. This is a peculiarity of null surfaces that was first encountered in chapter 6. In finding an expression for $\nabla\tau$ we have demonstrated how the particle proper time can be treated as a spacetime scalar field. Fields of this type are known as *adjunct* fields — they carry information, but do not exist in any physical sense.

To differentiate A we need an expression for $\nabla(X\cdot v)$. We find that

$$\begin{aligned}\nabla(X\cdot v) &= \dot{\nabla}(\dot{X})\cdot v + \nabla\tau\, X\cdot(\partial_\tau v)\\ &= v - \nabla\tau + \nabla\tau\, X\cdot\dot{v},\end{aligned} \tag{7.85}$$

where $\dot{v} = \partial_\tau v$. Provided X is defined in terms of the retarded time, $X\cdot v$ will always be positive and there is no need for the modulus in the denominator of equation (7.78). We are now in a position to evaluate ∇A. We find that

$$\begin{aligned}\nabla A &= \frac{q}{4\pi}\left(\frac{\nabla v}{X\cdot v} - \frac{1}{(X\cdot v)^2}\nabla(X\cdot v)v\right)\\ &= \frac{q}{4\pi}\left(\frac{X\dot{v}}{(X\cdot v)^2} - \frac{1}{(X\cdot v)^2} - \frac{(X\,X\cdot\dot{v} - X)v}{(X\cdot v)^3}\right)\\ &= \frac{q}{4\pi}\left(\frac{X\wedge\dot{v}}{(X\cdot v)^2} + \frac{X\wedge v - X\cdot\dot{v}\,X\wedge v}{(X\cdot v)^3}\right).\end{aligned} \tag{7.86}$$

The result is a pure bivector, so $\nabla\cdot A = 0$ and the A field of equation (7.78) is in the Lorentz gauge. This is to be expected, since the solution is obtained from the wave equation $\nabla^2 A = J$.

We can gain some insight into the expression for F by writing

$$X\cdot v\,X\wedge\dot{v} - X\cdot\dot{v}\,X\wedge v = -X\big(X\cdot(\dot{v}\wedge v)\big) = \tfrac{1}{2}X\dot{v}\wedge vX, \tag{7.87}$$

which uses the fact that $X^2 = 0$. Writing $\Omega_v = \dot{v}\wedge v$ for the acceleration bivector of the particle, we arrive at the compact formula

$$F = \frac{q}{4\pi}\frac{X\wedge v + \frac{1}{2}X\Omega_v X}{(X\cdot v)^3}. \tag{7.88}$$

One can proceed to show that, away from the worldline, F satisfies the free-field equation $\nabla F = 0$. The details are left as an exercise. The solution (7.88) displays a clean split into a velocity term proportional to $1/(\text{distance})^2$ and a long-range radiation term proportional to $1/(\text{distance})$. The term representing the distance is simply $X\cdot v$. This is just the distance between the events x and $x_0(\tau)$ as measured in the rest frame of the charge at its retarded position. The first term in equation (7.88) is the Coulomb field in the rest frame of the charge. The second, radiation, term:

$$F_{rad} = \frac{q}{4\pi}\frac{\frac{1}{2}X\Omega_v X}{(X\cdot v)^3}, \tag{7.89}$$

is proportional to the rest frame acceleration projected down the null vector X.

The fact that this term falls of as 1/(distance) implies that the energy-momentum tensor contains a term which falls of as the inverse square of distance. This gives a non-vanishing surface integral at infinity in equation (7.73) and describes how energy is carried away from the source.

7.3.2 Constant velocity

A charge with constant velocity v has the trajectory

$$x_0(\tau) = v\tau, \tag{7.90}$$

where we have chosen an origin so that the particle passes through this point at $\tau = 0$. The intersection of $x_0(\tau)$ with the past light-cone through x is determined by

$$(x - v\tau)^2 = 0 \quad \Rightarrow \tau = v\cdot x - \big((v\cdot x)^2 - x^2\big)^{1/2}. \tag{7.91}$$

We have chosen the earlier root to ensure that the intersection lies on the past light-cone. We now form $X\cdot v$ to find

$$X\cdot v = (x - v\tau)\cdot v = \big((v\cdot x)^2 - x^2\big)^{1/2}. \tag{7.92}$$

We can write this as $|x\wedge v|$ since

$$|x\wedge v|^2 = x\cdot\big(v\cdot(x\wedge v)\big) = (x\cdot v)^2 - x^2. \tag{7.93}$$

The acceleration bivector vanishes since v is constant, and $X\wedge v = x\wedge v$. It follows that the Faraday bivector is simply

$$F = \frac{q}{4\pi}\frac{x\wedge v}{|x\wedge v|^3}. \tag{7.94}$$

This is the Coulomb field solution with the velocity γ_0 replaced by v. This solution could be obtained by transforming the Coulomb field via

$$F \mapsto F' = RF(\tilde{R}xR)\tilde{R}, \tag{7.95}$$

where $v = R\gamma_0\tilde{R}$. Covariance of the field equations ensures that this process generates a new solution.

We next decompose F into electric and magnetic fields in the γ_0 frame. This requires the spacetime split

$$x\wedge v = \langle x\gamma_0\gamma_0 v\rangle_2 = \gamma\langle(t + \boldsymbol{r})(1 - \boldsymbol{v})\rangle_2 = \gamma(\boldsymbol{r} - \boldsymbol{v}t) - \gamma\boldsymbol{r}\wedge\boldsymbol{v}, \tag{7.96}$$

where $\boldsymbol{v}$ is the relative velocity and γ is the Lorentz factor. We now have

$$\boldsymbol{E} = \frac{q\gamma}{4\pi d^3}(\boldsymbol{r} - \boldsymbol{v}t), \qquad \boldsymbol{B} = \frac{q\gamma}{4\pi d^3}I\boldsymbol{r}\wedge\boldsymbol{v}. \tag{7.97}$$

Here, the effective distance d can be written

$$d^2 = \gamma^2(|\boldsymbol{v}|t - \boldsymbol{v}\cdot\boldsymbol{r}/|\boldsymbol{v}|)^2 + \boldsymbol{r}^2 - (\boldsymbol{r}\cdot\boldsymbol{v})^2/\boldsymbol{v}^2. \tag{7.98}$$

The electric field points towards the actual position of the charge at time t, and not its retarded position at time τ. The same is true of the advanced field, hence the retarded and advanced solutions are equal for charges with constant velocity.

7.3.3 Linear acceleration

Suppose that an accelerating charged particle follows the trajectory

$$x_0(\tau) = a\big(\sinh(g\tau)\gamma_0 + \cosh(g\tau)\gamma_3\big), \tag{7.99}$$

where $a = g^{-1}$ (see figure 7.3). The velocity is given by

$$v(\tau) = \cosh(g\tau)\gamma_0 + \sinh(g\tau)\gamma_3 = \mathrm{e}^{g\tau\boldsymbol{\sigma}_3}\gamma_0 \tag{7.100}$$

and the acceleration bivector is simply

$$\dot{v}v = g\boldsymbol{\sigma}_3. \tag{7.101}$$

The charge has constant (relativistic) acceleration in the γ_3 direction. We again seek the retarded solution of $X^2 = 0$. This is more conveniently expressed in a cylindrical polar coordinate system, with

$$\boldsymbol{r} = \rho(\cos(\phi)\,\boldsymbol{\sigma}_1 + \sin(\phi)\,\boldsymbol{\sigma}_2) + z\boldsymbol{\sigma}_3, \tag{7.102}$$

so that $r^2 = \rho^2 + z^2$. We then find the following equivalent expressions for the retarded proper time:

$$\begin{aligned} e^{g\tau} &= \frac{1}{2a(z-t)}\Big(a^2 + r^2 - t^2 - \big((a^2 + r^2 - t^2)^2 - 4a^2(z^2 - t^2)\big)^{1/2}\Big), \\ e^{-g\tau} &= \frac{1}{2a(z+t)}\Big(a^2 + r^2 - t^2 + \big((a^2 + r^2 - t^2)^2 - 4a^2(z^2 - t^2)\big)^{1/2}\Big). \end{aligned} \tag{7.103}$$

These equations have a solution provided $z + t > 0$. As the trajectory assumes that the charge has been accelerating for ever, a *horizon* is formed beyond which no effects of the charge are felt (figure 7.3). Constant eternal acceleration of this type is unphysical and in practice we only consider the acceleration taking place for a short period.

We can now calculate the radiation from the charge. First we need the effective distance

$$X\cdot v = \frac{\big((a^2 + r^2 - t^2)^2 - 4a^2(z^2 - t^2)\big)^{1/2}}{2a}. \tag{7.104}$$

This vanishes on the path of the particle ($\rho = 0$ and $z^2 - t^2 = a^2$), as required.

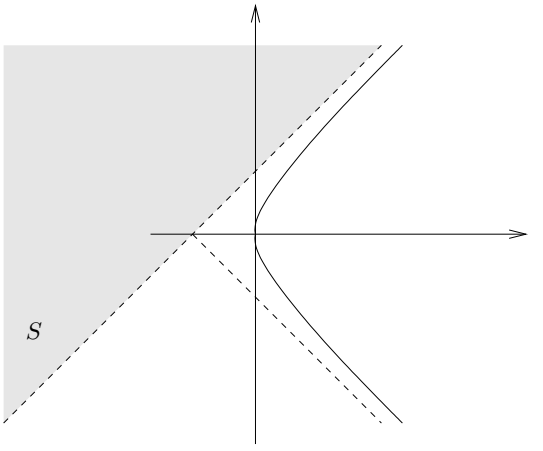

Figure 7.3 *Constant acceleration.* The spacetime trajectory of a particle with constant acceleration is a hyperbola. The asymptotes are null vectors and define future and past horizons. Any signal sent from within the shaded region S will never be received by the particle.

The remaining factor in F is

$$\begin{aligned} X \wedge v + \tfrac{1}{2} X \dot{v} v X &= x \wedge v - a\boldsymbol{\sigma}_3 + \frac{1}{2a}(x - x_0)\boldsymbol{\sigma}_3(x - x_0) \\ &= \frac{1}{2a} x \boldsymbol{\sigma}_3 x - \frac{a}{2}\boldsymbol{\sigma}_3 \\ &= \frac{1}{2a}(z^2 - \rho^2 - t^2 - a^2)\boldsymbol{\sigma}_3 + \frac{z\rho}{a}\boldsymbol{\sigma}_\rho + \frac{t\rho}{a} I \boldsymbol{\sigma}_\phi, \end{aligned} \tag{7.105}$$

where $\boldsymbol{\sigma}_\rho$ and $\boldsymbol{\sigma}_\phi$ are the unit spatial axial and azimuthal vectors respectively. An instructive way to display the information contained in the expression for F is to plot the field lines of $\boldsymbol{E}$ at a fixed time. We assume that the charge starts accelerating at $t = t_1$, and stops again at $t = t_2$. There are then discontinuities in the electric field line directions on the two appropriate light-spheres. In figure 7.4 the acceleration takes place for a short period of time, so that a pulse of radiation is sent outwards. In figure 7.5 the charge began accelerating from rest at $t = -10a$. The pattern is well developed, and shows clearly the refocusing of the field lines onto the 'image charge'. The image position corresponds to the place the charge would have reached had it not started accelerating. Of course, the image charge is not actually present, and the field lines diverge after they cross the light-sphere corresponding to the start of the acceleration.

For many applications we are only interested in the fields a long way from the source. In this region the fields can usually be approximated by simple dipole or higher order multipole fields. Suppose that the charge accelerates for a short

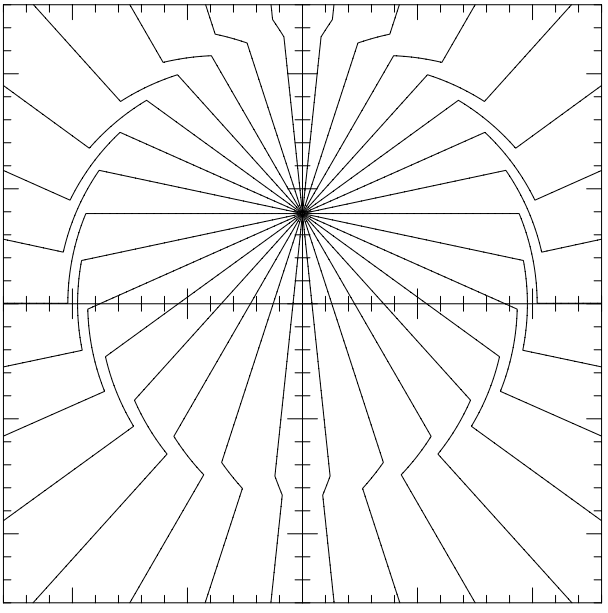

Figure 7.4 *Field lines from an accelerated charge I.* The charge accelerated for $-0.2a < t < 0.2a$, leaving an outgoing pulse of transverse radiation field. The field lines were computed at $t = 5a$.

period and emits a pulse of radiation. In the limit $r \gg a$ the pulse will arrive at some time which, to a good approximation, is centred around the time that minimises $X \cdot v$. This time is given by

$$t_0 = \sqrt{r^2 - a^2}. \tag{7.106}$$

At $t = t_0$ the proper distance $X \cdot v$ evaluates to ρ, the distance from the z axis. The point on the axis ρ away from the observer is where the charge would appear to be if it were not accelerating. For the large distance approximation to be valid we therefore also require that ρ is large, so that the proper distance from the source is large. (For small ρ and $z > a$ a different procedure can be used.) We can now obtain an approximate formula for the radiation field at a fixed location $\boldsymbol{r}$, with $r, \rho \gg a$, around $t = t_0$. For this we define

$$\delta_t = t - t_0 \tag{7.107}$$

so that the proper distance is approximated by

$$X \cdot v \approx \left(\rho^2 + r^2 \delta_t^2 / a^2\right)^{1/2}. \tag{7.108}$$

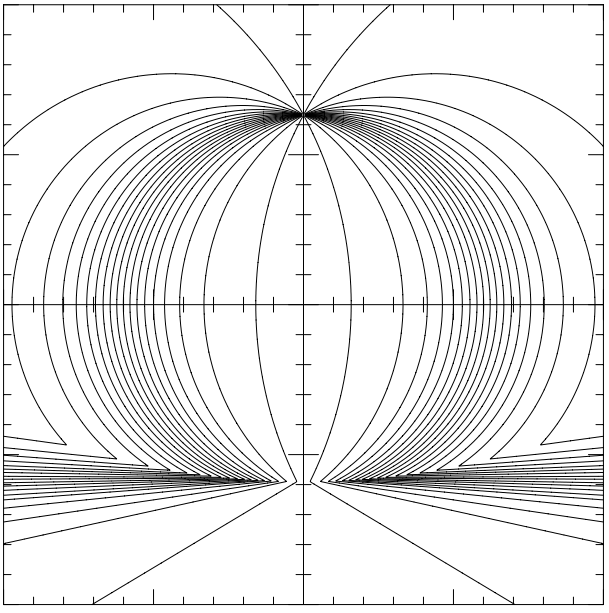

Figure 7.5 *Field lines from an accelerated charge II.* The charge began its acceleration at $t_1 = -10a$ and has thereafter accelerated uniformly. The field lines are plotted at $t = 3a$.

The remaining terms in F become

$$X \wedge v + \tfrac{1}{2} X \dot{v} v X \approx \frac{r\rho}{a} \left(\boldsymbol{\sigma}_\theta + I \boldsymbol{\sigma}_\phi \right), \tag{7.109}$$

where $\boldsymbol{\sigma}_\theta$ and $\boldsymbol{\sigma}_\phi$ are unit spherical-polar basis vectors. The final formula is

$$F \approx \frac{q}{4\pi} \frac{r\rho}{a} \left(\rho^2 + \frac{r^2 \delta_t^2}{a^2} \right)^{-3/2} \left(\boldsymbol{\sigma}_\theta + I \boldsymbol{\sigma}_\phi \right), \tag{7.110}$$

which describes a pure, outgoing radiation field a large distance from a linearly accelerating source. The magnitude of the acceleration is controlled by $g = a^{-1}$.

7.3.4 Circular orbits and synchrotron radiation

As a further application, consider a charge moving in a circular orbit. The worldline is defined by

$$x_0 = \tau \cosh(\alpha)\, \gamma_0 + a\big(\cos(\omega\tau)\gamma_1 + \sin(\omega\tau)\gamma_2\big), \tag{7.111}$$

where $a = \omega^{-1} \sinh(\alpha)$. The particle velocity is

$$v = \cosh(\alpha)\, \gamma_0 + \sinh(\alpha)\big(-\sin(\omega\tau)\gamma_1 + \cos(\omega\tau)\gamma_2\big) = R\gamma_0\tilde{R}, \tag{7.112}$$

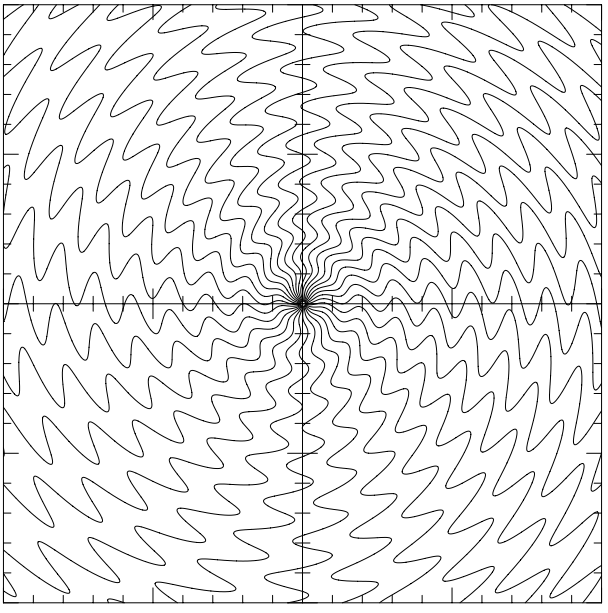

Figure 7.6 *Field lines from a rotating charge I.* The charge has $\alpha = 0.1$, which gives rise to a smooth, wavy pattern.

where the rotor R is given by

$$R = \mathrm{e}^{-\omega\tau I\boldsymbol{\sigma}_3/2}\mathrm{e}^{\alpha\boldsymbol{\sigma}_2/2}. \tag{7.113}$$

We must first locate the retarded null vector X. The equation $X^2 = 0$ reduces to

$$t = \tau\cosh(\alpha) + \left(r^2 + a^2 - 2a\rho\cos(\omega\tau - \phi)\right)^{1/2}, \tag{7.114}$$

which is an implicit equation for $\tau(x)$. No simple analytic solution exists, but a numerical solution is easy to achieve. This is aided by the observation that, for fixed $\boldsymbol{r}$, the mapping between t and τ is monotonic and τ is bounded by the conditions

$$t - \left(r^2 + 2a\rho + a^2\right)^{1/2} < \tau\cosh(\alpha) < t - \left(r^2 - 2a\rho + a^2\right)^{1/2}. \tag{7.115}$$

Once we have a satisfactory procedure for locating τ on the retarded light-cone, we can straightforwardly employ the formula for F in numerical simulations. The first term required is the effective distance $X \cdot v$, which is given by

$$X\cdot v = \cosh(\alpha)\left(r^2 + a^2 - 2a\rho\cos(\omega\tau - \phi)\right)^{1/2} + \rho\sinh(\alpha)\,\sin(\omega\tau - \phi). \tag{7.116}$$

The remaining term to compute, $X \wedge v + X\dot{v}vX/2$, is more complicated, as can be seen from the behaviour shown in figures 7.6, 7.7 and 7.8. They show the

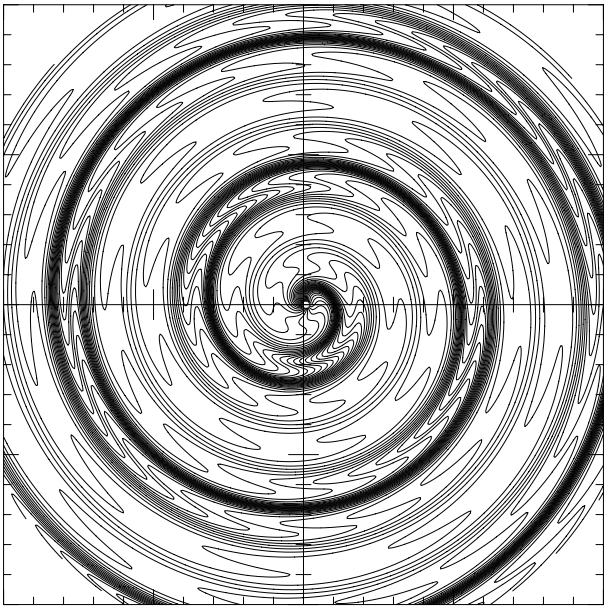

Figure 7.7 *Field lines from a rotating charge II.* The charge has an intermediate velocity, with $\alpha = 0.4$. Bunching of the field lines is clearly visible.

field lines in the equatorial plane of a rotating charge with $\omega = 1$. For 'low' speeds we get the gentle, wavy pattern of field lines shown in figure 7.6. The case displayed in figure 7.7 is for an intermediate velocity ($\alpha = 0.4$), and displays many interesting features. By $\alpha = 1$ (figure 7.8) the field lines have concentrated into synchrotron pulses, a pattern which continues thereafter.

Synchrotron radiation is important in many areas of physics, from particle physics through to radioastronomy. Synchrotron radiation from a radiogalaxy, for example, has $a \approx 10^8$ m and $r \approx 10^{25}$ m. A power-series expansion in a/r is therefore quite safe! Typical values of $\cosh(\alpha)$ are 10^4 for electrons producing radio emission. In the limit $r \gg a$, the relation between t and τ simplifies to

$$t - r \approx \tau \cosh(\alpha) - a \sin(\theta) \cos(\omega\tau - \phi). \tag{7.117}$$

The effective distance reduces to

$$X \cdot v \approx r \cosh(\alpha)\big(1 + \tanh(\alpha) \sin(\theta) \sin(\omega\tau - \phi)\big), \tag{7.118}$$

and the null vector X given by the simple expression

$$X \approx r(\gamma_0 + e_r). \tag{7.119}$$

In the expression for F of equation (7.88) we can ignore the $X \wedge v$ (Coulomb)

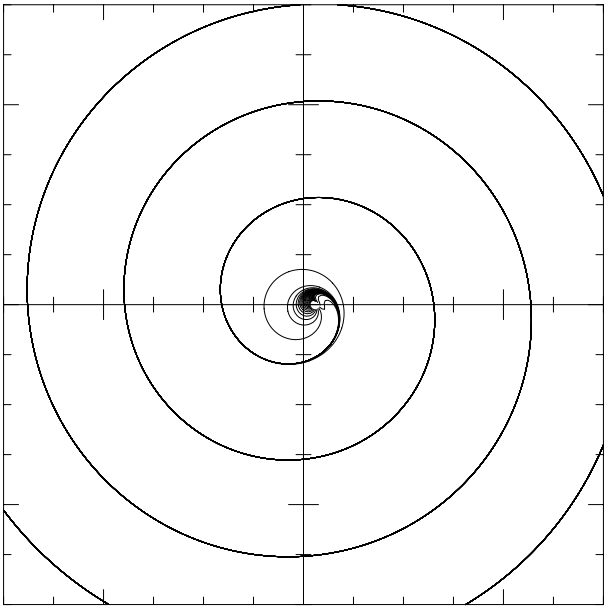

Figure 7.8 *Field lines from a rotating charge III.* The charge is moving at a highly relativistic velocity, with $\alpha = 1$. The field lines are concentrated into a series of synchrotron pulses.

term, which is negligible compared with the long-range radiation term. For the radiation term we need the acceleration bivector

$$\dot{v}v = -\omega \sinh(\alpha)\cosh(\alpha)\big(\cos(\omega\tau)\boldsymbol{\sigma}_1 + \sin(\omega\tau)\boldsymbol{\sigma}_2\big) + \omega \sinh^2(\alpha) I\boldsymbol{\sigma}_3. \quad (7.120)$$

The radiation term is governed by $X\Omega_v X/2$, which simplifies to

$$\begin{aligned}\tfrac{1}{2}X\dot{v}vX \approx{}& \omega r^2 \cosh(\alpha)\sinh(\alpha)(\cos(\theta)\cos(\omega\tau-\phi)\boldsymbol{\sigma}_\theta(1-\boldsymbol{\sigma}_r)\\ &+ \omega r^2 \sinh(\alpha)\big(\cosh(\alpha)\sin(\omega\tau-\phi) + \sinh(\alpha)\sin(\theta)\big)\boldsymbol{\sigma}_\phi(1-\boldsymbol{\sigma}_r). \quad (7.121)\end{aligned}$$

These formulae are sufficient to initiate studying synchrotron radiation. They contain a wealth of physical information, but a detailed study is beyond the scope of this book.

7.4 Electromagnetic waves

For many problems in electromagnetic theory it is standard practice to adopt a complex representation of the electromagnetic field, with the implicit assumption that only the real part represents the physical field. This is particularly convenient when discussing electromagnetic waves and diffraction, as studied in this and the following section. We have seen, however, that the field strength

F is equipped with a natural complex structure through the pseudoscalar I. We should therefore not be surprised to find that, in certain cases, the formal imaginary i plays the role of the pseudoscalar. This is indeed the case for circularly-polarised light. But one cannot always identify i with I, as is clear when handling plane-polarised light. The formal complexification retains its usefulness in such applications and we accordingly adopt it here. It is important to remember that this is a formal exercise, and that real parts must be taken before forming bilinear objects such as the energy-momentum tensor. The study of electromagnetic waves is an old and well-developed subject. Unfortunately, it suffers from the lack of a single, universal set of conventions. As far as possible, we have followed the conventions of Jackson (1999).

We seek vacuum solutions to the Maxwell equations which are purely oscillatory. We therefore start by writing

$$F = \mathrm{Re}\big(F_0 \mathrm{e}^{-ik\cdot x}\big). \tag{7.122}$$

The vacuum equation $\nabla F = 0$ then reduces to the algebraic equation

$$kF_0 = 0. \tag{7.123}$$

Pre-multiplying by k we immediately see that $k^2 = 0$, as expected of the wavevector. The constant bivector F_0 must contain a factor of k, as nothing else totally annihilates k. We therefore must have

$$F_0 = k\wedge n = kn, \tag{7.124}$$

where n is some vector satisfying $k\cdot n = 0$. We can always add a further multiple of k to n, since

$$k(n + \lambda k) = kn + \lambda k^2 = k\wedge n. \tag{7.125}$$

This freedom in n can be employed to ensure that n is perpendicular to the velocity vector of some chosen observer.

As an example, consider a wave travelling in the γ_3 direction with frequency ω as measured in the γ_0 frame. This implies that $\gamma_0\cdot k = \omega$, so the wavevector is given by

$$k = \omega(\gamma_0 + \gamma_3), \tag{7.126}$$

and the phase term is

$$-ik\cdot x = -i\omega(t - z). \tag{7.127}$$

The vector n can be chosen to just contain γ_1 and γ_2 components, so we can write

$$\begin{aligned} F &= -(\gamma_0 + \gamma_3)(\alpha_1\gamma_1 + \alpha_2\gamma_2)\cos(k\cdot x) \\ &= (1 + \boldsymbol{\sigma}_3)(\alpha_1\boldsymbol{\sigma}_1 + \alpha_2\boldsymbol{\sigma}_2)\cos(k\cdot x). \end{aligned} \tag{7.128}$$

This solution represents plane-polarised light, as both the $\boldsymbol{E}$ and $\boldsymbol{B}$ fields lie in fixed planes, 90° apart, and only their magnitudes oscillate in time.

An arbitrary phase can be added to the cosine term, so the most general solution for a wave travelling in the $+z$ direction is

$$F = (1+\boldsymbol{\sigma}_3)\big((\alpha_1\boldsymbol{\sigma}_1 + \alpha_2\boldsymbol{\sigma}_2)\cos(k\cdot x) + (\beta_1\boldsymbol{\sigma}_1 + \beta_2\boldsymbol{\sigma}_2)\sin(k\cdot x)\big), \qquad (7.129)$$

where the constants α_i and β_i, are all real. This general solution can describe all possible states of polarisation. A convenient representation is to introduce the complex coefficients

$$c_1 = \alpha_1 + i\beta_1, \quad c_2 = \alpha_2 + i\beta_2. \qquad (7.130)$$

These form the components of the complex *Jones vector* (c_1, c_2). In terms of these components we can write

$$F = \mathrm{Re}\big((1+\boldsymbol{\sigma}_3)(c_1\boldsymbol{\sigma}_1 + c_2\boldsymbol{\sigma}_2)\mathrm{e}^{-ik\cdot x}\big), \qquad (7.131)$$

and it is a straightforward matter to read off the separate $\boldsymbol{E}$ and $\boldsymbol{B}$ fields.

The multivector $(1+\boldsymbol{\sigma}_3)$ has a number of interesting properties. It absorbs factors of $\boldsymbol{\sigma}_3$, as can be seen from

$$\boldsymbol{\sigma}_3(1+\boldsymbol{\sigma}_3) = 1 + \boldsymbol{\sigma}_3. \qquad (7.132)$$

In addition, $(1+\boldsymbol{\sigma}_3)$ squares to give a multiple of itself,

$$(1+\boldsymbol{\sigma}_3)^2 = 1 + 2\boldsymbol{\sigma}_3 + \boldsymbol{\sigma}_3^2 = 2(1+\boldsymbol{\sigma}_3). \qquad (7.133)$$

This property implies that $(1+\boldsymbol{\sigma}_3)$ does not have an inverse, so in a multivector expression it acts as a projection operator. The combination $(1+\boldsymbol{\sigma}_3)/2$ has the particular property of squaring to give itself back again. Multivectors with this property are said to be *idempotent* and are important in the general classification of Clifford algebras and their spinor representations. In spacetime applications idempotents invariably originate from a null vector, in the manner that $(1+\boldsymbol{\sigma}_3)$ originates from a spacetime split of $\gamma_0 + \gamma_3$.

7.4.1 Circularly-polarised light

Many problems are more naturally studied using a basis of circularly-polarised states, as opposed to plane-polarised ones. These arise when c_1 and c_2 are $\pi/2$ out of phase. One form is given by $\alpha_1 = -\beta_2 = E_0$ and $\alpha_2 = \beta_1 = 0$, where E_0 denotes the magnitude of the electric field. For this solution we can write

$$\begin{aligned} F &= E_0(1+\boldsymbol{\sigma}_3)\big(\boldsymbol{\sigma}_1\cos(k\cdot x) - \boldsymbol{\sigma}_2\sin(k\cdot x)\big) \\ &= E_0(1+\boldsymbol{\sigma}_3)\boldsymbol{\sigma}_1\mathrm{e}^{-I\boldsymbol{\sigma}_3\omega(t-z)}. \end{aligned} \qquad (7.134)$$

In a plane of constant z (a wavefront) the $\boldsymbol{E}$ field rotates in a clockwise (negative) sense, when viewed looking back towards the source (figure 7.9). In the optics

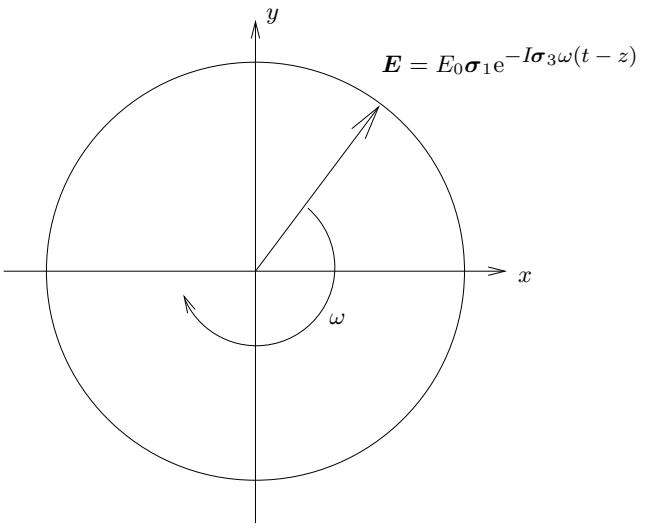

Figure 7.9 *Right-circularly-polarised light.* In the $z = 0$ plane the $\boldsymbol{E}$ vector rotates clockwise, when viewed from above. The wave vector points out of the page. In space, at constant time, the $\boldsymbol{E}$ field sweeps out a right-handed helix.

literature this is known as *right-circularly-polarised* light. The reason for this is that, at constant time, the $\boldsymbol{E}$ field sweeps out a helix in space which defines a right-handed screw. If you grip the helix in your right hand, your thumb points in the direction in which the helix advances if tracked along in the sense defined by your grip. This definition of handedness for a helix is independent of which way round you chose to grip it.

Left-circularly-polarised light has the $\boldsymbol{E}$ field rotating with the opposite sense. The general form of this solution is

$$F = (1 + \boldsymbol{\sigma}_3)(\alpha_1\boldsymbol{\sigma}_1 + \alpha_2\boldsymbol{\sigma}_2)\mathrm{e}^{I\boldsymbol{\sigma}_3 k\cdot x}. \tag{7.135}$$

Particle physicists prefer an alternative labelling scheme for circularly-polarised light. The scheme is based, in part, on the quantum definition of angular momentum. In the quantum theory, the total angular momentum consists of a spatial part and a spin component. Photons, the quanta of electromagnetic radiation, have spin-1. The spin vector for these can either point in the direction of propagation, or against it, depending on the orientation of rotation of the $\boldsymbol{E}$ field. It turns out that for *right*-circularly-polarised light the spin vector points against the direction of propagation, which is referred to as a state of *negative helicity*. Conversely, *left*-circularly-polarised light has *positive* helicity.

Equation (7.132) enables us to convert phase rotations with the bivector $I\boldsymbol{\sigma}_3$

into duality rotations governed by the pseudoscalar I. This relies on the relation

$$\begin{aligned}(1+\boldsymbol{\sigma}_3)\mathrm{e}^{I\boldsymbol{\sigma}_3\phi} &= (1+\boldsymbol{\sigma}_3)\big(\cos(\phi)+I\boldsymbol{\sigma}_3\sin(\phi)\big)\\ &= (1+\boldsymbol{\sigma}_3)\big(\cos(\phi)+I\sin(\phi)\big) = (1+\boldsymbol{\sigma}_3)\mathrm{e}^{I\phi}.\end{aligned} \tag{7.136}$$

The general solution for right-circularly-polarised light can now be written

$$\begin{aligned}F &= (1+\boldsymbol{\sigma}_3)\mathrm{e}^{I\boldsymbol{\sigma}_3 k\cdot x}(\alpha_1\boldsymbol{\sigma}_1+\alpha_2\boldsymbol{\sigma}_2)\\ &= (1+\boldsymbol{\sigma}_3)(\alpha_1\boldsymbol{\sigma}_1+\alpha_2\boldsymbol{\sigma}_2)\mathrm{e}^{Ik\cdot x}.\end{aligned} \tag{7.137}$$

In this case the complex structure is now entirely geometric, generated by the pseudoscalar. This means that there is no longer any need to take the real part of the solution, as the bivector is already entirely real. A similar trick can be applied to write the constant terms as

$$(1+\boldsymbol{\sigma}_3)(\alpha_1\boldsymbol{\sigma}_1+\alpha_2\boldsymbol{\sigma}_2) = (1+\boldsymbol{\sigma}_3)\boldsymbol{\sigma}_1(\alpha_1-I\alpha_2), \tag{7.138}$$

so that the coefficient also becomes 'complex' on the pseudoscalar. The general form for right-hand circularly-polarised light solution can now be written

$$F = (1+\boldsymbol{\sigma}_3)\boldsymbol{\sigma}_1\alpha_R\mathrm{e}^{Ik\cdot x}, \tag{7.139}$$

where α_R is a scalar + pseudoscalar combination. Left-hand circularly-polarised light is described by reversing the sign of the exponent to $-Ik\cdot x$. General polarisation states can be built up as linear combinations of these circularly polarised modes, so we can write

$$F = (1+\boldsymbol{\sigma}_3)\boldsymbol{\sigma}_1\big(\alpha_R\mathrm{e}^{Ik\cdot x}+\alpha_L\mathrm{e}^{-Ik\cdot x}\big). \tag{7.140}$$

Here both the coefficients α_L and α_R are scalar + pseudoscalar combinations. The complexification is now based on the pseudoscalar, and we can use α_R and α_L as alternative, geometrically meaningful, complex coefficients for describing general polarisation states. For completeness, the α_L and α_R parameters are related to the earlier plane-polarised coefficients α_i and β_i by

$$\begin{aligned}\alpha_R &= \tfrac{1}{2}(\alpha_1-\beta_2)+\tfrac{1}{2}(\alpha_2+\beta_1)I,\\ \alpha_L &= \tfrac{1}{2}(\alpha_1+\beta_2)+\tfrac{1}{2}(\alpha_2-\beta_1)I.\end{aligned} \tag{7.141}$$

The preceding solutions all assume that the wave vector is entirely in the $\boldsymbol{\sigma}_3$ direction. More generally, we can introduce a right-handed coordinate frame $\{\mathsf{e}_i\}$, with e_3 pointing along the direction of propagation. The solutions then all generalise straightforwardly. In more covariant notation the circularly-polarised modes can also be written

$$F = kn\big(\alpha_R\mathrm{e}^{Ik\cdot x}+\alpha_L\mathrm{e}^{-Ik\cdot x}\big), \tag{7.142}$$

where $k\cdot n = 0$.

7.4.2 Stokes parameters

A useful way of describing the state of polarisation in light emitted from some source is through the Stokes parameters. The general definition of these involves time averages of the fields, which we denote here with an overbar. To start with we assume that the light is coherent, so that all modes are in the same state. We first define the Stokes parameters in terms of the plane-polarised coefficients. The electric field is given by

$$\boldsymbol{E} = \mathrm{Re}\big((c_1\boldsymbol{\sigma}_1 + c_2\boldsymbol{\sigma}_2)\mathrm{e}^{-ik\cdot x}\big) = \mathrm{Re}(\mathcal{E}), \tag{7.143}$$

where $\mathcal{E}$ denotes the complex amplitude. The first Stokes parameter gives the magnitude of the electric field,

$$s_0 = 2\overline{\boldsymbol{E}^2} = \langle\mathcal{E}\mathcal{E}^*\rangle, \tag{7.144}$$

where the star denotes complex conjugation. This evaluates straightforwardly to

$$s_0 = |c_1|^2 + |c_2|^2. \tag{7.145}$$

The remaining three Stokes parameters describe the relative amounts of radiation present in various polarisation states. If we denote the real components of $\boldsymbol{E}$ by E_x and E_y the parameters are defined by

$$\begin{aligned} s_1 &= 2(\overline{E_x^2} - \overline{E_y^2}) = |c_1|^2 - |c_2|^2 \\ s_2 &= 4\overline{E_x E_y} = 2\mathrm{Re}(c_1 c_2^*) \\ s_3 &= 4\overline{E_x(t)E_y(t+\pi/(2\omega))} = -2\mathrm{Im}(c_1 c_2^*). \end{aligned} \tag{7.146}$$

The Stokes parameters can equally well be written in terms of the α_L and α_R coefficients of circularly-polarised modes:

$$\begin{aligned} s_0 &= 2(|\alpha_L|^2 + |\alpha_R|^2), \\ s_1 &= 4\langle\alpha_L\alpha_R\rangle, \\ s_2 &= -4\langle I\alpha_L\alpha_R\rangle, \\ s_3 &= 2(|\alpha_L^2| - |\alpha_R|^2). \end{aligned} \tag{7.147}$$

For coherent light the Stokes parameters are related by

$$s_0^2 = s_1^2 + s_2^2 + s_3^2. \tag{7.148}$$

The s_μ can therefore be viewed algebraically as the components of a null vector, though its direction in space has no physical significance. This representation for 'observables' in terms of a null vector is typical of a two-state quantum system. We can bring this out neatly in the spacetime algebra by introducing the three-dimensional rotor

$$\kappa = \langle\alpha_L\rangle + \langle I\alpha_L\rangle I\boldsymbol{\sigma}_3 - \langle\alpha_R\rangle I\boldsymbol{\sigma}_2 - \langle I\alpha_R\rangle I\boldsymbol{\sigma}_1. \tag{7.149}$$

The (quantum) origin of this object is explained in section 8.1. The rotor κ satisfies

$$\kappa\kappa^\dagger = \tfrac{1}{2}s_0, \qquad \kappa\boldsymbol{\sigma}_3\kappa^\dagger = \tfrac{1}{2}s_i\boldsymbol{\sigma}_i. \tag{7.150}$$

It follows that in spacetime

$$2\kappa(\gamma_0+\gamma_3)\tilde{\kappa} = 2\kappa(1+\boldsymbol{\sigma}_3)\kappa^\dagger\gamma_0 = s_0\gamma_0 + s_i\gamma_i, \tag{7.151}$$

and since we have rotated a null vector we automatically obtain a null vector. The unit spatial vector

$$\hat{\boldsymbol{s}} = \frac{\boldsymbol{s}}{s_0}, \qquad \boldsymbol{s} = s_i\boldsymbol{\sigma}_i \tag{7.152}$$

can be represented by a point on a sphere. For light polarisation states this is called the *Poincaré sphere*. For spin-1/2 systems the equivalent construction is known as the *Bloch sphere*. The construction is also useful for describing partially coherent light. In this case the light can be viewed as originating from a set of discrete (incoherent) sources. The single null vector is replaced by an average over the sources,

$$s = \sum_{k=1}^{n} s_k \tag{7.153}$$

and the unit vector $\hat{\boldsymbol{s}}$ is replaced by

$$\boldsymbol{s} = \frac{s\wedge\gamma_0}{s\cdot\gamma_0} = \sum_{k=1}^{n}\frac{\omega_k}{\omega}\hat{\boldsymbol{s}}_k, \qquad \omega = \sum_{k=1}^{n}\omega_k. \tag{7.154}$$

The resulting polarisation vector $\boldsymbol{s}$ has $\boldsymbol{s}^2 \leq 1$, so now defines a vector inside the Poincaré sphere. The length of this vector directly encodes the relative amounts of coherent and incoherent light present.

The preceding discussion also makes it a simple matter to compute how the Stokes parameters appear to observers moving at different velocities. Suppose that a second observer with velocity $v = \mathsf{e}_0$ sets up a frame $\{\mathsf{e}_\mu\}$. This is done in such a way that the wave vector still travels in the e_3 direction, which requires that

$$\mathsf{e}_3 = \frac{k - k\cdot v\, v}{k\cdot v}. \tag{7.155}$$

If the old and new frames are related by a rotor, $\mathsf{e}_\mu = R\gamma_\mu\tilde{R}$, then equation (7.155) restricts R to satisfy

$$Rk\tilde{R} = \lambda k. \tag{7.156}$$

Rather than work in the new frame, it is simpler to back-transform the field F and work in the original $\{\gamma_\mu\}$ frame. We define

$$F' = \tilde{R}F(Rx\tilde{R})R = \frac{1}{\lambda}kn'\left(\alpha_R e^{Ik\cdot x/\lambda} + \alpha_L e^{-Ik\cdot x/\lambda}\right), \tag{7.157}$$

where $n' = \tilde{R}nR$ and $k = \omega(\gamma_0+\gamma_3)$. We can again choose n' to be perpendicular to γ_0 by adding an appropriate multiple of k. It follows that the only change to the final vector n can be a rotation in the $I\boldsymbol{\sigma}_3$ plane. Performing a spacetime split on γ_0, and assuming that the original n was $-\gamma_1$, we obtain

$$F' = \frac{1}{\lambda}(1+\boldsymbol{\sigma}_3)\boldsymbol{\sigma}_1 \mathrm{e}^{-\phi I\boldsymbol{\sigma}_3}\left(\alpha_R \mathrm{e}^{Ik\cdot x/\lambda} + \alpha_L \mathrm{e}^{-Ik\cdot x/\lambda}\right), \tag{7.158}$$

where ϕ is the angle of rotation in the $I\boldsymbol{\sigma}_3$ plane. The rotation can again be converted to a phase factor on I, so the overall change is that α_R and α_L are multiplied by $\lambda^{-1}\exp(I\phi)$. The rescaling has no effect on the unit vector on the Poincaré sphere, so the only change is a rotation through 2ϕ in the $I\boldsymbol{\sigma}_3$ plane. This implies that the $\boldsymbol{\sigma}_3$ component of the vector on the Poincaré sphere is constant, which is sensible. This component determines the relative amounts of left and right-circularly-polarised light present, and this ratio is independent of which observer measures it. Similar arguments apply to the case of partially coherent light.

7.5 Scattering and diffraction

We turn now to the related subjects of the scattering and diffraction of electromagnetic waves. This is an enormous subject and our aim here is to provide little more than an introduction, highlighting in particular a unified approach based on the free-space multivector Green's function. This provides a first-order formulation of the scattering problem, which is valuable in numerical computation. We continue to adopt a complex representation for the electromagnetic field, and will concentrate on waves of a single frequency. The time dependence is then expressed via

$$F(x) = F(\boldsymbol{r})\mathrm{e}^{-i\omega t}, \tag{7.159}$$

so that the Maxwell equations reduce to

$$\boldsymbol{\nabla}F - i\omega F = 0. \tag{7.160}$$

This is the first-order equivalent of the vector Helmholtz equation. Throughout this section we work with the full, complex quantities, and suppress all factors of $\exp(i\omega t)$. All quadratic quantities are assumed to be time averaged.

If sources are present the Maxwell equations become

$$(\boldsymbol{\nabla} - i\omega)F = \rho - \boldsymbol{J}. \tag{7.161}$$

Current conservation tells us that the (complex) current satisfies

$$i\omega\rho = \boldsymbol{\nabla}\cdot\boldsymbol{J}. \tag{7.162}$$

Provided that all the sources are localised in some region in space, there can be

no electric monopole term present. This follows because

$$Q = \int |dX|\rho = \frac{1}{i\omega} \oint \boldsymbol{J}\cdot\boldsymbol{n}\,|dA|, \tag{7.163}$$

where $\boldsymbol{n}$ is the outward normal. Taking the surface to totally enclose the sources, so that $\boldsymbol{J}$ vanishes over the surface of integration, we see that $Q = 0$.

7.5.1 First-order Green's function

The main result we employ in this section is Green's theorem in three dimensions in the general form

$$\int_V (\dot{G}\dot{\boldsymbol{\nabla}}\, F + G\,\boldsymbol{\nabla} F)\,|dX| = \oint_{\partial V} G\boldsymbol{n}F\,dA \tag{7.164}$$

where $\boldsymbol{n}$ is the outward-pointing normal vector over the surface ∂V. If F satisfies the vacuum Maxwell equations, we have

$$\oint_{\partial V} G\boldsymbol{n}F\,dA = \int_V (\dot{G}\dot{\boldsymbol{\nabla}} + i\omega G)F\,|dX|. \tag{7.165}$$

We therefore seek a Green's function satisfying

$$\dot{G}\dot{\boldsymbol{\nabla}} + i\omega G = \delta(\boldsymbol{r}). \tag{7.166}$$

It will turn out that G only contains (complex) scalar and vector terms, so (by reversing both sides) this equation is equivalent to

$$(\boldsymbol{\nabla} + i\omega)G = \delta(\boldsymbol{r}). \tag{7.167}$$

The Green's function is easily found from the Green's function for the (scalar) Helmholtz equation,

$$\phi(\boldsymbol{r}) = -\frac{1}{4\pi r}\mathrm{e}^{i\omega r}. \tag{7.168}$$

This is appropriate for *outgoing* radiation. Choosing the outgoing Green's function is equivalent to imposing causality by working with retarded fields. The function ϕ satisfies

$$(\boldsymbol{\nabla}^2 + \omega^2)\phi = \delta(\boldsymbol{r}) = (\boldsymbol{\nabla} + i\omega)(\boldsymbol{\nabla} - i\omega)\phi. \tag{7.169}$$

We therefore see that the required first-order Green's function is

$$\begin{aligned} G(\boldsymbol{r}) &= (\boldsymbol{\nabla} - i\omega)\phi \\ &= \frac{\mathrm{e}^{i\omega r}}{4\pi}\left(\frac{i\omega}{r}(1 - \boldsymbol{\sigma}_r) + \frac{\boldsymbol{r}}{r^3}\right), \end{aligned} \tag{7.170}$$

where $\boldsymbol{\sigma}_r = \boldsymbol{r}/r$ is the unit vector in the direction of $\boldsymbol{r}$. This Green's function is the key to much of scattering theory. With a general argument it satisfies

$$(\boldsymbol{\nabla} + i\omega)G(\boldsymbol{r} - \boldsymbol{r}') = \delta(\boldsymbol{r} - \boldsymbol{r}') \tag{7.171}$$

or, equivalently,

$$(\boldsymbol{\nabla}' - i\omega)G(\boldsymbol{r} - \boldsymbol{r}') = -\delta(\boldsymbol{r} - \boldsymbol{r}'), \tag{7.172}$$

where $\boldsymbol{\nabla}'$ denotes the vector derivative with respect to $\boldsymbol{r}'$.

7.5.2 Radiation and multipole fields

As a first application, suppose that a localised system of charges in free space, with sinusoidal time dependence, generates outgoing radiation fields. We could find these by generalising our point source solutions of section 7.3, but here we wish to exploit our new Green's function. We can now immediately write down the solution

$$F(\boldsymbol{r}) = -\int_V G(\boldsymbol{r}' - \boldsymbol{r})\big(\rho(\boldsymbol{r}') - \boldsymbol{J}(\boldsymbol{r}')\big)\,|dX'|, \tag{7.173}$$

where the integral is over a volume enclosing all of the sources. Equation (7.172) guarantees that this equation solves the Maxwell equations (7.161), subject to the boundary condition that only outgoing waves are present at large distances. It is worth stressing that the geometric algebra formulation is crucial to the way we have a single integral yielding both the electric and magnetic fields.

Often, one is mainly interested in the radiation fields present at large distances from the source. These are the contributions to F which fall off as $1/r$. To isolate these terms we use the expansion

$$\mathrm{e}^{i\omega|\boldsymbol{r} - \boldsymbol{r}'|} = \mathrm{e}^{i\omega r}\mathrm{e}^{-i\omega\boldsymbol{\sigma}_r\cdot\boldsymbol{r}'} + \mathrm{O}(r^{-1}), \tag{7.174}$$

so that the Green's function satisfies

$$\lim_{r\mapsto\infty} G(\boldsymbol{r}' - \boldsymbol{r}) = \frac{i\omega}{4\pi r}\mathrm{e}^{i\omega r}(1 + \boldsymbol{\sigma}_r)\mathrm{e}^{-i\omega\boldsymbol{\sigma}_r\cdot\boldsymbol{r}'}. \tag{7.175}$$

We therefore find that the limiting form of F can be written

$$F(\boldsymbol{r}) = -\frac{i\omega}{4\pi r}\mathrm{e}^{i\omega r}(1 + \boldsymbol{\sigma}_r)\int \mathrm{e}^{-i\omega\boldsymbol{\sigma}_r\cdot\boldsymbol{r}'}\big(\rho(\boldsymbol{r}') - \boldsymbol{J}(\boldsymbol{r}')\big)\,|dX'|. \tag{7.176}$$

As expected, the multivector is controlled by the idempotent term $(1 + \boldsymbol{\sigma}_r) = (\gamma_0 + e_r)\gamma_0$, appropriate for outgoing radiation.

A multipole expansion of the radiation field is achieved by expanding (7.176) in a series in ωd, where d is the dimension of the source. To leading order, and recalling that no monopole term is present, we find that

$$\begin{aligned}\int \mathrm{e}^{-i\omega\boldsymbol{\sigma}_r\cdot\boldsymbol{r}'}\big(\rho(\boldsymbol{r}') - \boldsymbol{J}(\boldsymbol{r}')\big)\,|dX'| &\approx \int(-\boldsymbol{J} - i\omega\rho\boldsymbol{\sigma}_r\cdot\boldsymbol{r}')|dX'| \\ &= \int(-\boldsymbol{J} + \boldsymbol{\sigma}_r\cdot\boldsymbol{J})\,|dX'|,\end{aligned} \tag{7.177}$$

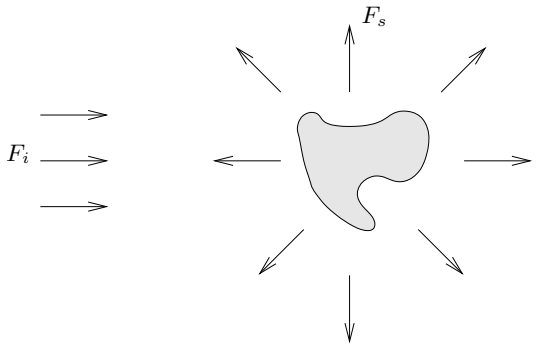

Figure 7.10 *Scattering by a localised object.* The incident field F_i sets up oscillating currents in the object, which generate an outgoing radiation field F_s.

where we have integrated by parts to obtain the final expression. This result is more commonly expressed in terms of the *electric dipole moment* $\boldsymbol{p}$, via

$$\int \boldsymbol{J}\,|dX| = -\int \boldsymbol{r}\,\boldsymbol{\nabla}\cdot\boldsymbol{J}\,|dX| = -i\omega\int \boldsymbol{r}\rho(\boldsymbol{r})\,|dX| = -i\omega\boldsymbol{p}. \tag{7.178}$$

The result is that the F field is given by

$$F(\boldsymbol{r}) = \frac{\omega^2}{4\pi r}\mathrm{e}^{i\omega r}(1+\boldsymbol{\sigma}_r)(\boldsymbol{p}-\boldsymbol{\sigma}_r\cdot\boldsymbol{p}). \tag{7.179}$$

An immediate check is that the scalar term in F vanishes, as it must. The electric and magnetic dipole fields can be read off easily now as

$$\boldsymbol{E} = \frac{\omega^2}{4\pi r}\mathrm{e}^{i\omega r}\boldsymbol{\sigma}_r\,\boldsymbol{\sigma}_r\wedge\boldsymbol{p}, \qquad I\boldsymbol{B} = \frac{\omega^2}{4\pi r}\mathrm{e}^{i\omega r}\boldsymbol{\sigma}_r\wedge\boldsymbol{p}. \tag{7.180}$$

These formulae are quite general for any (classical) radiating object.

7.6 Scattering

The geometry of a basic scattering problem is illustrated in figure 7.10. A (known) field F_i is incident on a localised object. Usually the incident radiation is taken to be a plane wave. This radiation sets up oscillating currents in the scatterer, which in turn generate a scattered field F_s. The total field F is given by

$$F = F_i + F_s, \tag{7.181}$$

and both F_i and F_s satisfy the vacuum Maxwell equations away from the scatterers.

The essential difficulty is how to solve for the currents set up by the incident

radiation. This is extremely complex and a number of distinct approaches are described in the literature. One straightforward result is for scattering from a small uniform dielectric sphere. For this situation we have

$$\boldsymbol{p} = 4\pi a^3 \frac{\epsilon_r - 1}{\epsilon_r + 2} \boldsymbol{E}_i, \tag{7.182}$$

where a is the radius of the sphere. From equation (7.180) we see that the ratio of incident to scattered radiation is controlled by ω^2. This ratio determines the differential cross section via

$$\frac{d\sigma}{d\Omega} = r^2 \frac{|\boldsymbol{e}^* \cdot \boldsymbol{E}_s|^2}{|\boldsymbol{e}^* \cdot \boldsymbol{E}_i|^2}, \tag{7.183}$$

where the complex vector $\boldsymbol{e}$ determines the polarisation. The cross section clearly depends of the polarisation of the incident wave. Summing over polarisations the differential cross section is

$$\frac{d\sigma}{d\Omega} = \omega^4 a^6 \left(\frac{\epsilon_r - 1}{\epsilon_r + 2} \right)^2 \frac{1 + \cos^2(\theta)}{2}. \tag{7.184}$$

The factor of $\omega^4 = \lambda^{-4}$ is typical of Rayleigh scattering. These results are central to Rayleigh's explanation of blue skies and red sunsets.

Suppose now that we know the fields over a closed surface enclosing a volume V. Provided that F satisfies the vacuum Maxwell equations throughout V we can compute F_s directly from

$$F_s(\boldsymbol{r}') = \oint_{\partial V} G(\boldsymbol{r} - \boldsymbol{r}') \boldsymbol{n} F_s(\boldsymbol{r}) \, |dS|. \tag{7.185}$$

We take the volume V to be bounded by two surfaces, S_1 and S_2, as shown in figure 7.11. The surface S_1 is assumed to lie just outside the scatterers, so that $J = 0$ over S_1. The surface S_2 is assumed to be spherical, and is taken out to infinity. In this limit only the $1/r$ terms in G and F can contribute to the surface integral over S_2. But from equation (7.175) we know that

$$\lim_{r \mapsto \infty} G(\boldsymbol{r} - \boldsymbol{r}') = \frac{i\omega}{4\pi r} \mathrm{e}^{i\omega r} (1 - \boldsymbol{\sigma}_r) \mathrm{e}^{-i\omega \boldsymbol{\sigma}_r \cdot \boldsymbol{r}'}, \tag{7.186}$$

whereas F_s contains a factor of $(1 + \boldsymbol{\sigma}_r)$. It follows that the integrand $G\boldsymbol{n}F_s$ contains the term

$$(1 - \boldsymbol{\sigma}_r)\boldsymbol{\sigma}_r(1 + \boldsymbol{\sigma}_r) = 0. \tag{7.187}$$

This is identically zero, so there is no contribution from the surface at infinity. The result is that the scattered field is given by

$$F_s(\boldsymbol{r}) = \frac{1}{4\pi} \oint_{S_1} \mathrm{e}^{i\omega d} \left(\frac{i\omega}{d} + \frac{i\omega(\boldsymbol{r} - \boldsymbol{r}')}{d^2} - \frac{\boldsymbol{r} - \boldsymbol{r}'}{d^3} \right) \boldsymbol{n}' F_s(\boldsymbol{r}') \, |dS(\boldsymbol{r}')|, \tag{7.188}$$

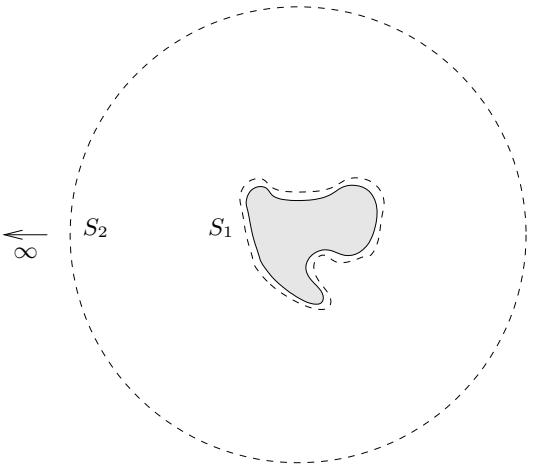

Figure 7.11 *Surfaces for Green's theorem.* The surface S_2 can be taken out to infinity, and S_1 lies just outside the scattering surface.

where

$$d = |\boldsymbol{r} - \boldsymbol{r}'|. \tag{7.189}$$

Since $\boldsymbol{n}$ is the outward pointing normal to the volume, this points *into* the scatterers. This result contains all the necessary polarisation and obliquity factors, often derived at great length in standard optics texts.

A significant advantage of this first-order approach is that it clearly embodies Huygens' principle. The scattered field F_s is propagated into the interior simply by multiplying it by a Green's function. This accords with Huygen's original idea of reradiation of wavelets from any given wavefront. Two significant problems remain, however. The first is how to specify F_s over the surface of integration. This requires detailed modelling of the polarisation currents set up by the incident radiation. A subtlety here is that we do not have complete freedom to specify F over the surface. The equation $\boldsymbol{\nabla} F = i\omega F$ implies that the components of $\boldsymbol{E}$ and $\boldsymbol{B}$ perpendicular to the boundary surface are determined by the derivatives of the components in the surface. This reduces the number of degrees of freedom in the problem from six to four, as is required for electromagnetism.

A further problem is that, even if F_s has been found, the integrals in equation (7.188) cannot be performed analytically. One can approximate to the large r regime and, after various approximations, recover Fraunhofer and Fresnel optics. Alternatively, equation (7.188) can be used as the basis for numerical simulations of scattered fields. Figure 7.12 shows the type of detailed patterns that can emerge. The plot was calculated using the two-dimensional equivalent of equation (7.188). The total energy density is shown, where the scattering

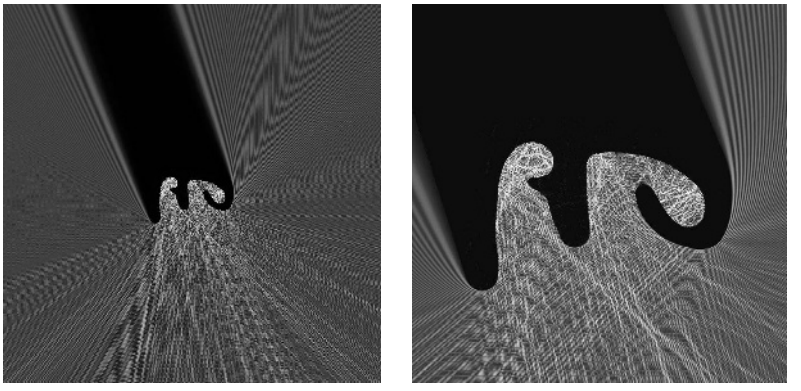

Figure 7.12 *Scattering in two dimensions.* The plots show the intensity of the electric field, with higher intensity coloured lighter. The incident radiation enters from the bottom right of the diagram and scatters off a conductor with complicated surface features. The conductor is closed in the shadow region. Various diffraction effects are clearly visible. The right-hand plot is a close-up near the surface and shows the complicated pattern of hot and cold regions that can develop.

is performed by a series of perfect conductors. A good check that the calculations have been performed correctly is that all the expected shadowing effects are present.

7.7 Notes

There is a vast literature on electromagnetism and electrodynamics. For this chapter we particularly made use of the classic texts by Jackson (1999) and Schwinger et al. (1998), both entitled *Classical Electrodynamics*. The former of these also contains an exhaustive list of further references. Applications of geometric algebra to electromagnetism are discussed in the book *Multivectors and Clifford Algebra in Electrodynamics* by Jancewicz (1989). This is largely an introductory text and stops short of tackling the more advanced applications.

We are grateful to Stephen Gull for producing the figures in section 7.3 and for stimulating much of the work described in this chapter. Further material can be found in the Banff series of lectures by Doran et al (1996a). Readers interested in the action at a distance formalism of Wheeler and Feynman can do no better than return to their original 1949 paper. It is a good exercise to convert their arguments into a more streamlined geometric algebra notation!

7.8 Exercises

7.1 A circular current loop has radius a and lies in the $z = 0$ plane with its centre at the origin. The loop carries a current J. Write down an integral expression for the $\boldsymbol{B}$ field, and show that on the z axis,

$$\boldsymbol{B} = \frac{\mu_0 J a^2}{2(a^2+z^2)^{3/2}}\boldsymbol{\sigma}_3.$$

7.2 An extension to the Maxwell equations which is regularly discussed is how they are modified in the presence of magnetic monopoles. If ρ_m and $\boldsymbol{J}_m$ denote magnetic charges and currents, the relevant equations are

$$\boldsymbol{\nabla}\cdot\boldsymbol{D} = \rho_e, \qquad \boldsymbol{\nabla}\cdot\boldsymbol{B} = \rho_m,$$
$$-\boldsymbol{\nabla}\times\boldsymbol{E} = \frac{\partial}{\partial t}\boldsymbol{B} + \boldsymbol{J}_m, \qquad \boldsymbol{\nabla}\times\boldsymbol{H} = \frac{\partial}{\partial t}\boldsymbol{D} + \boldsymbol{J}_e.$$

Prove that in free space these can be written

$$\nabla F = J_e + J_m I,$$

where $J_m = (\rho_m + \boldsymbol{J}_m)\gamma_0$. A duality transformation of the $\boldsymbol{E}$ and $\boldsymbol{B}$ fields is defined by

$$\boldsymbol{E}' = \boldsymbol{E}\cos(\alpha) + \boldsymbol{B}\sin(\alpha), \qquad \boldsymbol{B}' = \boldsymbol{B}\cos(\alpha) - \boldsymbol{B}\sin(\alpha).$$

Prove that this can be written compactly as $F' = F\mathrm{e}^{-I\alpha}$. Hence find the equivalent transformation law for the source terms such that the equations remain invariant, and prove that the electromagnetic energy-momentum tensor is also invariant under a duality transformation.

7.3 A particle follows the trajectory $x_0(\tau)$, with velocity $v = \dot{x}$ and acceleration $\dot{v}$. If X is the retarded null vector connecting the point x to the worldline, show that the electromagnetic field at x is given by

$$F = \frac{q}{4\pi}\frac{X\wedge v + \frac{1}{2}X\Omega_v X}{(X\cdot v)^3},$$

where $\Omega_v = \dot{v}\wedge v$. Prove directly that F satisfies $\nabla F = 0$ off the particle worldline.

7.4 Prove the following formulae relating the retarded A and F fields for a point charge to the null vector X:

$$A = -\frac{q}{8\pi\epsilon_0}\nabla^2 X, \qquad F = -\frac{q}{8\pi\epsilon_0}\nabla^3 X.$$

These expressions are of interest in the 'action at a distance' formulation of electrodynamics, as discussed by Wheeler and Feynman (1949).

7.5 Confirm that, at large distances for the source, the radiation fields due to both linearly and circularly accelerating charges go as

$$F_{rad} \approx \frac{1}{r}(1+\boldsymbol{\sigma}_r)\boldsymbol{a},$$

where $\boldsymbol{\sigma}_r \cdot \boldsymbol{a} = 0$.

7.6 From the solution for the fields due to a point charge in a circular orbit (section 7.3.4), explain why synchrotron radiation arrives in pulses.

7.7 For the κ defined in equation (7.149), verify that $\kappa\boldsymbol{\sigma}_3\kappa^\dagger = s_i\boldsymbol{\sigma}_i$, where s_i are Stokes parameters.

7.8 A rotor R relates two frames by $\mathsf{e}'_\mu = R\mathsf{e}_\mu\tilde{R}$. In both frames the vector e_3 vector is defined by

$$\mathsf{e}_3 = \mathsf{e}'_3 = \frac{k - k\cdot\mathsf{e}_0\,\mathsf{e}_0}{k\cdot\mathsf{e}_0},$$

where k is a fixed null vector. Prove that for this relation to be valid for both frames we must have

$$Rk\tilde{R} = \lambda k.$$

How many degrees of freedom are left in the rotor R if this equation holds?

7.9 In optical problems we are regularly interested in the effects of a planar aperture on incident plane waves. Suppose that the aperture lies in the $z = 0$ plane, and we are interested in the fields in the region $z > 0$. By introducing the Green's function

$$G'(\boldsymbol{r};\boldsymbol{r}') = G(\boldsymbol{r}-\boldsymbol{r}') - G(\boldsymbol{r}-\bar{\boldsymbol{r}}'),$$

where $\bar{\boldsymbol{r}} = -\boldsymbol{\sigma}_3\boldsymbol{r}\boldsymbol{\sigma}_3$, prove that the field in the region $z > 0$ is given by

$$F_s(\boldsymbol{r}') = \int dx\,dy\,\frac{z'\mathrm{e}^{I\omega d}}{2\pi d^3}(1-i\omega d)F_s(x,y,0), \qquad \text{(E7.1)}$$

where $d = |\boldsymbol{r}-\boldsymbol{r}'|$. In the Kirchoff approximation we assume that F_s over the aperture can be taken as the incident plane wave. By working in the large r and small angle limit, prove the Fraunhofer result that the transmitted amplitude is controlled by the Fourier transform of the aperture function.

7.10 Repeat the analysis of the previous question for a two-dimensional arrangement. You will need to understand some of the properties of Hankel functions.

8

Quantum theory and spinors

In this chapter we study the application of geometric algebra to both non-relativistic and relativistic quantum mechanics. We concentrate on the quantum theory of spin-1/2 particles, whose dynamics is described by the Pauli and Dirac equations. For interactions where spin and relativity are not important the dynamics reduces to that of the Schrödinger equation. There are many good textbooks describing this topic and we will make no attempt to cover it here. We assume, furthermore, that most readers have a basic understanding of quantum mechanics, and are familiar with the concepts of states and operators.

Both the Pauli and Dirac matrices arise naturally as representations of the geometric algebras of space and spacetime. It is no surprise, then, that much of quantum theory finds a natural expression within geometric algebra. To achieve this, however, one must reconsider the standard interpretation of the quantum spin operators. Like much discussion of the interpretation of quantum theory, certain issues raised here are controversial. There is no question about the validity of our algebraic approach, however, and little doubt about its advantages. Whether the algebraic simplifications obtained here are indicative of a deeper structure embedded in quantum mechanics is an open question.

In this chapter we only consider the quantum theory of single particles in background fields. Multiparticle systems are considered in the following chapter. Amongst the results discussed in this section are the angular separation of the Dirac equation, and a method of calculating cross sections that avoids the need for spin sums. Both of these results are used in chapter 14 for studying the behaviour of fermions in gravitational backgrounds.

8.1 Non-relativistic quantum spin

The Stern–Gerlach experiment was the first to demonstrate the quantum nature of the magnetic moment. In this experiment a beam of particles passes through

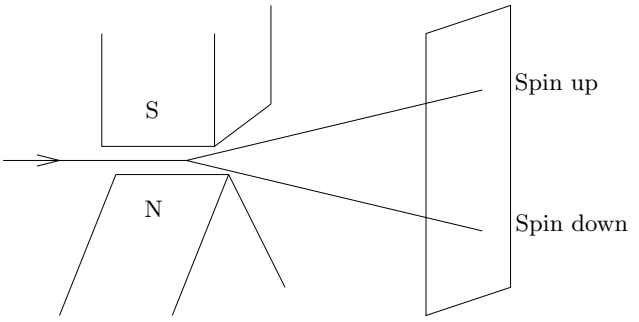

Figure 8.1 *The Stern–Gerlach experiment.* A particle beam is sent through a highly non-uniform $\boldsymbol{B}$ field. What emerges is a set of discrete, evenly-spaced beams.

a non-uniform magnetic field $\boldsymbol{B}$. Classically, one would expect the force on each particle to be governed by the equation

$$\boldsymbol{f} = \boldsymbol{\mu}\cdot\boldsymbol{\nabla}\boldsymbol{B}, \tag{8.1}$$

where $\boldsymbol{\mu}$ is the magnetic moment. This would give rise to a continuous distribution after passing through the field. Instead, what is observed is a number of evenly-spaced discrete bands (figure 8.1). The magnetic moment is quantised in the same manner as angular momentum.

When silver atoms are used to make up the beam there is a further surprise: only two beams emerge on the far side. Silver atoms contain a single electron in their outermost shell, so it looks as if electrons have an intrinsic angular momentum which can take only two values. This is known as its *spin*, though no classical picture should be inferred from this name. The double-valued nature of the spin suggests that the electron's wavefunction should contain two terms, representing a superposition of the possible spin states,

$$|\psi\rangle = \alpha|\uparrow\rangle + \beta|\downarrow\rangle, \tag{8.2}$$

where α and β are complex numbers. Such a state can be represented in matrix form as the *spinor*

$$|\psi\rangle = \begin{pmatrix}\alpha\\ \beta\end{pmatrix}. \tag{8.3}$$

If we align the z axis with the spin-up direction, then the operator returning the spin along the z axis must be

$$\hat{s}_3 = \lambda\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}, \tag{8.4}$$

where λ is to be determined. The spin is added to the orbital angular momentum

to give a conserved total angular momentum operator $\hat{\boldsymbol{j}} = \hat{\boldsymbol{l}} + \hat{\boldsymbol{s}}$. For this to make sense the spin operators should have the same commutation relations as the angular momentum operators $\hat{l}_i$,

$$\hat{l}_i = -i\hbar\epsilon_{ijk}x_j\partial_k, \quad [\hat{l}_i, \hat{l}_j] = i\hbar\epsilon_{ijk}\hat{l}_k. \tag{8.5}$$

This is sufficient to specify the remaining operators, up to an arbitrary phase (see exercise 8.1). The result is that the spin operators are given by

$$\hat{s}_k = \tfrac{1}{2}\hbar\hat{\sigma}_k, \tag{8.6}$$

where the $\hat{\sigma}_k$ are the familiar *Pauli matrices*

$$\hat{\sigma}_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \hat{\sigma}_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \hat{\sigma}_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{8.7}$$

The 'hat' notation is used to record the fact that these are viewed explicitly as matrix operators, rather than as elements of a geometric algebra. The Pauli matrices satisfy the commutation relations,

$$[\hat{\sigma}_i, \hat{\sigma}_j] = 2i\epsilon_{ijk}\hat{\sigma}_k. \tag{8.8}$$

They also have the property that two different matrices *anticommute*,

$$\hat{\sigma}_1\hat{\sigma}_2 + \hat{\sigma}_2\hat{\sigma}_1 = 0, \quad \text{etc.} \tag{8.9}$$

and all of the matrices square to the identity matrix,

$$\hat{\sigma}_1^2 = \hat{\sigma}_2^2 = \hat{\sigma}_3^2 = \mathsf{I}. \tag{8.10}$$

These are precisely the relations obeyed by a set of orthonormal vectors in space. We denote such a set by $\{\boldsymbol{\sigma}_k\}$. The crucial distinction is that the Pauli matrices are operators in quantum isospace, whereas the $\{\boldsymbol{\sigma}_k\}$ are vectors in real space.

The $\hat{\sigma}_k$ operators act on two-component complex spinors as described in equation (8.3). Spinors belong to two-dimensional complex vector space, so have four real degrees of freedom. A natural question to ask is whether an equivalent representation can be found in terms of real multivectors, such that the matrix action is replaced by multiplication by the $\{\boldsymbol{\sigma}_k\}$ vectors. To find a natural way to do this we consider the observables of a spinor. These are the eigenvalues of Hermitian operators and, for two-state systems, the relevant operators are the Pauli matrices. We therefore form the three observables

$$s_k = \tfrac{1}{2}\hbar n_k = \langle\psi|\hat{s}_k|\psi\rangle. \tag{8.11}$$

The n_k are the components of a single vector in the quantum theory of spin. Focusing attention on the components of this vector, we have

$$\begin{aligned} n_1 &= \langle\psi|\hat{\sigma}_1|\psi\rangle = \alpha\beta^* + \alpha^*\beta, \\ n_2 &= \langle\psi|\hat{\sigma}_2|\psi\rangle = i(\alpha\beta^* - \alpha^*\beta), \\ n_3 &= \langle\psi|\hat{\sigma}_3|\psi\rangle = \alpha\alpha^* - \beta\beta^*. \end{aligned} \tag{8.12}$$

The magnitude of the vector with components n_k is

$$\begin{aligned} |\boldsymbol{n}|^2 &= (\alpha\beta^* + \alpha^*\beta)^2 - (\alpha\beta^* - \alpha^*\beta)^2 + (\alpha\alpha^* - \beta\beta^*)^2 \\ &= (|\alpha|^2 + |\beta|^2)^2 = \langle\psi|\psi\rangle^2. \end{aligned} \tag{8.13}$$

So, provided the state is normalised to 1, the vector $\boldsymbol{n}$ must have unit length. We can therefore introduce polar coordinates and write

$$\begin{aligned} n_1 &= \sin(\theta)\cos(\phi), \\ n_2 &= \sin(\theta)\sin(\phi), \\ n_3 &= \cos(\theta). \end{aligned} \tag{8.14}$$

Comparing equation (8.14) with equation (8.12) we see that we must have

$$\alpha = \cos(\theta/2)\mathrm{e}^{i\gamma}, \qquad \beta = \sin(\theta/2)\mathrm{e}^{i\delta} \tag{8.15}$$

where $\delta - \gamma = \phi$. It follows that the spinor can be written in terms of the polar coordinates of the vector observable as

$$|\psi\rangle = \begin{pmatrix} \cos(\theta/2)\mathrm{e}^{-i\phi/2} \\ \sin(\theta/2)\mathrm{e}^{i\phi/2} \end{pmatrix} \mathrm{e}^{i(\gamma+\delta)/2}. \tag{8.16}$$

The overall phase factor can be ignored, and what remains is a description in terms of half-angles. This suggests a strong analogy with *rotors*. To investigate this analogy, we use the idea that polar coordinates can be viewed as part of an instruction to rotate the 3 axis onto the chosen vector. To expose this we write the vector $\boldsymbol{n}$ as

$$\boldsymbol{n} = \sin(\theta)\big(\cos(\phi)\boldsymbol{\sigma}_1 + \sin(\phi)\boldsymbol{\sigma}_2\big) + \cos(\theta)\boldsymbol{\sigma}_3. \tag{8.17}$$

This can be written

$$\boldsymbol{n} = R\boldsymbol{\sigma}_3 R^\dagger, \tag{8.18}$$

where

$$R = \mathrm{e}^{-\phi I\boldsymbol{\sigma}_3/2}\mathrm{e}^{-\theta I\boldsymbol{\sigma}_2/2}. \tag{8.19}$$

This suggests that there should be a natural map between the normalised spinor of equation (8.16) and the rotor R. Both belong to linear spaces of real dimension four and both are normalised. Expanding out the rotor R the following one-to-one map is found:

$$|\psi\rangle = \begin{pmatrix} a^0 + ia^3 \\ -a^2 + ia^1 \end{pmatrix} \leftrightarrow \psi = a^0 + a^k I\boldsymbol{\sigma}_k. \tag{8.20}$$

This map will enable us to perform all operations involving spinors without leaving the geometric algebra of space. Throughout this chapter we use the $\leftrightarrow$ symbol to denote a one-to-one map between conventional quantum mechanics and the multivector equivalent. We will continue to refer to the multivector ψ

as a spinor. On this scheme the spin-up and spin-down basis states $|\uparrow\rangle$ and $|\downarrow\rangle$ become

$$|\uparrow\rangle \leftrightarrow 1 \qquad |\downarrow\rangle \leftrightarrow -I\boldsymbol{\sigma}_2. \tag{8.21}$$

One can immediately see for these that the vectors of observables have components $(0, 0, \pm 1)$, as required.

8.1.1 Pauli operators

Now that a suitable one-to-one map has been found, we need to find a representation for Pauli operators acting on the multivector version of a spinor. It turns out that the action of the quantum $\hat{\sigma}_k$ operators on a state $|\psi\rangle$ is equivalent to the following operation on ψ:

$$\hat{\sigma}_k|\psi\rangle \leftrightarrow \boldsymbol{\sigma}_k\psi\boldsymbol{\sigma}_3 \quad (k = 1, 2, 3). \tag{8.22}$$

The $\boldsymbol{\sigma}_3$ on the right-hand side ensures that the multivector remains in the even subalgebra. The choice of vector does not break rotational covariance, in the same way that choosing the $\hat{\sigma}_3$ matrix to be diagonal does not alter the rotational covariance of the Pauli theory. One can explicitly verify that the translation procedure of equation (8.20) and equation (8.22) is consistent by routine computation; for example

$$\hat{\sigma}_1|\psi\rangle = \begin{pmatrix} -a^2 + ia^1 \\ a^0 + ia^3 \end{pmatrix} \leftrightarrow -a^2 + a^1 I\boldsymbol{\sigma}_3 - a^0 I\boldsymbol{\sigma}_2 + a^3 I\boldsymbol{\sigma}_1 = \boldsymbol{\sigma}_1\psi\boldsymbol{\sigma}_3. \tag{8.23}$$

The remaining cases, for $\hat{\sigma}_2$ and $\hat{\sigma}_3$ can be checked equally easily.

Now that we have a translation for the action of the Pauli matrices, we can find the equivalent of multiplying by the unit imaginary i. To find this we note that

$$\hat{\sigma}_1\hat{\sigma}_2\hat{\sigma}_3 = \begin{pmatrix} i & 0 \\ 0 & i \end{pmatrix}, \tag{8.24}$$

so multiplication of both components of $|\psi\rangle$ by i can be achieved by multiplying by the product of the three matrix operators. We therefore arrive at the translation

$$i|\psi\rangle \leftrightarrow \boldsymbol{\sigma}_1\boldsymbol{\sigma}_2\boldsymbol{\sigma}_3\psi(\boldsymbol{\sigma}_3)^3 = \psi I\boldsymbol{\sigma}_3. \tag{8.25}$$

So, on this scheme, the unit imaginary of quantum theory is replaced by right multiplication by the *bivector* $I\boldsymbol{\sigma}_3$. This is certainly suggestive, though it should be borne in mind that this conclusion is a feature of our chosen representation. The appearance of the bivector $I\boldsymbol{\sigma}_3$ is to be expected, since the vector of observables $\boldsymbol{s} = s_k\boldsymbol{\sigma}_k$ was formed by rotating the $\boldsymbol{\sigma}_3$ vector. This vector is unchanged by rotations in the $I\boldsymbol{\sigma}_3$ plane, which provides a geometric picture of phase invariance.

8.1.2 Observables in the Pauli theory

We next need to establish the quantum inner product for our multivector form of a spinor. We first note that the Hermitian adjoint operation has $\hat{\sigma}_k^\dagger = \hat{\sigma}_k$, and reverses the order of all products. This is precisely the same as the reversion operation for multivectors in three dimensions, so the dagger symbol can be used consistently for both operations. The quantum inner product is

$$\langle\psi|\phi\rangle = (\psi_1^*, \psi_2^*)\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix} = \psi_1^*\phi_1 + \psi_2^*\phi_2, \tag{8.26}$$

where we ignore spatial integrals. For a wide range of problems the spatial and spin components of the wave function can be separated. If this is not the case then the quantum inner product should also contain an integral over all space. The result of the real part of the inner product is reproduced by

$$\mathrm{Re}\langle\psi|\phi\rangle \;\leftrightarrow\; \langle\psi^\dagger\phi\rangle, \tag{8.27}$$

so that, for example,

$$\langle\psi|\psi\rangle \;\leftrightarrow\; \langle\psi^\dagger\psi\rangle = \langle(a^0 - a^j I\boldsymbol{\sigma}_j)(a^0 + a^k I\boldsymbol{\sigma}_k)\rangle = \sum_{\alpha=0}^{3} a^\alpha a^\alpha. \tag{8.28}$$

Since

$$\langle\psi|\phi\rangle = \mathrm{Re}\langle\psi|\phi\rangle - i\mathrm{Re}\langle\psi|i\phi\rangle, \tag{8.29}$$

the full inner product can be written

$$\langle\psi|\phi\rangle \;\leftrightarrow\; \langle\psi^\dagger\phi\rangle - \langle\psi^\dagger\phi I\boldsymbol{\sigma}_3\rangle I\boldsymbol{\sigma}_3. \tag{8.30}$$

The right-hand side projects out the 1 and $I\boldsymbol{\sigma}_3$ components from the geometric product $\psi^\dagger\phi$. The result of this projection on a multivector A is written $\langle A\rangle_q$. For even-grade multivectors in three dimensions this projection has the simple form

$$\langle A\rangle_q = \tfrac{1}{2}(A + \boldsymbol{\sigma}_3 A\boldsymbol{\sigma}_3). \tag{8.31}$$

If the result of an inner product is used to multiply a second multivector, one has to remember to keep the terms in $I\boldsymbol{\sigma}_3$ to the right of the multivector. This might appear a slightly clumsy procedure at first, but it is easy to establish conventions so that manipulations are just as efficient as in the standard treatment. Furthermore, the fact that all manipulations are now performed within the geometric algebra framework offers a number of new ways to simplify the analysis of a range of problems.

8.1.3 The spin vector

As a check on the consistency of our scheme, we return to the expectation value of the spin in the k-direction, $\langle\psi|\hat{s}_k|\psi\rangle$. For this we require

$$\langle\psi|\hat{\sigma}_k|\psi\rangle \leftrightarrow \langle\psi^\dagger\boldsymbol{\sigma}_k\psi\boldsymbol{\sigma}_3\rangle - \langle\psi^\dagger\boldsymbol{\sigma}_k\psi I\rangle I\boldsymbol{\sigma}_3. \tag{8.32}$$

Since $\psi^\dagger I\boldsymbol{\sigma}_k\psi$ reverses to give minus itself it has zero scalar part, so the final term on the right-hand side vanishes. This is to be expected, as the $\hat{\sigma}_k$ are Hermitian operators. For the remaining term we note that in three dimensions $\psi\boldsymbol{\sigma}_3\psi^\dagger$ is both odd-grade and reverses to itself, so is a pure vector. We therefore define the spin vector

$$\boldsymbol{s} = \tfrac{1}{2}\hbar\psi\boldsymbol{\sigma}_3\psi^\dagger. \tag{8.33}$$

The quantum expectation now reduces to

$$\langle\psi|\hat{s}_k|\psi\rangle = \tfrac{1}{2}\hbar\langle\boldsymbol{\sigma}_k\psi\boldsymbol{\sigma}_3\psi^\dagger\rangle = \boldsymbol{\sigma}_k\cdot\boldsymbol{s}. \tag{8.34}$$

This new expression has a rather different interpretation to that usually encountered in quantum theory. Rather than forming the expectation value of a quantum operator, we are simply projecting out the kth component of the vector $\boldsymbol{s}$. Working with the vector $\boldsymbol{s}$ may appear to raise questions about whether we are free to talk about all three components of the spin vector. This is in fact consistent with the results of spin measurements, if we view the spin measurement apparatus as acting more as a spin polariser. This is discussed in Doran *et al.* (1996b).

The rotor description introduced at the start of this section is recovered by first defining the scalar

$$\rho = \psi\psi^\dagger. \tag{8.35}$$

The spinor ψ then decomposes into

$$\psi = \rho^{1/2}R, \tag{8.36}$$

where $R = \rho^{-1/2}\psi$. The multivector R satisfies $RR^\dagger = 1$, so is a rotor. In this approach, Pauli spinors are nothing but unnormalised rotors. The spin vector $\boldsymbol{s}$ can now be written as

$$\boldsymbol{s} = \tfrac{1}{2}\hbar\rho R\boldsymbol{\sigma}_3R^\dagger, \tag{8.37}$$

which recovers the form of equation (8.18).

The double-sided construction of the expectation value of equation (8.32) contains an instruction to rotate the fixed $\boldsymbol{\sigma}_3$ axis into the spin direction and dilate it. It might appear here that we are singling out some preferred direction in space. But in fact all we are doing is utilising an idea from rigid-body dynamics, as discussed in section 3.4.3. The $\boldsymbol{\sigma}_3$ on the right of ψ represents a vector in a 'reference' frame. All physical vectors, like $\boldsymbol{s}$, are obtained by rotating this frame

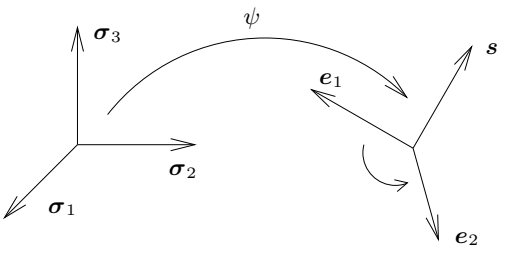

Figure 8.2 *The spin vector.* The normalised spinor ψ transforms the initial reference frame onto the frame $\{\boldsymbol{e}_k\}$. The vector $\boldsymbol{e}_3$ is the spin vector. A phase transformation of ψ generates a rotation in the $\boldsymbol{e}_1\boldsymbol{e}_2$ plane. Such a transformation is unobservable, so the $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ vectors are also unobservable.

onto the physical values (see figure 8.2). There is nothing special about $\boldsymbol{\sigma}_3$ — one can choose any (constant) reference frame and use the appropriate rotation onto $\boldsymbol{s}$, in the same way that there is nothing special about the orientation of the reference configuration of a rigid body. In rigid-body mechanics this freedom is usually employed to align the reference configuration with the initial state of the body. In quantum theory the convention is to work with the z axis as the reference vector.

8.1.4 Rotating spinors

Suppose that the vector $\boldsymbol{s}$ is to be rotated to a new vector $R_0\boldsymbol{s}R_0^\dagger$. To achieve this the spinor ψ must transform according to

$$\psi \mapsto R_0\psi. \tag{8.38}$$

Now suppose that for R_0 we use the rotor R_θ,

$$R_\theta = \exp(-\hat{B}\theta/2), \tag{8.39}$$

where $\hat{B}^2 = -1$ is a constant bivector. The resulting spinor is

$$\psi' = R_\theta\psi = \mathrm{e}^{-\hat{B}\theta/2}\psi. \tag{8.40}$$

We now start to increase θ from 0 through to 2π, so that $\theta = 2\pi$ corresponds to a 2π rotation, bringing all observables back to their original values. But under this we see that ψ transforms to

$$\psi' = \mathrm{e}^{-\hat{B}\pi}\psi = \big(\cos(\pi) - \hat{B}\sin(\pi)\big)\psi = -\psi. \tag{8.41}$$

The spinor changes sign! If a spin vector is rotated through 2π, the wavefunction does not come back to itself, but instead transforms to minus its original value.

This change of sign of a state vector under 2π rotations is the distinguishing property of spin-1/2 fermions in quantum theory. Once one sees the rotor derivation of this result, however, it is rather less mysterious. Indeed, there are classical phenomena involving systems of linked rotations that show precisely the same property. One example is the 4π symmetry observed when rotating an arm holding a tray. For a more detailed discussion if this point, see chapter 41 of *Gravitation* by Misner, Thorne & Wheeler (1973). A linear space which is acted on in a single-sided manner by rotors forms a carrier space for a spin representation of the rotation group. Elements of such a space are generally called spinors, which is why that name is adopted for our representation in terms of even multivectors.

8.1.5 Quantum particles in a magnetic field

Particles with non-zero spin also have a magnetic moment which is proportional to the spin. This is expressed as the operator relation

$$\hat{\mu}_k = \gamma \hat{s}_k, \tag{8.42}$$

where $\hat{\mu}_k$ is the magnetic moment operator, γ is the gyromagnetic ratio and $\hat{s}_k$ is the spin operator. The gyromagnetic ratio is usually written in the form

$$\gamma = g\frac{q}{2m}, \tag{8.43}$$

where m is the particle mass, q is the charge and g is the reduced gyromagnetic ratio. The reduced gyromagnetic ratios are determined experimentally to be

electron	$g_e = 2$	(actually $2(1 + \alpha/2\pi + \cdots)$),
proton	$g_p = 5.587$,	
neutron	$g_n = -3.826$	(using proton charge).

The value for the neutron is negative because its spin and magnetic moment are antiparallel. All of the above are spin-1/2 particles for which we have $\hat{s}_k = (\hbar/2)\hat{\sigma}_k$.

Now suppose that the particle is placed in a magnetic field, and that all of the spatial dynamics has been separated out. We introduce the Hamiltonian operator

$$\hat{H} = -\tfrac{1}{2}\gamma\hbar B_k\hat{\sigma}_k = -\hat{\mu}_k B_k. \tag{8.44}$$

The spin state at time t is then written as

$$|\psi(t)\rangle = \alpha(t)|\uparrow\rangle + \beta(t)|\downarrow\rangle, \tag{8.45}$$

with α and β general complex coefficients. The dynamical equation for these coefficients is given by the time-dependent Schrödinger equation

$$\hat{H}|\psi\rangle = i\hbar\frac{d|\psi\rangle}{dt}. \tag{8.46}$$

This equation can be hard to analyse, conventionally, because it involves a pair of coupled differential equations for α and β. Instead, let us see what the Schrödinger equation looks like in the geometric algebra formulation. We first write the equation in the form

$$\frac{d|\psi\rangle}{dt} = \tfrac{1}{2}\gamma i B_k\hat{\sigma}_k|\psi\rangle. \tag{8.47}$$

Now replacing $|\psi\rangle$ by the multivector ψ we see that the left-hand side is simply $\dot{\psi}$, where the dot denotes the time derivative. The right-hand side involves multiplication of the spinor $|\psi\rangle$ by $i\hat{\sigma}_k$, which we replace by

$$i\hat{\sigma}_k|\psi\rangle \leftrightarrow \boldsymbol{\sigma}_k\psi\boldsymbol{\sigma}_3(I\boldsymbol{\sigma}_3) = I\boldsymbol{\sigma}_k\psi. \tag{8.48}$$

The Schrödinger equation (8.46) is therefore simply

$$\dot{\psi} = \tfrac{1}{2}\gamma B_k I\boldsymbol{\sigma}_k\psi = \tfrac{1}{2}\gamma I\boldsymbol{B}\psi, \tag{8.49}$$

where $\boldsymbol{B} = B_k\boldsymbol{\sigma}_k$. If we now decompose ψ into $\rho^{1/2}R$ we see that

$$\dot{\psi}\psi^\dagger = \tfrac{1}{2}\dot{\rho} + \rho\dot{R}R^\dagger = \tfrac{1}{2}\rho\gamma I\boldsymbol{B}. \tag{8.50}$$

The right-hand side is a bivector, so ρ must be constant. This is to be expected, as the evolution should be unitary. The dynamics now reduces to

$$\dot{R} = \tfrac{1}{2}\gamma I\boldsymbol{B}R, \tag{8.51}$$

so the quantum theory of a spin-1/2 particle in a magnetic field reduces to a simple rotor equation. This is very natural, if one thinks about the behaviour of particles in magnetic fields, and is an important justification for our approach.

Recovering a rotor equation explains the difficulty of the traditional analysis based on a pair of coupled equations for the components of $|\psi\rangle$. This approach fails to capture the fact that there is a rotor underlying the dynamics, and so carries along a redundant degree of freedom in the normalisation. In addition, the separation of a rotor into a pair of components is far from natural. For example, suppose that $\boldsymbol{B}$ is a constant field. The rotor equation integrates immediately to give

$$\psi(t) = \mathrm{e}^{\gamma I\boldsymbol{B}t/2}\psi_0. \tag{8.52}$$

The spin vector $\boldsymbol{s}$ therefore just precesses in the $I\boldsymbol{B}$ plane at a rate $\omega_0 = \gamma|\boldsymbol{B}|$. Even this simple result is rather more difficult to establish when working with the components of $|\psi\rangle$.

8.1.6 NMR and magnetic resonance imaging

A more interesting example of a particle in a magnetic field is provided by nuclear magnetic resonance, or NMR. Suppose that the $\boldsymbol{B}$ field includes an oscillatory field $(B_1\cos(\omega t), B_1\sin(\omega t), 0)$ together with a constant field along the z axis. This oscillatory field induces transitions (spin-flips) between the up and down states, which differ in energy because of the constant component of the field. This is a very interesting system of great practical importance. It is the basis of magnetic resonance imaging and Rabi molecular beam spectroscopy.

To study this system we first write the $\boldsymbol{B}$ field as

$$B_1\big(\cos(\omega t)\boldsymbol{\sigma}_1 + \sin(\omega t)\boldsymbol{\sigma}_2\big) + B_0\boldsymbol{\sigma}_3 = S(B_1\boldsymbol{\sigma}_1 + B_0\boldsymbol{\sigma}_3)S^\dagger, \tag{8.53}$$

where

$$S = \mathrm{e}^{-\omega t I\boldsymbol{\sigma}_3/2}. \tag{8.54}$$

We now define

$$\boldsymbol{B}_c = B_1\boldsymbol{\sigma}_1 + B_0\boldsymbol{\sigma}_3 \tag{8.55}$$

so that we can write $\boldsymbol{B} = S\boldsymbol{B}_cS^\dagger$. The rotor equation now simplifies to

$$S^\dagger\dot{\psi} = \tfrac{1}{2}\gamma I\boldsymbol{B}_cS^\dagger\psi, \tag{8.56}$$

where we have pre-multiplied by $S^\dagger$, and we continue to use ψ for the normalised rotor. Now noting that

$$\dot{S}^\dagger = \tfrac{1}{2}\omega I\boldsymbol{\sigma}_3S^\dagger \tag{8.57}$$

we see that

$$\frac{d}{dt}(S^\dagger\psi) = \tfrac{1}{2}(\gamma I\boldsymbol{B}_c + \omega I\boldsymbol{\sigma}_3)S^\dagger\psi. \tag{8.58}$$

It is now $S^\dagger\psi$ that satisfies a rotor equation with a constant field. The solution is straightforward:

$$S^\dagger\psi(t) = \exp\Big(\tfrac{1}{2}\gamma t\, I\boldsymbol{B}_c + \tfrac{1}{2}\omega t\, I\boldsymbol{\sigma}_3\Big)\psi_0, \tag{8.59}$$

and we arrive at

$$\psi(t) = \exp\Big(-\tfrac{1}{2}\omega t\, I\boldsymbol{\sigma}_3\Big)\exp\Big(\tfrac{1}{2}(\omega_0+\omega)t\, I\boldsymbol{\sigma}_3 + \tfrac{1}{2}\omega_1 t\, I\boldsymbol{\sigma}_1\Big)\psi_0, \tag{8.60}$$

where $\omega_1 = \gamma B_1$. There are three separate frequencies in this solution, which contains a wealth of interesting physics.

To complete our analysis we must relate our solution to the results of experiments. Suppose that at time $t = 0$ we switch on the oscillating field. The particle is initially in a spin-up state, so $\psi_0 = 1$, which also ensures that the state is normalised. The probability that at time t the particle is in the spin-down state is

$$P_\downarrow = |\langle\downarrow|\psi(t)\rangle|^2. \tag{8.61}$$

We therefore need to form the inner product

$$\langle\downarrow|\psi(t)\rangle \leftrightarrow \langle I\boldsymbol{\sigma}_2\psi\rangle_q = \langle I\boldsymbol{\sigma}_2\psi\rangle - I\boldsymbol{\sigma}_3\langle I\boldsymbol{\sigma}_1\psi\rangle. \tag{8.62}$$

To find this inner product we write

$$\psi(t) = \mathrm{e}^{-\omega t I\boldsymbol{\sigma}_3/2}\big(\cos(\alpha t/2) + I\hat{\boldsymbol{B}}\sin(\alpha t/2)\big), \tag{8.63}$$

where

$$\hat{\boldsymbol{B}} = \frac{(\omega_0+\omega)\boldsymbol{\sigma}_3 + \omega_1\boldsymbol{\sigma}_1}{\alpha} \quad \text{and} \quad \alpha = \sqrt{(\omega+\omega_0)^2+\omega_1^2}. \tag{8.64}$$

The only term giving a contribution in the $I\boldsymbol{\sigma}_1$ and $I\boldsymbol{\sigma}_2$ planes is that in $\omega_1 I\boldsymbol{\sigma}_1/\alpha$. We therefore have

$$\langle I\boldsymbol{\sigma}_2\psi\rangle_q = \frac{\omega_1\sin(\alpha t/2)}{\alpha}\,\mathrm{e}^{-\omega t I\boldsymbol{\sigma}_3/2} I\boldsymbol{\sigma}_3 \tag{8.65}$$

and the probability is immediately

$$P_\downarrow = \left(\frac{\omega_1\sin(\alpha t/2)}{\alpha}\right)^2. \tag{8.66}$$

The maximum value is at $\alpha t = \pi$, and the probability at this time is maximised by choosing α as small as possible. This is achieved by setting $\omega = -\omega_0 = -\gamma B_0$. This is the *spin resonance condition* which is the basis of NMR spectroscopy.

8.2 Relativistic quantum states

The relativistic quantum dynamics of a spin-1/2 particle is described by the Dirac theory. The Dirac matrix operators are

$$\hat{\gamma}_0 = \begin{pmatrix} \mathsf{I} & 0 \\ 0 & -\mathsf{I} \end{pmatrix}, \quad \hat{\gamma}_k = \begin{pmatrix} 0 & -\hat{\sigma}_k \\ \hat{\sigma}_k & 0 \end{pmatrix}, \quad \hat{\gamma}_5 = \begin{pmatrix} 0 & \mathsf{I} \\ \mathsf{I} & 0 \end{pmatrix}, \tag{8.67}$$

where $\hat{\gamma}_5 = -i\hat{\gamma}_0\hat{\gamma}_1\hat{\gamma}_2\hat{\gamma}_3$ and I is the 2×2 identity matrix. These matrices act on Dirac spinors, which have four complex components (eight real degrees of freedom). We follow an analogous procedure to the Pauli case and map these spinors onto elements of the eight-dimensional even subalgebra of the spacetime algebra. Dirac spinors can be visualised as decomposing into 'upper' and 'lower' components,

$$|\psi\rangle = \begin{pmatrix} |\phi\rangle \\ |\eta\rangle \end{pmatrix}, \tag{8.68}$$

where $|\phi\rangle$ and $|\eta\rangle$ are a pair of two-component spinors. We already know how to represent these as multivectors ϕ and η, which lie in the space of scalars +

relative bivectors. Our map from the Dirac spinor onto an element of the full eight-dimensional subalgebra is simply

$$|\psi\rangle = \begin{pmatrix} |\phi\rangle \\ |\eta\rangle \end{pmatrix} \leftrightarrow \psi = \phi + \eta\boldsymbol{\sigma}_3. \tag{8.69}$$

The action of the Dirac matrix operators now becomes,

$$\begin{aligned} \hat{\gamma}_\mu|\psi\rangle &\leftrightarrow \gamma_\mu\psi\gamma_0 \quad (\mu = 0, \ldots, 3), \\ i|\psi\rangle &\leftrightarrow \psi\, I\boldsymbol{\sigma}_3, \\ \hat{\gamma}_5|\psi\rangle &\leftrightarrow \psi\boldsymbol{\sigma}_3. \end{aligned} \tag{8.70}$$

Again, verifying the details of this map is a matter of routine computation. One feature is that we now have two 'reference' vectors that can appear on the right-hand side of ψ: γ_0 and γ_3. That is, the relative vector $\boldsymbol{\sigma}_3$ used in the Pauli theory has been decomposed into a spacelike and a timelike direction. As in the Pauli theory, these reference vectors multiplying ψ from the right do not break Lorentz covariance, as all observables are formed by rotating these reference vectors onto the frame of observables. Since $I\boldsymbol{\sigma}_3$ and γ_0 commute, our use of right-multiplication by $I\boldsymbol{\sigma}_3$ for the complex structure remains consistent.

The goal of our approach is to perform all calculations without ever having to introduce an explicit matrix representation. The explicit map of equation (8.69) is for column spinors written in the Dirac–Pauli representation, but it is a simple matter to establish similar maps for other representations. All one needs to do is find the unitary matrix which transforms the second representation into the Dirac–Pauli one, and then apply the map of equation (8.69). All of the matrix operators are then guaranteed to have the equivalence defined in equation (8.70). Certain other operations, such as complex conjugation, depend on the particular representation. But rather than think of these as the same operation in different representations, it is simpler to view them as different operations which can be applied to the multivector ψ.

In order to discuss the observables of the Dirac theory, we must first distinguish between the Hermitian and Dirac adjoints. The Hermitian adjoint is written as usual as $\langle\psi|$. The Dirac adjoint is written as $\langle\bar{\psi}|$ and is defined by

$$\langle\bar{\psi}| = (\langle\psi_u|, -\langle\psi_l|), \tag{8.71}$$

where the subscripts u and l refer to the upper and lower components. It is the Dirac adjoint which gives Lorentz-covariant observables. The Dirac inner product decomposes into

$$\langle\bar{\psi}|\phi\rangle = \langle\psi_u|\phi_u\rangle - \langle\psi_l|\phi_l\rangle. \tag{8.72}$$

This has the equivalent form

$$\langle\psi_u^\dagger\phi_u\rangle_q - \langle\psi_l^\dagger\phi_l\rangle_q = \langle(\psi_u^\dagger - \boldsymbol{\sigma}_3\psi_l^\dagger)(\phi_u + \phi_l\boldsymbol{\sigma}_3)\rangle_q = \langle\tilde{\psi}\phi\rangle_q. \tag{8.73}$$

So the Dirac adjoint is replaced by the manifestly covariant operation of spacetime reversion in the spacetime algebra formulation. The Hermitian adjoint now becomes

$$\langle\psi| \leftrightarrow \psi^\dagger = \gamma_0\tilde{\psi}\gamma_0, \tag{8.74}$$

which defines the meaning of the dagger symbol in the full spacetime algebra. Clearly, this operation requires singling out a preferred timelike vector, so is not covariant. In the relative space defined by γ_0, the Hermitian adjoint reduces to the non-relativistic reverse operation, so our notation is consistent with the use of the dagger for the reverse in three-dimensional space.

We can now look at the main observables formed from a Dirac spinor. The first is the current

$$J_\mu = \langle\bar{\psi}|\hat{\gamma}_\mu|\psi\rangle \leftrightarrow \langle\tilde{\psi}\gamma_\mu\psi\gamma_0\rangle - \langle\tilde{\psi}\gamma_\mu\psi I\gamma_3\rangle I\boldsymbol{\sigma}_3. \tag{8.75}$$

The final term contains $\langle\gamma_\mu\psi I\gamma_3\tilde{\psi}\rangle$. This vanishes because $\psi I\gamma_3\tilde{\psi}$ is odd-grade and reverses to minus itself, so is a pure trivector. Similarly, $\psi\gamma_0\tilde{\psi}$ is a pure vector, and we are left with

$$J_\mu = \langle\bar{\psi}|\hat{\gamma}_\mu|\psi\rangle \leftrightarrow \gamma_\mu\cdot(\psi\gamma_0\tilde{\psi}). \tag{8.76}$$

As with the Pauli theory, the operation of taking the expectation value of a matrix operator is replaced by that of picking out a component of a vector. We can therefore reconstitute the full vector J and write

$$J = \psi\gamma_0\tilde{\psi} \tag{8.77}$$

for the first of our observables.

To gain some further insight into the form of J, and its formation from ψ, we introduce the scalar + pseudoscalar quantity $\psi\tilde{\psi}$ as

$$\psi\tilde{\psi} = \rho\mathrm{e}^{I\beta}. \tag{8.78}$$

Factoring this out from ψ, we define the spacetime rotor R:

$$R = \psi\rho^{-1/2}\mathrm{e}^{-I\beta/2}, \qquad R\tilde{R} = 1. \tag{8.79}$$

(If $\rho = 0$ a slightly different procedure can be used.) We have now decomposed the spinor ψ into

$$\psi = \rho^{1/2}\mathrm{e}^{I\beta/2}R, \tag{8.80}$$

which separates out a density ρ and the rotor R. The remaining factor of β is curious. It turns out that plane-wave particle states have $\beta = 0$, whereas antiparticle states have $\beta = \pi$. The picture for bound state wavefunctions is more complicated, however, and β appears to act as a remnant of multiparticle

Bilinear covariant	Standard form	STA equivalent	Frame-free form
Scalar	$\langle\bar{\psi}\vert\psi\rangle$	$\langle\psi\tilde{\psi}\rangle$	$\rho\cos(\beta)$
Vector	$\langle\bar{\psi}\vert\hat{\gamma}_\mu\vert\psi\rangle$	$\gamma_\mu\cdot(\psi\gamma_0\tilde{\psi})$	$\psi\gamma_0\tilde{\psi}=J$
Bivector	$\langle\bar{\psi}\vert i\hat{\gamma}_{\mu\nu}\vert\psi\rangle$	$(\gamma_\mu\wedge\gamma_\nu)\cdot(\psi I\boldsymbol{\sigma}_3\tilde{\psi})$	$\psi I\boldsymbol{\sigma}_3\tilde{\psi}=S$
Pseudovector	$\langle\bar{\psi}\vert\hat{\gamma}_\mu\hat{\gamma}_5\vert\psi\rangle$	$\gamma_\mu\cdot(\psi\gamma_3\tilde{\psi})$	$\psi\gamma_3\tilde{\psi}=s$
Pseudoscalar	$\langle\bar{\psi}\vert i\hat{\gamma}_5\vert\psi\rangle$	$\langle\psi\tilde{\psi}I\rangle$	$-\rho\sin(\beta)$

Table 8.1 *Observables in the Dirac theory.* The standard expressions for the bilinear covariants are shown, together with their spacetime algebra (STA) equivalents.

effects from the full quantum field theory. With this decomposition of ψ, the current becomes

$$J=\psi\gamma_0\tilde{\psi}=\rho\mathrm{e}^{I\beta/2}R\gamma_0\tilde{R}\mathrm{e}^{I\beta/2}=\rho R\gamma_0\tilde{R}. \tag{8.81}$$

So the rotor is now an instruction to rotate γ_0 onto the direction of the current. This is precisely the picture we adopted in section 5.5 for studying the dynamics of a relativistic point particle.

A similar picture emerges for the spin. In relativistic mechanics angular momentum is a bivector quantity. Accordingly, the spin observables form a rank-2 antisymmetric tensor, with components given by

$$\langle\bar{\psi}\vert i\tfrac{1}{2}(\hat{\gamma}_\mu\hat{\gamma}_\nu-\hat{\gamma}_\nu\hat{\gamma}_\mu)\vert\psi\rangle\ \leftrightarrow\ \langle\tilde{\psi}\gamma_\mu\wedge\gamma_\nu\psi I\boldsymbol{\sigma}_3\rangle_q=\langle\gamma_\mu\wedge\gamma_\nu\psi I\boldsymbol{\sigma}_3\tilde{\psi}\rangle, \tag{8.82}$$

where again there is no imaginary component. This time we are picking out the components of the spin bivector S, given by

$$S=\psi I\boldsymbol{\sigma}_3\tilde{\psi}. \tag{8.83}$$

This is the natural spacetime generalisation of the Pauli result of equation (8.18). (Factors of $\hbar/2$ can always be inserted when required.) There are five such observables in all, which are summarised in Table 8.1. Of particular interest is the spin vector $s=\rho R\gamma_3\tilde{R}$. This justifies the classical model of spin introduced in section 5.5.6, where it was shown that the rotor form of the Lorentz force law naturally gives rise to a reduced gyromagnetic ratio of $g=2$.

8.3 The Dirac equation

While much of the preceding discussion is both suggestive about the role of spinors in quantum theory, and algebraically very useful, one has to remember that quantum mechanics deals with *wave equations*. We therefore need to

construct a relativistic wave equation for our Dirac spinor ψ, where ψ is an element of the eight-dimensional even subalgebra of the spacetime algebra. The relativistic wave equation for a spin-1/2 particle is the *Dirac equation*. This is a first-order wave equation, which is both Lorentz-invariant and has a future-pointing conserved current.

Like Pauli spinors, ψ is also subject to a single-sided rotor transformation law, $\psi \mapsto R\psi$, where R is a Lorentz rotor. To write down a covariant equation, we can therefore only place other covariant objects on the left of ψ. The available objects are any scalar or pseudoscalar, the vector derivative ∇ and any gauge fields describing interactions. On the right of ψ we can place combinations of γ_0, γ_3 and $I\boldsymbol{\sigma}_3$. The first equation we could write down is simply

$$\nabla\psi = 0. \tag{8.84}$$

This is the spacetime generalisation of the Cauchy–Riemann equations, as described in section 6.3. Remarkably, this equation does describe the behaviour of fermions — it is the wave equation for a (massless) *neutrino*. Any solution to this decomposes into two separate solutions by writing

$$\psi = \psi\tfrac{1}{2}(1 + \boldsymbol{\sigma}_3) + \psi\tfrac{1}{2}(1 - \boldsymbol{\sigma}_3) = \psi_+ + \psi_-. \tag{8.85}$$

The separate solutions ψ_+ and ψ_- are the right-handed and left-handed helicity eigenstates. For neutrinos, nature only appears to make use of the left-handed solutions. A more complete treatment of this subject involves the *electroweak* theory. (In fact, recent experiments point towards neutrinos carrying a small mass, whose origin can be explained by an interaction with the Higgs field.)

The formal operator identification of $i\partial_\mu$ with p_μ tells us that any wavefunction for a free massive particle should satisfy the Klein–Gordon equation $\nabla^2\psi = -m^2\psi$. We therefore need to add to the right-hand side of equation (8.84) a term that is linear in the particle mass m and that generates $-m^2\psi$ on squaring the operator. The natural covariant vector to form on the left of ψ is the momentum $\gamma^\mu p_\mu$. In terms of this operator we are led to an equation of the form

$$p\psi = m\psi a_0, \tag{8.86}$$

where a_0 is some multivector to be determined. It is immediately clear that a_0 must have odd grade, and must square to $+1$. The obvious candidate is γ_0, so that ψ contains a rotor to transform γ_0 to the velocity p/m. We therefore arrive at the equation

$$\nabla\psi I\boldsymbol{\sigma}_3 = m\psi\gamma_0. \tag{8.87}$$

This is the *Dirac equation* in its spacetime algebra form. This is easily seen to be equivalent to the matrix form of the equation

$$\hat{\gamma}^u\mu(\partial_\mu - ieA_\mu)|\psi\rangle = m|\psi\rangle, \tag{8.88}$$

where the electromagnetic vector potential has been included. The full Dirac equation is now

$$\nabla\psi I\boldsymbol{\sigma}_3 - eA\psi = m\psi\gamma_0. \tag{8.89}$$

A remarkable feature of this formulation is that the equation and all of its observables have been captured in the *real* algebra of spacetime, with no need for a unit imaginary. This suggests that interpretations of quantum mechanics that place great significance in the need for complex numbers are wide off the mark.

8.3.1 Symmetries and currents

The subject of the symmetries of the Dirac equation, and their conjugate currents, is discussed more fully in chapter 12. Here we highlight the main results. There are three important discrete symmetry operations: charge conjugation, parity and time reversal, denoted C, P and T respectively. Following the conventions of Bjorken & Drell (1964) we find that

$$\begin{aligned}
\hat{P}|\psi\rangle &\leftrightarrow \gamma_0\psi(\bar{x})\gamma_0, \\
\hat{C}|\psi\rangle &\leftrightarrow \psi\boldsymbol{\sigma}_1, \\
\hat{T}|\psi\rangle &\leftrightarrow I\gamma_0\psi(-\bar{x})\gamma_1,
\end{aligned} \tag{8.90}$$

where $\bar{x} = \gamma_0 x \gamma_0$ is (minus) the reflection of x in the timelike γ_0 axis. The combined CPT symmetry corresponds to

$$\psi \mapsto -I\psi(-x) \tag{8.91}$$

so that CPT symmetry does not require singling out a preferred timelike vector.

Amongst the continuous symmetries of the Dirac equation, the most significant is local electromagnetic gauge invariance. The equation is unchanged in physical content if we make the simultaneous replacements

$$\psi \mapsto \psi \mathrm{e}^{\alpha I\boldsymbol{\sigma}_3}, \qquad eA \mapsto eA - \nabla\alpha. \tag{8.92}$$

The conserved current conjugate to this symmetry is the Dirac current $J = \psi\gamma_0\tilde{\psi}$. This satisfies

$$\begin{aligned}
\nabla\cdot J &= \langle\nabla\psi\gamma_0\tilde{\psi}\rangle + \langle\psi\gamma_0\dot{\tilde{\psi}}\dot{\nabla}\rangle \\
&= -2\langle(eA\psi\gamma_0 + m\psi)I\boldsymbol{\sigma}_3\tilde{\psi}\rangle \\
&= 0
\end{aligned} \tag{8.93}$$

and so is conserved even in the presence of a background field. This is important. It means that single fermions cannot be created or destroyed. This feature was initially viewed as a great strength of the Dirac equation, though ultimately it is its biggest weakness. Fermion pairs, such as an electron and a positron, can be created and destroyed — a process which cannot be explained by the Dirac

equation alone. These are many-body problems and are described by *quantum field theory*.

The timelike component of J in the γ_0 frame, say, is

$$J_0 = \gamma_0 \cdot J = \langle \gamma_0 \tilde{\psi} \gamma_0 \psi \rangle = \langle \psi^\dagger \psi \rangle > 0, \tag{8.94}$$

which is *positive definite*. This is interpreted as a probability density, and localised wave functions are usually normalised such that

$$\int d^3x \, J_0 = 1. \tag{8.95}$$

Arriving at a relativistic theory with a consistent probabilistic interpretation was Dirac's original goal.

8.3.2 Plane-wave states

A positive energy plane-wave state is defined by

$$\psi = \psi_0 \mathrm{e}^{-I\boldsymbol{\sigma}_3 p \cdot x}, \tag{8.96}$$

where ψ_0 is a constant spinor. The Dirac equation (8.87) tells us that ψ_0 satisfies

$$p\psi_0 = m\psi_0\gamma_0, \tag{8.97}$$

and post-multiplying by $\tilde{\psi}_0$ we see that

$$p\psi_0\tilde{\psi}_0 = mJ. \tag{8.98}$$

Recalling that we have $\psi\tilde{\psi} = \rho \mathrm{e}^{i\beta}$, and noting that both p and J are vectors, we see that we must have $\exp(i\beta) = \pm 1$. For positive energy states the timelike component of p is positive, as is the timelike component of J, so we take the positive solution $\beta = 0$. It follows that ψ_0 is then simply a rotor with a normalisation constant. The proper boost L taking $m\gamma_0$ onto the momentum has

$$p = mL\gamma_0\tilde{L} = mL^2\gamma_0, \tag{8.99}$$

and from section 5.4.4 the solution is

$$L = \frac{m + p\gamma_0}{[2m(m + p\cdot\gamma_0)]^{1/2}} = \frac{E + m + \boldsymbol{p}}{[2m(E+m)]^{1/2}}, \tag{8.100}$$

where $p\gamma_0 = E + \boldsymbol{p}$. The full spinor ψ_0 is LU, where U is a spatial rotor in the γ_0 frame, so is a Pauli spinor.

Negative-energy solutions have a phase factor of $\exp(+I\boldsymbol{\sigma}_3 p\cdot x)$, with $E = \gamma_0 \cdot p > 0$. For these we have $-p\psi\tilde{\psi} = mJ$ so it is clear that we now need $\beta = \pi$. Positive and negative energy plane wave states can therefore be summarised by

$$\begin{aligned} \text{positive energy:} \quad & \psi^{(+)}(x) = L(p)U_r \mathrm{e}^{-I\boldsymbol{\sigma}_3 p\cdot x}, \\ \text{negative energy:} \quad & \psi^{(-)}(x) = L(p)U_r I \mathrm{e}^{I\boldsymbol{\sigma}_3 p\cdot x}, \end{aligned} \tag{8.101}$$