4.1 Axiomatic development

We should now have an intuitive feel for the elements of a geometric algebra — the multivectors — and some of their multiplicative properties. The next step is to define a set of axioms and conventions which enable us to efficiently manipulate them. Geometric algebra can be defined using a number of axiomatic frameworks, all of which give rise to the same final algebra. In the main we will follow the approach first developed by Hestenes and Sobczyk and raise the geometric product to primary status in the algebra. The properties of the inner and outer products are then inherited from the full geometric product, and this simplifies proofs of a number of important results.

Our starting point is the vector space from which the entire algebra will be generated. Vectors (i.e. grade-1 multivectors) have a special status in the algebra, as the grading of the algebra is determined by them. Three main axioms govern the properties of the geometric product for vectors.

(i) The geometric product is associative:

$$a(bc) = (ab)c = abc. \tag{4.1}$$

(ii) The geometric product is distributive over addition:

$$a(b + c) = ab + ac. \tag{4.2}$$

(iii) The square of any vector is a real scalar: $a^2 \in \Re$.

The final axiom is the key one which distinguishes a geometric algebra from a general associative algebra. We do not force the scalar to be positive, so we can incorporate Minkowski spacetime without modification of our axioms. Nothing is assumed about the commutation properties of the geometric product — matrix multiplication is one picture to keep in mind. Indeed, one can always represent the geometric product in terms of products of suitably chosen matrices, but this does not bring any new insights into the properties of the geometric product.

By successively multiplying together vectors we generate the complete algebra. Elements of this algebra are called multivectors and are usually written in upper-case italic font. The space of multivectors is *linear over the real numbers*, so if λ and μ are scalars and A and B are multivectors $\lambda A + \mu B$ is also a multivector. We only consider the algebra over the reals as most occurrences of complex numbers in physics turn out to have a geometric origin. This geometric meaning can be lost if we admit a scalar unit imaginary. Any multivector can be written as a sum of geometric products of vectors. They too can be multiplied using the geometric product and this product inherits properties (i) and (ii) above. So, for multivectors A, B and C, we have

$$(AB)C = A(BC) = ABC \tag{4.3}$$

and

$$A(B + C) = AB + AC. \tag{4.4}$$

If we now form the square of the vector $a + b$ we find that

$$(a + b)^2 = (a + b)(a + b) = a^2 + ab + ba + b^2. \tag{4.5}$$

It follows that the symmetrised product of two vectors can be written

$$ab + ba = (a + b)^2 - a^2 - b^2, \tag{4.6}$$

and so must also be a scalar, by axiom (iii). We therefore *define* the inner product for vectors by

$$a \cdot b = \frac{1}{2}(ab + ba). \tag{4.7}$$

The remaining, antisymmetric part of the geometric product is defined as the exterior product and returns a bivector,

$$a \wedge b = \frac{1}{2}(ab - ba). \tag{4.8}$$

These definitions combine to give the familiar result

$$ab = a \cdot b + a \wedge b. \tag{4.9}$$

In forming this decomposition we have defined both the inner and outer products of vectors in terms of the geometric product. This contrasts with the common alternative of defining the geometric product in terms of separate inner and outer products. Some authors prefer this alternative because the (less familiar) geometric product is defined in terms of more familiar objects. The main drawback, however, is that work still remains to establish the main properties of the geometric product. In particular, it is far from obvious that the product is associative, which is invaluable for its use.

4.1.1 The outer product, grading and bases

In the preceding we defined the outer product of two vectors and asserted that this returns a bivector (a grade-2 multivector). This is the key to defining the grade operation for the entire algebra. To do this we first extend the definition of the outer product to arbitrary numbers of vectors. The outer (exterior) product of the vectors $a_1, \ldots, a_r$ is denoted by $a_1 \wedge a_2 \wedge \cdots \wedge a_r$ and is defined as the totally antisymmetrised sum of all geometric products:

$$a_1 \wedge a_2 \wedge \cdots \wedge a_r = \frac{1}{r!} \sum (-1)^\epsilon a_{k_1} a_{k_2} \cdots a_{k_r}. \tag{4.10}$$

The sum runs over every permutation $k_1, \ldots, k_r$ of $1, \ldots, r$, and $(-1)^\epsilon$ is $+1$ or -1 as the permutation $k_1, \ldots, k_r$ is even or odd respectively. So, for example,

$$a_1 \wedge a_2 = \frac{1}{2!}(a_1 a_2 - a_2 a_1) \tag{4.11}$$

as required.

The antisymmetry of the outer product ensures that it vanishes if any two vectors are the same. It follows that the outer product vanishes if the vectors are linearly dependent, since in this case one vector can be written as a linear combination of the remaining vectors. The outer product therefore records the dimensionality of the object formed from a set of vectors. This is precisely what we mean by *grade*, so we define the outer product of r vectors as having grade r. Any multivector which can be written purely as the outer product of a set of vectors is called a *blade*. Any multivector can be expressed as a sum of blades, as can be verified by introducing an explicit basis. These blades all have definite grade and in turn define the grade or grades of the multivector.

We rarely need the full antisymmetrised expression when studying blades. Instead we can employ the result that *every blade can be written as a geometric product of orthogonal, anticommuting vectors.* The anticommutation of orthogonal vectors then takes care of the antisymmetry of the product. In Euclidean space this result is simple to prove using a form of Gram–Schmidt orthogonalisation. Given two vectors a and b we form

$$b' = b - \lambda a. \tag{4.12}$$

We then see that

$$a \wedge (b - \lambda a) = a \wedge b - \lambda a \wedge a = a \wedge b. \tag{4.13}$$

So the same bivector is obtained, whatever the value of λ (figure 4.1). The bivector encodes an oriented plane with magnitude determined by the area. Interchanging b and b' changes neither the orientation nor the magnitude, so returns the same bivector. We now form

$$a \cdot b' = a \cdot (b - \lambda a) = a \cdot b - \lambda a^2. \tag{4.14}$$

So if we set $\lambda = a \cdot b / a^2$ we have $a \cdot b' = 0$ and can write

$$a \wedge b = a \wedge b' = ab'. \tag{4.15}$$

One can continue in this manner and construct a complete set of orthogonal vectors generating the same outer product.

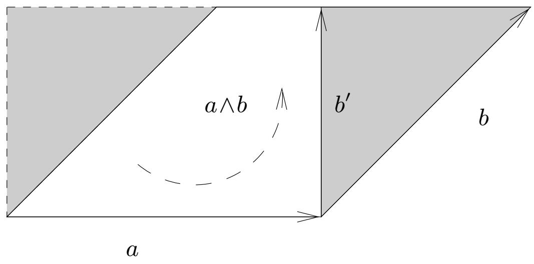


Figure 4.1 *The Gram–Schmidt process.* The outer product $a \wedge b$ is independent of shape of the parallelogram formed by a and b. The only information contained in $a \wedge b$ is the oriented plane and a magnitude. The vectors b and b' generate the same bivector, so we can choose b' orthogonal to a and write $a \wedge b = ab'$.

An alternative form for b' is quite revealing. We write

$$\begin{aligned} b' &= b - a^{-1} a \cdot b \\ &= a^{-1}(ab - a \cdot b) \\ &= a^{-1}(a \wedge b). \end{aligned} \tag{4.16}$$

This shows that b' is formed by rotating a through 90° in the $a \wedge b$ plane, and dilating by the appropriate amount. The algebraic form also makes it clear why $ab' = a \wedge b$, and gives a formula that extends simply to higher grades.

The above argument is fine for Euclidean space, but breaks down for spaces of mixed signature. The inverse $a^{-1} = a/a^2$ is not defined when a is null ($a^2 = 0$), so an alternative procedure is required. Fortunately this is a relatively straightforward exercise. We start with the set of r independent vectors $a_1, \ldots, a_r$ and form the $r \times r$ symmetric matrix

$$\mathsf{M}_{ij} = a_i \cdot a_j. \tag{4.17}$$

The symmetry of this matrix ensures that it can always be diagonalised with an orthogonal matrix R_{ij},

$$\mathsf{R}_{ik} \mathsf{M}_{kl} \mathsf{R}^{\mathrm{t}}_{lj} = \mathsf{R}_{ik} \mathsf{R}_{jl} \mathsf{M}_{kl} = \Lambda_{ij}. \tag{4.18}$$

Here Λ_{ij} is diagonal and, unless stated otherwise, the summation convention is employed. The matrix R_{ij} defines a new set of vectors via

$$e_i = \mathsf{R}_{ij} a_j. \tag{4.19}$$

These satisfy

$$\begin{aligned} e_i \cdot e_j &= (\mathsf{R}_{ik} a_k) \cdot (\mathsf{R}_{jl} a_l) \\ &= \mathsf{R}_{ik} \mathsf{R}_{jl} \mathsf{M}_{kl} \\ &= \Lambda_{ij}. \end{aligned} \tag{4.20}$$

The vectors $e_1, \ldots, e_r$ are therefore orthogonal and hence all anticommute. Their geometric product is therefore totally antisymmetric, and we have

$$\begin{aligned} e_1 e_2 \cdots e_r &= e_1 \wedge \cdots \wedge e_r \\ &= (\mathsf{R}_{1i} a_i) \wedge \cdots (\mathsf{R}_{rk} a_k) \\ &= \det\,(\mathsf{R}_{ij})\, a_1 \wedge a_2 \wedge \cdots \wedge a_r. \end{aligned} \tag{4.21}$$

The determinant appears here because of the total antisymmetry of the expression (see section 4.5.2). But since R_{ij} is an orthogonal matrix it has determinant ± 1, and by choosing the order of the $\{e_i\}$ vectors appropriately we can set the determinant of R_{ij} to 1. This ensures that we can always find a set of vectors such that

$$a_1 \wedge a_2 \wedge \cdots \wedge a_r = e_1 e_2 \cdots e_r. \tag{4.22}$$

This result will simplify the proofs of a number of results in this chapter.

For a given vector space, an orthonormal frame $\{\mathsf{e}_i\}, i = 1, \ldots, n$ provides a natural way to view the entire geometric algebra. We denote this algebra $\mathcal{G}_n$. Most of the results derived in this chapter are independent of signature, so in the following we let $\mathcal{G}_n$ denote the geometric algebra of a space of dimension n with arbitrary (non-degenerate) signature. One can also consider the degenerate case where some of the basis vectors are null, though we will not need such algebras in this book. The basis vectors build up to form a basis for the entire algebra as

$$1, \quad \mathsf{e}_i, \quad \mathsf{e}_i \mathsf{e}_j \ (i < j), \quad \mathsf{e}_i \mathsf{e}_j \mathsf{e}_k \ (i < j < k), \qquad \ldots . \tag{4.23}$$

The fact that the basis vectors anticommute ensures that each product in the basis set is totally antisymmetric. The product of r distinct basis vectors is then, by definition, a grade-r multivector. The basis (4.23) therefore naturally defines a basis for each of the grade-r subspaces of $\mathcal{G}_n$. We denote each of these subspaces by $\mathcal{G}_n^r$. The size of each subspace is given by the number of distinct combinations of r objects from a set of n. (The order is irrelevant, because of the total antisymmetry.) These are given by the binomial coefficients, so

$$\dim\big(\mathcal{G}_n^r\big) = \binom{n}{r}. \tag{4.24}$$

For example, we have already seen that in two dimensions the algebra contains terms of grade $0, 1, 2$ with each space having dimension $1, 2, 1$ respectively. Similarly in three dimensions the separate graded subspaces have dimension $1, 3, 3, 1$. The binomial coefficients always exhibit a mirror symmetry between the r and $n - r$ terms. This gives rise to the notion of duality, which is explained in section 4.1.4 where we explore the properties of the highest grade element of the algebra — the pseudoscalar.

The total dimension of the algebra is

$$\dim(\mathcal{G}_n) = \sum_{r=0}^{n} \dim(\mathcal{G}_n^r) = \sum_{r=0}^{n} \binom{n}{r} = (1+1)^n = 2^n. \tag{4.25}$$

One can see that the total size of the algebra quickly becomes very large. If one wanted to find a matrix representation of the algebra, the matrices would have to be of the order of $2^{n/2} \times 2^{n/2}$. For all but the lowest values of n these matrices become totally impractical for computations. This is one reason why matrix representations do not help much with understanding and using geometric algebra.

We have now defined the grade operation for our linear space $\mathcal{G}_n$. An arbitrary multivector A can be decomposed into a sum of pure grade terms

$$A = \langle A\rangle_0 + \langle A\rangle_1 + \cdots = \sum_r \langle A\rangle_r. \tag{4.26}$$

The operator $\langle\ \rangle_r$ projects onto the grade-r terms in the argument, so $\langle A\rangle_r$ returns the grade-r components in A. Multivectors containing terms of only one grade are called *homogeneous*. They are often written as A_r, so

$$\langle A_r\rangle_r = A_r. \tag{4.27}$$

Take care not to confuse the grading subscript in A_r with frame indices in expressions like $\{\mathsf{e}_k\}$. The context should always make clear which is intended. The grade-0 terms in $\mathcal{G}_n$ are the real scalars and commute with all other elements. We continue to employ the useful abbreviation

$$\langle A\rangle = \langle A\rangle_0 \tag{4.28}$$

for the operation of taking the scalar part.

An important feature of a geometric algebra is that *not all homogeneous multivectors are pure blades.* This is confusing at first, because we have to go to four dimensions before we reach our first counterexample. Suppose that $\{\mathsf{e}_1, \ldots, \mathsf{e}_4\}$ form an orthonormal basis for the Euclidean algebra $\mathcal{G}_4$. There are six independent basis bivectors in this algebra, and from these we can construct terms like

$$B = \alpha \mathsf{e}_1 \wedge \mathsf{e}_2 + \beta \mathsf{e}_3 \wedge \mathsf{e}_4, \tag{4.29}$$

where α and β are scalars. B is a pure bivector, so is homogeneous, but it cannot be reduced to a blade. That is, we cannot find two vectors a and b such that $B = a \wedge b$. The reason is that $\mathsf{e}_1 \wedge \mathsf{e}_2$ and $\mathsf{e}_3 \wedge \mathsf{e}_4$ do not share a common vector. This is not possible in three dimensions, because any two planes with a common origin share a common line. A four-dimensional bivector like B is therefore hard for us to visualise. There is a way to visualise B in three dimensions, however, and it is provided by *projective geometry*. This is described in chapter 10.

4.1.2 Further properties of the geometric product

The decomposition of the geometric product of two vectors into a scalar term and a bivector term has a natural extension to general multivectors. To establish the results of this section we make repeated use of the formula

$$ab = 2a\cdot b - ba \tag{4.30}$$

which we use to reorder expressions. As a first example, consider the case of a geometric product of vectors. We find that

$$\begin{aligned} aa_1a_2\cdots a_r &= 2a\cdot a_1\, a_2\cdots a_r - a_1aa_2\cdots a_r \\ &= 2a\cdot a_1\, a_2\cdots a_r - 2a\cdot a_2\, a_1a_3\cdots a_r + a_1a_2aa_3\cdots a_r \\ &= 2\sum_{k=1}^{r}(-1)^{k+1}a\cdot a_k\, a_1a_2\cdots \check{a}_k\cdots a_r + (-1)^r a_1a_2\cdots a_r a, \end{aligned} \tag{4.31}$$

where the check on $\check{a}_k$ denotes that this term is missing from the series. We continue to follow the conventions introduced in chapter 2 so, in the absence of brackets, inner products are performed before outer products, and both are performed before geometric products.

Suppose now that the vectors $a_1, \ldots, a_r$ are replaced by a set of anticommuting vectors $e_1, \ldots, e_r$. We find that

$$\frac{1}{2}\Big(ae_1e_2\cdots e_r - (-1)^r e_1e_2\cdots e_r a\Big) = \sum_{k=1}^{r}(-1)^{k+1}a\cdot e_k\, e_1e_2\cdots \check{e}_k\cdots e_r. \tag{4.32}$$

The right-hand side contains a sum of terms formed from the product of $r-1$ anticommuting vectors, so has grade $r-1$. Since any grade-r multivector can be written as a sum of terms formed from anticommuting vectors, the combination on the left-hand side will always return a multivector of grade $r-1$. We therefore define the inner product between a vector a and a grade-r multivector A_r by

$$a\cdot A_r = \frac{1}{2}\Big(aA_r - (-1)^r A_r a\Big). \tag{4.33}$$

The inner product of a vector and a grade-r multivector results in a multivector with grade reduced by one.

The main work of this section is in establishing the properties of the remaining part of the product aA_r. For the case where A_r is a vector, the remaining term is the antisymmetric product, and so is a bivector. This turns out to be true in general — the remaining part of the geometric product returns the exterior product,

$$\frac{1}{2}\Big(a(a_1\wedge a_2\wedge\cdots\wedge a_r) + (-1)^r(a_1\wedge a_2\wedge\cdots\wedge a_r)a\Big) = a\wedge a_1\wedge a_2\wedge\cdots\wedge a_r. \tag{4.34}$$

We will prove this important result by induction. First, we write the blade as a geometric product of anticommuting vectors, so that the result we will establish becomes

$$\frac{1}{2}\Big(ae_1e_2\cdots e_r + (-1)^r e_1e_2\cdots e_r a\Big) = a\wedge e_1\wedge e_2\wedge\cdots\wedge e_r. \tag{4.35}$$

For $r = 1$ the result is true as the right-hand side defines the bivector $a\wedge e_1$. For $r > 1$ we proceed by writing

$$\begin{aligned} a\wedge e_1\wedge e_2\wedge\cdots\wedge e_r = {}& \frac{1}{r+1} ae_1e_2\cdots e_r \\ &+ \frac{1}{r+1}\sum_{k=1}^{r}(-1)^k e_k(a\wedge e_1\wedge\cdots\wedge\check{e}_k\wedge\cdots\wedge e_r). \end{aligned} \tag{4.36}$$

This result is easily established by writing out all terms in the full antisymmetric product and gathering together the terms which start with the same vector. Next we assume that equation (4.35) holds for the case of an $r-1$ blade, and expand the term inside the sum as follows:

$$\begin{aligned} &\sum_{k=1}^{r}(-1)^k e_k(a\wedge e_1\wedge\cdots\wedge\check{e}_k\wedge\cdots\wedge e_r) \\ &\quad= \frac{1}{2}\sum_{k=1}^{r}(-1)^k e_k\Big(ae_1\cdots\check{e}_k\cdots e_r + (-1)^{r-1}e_1\cdots\check{e}_k\cdots e_r a\Big) \\ &\quad= \frac{1}{2}\sum_{k=1}^{r}(-1)^k e_k a e_1\cdots\check{e}_k\cdots e_r + \frac{r}{2}(-1)^r e_1\cdots e_r a \\ &\quad= \sum_{k=1}^{r}(-1)^k (e_k\cdot a) e_1\cdots\check{e}_k\cdots e_r + \frac{r}{2}\Big(ae_1\cdots e_r + (-1)^r e_1\cdots e_r a\Big) \\ &\quad= \frac{r-1}{2} ae_1\cdots e_r + \frac{r+1}{2}(-1)^r e_1\cdots e_r a, \end{aligned} \tag{4.37}$$

where we have used equation (4.32). Substituting this result into equation (4.36) then proves equation (4.35) for a grade-r blade, assuming it is true for a blade of grade $r-1$. Since the result is already established for $r = 1$, equation (4.34) holds for all blades and hence all multivectors.

We extend the definition of the wedge symbol by writing

$$a\wedge A_r = \frac{1}{2}\Big(aA_r + (-1)^r A_r a\Big). \tag{4.38}$$

With this definition we now have

$$aA_r = a\cdot A_r + a\wedge A_r, \tag{4.39}$$

which extends the decomposition of the geometric product in precisely the desired way. In equation (4.38) one can see how the geometric product can simplify many calculations. The left-hand side would, in general, require totally antisymmetrising all possible products. But the right-hand side only requires evaluating two products — an enormous saving! As we have established the grades of the separate inner and outer products, we also have

$$aA_r = \langle aA_r\rangle_{r-1} + \langle aA_r\rangle_{r+1}, \tag{4.40}$$

where

$$a\cdot A_r = \langle aA_r\rangle_{r-1}, \quad a\wedge A_r = \langle aA_r\rangle_{r+1}. \tag{4.41}$$

So, as expected, multiplication by a vector raises and lowers the grade of a multivector by 1.

A homogeneous multivector can be written as a sum of blades, and each blade can be written as a geometric product of anticommuting vectors. Applying the preceding decomposition, we establish that the product of two homogeneous multivectors decomposes as

$$A_rB_s = \langle A_rB_s\rangle_{|r-s|} + \langle A_rB_s\rangle_{|r-s|+2} + \cdots + \langle A_rB_s\rangle_{r+s}. \tag{4.42}$$

We retain the $\cdot$ and $\wedge$ symbols for the lowest and highest grade terms in this series:

$$\begin{aligned} A_r\cdot B_s &= \langle A_rB_s\rangle_{|r-s|}, \\ A_r\wedge B_s &= \langle A_rB_s\rangle_{r+s}. \end{aligned} \tag{4.43}$$

This is the most general use of the wedge symbol, and is consistent with the earlier definition as the antisymmetrised product of a set of vectors. We can check that the outer product is associative by forming

$$(A_r\wedge B_s)\wedge C_t = \langle A_rB_s\rangle_{r+s}\wedge C_t = \langle (A_rB_s)C_t\rangle_{r+s+t}. \tag{4.44}$$

Associativity of the outer product then follows from the fact that the geometric product is associative:

$$\langle (A_rB_s)C_t\rangle_{r+s+t} = \langle A_rB_sC_t\rangle_{r+s+t} = A_r\wedge B_s\wedge C_t. \tag{4.45}$$

In equation (4.32) we established a formula for the result for the inner product of a vector and a blade formed from orthogonal vectors. We now extend this to a more general result that is extremely useful in practice. We start by writing

$$a\cdot(a_1\wedge a_2\wedge\cdots\wedge a_r) = a\cdot\langle a_1a_2\cdots a_r\rangle_r, \tag{4.46}$$

where $a_1, \ldots, a_r$ are a general set of vectors. The geometric product $a_1a_2\cdots a_r$ can only contain terms of grade r, $r-2$, $\ldots$, so

$$\begin{aligned} &\frac{1}{2}\Big(aa_1a_2\cdots a_r - (-1)^r a_1a_2\cdots a_ra\Big) \\ &\qquad\qquad = a\cdot\langle a_1a_2\cdots a_r\rangle_r + a\cdot\langle a_1a_2\cdots a_r\rangle_{r-2} + \cdots . \end{aligned} \tag{4.47}$$

The term we are after is the $r-1$ grade part, so we have

$$a\cdot(a_1\wedge a_2\wedge\cdots\wedge a_r) = \frac{1}{2}\langle aa_1a_2\cdots a_r - (-1)^r a_1a_2\cdots a_ra\rangle_{r-1}. \tag{4.48}$$

We can now apply equation (4.31) inside the grade projection operator to form

$$a\cdot(a_1\wedge a_2\wedge\cdots\wedge a_r) = \sum_{k=1}^{r}(-1)^{k+1}a\cdot a_k \langle a_1\cdots\check{a}_k\cdots a_r\rangle_{r-1}$$
$$= \sum_{k=1}^{r}(-1)^{k+1}a\cdot a_k\, a_1\wedge\cdots\wedge\check{a}_k\wedge\cdots\wedge a_r. \tag{4.49}$$

The first two cases illustrate how the general formula behaves:

$$\begin{aligned} a\cdot(a_1\wedge a_2) &= a\cdot a_1\, a_2 - a\cdot a_2\, a_1, \\ a\cdot(a_1\wedge a_2\wedge a_3) &= a\cdot a_1\, a_2\wedge a_3 - a\cdot a_2\, a_1\wedge a_3 + a\cdot a_3\, a_1\wedge a_2. \end{aligned} \tag{4.50}$$

The first case was established in chapter 2, where it was used to replace the formula for the double cross product of vectors in three dimensions.

4.1.3 The reverse, the scalar and the commutator product

Now that the grading is established, we can establish some general properties of the reversion operator, which was first introduced in chapter 2. The reverse of a product of vectors is defined by

$$(ab\cdots c)^\dagger = c\cdots ba. \tag{4.51}$$

For a blade the reverse can be formed by a series of swaps of anticommuting vectors, each resulting in a minus sign. The first vector has to swap past $r-1$ vectors, the second past $r-2$, and so on. This demonstrates that

$$A_r^\dagger = (-1)^{r(r-1)/2}A_r. \tag{4.52}$$

If we now consider the scalar part of a geometric product of two grade-r multivectors we find that

$$\langle A_rB_r\rangle = \langle A_rB_r\rangle^\dagger = \langle B_r^\dagger A_r^\dagger\rangle = (-1)^{r(r-1)}\langle B_rA_r\rangle = \langle B_rA_r\rangle, \tag{4.53}$$

so, for general A and B,

$$\langle AB\rangle = \langle BA\rangle. \tag{4.54}$$

It follows that

$$\langle A\cdots BC\rangle = \langle CA\cdots B\rangle. \tag{4.55}$$

This cyclic reordering property is frequently useful for manipulating expressions. The product in equation (4.54) is sometimes given the symbol $*$, so we write

$$A{*}B = \langle AB\rangle. \tag{4.56}$$

A further product of considerable importance in geometric algebra is the commutator product of two multivectors. This is denoted with a cross, $\times$, and is defined by

$$A\times B = \frac{1}{2}(AB - BA). \tag{4.57}$$

Care must be taken to include the factor of one-half, which is different to the standard commutator of two operators in quantum mechanics. The commutator product satisfies the *Jacobi identity*

$$A\times(B\times C) + B\times(C\times A) + C\times(A\times B) = 0, \tag{4.58}$$

which is easily seen by expanding out the products.

The commutator arises most frequently in equations involving bivectors. Given a bivector B and a vector a we have

$$B\times a = \frac{1}{2}(Ba - aB) = B\cdot a, \tag{4.59}$$

which therefore results in a second vector. Now consider the product of a bivector and a blade formed from anticommuting vectors. We have

$$\begin{aligned} B(e_1 e_2 \cdots e_r) &= 2(B\times e_1)e_2\cdots e_r + e_1 B e_2 \cdots e_r \\ &= 2(B\times e_1)e_2\cdots e_r + \cdots + 2e_1\cdots(B\times e_r) + e_1 e_2\cdots e_r B. \end{aligned} \tag{4.60}$$

It follows that

$$B\times(e_1 e_2\cdots e_r) = \sum_{i=1}^{r} e_1\cdots(B\cdot e_i)\cdots e_r. \tag{4.61}$$

The sum involves a series of terms which can only contain grades r and $r-2$. But if we form the reverse of the commutator product between a bivector and a homogeneous multivector, we find that

$$\begin{aligned} (B\times A_r)^\dagger &= \frac{1}{2}(BA_r - A_r B)^\dagger \\ &= \frac{1}{2}(-A_r^\dagger B + BA_r^\dagger) \\ &= (-1)^{r(r-1)/2} B\times A_r. \end{aligned} \tag{4.62}$$

It follows that $B\times A_r$ has the same properties under reversion as A_r. But multivectors of grade r and $r-2$ always behave differently under reversion. The commutator product in equation (4.61) must therefore result in a grade-r multivector. Since this is true of any grade-r basis element, it must be true of any homogeneous multivector. That is,

$$B\times A_r = \langle B\times A_r\rangle_r. \tag{4.63}$$

The commutator of a multivector with a bivector therefore preserves the grade of the multivector. Furthermore, the commutator of two bivectors must result in a third bivector. This is the basis for incorporating the theory of Lie groups into geometric algebra.

A similar argument to the preceding one shows that the symmetric product with a bivector must raise or lower the grade by 2. We can summarise this by writing

$$\begin{aligned} BA_r &= \langle BA_r\rangle_{r-2} + \langle BA_r\rangle_r + \langle BA_r\rangle_{r+2} \\ &= B\cdot A_r + B\times A_r + B\wedge A_r, \end{aligned} \tag{4.64}$$

where

$$\frac{1}{2}(BA_r - A_rB) = B\times A_r \tag{4.65}$$

and

$$\frac{1}{2}(BA_r + A_rB) = B\cdot A_r + B\wedge A_r. \tag{4.66}$$

It is assumed in these formulae that A_r has grade $r > 1$.

4.1.4 Pseudoscalars and duality

The exterior product of n vectors defines a grade-n blade. For a given vector space the highest grade element is unique, up to a magnitude. The outer product of n vectors is therefore a multiple of the unique *pseudoscalar* for $\mathcal{G}_n$. This is denoted I, and has two important properties. The first is that I is normalised to

$$|I^2| = 1. \tag{4.67}$$

The sign of I^2 depends on the size of space and the signature. It turns out that the pseudoscalar squares to -1 for the three algebras of most use in this book — those of the Euclidean plane and space, and of spacetime. But this is in no way a general property.

The second property of the pseudoscalar I is that it defines an *orientation*. For any ordered set of n vectors, their outer product will either have the same sign as I, or the opposite sign. Those with the same sign are assigned a positive orientation, and those with opposite sign have a negative orientation. The orientation is swapped by interchanging any pair of vectors. In three dimensions we always choose the pseudoscalar I such that it has the orientation specified by a right-handed set of vectors. In other spaces one just asserts a choice of I and then sticks to that choice consistently.

The product of the grade-n pseudoscalar I with a grade-r multivector A_r is a grade $n-r$ multivector. This operation is called a *duality* transformation. If A_r is a blade, IA_r returns the *orthogonal complement* of A_r. That is, the blade formed from the space of vectors not contained in A_r. It is clear why this has grade $n-r$. Every blade acts as a pseudoscalar for the space spanned by its generating vectors. So, even if we are working in three dimensions, we can treat the bivector $e_1 e_2$ as a pseudoscalar for any manipulation taking place entirely in the $e_1 e_2$ plane. This is often a very helpful idea.

In spaces of odd dimension, I commutes with all vectors and so commutes with all multivectors. In spaces of even dimension, I anticommutes with vectors and so anticommutes with all odd-grade multivectors. In all cases the pseudoscalar commutes with all even-grade multivectors in its algebra. We summarise this by

$$IA_r = (-1)^{r(n-1)} A_r I. \tag{4.68}$$

An important use of the pseudoscalar is for interchanging inner and outer products. For example, we have

$$\begin{aligned} a\cdot(A_r I) &= \frac{1}{2}\Big(aA_r I - (-1)^{n-r} A_r I a\Big) \\ &= \frac{1}{2}\Big(aA_r I - (-1)^{n-r}(-1)^{n-1} A_r a I\Big) \\ &= \frac{1}{2}\Big(aA_r + (-1)^{r} A_r a\Big) I \\ &= a\wedge A_r\, I. \end{aligned} \tag{4.69}$$

More generally, we can take two multivectors A_r and B_s, with $r+s \leq n$, and form

$$\begin{aligned} A_r\cdot(B_s I) &= \langle A_r B_s I\rangle_{|r-(n-s)|} \\ &= \langle A_r B_s I\rangle_{n-(r+s)} \\ &= \langle A_r B_s\rangle_{r+s} I \\ &= A_r\wedge B_s\, I. \end{aligned} \tag{4.70}$$

This type of interchange is very common in applications. Note how simple this proof is made by the application of the geometric product in the intermediate steps.

4.2 Rotations and reflections

In chapter 2 we showed that in three dimensions a reflection in the plane perpendicular to the unit vector n is performed by

$$a \mapsto a' = -nan. \tag{4.71}$$

This formula holds in arbitrary numbers of dimensions. Provided $n^2 = 1$, we see that n is transformed to

$$n \mapsto -nnn = -n, \tag{4.72}$$

whereas any vector $a_\perp$ perpendicular to n is mapped to

$$a_\perp \mapsto -na_\perp n = a_\perp nn = a_\perp. \tag{4.73}$$

So, for a vector a, the component parallel to n has its sign reversed, whereas the component perpendicular to n is unchanged. This is what we mean by a reflection in the hyperplane perpendicular to n.

Two successive reflections in the hyperplanes perpendicular to m and n result in a rotation in the $m \wedge n$ plane. This is encoded in the rotor

$$R = nm = \exp(-\hat{B}\theta/2) \tag{4.74}$$

where

$$\cos(\theta/2) = n\cdot m, \quad \hat{B} = \frac{m\wedge n}{\sin(\theta/2)}. \tag{4.75}$$

The rotor R generates a rotation through the by now familiar formula

$$a \mapsto a' = RaR^\dagger. \tag{4.76}$$

Rotations form a group, as the result of combining two rotations is a third rotation. The same must therefore be true of rotors. Suppose that R_1 and R_2 generate two distinct rotations. The combined rotations take a to

$$a \mapsto R_2(R_1 a R_1^\dagger)R_2^\dagger = R_2 R_1 a R_1^\dagger R_2^\dagger. \tag{4.77}$$

We therefore define the product rotor

$$R = R_2 R_1, \tag{4.78}$$

so that the result of the composite rotation is described by $RaR^\dagger$, as usual. The product R is a new rotor, and in general it will consist of geometric products of an even number of unit vectors,

$$R = lk \cdots nm. \tag{4.79}$$

We will adopt this as our definition of a rotor. The reversed rotor is

$$R^\dagger = mn \cdots kl. \tag{4.80}$$

The result of the map $a \mapsto RaR^\dagger$ returns a vector for any vector a, since

$$RaR^\dagger = lk \cdots \big(n(mam)n\big) \cdots kl \tag{4.81}$$

and each successive sandwich between a vector returns a new vector.

We can immediately establish the normalisation condition

$$RR^\dagger = lk \cdots nmmn \cdots kl = 1 = R^\dagger R. \tag{4.82}$$

In Euclidean spaces, where every vector has a positive square, this normalisation is automatic. In mixed signature spaces, like Minkowski spacetime, unit vectors can have $n^2 = \pm 1$. In this case the condition $RR^\dagger = 1$ is taken as a further condition satisfied by a rotor. In the case where R is the product of two rotors we can easily confirm that

$$RR^\dagger = R_2 R_1 (R_2 R_1)^\dagger = R_2 R_1 R_1^\dagger R_2^\dagger = 1. \tag{4.83}$$

The set of rotors therefore forms a *group*, called a rotor group. This is similar to the group of rotation matrices, though not identical due to the two-to-one map between rotors and rotation matrices. We will have more to say about the group properties of rotors in chapter 11.

In Euclidean spaces every rotor can be written as the exponential of a bivector,

$$R = \exp(-B/2). \tag{4.84}$$

The bivector B defines the plane or planes in which the rotation takes place. The sign ensures that the rotation has the orientation defined by B. In mixed signature spaces one can always write a rotor as $\pm\exp(B)$. In either case the effect of the rotor R on the vector a is

$$a \mapsto \exp(-B/2) a \exp(B/2). \tag{4.85}$$

We can prove that the right-hand side always returns a vector by considering a Taylor expansion of

$$a(\lambda) = \exp(-\lambda B/2) a \exp(\lambda B/2). \tag{4.86}$$

Differentiating the expression on the right produces the power series expansion

$$a(\lambda) = a + \lambda a\cdot B + \frac{\lambda^2}{2!}(a\cdot B)\cdot B + \cdots . \tag{4.87}$$

Since the inner product of a vector and a bivector always results in a new vector, each term in this expansion is a vector. Setting $\lambda = 1$ then demonstrates that equation (4.85) results in a new vector, defined by

$$\exp(-B/2) a \exp(B/2) = a + a\cdot B + \frac{1}{2!}(a\cdot B)\cdot B + \cdots . \tag{4.88}$$

4.2.1 Multivector transformations

Suppose now that every vector in a blade undergoes the same rotation. This is the sort of transformation implied if a plane or volume element is to be rotated. The r-blade A_r can be written

$$A_r = a_1 \wedge \cdots \wedge a_r = \frac{1}{r!}\sum (-1)^\epsilon a_{k_1} a_{k_2} \cdots a_{k_r}, \tag{4.89}$$

with the sum running over all permutations. If each vector in a geometric product is rotated, the result is the multivector

$$\begin{aligned}(Ra_1R^\dagger)(Ra_2R^\dagger)\cdots(Ra_rR^\dagger) &= Ra_1R^\dagger Ra_2R^\dagger \cdots Ra_rR^\dagger \\ &= Ra_1a_2\cdots a_rR^\dagger.\end{aligned} \tag{4.90}$$

This holds for each term in the antisymmetrised sum, so the transformation law for the blade A_r is simply

$$A_r \mapsto A_r' = RA_rR^\dagger. \tag{4.91}$$

Blades transform with the same simple law as vectors! All multivectors share the same transformation law regardless of grade when each component vector is rotated. This is one reason why the rotor formulation is so powerful. The alternative, tensor form would require an extra matrix for each additional vector.

4.3 Bases, frames and components

Any set of linearly independent vectors form a basis for the vectors in a geometric algebra. Such a set is often referred to as a *frame*. Repeated use of the outer product then builds up a basis for the entire algebra. In this section we use the symbols $\mathsf{e}_1, \dots, \mathsf{e}_n$ or $\{\mathsf{e}_k\}$ to denote a frame for n-dimensional space. We do not restrict the frame to be orthonormal, so the $\{\mathsf{e}_k\}$ do not necessarily anticommute. The reason for the change of font for frame vectors, as opposed to general sets of vectors, is that use of frames nearly always implies reference to coordinates. It is natural write the coordinates of the vector a as a_i or a^i so, to avoid confusion with a set of vectors, we write the frame vectors in a different font.

The volume element for the $\{\mathsf{e}_k\}$ frame is defined by

$$E_n \equiv \mathsf{e}_1 \wedge \mathsf{e}_2 \wedge \cdots \wedge \mathsf{e}_n. \tag{4.92}$$

The grade-n multivector E_n is a multiple of the pseudoscalar for the space spanned by the $\{\mathsf{e}_k\}$. The fact that the vectors are independent guarantees that $E_n \neq 0$. Associated with any arbitrary frame is a reciprocal frame $\{\mathsf{e}^k\}$ defined by the property

$$\mathsf{e}^i \cdot \mathsf{e}_j = \delta^i_j, \quad \forall i, j = 1 \dots n. \tag{4.93}$$

The 'Kronecker δ', δ^i_j, has value $+1$ if $i = j$ and is zero otherwise. The reciprocal frame is constructed as follows:

$$\mathsf{e}^j = (-1)^{j-1} \mathsf{e}_1 \wedge \mathsf{e}_2 \wedge \cdots \wedge \check{\mathsf{e}}_j \wedge \cdots \wedge \mathsf{e}_n \, E_n^{-1}, \tag{4.94}$$

where as usual the check on $\check{\mathsf{e}}_j$ denotes that this term is missing from the expression. The formula for e^j has a simple interpretation. The vector e^j must be perpendicular to all the vectors $\{\mathsf{e}_i, i \neq j\}$. To find this we form the exterior product of the $n-1$ vectors $\{\mathsf{e}_i, i \neq j\}$. The dual of this returns a vector perpendicular to all vectors in the subspace, and this duality is achieved by the factor of E_n. All that remains is to fix up the normalisation. For this we recall the duality results of section 4.1.4 and form

$$\mathsf{e}_1 \cdot \mathsf{e}^1 = \mathsf{e}_1 \cdot (\mathsf{e}_2 \wedge \cdots \wedge \mathsf{e}_n \, E_n^{-1}) = (\mathsf{e}_1 \wedge \mathsf{e}_2 \wedge \cdots \wedge \mathsf{e}_n) E_n^{-1} = 1. \tag{4.95}$$

This confirms that the formula for the reciprocal frame is correct.

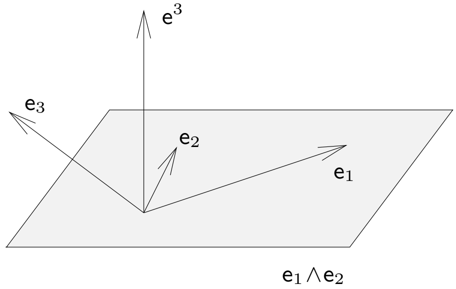


Figure 4.2 *The reciprocal frame.* The vectors e_1, e_2 and e_3 form a non-orthonormal frame for three-dimensional space. The vector e^3 is formed by constructing the $e_1 \wedge e_2$ plane, and forming the vector perpendicular to this plane. The length is fixed by demanding $e^3 \cdot e_3 = 1$.

4.3.1 Application — crystallography

An important application of the formula for a reciprocal frame is in crystallography. If a crystal contains some repeated structure defined by the vectors e_1, e_2, e_3, then constructive interference occurs for wavevectors whose difference satisfies

$$\Delta k = 2\pi(n_1 e^1 + n_2 e^2 + n_3 e^3), \tag{4.96}$$

where n_1, n_2, n_3 are integers. The reciprocal frame is defined by

$$e^1 = \frac{e_2 \wedge e_3}{e_1 \wedge e_2 \wedge e_3}, \quad e^2 = \frac{e_3 \wedge e_1}{e_1 \wedge e_2 \wedge e_3}, \quad e^3 = \frac{e_1 \wedge e_2}{e_1 \wedge e_2 \wedge e_3}. \tag{4.97}$$

If we write

$$e_1 \wedge e_2 \wedge e_3 = [e_1, e_2, e_3] I, \tag{4.98}$$

where I is the three-dimensional pseudoscalar and $[e_1, e_2, e_3]$ denotes the scalar triple product, we arrive at the standard formula

$$e^1 = \frac{(e_2 \wedge e_3) I^{-1}}{[e_1, e_2, e_3]} = \frac{e_2 \times e_3}{[e_1, e_2, e_3]}, \tag{4.99}$$

with similar results holding for e^2 and e^3. Here the bold cross $\times$ denotes the vector cross product, not to be confused with the commutator product. Figure 4.2 illustrates the geometry involved in defining the reciprocal frame.

4.3.2 Components

The basis vectors $\{e_k\}$ are linearly independent, so any vector a can be written uniquely in terms of this set as

$$a = a^i e_i = a_i e^i. \tag{4.100}$$

We continue to employ the summation convention and summed indices appear once as a superscript and once as a subscript. The set of scalars $(a^1, \ldots, a^n)$ are the *components* of the vector a in the $\{\mathsf{e}_k\}$ frame. To find the components we form

$$a \cdot \mathsf{e}^i = a^j \mathsf{e}_j \cdot \mathsf{e}^i = a^j \delta_j^i = a^i \tag{4.101}$$

and

$$a \cdot \mathsf{e}_i = a_j \mathsf{e}^j \cdot \mathsf{e}_i = a_j \delta_i^j = a_i. \tag{4.102}$$

These formulae explain the labelling scheme for the components. In many applications we are only interested in orthonormal frames in Euclidean space. In this case the frame and its reciprocal are equivalent, and there is no need for the distinct subscript and superscript indices. The notation is unavoidable in mixed signature spaces, however, and is very useful in differential geometry, so it is best to adopt it at the outset.

Combining the equations (4.100), (4.101) and (4.102) we see that

$$a \cdot \mathsf{e}_i \, \mathsf{e}^i = a \cdot \mathsf{e}^i \, \mathsf{e}_i = a. \tag{4.103}$$

This holds for any vector a in the space spanned by the $\{\mathsf{e}_k\}$. This result generalises simply to arbitrary multivectors. First, for the bivector $a \wedge b$ we have

$$\mathsf{e}_i \, \mathsf{e}^i \cdot (a \wedge b) = \mathsf{e}_i \, \mathsf{e}^i \cdot a \, b - \mathsf{e}_i \, \mathsf{e}^i \cdot b \, a = ab - ba = 2a \wedge b. \tag{4.104}$$

This extends for an arbitrary grade-r multivector A_r to give

$$\mathsf{e}_i \, \mathsf{e}^i \cdot A_r = r A_r. \tag{4.105}$$

Since $\mathsf{e}_i \mathsf{e}^i = n$, we also see that

$$\mathsf{e}_i \, \mathsf{e}^i \wedge A_r = \mathsf{e}_i (\mathsf{e}^i A_r - \mathsf{e}^i \cdot A_r) = (n - r) A_r. \tag{4.106}$$

Subtracting the two preceding results we obtain,

$$\mathsf{e}_i A_r \mathsf{e}^i = (-1)^r (n - 2r) A_r. \tag{4.107}$$

The $\{\mathsf{e}_k\}$ basis extends easily to provide a basis for the entire algebra generated by the basis vectors. We can then decompose any multivector A into a set of components through

$$A_{i \cdots jk} = \langle (\mathsf{e}_k \wedge \mathsf{e}_j \cdots \wedge \mathsf{e}_i) A \rangle. \tag{4.108}$$

and

$$A = \sum_{i < j \cdots < k} A_{ij \cdots k} \mathsf{e}^i \wedge \cdots \wedge \mathsf{e}^j \wedge \mathsf{e}^k. \tag{4.109}$$

The components $A_{ij \cdots k}$ are totally antisymmetric on all indices and are usually referred to as the components of an *antisymmetric tensor*. We shall have more to say about tensors in following sections.

4.3.3 Application — recovering a rotor

As an application of the preceding results, suppose that we have two sets of vectors in three dimensions $\{\mathsf{e}_k\}$ and $\{\mathsf{f}_k\}$, $k = 1, 2, 3$. The vectors need not be orthonormal, but we know that the two sets are related by a rotation. The rotation is governed by the formula

$$\mathsf{f}_k = R\mathsf{e}_k R^\dagger \tag{4.110}$$

and we seek a simple expression for the rotor R. In three dimensions the rotor R can be written as

$$R = \exp(-B/2) = \alpha - \beta B, \tag{4.111}$$

where

$$\alpha = \cos(|B|/2), \quad \beta = \frac{\sin(|B|/2)}{|B|}. \tag{4.112}$$

The reverse is

$$R^\dagger = \exp(B/2) = \alpha + \beta B. \tag{4.113}$$

We therefore find that

$$\begin{aligned} \mathsf{e}_k R^\dagger \mathsf{e}^k &= \mathsf{e}_k(\alpha + \beta B)\mathsf{e}^k \\ &= 3\alpha - \beta B \\ &= 4\alpha - R^\dagger. \end{aligned} \tag{4.114}$$

We now form

$$\mathsf{f}_k \mathsf{e}^k = R\mathsf{e}_k R^\dagger \mathsf{e}^k = 4\alpha R - 1. \tag{4.115}$$

It follows that R is a scalar multiple of $1 + \mathsf{f}_k\mathsf{e}^k$. We therefore establish the simple formula

$$R = \frac{1 + \mathsf{f}_k\mathsf{e}^k}{|1 + \mathsf{f}_k\mathsf{e}^k|} = \frac{\psi}{\sqrt{(\psi\tilde{\psi})}}, \tag{4.116}$$

where $\psi = 1 + \mathsf{f}_k\mathsf{e}^k$. This compact formula recovers the rotor directly from the frame vectors. A problem arises if the rotation is through precisely 180°, in which case ψ vanishes. This case can be dealt with simply enough by considering the image of two of the three vectors.

4.4 Linear algebra

Many key relations in physics involve linear mappings between two, sometimes different, spaces. These are the subject of tensor analysis in the standard literature. Examples include the stress and strain tensors of elasticity, the conductivity tensor of electromagnetism and the inertia tensor of dynamics. If one has only met the study of linear transformations through tensor analysis, one could be forgiven for thinking that the subject cannot be discussed without a large dose of index notation. The indices refer to components of tensors in some frame, though the essence of tensor analysis is to establish a set of results which are independent of the choice of frame. In our opinion, this subject is much more simply dealt with if one can avoid specifying a frame until it is absolutely necessary. Perhaps unsurprisingly, it is geometric algebra that provides precisely the tools necessary to achieve such a development.

In this section we use capital, sans-serif symbols for linear functions. This helps to distinguish functions from their multivector argument. The dimension and signature of the vector space is arbitrary unless otherwise specified. We assume that readers are familiar with the basic properties of linear transformations in the guise of matrices. Suppose, then, that we are interested in a quantity F which maps vectors to vectors linearly in the same space. That is, if a is a vector in the space acted on by F, then $\mathsf{F}(a)$ lies in the same space. The linearity of F is expressed by

$$\mathsf{F}(\lambda a + \mu b) = \lambda \mathsf{F}(a) + \mu \mathsf{F}(b), \tag{4.117}$$

for scalars λ and μ and vectors a and b. Geometrically, we can think of F as an instruction to take a vector and rotate/dilate it to a new vector. No frame or components are required for such a picture. A simple example is provided by a rotation, which can be written as

$$\mathsf{R}(a) = RaR^{\dagger}, \tag{4.118}$$

where R is a rotor. It is a simple matter to confirm that this map is linear.

4.4.1 Extension to multivectors

Once one has formulated the action of a linear function on a vector, the obvious next step is to let the function act on a multivector. In this way we extend the action of a linear function to the full geometric algebra defined by the underlying vector space. Suppose that two vectors a and b are acted on by the linear function F. The bivector $a \wedge b$ then transforms to $\mathsf{F}(a) \wedge \mathsf{F}(b)$. We take this as the definition for the action of F on a bivector blade:

$$\mathsf{F}(a{\wedge}b) = \mathsf{F}(a){\wedge}\mathsf{F}(b). \tag{4.119}$$

Since the right-hand side is the outer product of two vectors, it is also a bivector blade (see figure 4.3). The action on sums of blades is defined by the linearity of F:

$$\mathsf{F}(a{\wedge}b + c{\wedge}d) = \mathsf{F}(a{\wedge}b) + \mathsf{F}(c{\wedge}d). \tag{4.120}$$

Continuing in this manner, we define the action of F on an arbitrary blade by

$$\mathsf{F}(a{\wedge}b{\wedge}\cdots{\wedge}c) = \mathsf{F}(a){\wedge}\mathsf{F}(b){\wedge}\cdots{\wedge}\mathsf{F}(c). \tag{4.121}$$

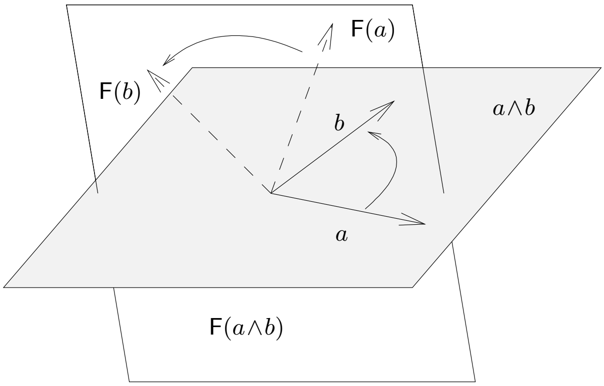


Figure 4.3 *The extended linear function.* The action of F on the bivector $a \wedge b$ results in the new plane $\mathsf{F}(a) \wedge \mathsf{F}(b)$. This is the definition of $\mathsf{F}(a \wedge b)$.

Extension by linearity then defines the action of F on arbitrary multivectors. By construction, F is both linear over multivectors,

$$\mathsf{F}(\lambda A + \mu B) = \lambda \mathsf{F}(A) + \mu \mathsf{F}(B), \tag{4.122}$$

and grade-preserving,

$$\mathsf{F}(A_r) = \langle \mathsf{F}(A_r) \rangle_r, \tag{4.123}$$

where A_r is a grade-r multivector. A simple example is provided by rotations. We have already established a formula for the result of rotating all of the vectors in a blade. For the extension of a rotation we therefore have

$$\begin{aligned} \mathsf{R}(a \wedge b \wedge \cdots \wedge c) &= (RaR^\dagger) \wedge (RbR^\dagger) \wedge \cdots \wedge (RcR^\dagger) \\ &= R\, a \wedge b \wedge \cdots \wedge c\, R^\dagger. \end{aligned} \tag{4.124}$$

It follows that acting on an arbitrary multivector A we have

$$\mathsf{R}(A) = RAR^\dagger. \tag{4.125}$$

Again, it is simple to confirm that this has the expected properties.

4.4.2 The product

The product of two linear functions is formed by letting a second function act on the result of the first function. Thus the action of the product of F and G is defined by

$$(\mathsf{FG})(a) = \mathsf{F}\big(\mathsf{G}(a)\big) = \mathsf{FG}(a). \tag{4.126}$$

The final expression enables us to remove some brackets without any ambiguity. A price to pay for removing indices is that brackets are often required to show

how calculations are ordered. Any convention that enables brackets to be systematically dropped is then well worth adopting. It is straightforward to show that FG is a linear function if F and G are both linear:

$$\mathsf{FG}(\lambda a + \mu b) = \mathsf{F}(\lambda \mathsf{G}(a) + \mu \mathsf{G}(b)) = \lambda \mathsf{FG}(a) + \mu \mathsf{FG}(b). \tag{4.127}$$

Next we form the extension of a product function. Suppose that H is given by the product of F and G:

$$\mathsf{H}(a) = \mathsf{F}\big(\mathsf{G}(a)\big) = \mathsf{FG}(a). \tag{4.128}$$

It follows that

$$\begin{aligned} \mathsf{H}(a\wedge b\wedge\cdots\wedge c) &= \mathsf{F}\big(\mathsf{G}(a)\big)\wedge\mathsf{F}\big(\mathsf{G}(b)\big)\wedge\cdots\wedge\mathsf{F}\big(\mathsf{G}(c)\big) \\ &= \mathsf{F}\big(\mathsf{G}(a)\wedge\mathsf{G}(b)\wedge\cdots\wedge\mathsf{G}(c)\big) \\ &= \mathsf{F}\big(\mathsf{G}(a\wedge b\wedge\cdots\wedge c)\big), \end{aligned} \tag{4.129}$$

so the multilinear action of the product of two linear functions is the product of their exterior actions. In dealing with combinations of linear functions we can therefore write

$$\mathsf{H}(A) = \mathsf{FG}(A), \tag{4.130}$$

since the meaning of the right-hand side is unambiguous.

4.4.3 The adjoint

Given a linear function F, the adjoint, or transpose, $\bar{\mathsf{F}}$ is defined so that

$$a\cdot\bar{\mathsf{F}}(b) = \mathsf{F}(a)\cdot b, \tag{4.131}$$

for all vectors a and b. If F is a mapping from one vector space to another, then the adjoint function maps from the second space back to the first. In terms of an arbitrary frame $\{\mathsf{e}_k\}$ we have

$$\mathsf{e}_i\cdot\bar{\mathsf{F}}(a) = a\cdot\mathsf{F}(\mathsf{e}_i), \tag{4.132}$$

so we can construct the adjoint using

$$\mathrm{ad}(\mathsf{F})(a) = \bar{\mathsf{F}}(a) = \mathsf{e}^i\, a\cdot\mathsf{F}(\mathsf{e}_i). \tag{4.133}$$

The notation of a bar for the adjoint, rather than a superscript T or †, is slightly unconventional, though it does agree with that of Hestenes & Sobczyk (1984). The notation is very useful in handwritten work, where it is also convenient to denote the linear function with an underline. Some formulae relating functions and their adjoints have a neat symmetry when this overbar/underbar convention is followed.

The operation of taking the adjoint of the adjoint of a function returns the original function. This is verified by forming

$$\mathrm{ad}(\bar{\mathsf{F}})(a) = \mathsf{e}^i a\cdot\bar{\mathsf{F}}(\mathsf{e}_i) = \mathsf{e}^i\, \mathsf{e}_i\cdot\mathsf{F}(a) = \mathsf{F}(a). \tag{4.134}$$

The adjoint of a product of two functions is found as follows:

$$\begin{aligned}\mathrm{ad}(\mathsf{FG})(a) &= \mathsf{e}^i\, a\cdot\mathsf{FG}(\mathsf{e}_i) = \bar{\mathsf{F}}(a)\cdot\mathsf{G}(\mathsf{e}_i)\,\mathsf{e}^i \\ &= \bar{\mathsf{G}}\bar{\mathsf{F}}(a)\cdot\mathsf{e}_i\,\mathsf{e}^i = \bar{\mathsf{G}}\bar{\mathsf{F}}(a).\end{aligned} \tag{4.135}$$

The operation of taking the adjoint of a product therefore reverses the order in which the linear functions act. A *symmetric* function is one which is equal to its own adjoint, $\bar{\mathsf{F}} = \mathsf{F}$. Two particularly significant examples of symmetric functions are the functions $\mathsf{F}\bar{\mathsf{F}}$ and $\bar{\mathsf{F}}\mathsf{F}$. To verify that these are symmetric we form

$$\mathrm{ad}(\mathsf{F}\bar{\mathsf{F}}) = \mathrm{ad}(\bar{\mathsf{F}})\mathrm{ad}(\mathsf{F}) = \mathsf{F}\bar{\mathsf{F}}, \tag{4.136}$$

with a similar derivation holding for $\bar{\mathsf{F}}\mathsf{F}$. These functions will be met again later in this chapter.

The adjoint is still a linear function, so its extension to arbitrary multivectors is precisely as expected:

$$\bar{\mathsf{F}}(a\wedge b\wedge\cdots\wedge c) = \bar{\mathsf{F}}(a)\wedge\bar{\mathsf{F}}(b)\wedge\cdots\wedge\bar{\mathsf{F}}(c). \tag{4.137}$$

If we now consider two bivectors $a_1 \wedge a_2$ and $b_1 \wedge b_2$, we find that

$$\begin{aligned}(a_1\wedge a_2)\cdot\mathsf{F}(b_1\wedge b_2) &= a_1\cdot\mathsf{F}(b_2)\, a_2\cdot\mathsf{F}(b_1) - a_1\cdot\mathsf{F}(b_1)\, a_2\cdot\mathsf{F}(b_2) \\ &= \bar{\mathsf{F}}(a_1)\cdot b_2\, \bar{\mathsf{F}}(a_2)\cdot b_1 - \bar{\mathsf{F}}(a_1)\cdot b_1\, \bar{\mathsf{F}}(a_2)\cdot b_2 \\ &= \bar{\mathsf{F}}(a_1\wedge a_2)\cdot(b_1\wedge b_2).\end{aligned} \tag{4.138}$$

It follows that for two bivectors B_1 and B_2

$$B_1\cdot\bar{\mathsf{F}}(B_2) = \mathsf{F}(B_1)\cdot B_2. \tag{4.139}$$

This result extends for arbitrary multivectors to give

$$\langle A\bar{\mathsf{F}}(B)\rangle = \langle\mathsf{F}(A)B\rangle. \tag{4.140}$$

This is a special case of an even more general and powerful result. Consider the expression

$$\begin{aligned}\mathsf{F}(a\wedge b)\cdot c &= \mathsf{F}(a)\,\mathsf{F}(b)\cdot c - \mathsf{F}(b)\,\mathsf{F}(a)\cdot c \\ &= \mathsf{F}\big(a\, b\cdot\bar{\mathsf{F}}(c) - b\, a\cdot\bar{\mathsf{F}}(c)\big) \\ &= \mathsf{F}\big((a\wedge b)\cdot\bar{\mathsf{F}}(c)\big).\end{aligned} \tag{4.141}$$

Building up in this way we establish the useful results:

$$\begin{aligned}A_r\cdot\bar{\mathsf{F}}(B_s) &= \bar{\mathsf{F}}\big(\mathsf{F}(A_r)\cdot B_s\big) \qquad r \le s, \\ \mathsf{F}(A_r)\cdot B_s &= \mathsf{F}\big(A_r\cdot\bar{\mathsf{F}}(B_s)\big) \qquad r \ge s.\end{aligned} \tag{4.142}$$

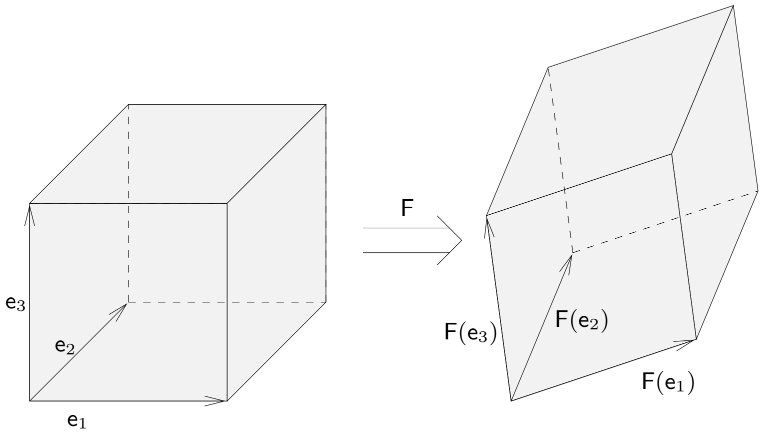


Figure 4.4 *The determinant.* The unit cube is transformed to a parallelepiped with sides $\mathsf{F}(\mathsf{e}_1)$, $\mathsf{F}(\mathsf{e}_2)$ and $\mathsf{F}(\mathsf{e}_3)$. The determinant is the volume scale factor, so is given by the volume of the parallelepiped, $\mathsf{F}(\mathsf{e}_1)\wedge\mathsf{F}(\mathsf{e}_2)\wedge\mathsf{F}(\mathsf{e}_3) = \mathsf{F}(I)$.

These reduce to equation (4.140) in the case when $r = s$. One way to think of these formulae is as follows. In the expression $\mathsf{F}(A_r) \cdot B_s$, with $r \geq s$, there are r separate applications of the function F on vectors. When the result is contracted with B_s, s of these applications are converted to adjoint functions $\bar{\mathsf{F}}$. The remaining $r - s$ applications act on the multivector $A_r \cdot \bar{\mathsf{F}}(B_s)$, which has grade $r - s$.

4.4.4 The determinant

Now that we have seen how a linear function defines an action on the entire geometric algebra, we can give a very compact definition of the determinant. The pseudoscalar for any space is unique up to scaling, and linear functions are grade-preserving, so we define

$$\mathsf{F}(I) = \det(\mathsf{F})\, I. \tag{4.143}$$

It should be immediately apparent that this definition of the determinant is much more compact and intuitive than the matrix definition (discussed later). The definition (4.143) shows clearly that the determinant is the volume scale factor for the operation F. In particular, acting on the unit hypercube, the result $\mathsf{F}(I)$ returns the directed volume of the resultant object (see figure 4.4).

As an example of the power of the geometric algebra definition, consider the product of two functions, F and G. From equation (4.130) it follows that

$$\det(\mathsf{FG})I = \mathsf{FG}(I) = \det(\mathsf{G})\,\mathsf{F}(I) = \det(\mathsf{F})\det(\mathsf{G})\, I, \tag{4.144}$$

which establishes that the determinant of the product of two functions is the product of their determinants. This is one of the key properties of the determinant, yet in conventional developments it is hard to prove. By contrast, the geometric algebra approach establishes the result in a few lines. Similarly, one can easily establish that the determinant of the adjoint is the same as that of the original function,

$$\det(\mathsf{F}) = \langle \mathsf{F}(I) I^{-1} \rangle = \langle I \bar{\mathsf{F}}(I^{-1}) \rangle = \det(\bar{\mathsf{F}}). \tag{4.145}$$

Example 4.1

Consider the linear function

$$\mathsf{F}(a) = a + \alpha a \cdot f_1 \, f_2, \tag{4.146}$$

where α is a scalar and f_1 and f_2 are a pair of arbitrary vectors. Construct the action of F on a general multivector and find its determinant.

We start by forming

$$\begin{aligned} \mathsf{F}(a \wedge b) &= (a + \alpha a \cdot f_1 f_2) \wedge (b + \alpha b \cdot f_1 f_2) \\ &= a \wedge b + \alpha (b \cdot f_1 a - a \cdot f_1 b) \wedge f_2 \\ &= a \wedge b + \alpha \big((a \wedge b) \cdot f_1\big) \wedge f_2. \end{aligned} \tag{4.147}$$

It follows that

$$\mathsf{F}(A) = A + \alpha (A \cdot f_1) \wedge f_2. \tag{4.148}$$

The determinant is now calculated as follows:

$$\begin{aligned} \mathsf{F}(I) &= I + \alpha (I \cdot f_1) \wedge f_2 \\ &= I + \alpha f_1 \cdot f_2 \, I, \end{aligned} \tag{4.149}$$

hence $\det(\mathsf{F}) = 1 + \alpha f_1 \cdot f_2$.

4.4.5 The inverse

We now construct a simple, explicit formula for the inverse of a linear function. We start by considering a multivector B, lying entirely in the algebra defined by the pseudoscalar I. For these we have

$$\det(\mathsf{F}) I B = \mathsf{F}(I) B = \mathsf{F}\big(I \bar{\mathsf{F}}(B)\big), \tag{4.150}$$

where we have used the adjoint formulae of equation (4.142). The inner product with a pseudoscalar is replaced with a geometric product, since no other grades are present in the full product. Replacing IB by A we find that

$$\det(\mathsf{F}) A = \mathsf{F}\big(I \bar{\mathsf{F}}(I^{-1} A)\big) \tag{4.151}$$

with a similar result holding for the adjoint. It follows that

$$\begin{aligned} \mathsf{F}^{-1}(A) &= I\bar{\mathsf{F}}(I^{-1}A)\,\det\,(\mathsf{F})^{-1}, \\ \bar{\mathsf{F}}^{-1}(A) &= I\mathsf{F}(I^{-1}A)\,\det\,(\mathsf{F})^{-1}. \end{aligned} \tag{4.152}$$

These relations provide simple, explicit formulae for the inverse of a function. The derivation of these formulae is considerably quicker than anything available in traditional matrix/tensor analysis.

Example 4.2

Find the inverse of the function defined in equation (4.146).

With

$$\mathsf{F}(A) = A + \alpha(A\cdot f_1)\wedge f_2 \tag{4.153}$$

we have

$$\begin{aligned} \langle A_r \mathsf{F}(B_r)\rangle &= \langle A_r B_r\rangle + \alpha\langle A_r(B_r\cdot f_1)\wedge f_2\rangle \\ &= \langle A_r B_r\rangle + \alpha\langle f_2\cdot A_r B_r f_1\rangle, \end{aligned} \tag{4.154}$$

hence

$$\bar{\mathsf{F}}(A) = A + \alpha f_1\wedge(f_2\cdot A). \tag{4.155}$$

It follows that

$$\begin{aligned} \mathsf{F}^{-1}(A) &= \big(I^{-1}A + \alpha f_1\wedge(f_2\cdot(I^{-1}A))\big)(1+\alpha f_1\cdot f_2)^{-1} \\ &= (A + \alpha f_1\cdot(f_2\wedge A))(1+\alpha f_1\cdot f_2)^{-1} \\ &= A - \frac{\alpha}{1+\alpha f_1\cdot f_2} f_2\wedge(f_1\cdot A). \end{aligned} \tag{4.156}$$

Example 4.3

Find the inverse of the rotation

$$\mathsf{R}(a) = RaR^\dagger, \tag{4.157}$$

where R is a rotor.

We have already seen that the action of R on a general multivector is

$$\mathsf{R}(A) = RAR^\dagger \quad \text{and} \quad \bar{\mathsf{R}}(A) = R^\dagger AR \tag{4.158}$$

Hence

$$\det\,(\mathsf{R})I = RIR^\dagger = IRR^\dagger = I, \tag{4.159}$$

so $\det\,(\mathsf{R}) = 1$. It follows that

$$\mathsf{R}^{-1}(A) = IR^\dagger I^{-1}AR = R^\dagger AR = \bar{\mathsf{R}}(a), \tag{4.160}$$

so, as expected, the inverse of a rotation is the same as the adjoint. This is the definition of an orthogonal transformation.

4.4.6 Eigenvectors and eigenblades

We assume that readers are familiar with the concept of an eigenvalue and eigenvector of a matrix. All of the standard results for these have obvious counterparts in the geometric algebra framework. This subject will be explored more thoroughly in chapter 11. Here we give a simple outline, concentrating on the new concepts that geometric algebra offers. A linear function F has an eigenvector e if

$$\mathsf{F}(e) = \lambda e. \tag{4.161}$$

The scalar λ is the associated eigenvalue. It follows that

$$\det\,(\mathsf{F} - \lambda \mathsf{I}) = 0, \tag{4.162}$$

which defines a polynomial equation for λ. Techniques for finding eigenvalues and eigenvectors are discussed widely in the literature.

In general, the polynomial equation for λ will have complex roots. Traditional developments of the subject usually allow these and consider linear superpositions over the complex field. But if one starts with a real mapping between real vectors it is not clear that this formal complexification is useful. What one would like would be a more geometric classification of a general linear transformation. This is provided by the notion of an *eigenblade*. We extend the notion of an eigenvector to that of an eigenblade A_r satisfying

$$\mathsf{F}(A_r) = \lambda A_r, \tag{4.163}$$

where A_r is a grade-r blade and λ is real. One immediate example is the pseudoscalar, for which $\lambda = \det\,(\mathsf{F})$. More generally, each eigenblade determines an invariant subspace of the transformation.

As an example of the geometric clarity of the eigenblade concept, consider a function satisfying

$$\mathsf{F}(e_1) = \lambda e_2, \quad \mathsf{F}(e_2) = -\lambda e_1. \tag{4.164}$$

Traditionally, one might write that $e_1 \pm ie_2$ are eigenvectors with eigenvalues $\mp i\lambda$, where i is the unit imaginary. But the identity

$$\mathsf{F}(e_1 \wedge e_2) = \lambda^2 e_1 \wedge e_2 \tag{4.165}$$

identifies the plane $e_1 \wedge e_2$ as an eigenbivector of F. The role of the complex structure inherent in F is played by the unit bivector $e_1 \wedge e_2$. A linear function can have many distinct eigenbivectors, each acting as a distinct imaginary for its own plane. Replacing all of these by a single scalar imaginary throws away a considerable amount of useful information.

4.4.7 Symmetric and antisymmetric functions

An important aspect of the theory of linear functions is finding natural, *canonical*† expressions for a function. For symmetric functions in Euclidean space this form is via its spectral decomposition. If e_i and e_j are eigenvectors of a function, with eigenvalues λ_i and λ_j, we have (no sums implied)

$$e_i \cdot \mathsf{F}(e_j) = e_i \cdot (\lambda_j e_j) = \lambda_j e_i \cdot e_j. \tag{4.166}$$

But if F is symmetric, this also equals

$$\bar{\mathsf{F}}(e_i) \cdot e_j = \mathsf{F}(e_i) \cdot e_j = (\lambda_i e_i) \cdot e_j = \lambda_i e_i \cdot e_j. \tag{4.167}$$

It follows that

$$(\lambda_i - \lambda_j) e_i \cdot e_j = 0, \tag{4.168}$$

so eigenvectors of a symmetric function with distinct eigenvalues must be orthogonal.

If we admit the existence of complex eigenvectors and eigenvalues we also find that (no sums)

$$e^* \cdot \mathsf{F}(e) = \lambda e^* \cdot e = \mathsf{F}(e^*) \cdot e = \lambda^* e^* \cdot e. \tag{4.169}$$

So for any symmetric function we also have

$$(\lambda - \lambda^*) e^* \cdot e = 0. \tag{4.170}$$

Provided $e^* \cdot e \neq 0$ we can conclude that the eigenvalue, and hence the eigenvector, is real. In Euclidean space this inequality is always satisfied, and every symmetric function on an n-dimensional space has a spectral decomposition of the form

$$\mathsf{F}(a) = \lambda_1 \mathsf{P}_1(a) + \lambda_2 \mathsf{P}_2(a) + \cdots + \lambda_m \mathsf{P}_m(a). \tag{4.171}$$

Here $\lambda_1 < \lambda_2 < \cdots < \lambda_m$ are the m distinct eigenvalues ($m \leq n$) and the P_i are projections onto each of the invariant subspaces defined by the eigenvectors. For the case of a projection onto a one-dimensional space we have simply

$$\mathsf{P}_i(a) = a \cdot e_i \, e_i. \tag{4.172}$$

The eigenvectors form an orthonormal frame, which is the natural frame in which to study the linear function. If two eigenvalues are the same, it is always possible to choose the eigenvectors so that they remain orthogonal. In non-Euclidean spaces, such as spacetime, one has to be careful due to the possibility of complex null vectors. These can have $e^* \cdot e = 0$, so the above reasoning breaks down and

† The origin of the use of the word *canonical* is obscure — see for example the comments in Goldstein (1950). In mathematical physics, a canonical form usually refers to a standard way of simplifying an expression without altering its meaning.

one cannot guarantee the existence of an orthonormal frame of eigenvectors. We will encounter examples of this when we study gravitation.

Antisymmetric functions have $\bar{\mathsf{F}}(a) = -\mathsf{F}(a)$. It follows that

$$a\cdot\mathsf{F}(a) = \bar{\mathsf{F}}(a)\cdot a = -\mathsf{F}(a)\cdot a = 0. \tag{4.173}$$

The natural way to study antisymmetric functions is through the bivector

$$F = \frac{1}{2}\mathsf{e}^i \wedge \mathsf{F}(\mathsf{e}_i), \tag{4.174}$$

where the $\{\mathsf{e}_k\}$ are an arbitrary frame for the space acted on by F. The bivector F is independent of the choice of frame, so is an invariant quantity. One can easily confirm that the bivector F has the same number of degrees of freedom as F. If we now form $2a\cdot F$ we find that

$$\begin{aligned}2a\cdot F &= a\cdot\big(\mathsf{e}^i \wedge \mathsf{F}(\mathsf{e}_i)\big)\\ &= a\cdot\mathsf{e}^i\, \mathsf{F}(\mathsf{e}_i) - \mathsf{e}^i a\cdot\mathsf{F}(\mathsf{e}_i)\\ &= \mathsf{F}(a\cdot\mathsf{e}^i\, \mathsf{e}_i) + \mathsf{e}^i\, \mathsf{e}_i\cdot\mathsf{F}(a)\\ &= 2\mathsf{F}(a).\end{aligned} \tag{4.175}$$

The action of an antisymmetric function therefore reduces to contracting with the *characteristic bivector* F:

$$\mathsf{F}(a) = a\cdot F. \tag{4.176}$$

The problem of reducing an antisymmetric function to its simplest form reduces to that of splitting F into a set of commuting blades:

$$F = \lambda_1 \hat{F}_1 + \cdots + \lambda_k \hat{F}_k, \tag{4.177}$$

where $k \leq n/2$ and each of the $\hat{F}_i$ is a unit blade. This decomposition is always possible in Euclidean space, though the answer is only unique if the blades all have different magnitudes. Each component blade of F is an eigenblade of F and determines an invariant subspace. Within this subspace the effect of F is simply to rotate all vectors by $\pm 90°$, and to scale the result by the magnitude of the eigenblade. In non-Euclidean spaces such a decomposition is not always possible.

4.4.8 The singular value decomposition

For linear functions of no symmetry a number of alternative canonical forms can be found. Among these, perhaps the most useful is the singular value decomposition. We start with an arbitrary function F and restrict the discussion to the case where F acts on an n-dimensional Euclidean space. We also suppose that $\det(\mathsf{F}) \neq 0$; the case of $\det(\mathsf{F}) = 0$ is easily dealt with by separating out the space which is mapped onto the origin, and working with a reduced function acting in the subspace over which F is non-singular. We next form the function D by

$$\mathsf{D}(a) = \bar{\mathsf{F}}\mathsf{F}(a). \tag{4.178}$$

This function is symmetric and has n orthogonal eigenvectors with real, positive eigenvalues. The fact that the eigenvalues are positive follows from

$$\bar{\mathsf{F}}\mathsf{F}(e) = \lambda e \quad \Rightarrow \quad \mathsf{F}(e)\cdot\mathsf{F}(e) = \lambda e^2. \tag{4.179}$$

Since (in Euclidean space) the square of any vector is a positive scalar we see that λ must be positive. The assumption that det $(\mathsf{F}) \neq 0$ rules out the possibility of any eigenvalues being zero. It follows that we can write

$$\mathsf{D}(a) = \sum_{i=1}^{n} \lambda_i a\cdot\mathsf{e}_i\, \mathsf{e}_i, \tag{4.180}$$

where the $\{\mathsf{e}_i\}$ are the *orthonormal* frame of eigenvectors. Degenerate eigenvalues are dealt with by picking a set of arbitrary orthonormal vectors in the invariant subspace.

The linear function D has a simple (positive) square root,

$$\mathsf{D}^{1/2} = \sum_{i=1}^{n} \lambda_i^{1/2} a\cdot\mathsf{e}_i\, \mathsf{e}_i \tag{4.181}$$

and this is also invertible,

$$\mathsf{D}^{-1/2} = \sum_{i=1}^{n} \lambda_i^{-1/2} a\cdot\mathsf{e}_i\, \mathsf{e}_i. \tag{4.182}$$

We now set

$$\mathsf{S} = \mathsf{F}\mathsf{D}^{-1/2}. \tag{4.183}$$

This satisfies

$$\bar{\mathsf{S}}\mathsf{S} = \mathsf{D}^{-1/2}\bar{\mathsf{F}}\mathsf{F}\mathsf{D}^{-1/2} = \mathsf{D}^{-1/2}\mathsf{D}\,\mathsf{D}^{-1/2} = \mathsf{I}, \tag{4.184}$$

where I is the identity function. It follows that S is an orthogonal function. The function F can now be written

$$\mathsf{F} = \mathsf{S}\mathsf{D}^{1/2}. \tag{4.185}$$

This represents a series of dilations along the eigendirections of D, followed by a rotation.

If the linear function F is presented as an $n \times n$ matrix of components in some frame, then one usually includes a further rotation R to align this arbitrary frame with the frame of eigenvectors. In this case one writes

$$\mathsf{F} = \mathsf{S}\Lambda^{1/2}\bar{\mathsf{R}}, \tag{4.186}$$

where Λ is a diagonal matrix in the arbitrary coordinate frame. This writes a matrix as a dilation sandwiched between two rotations, and is called the singular value decomposition of the matrix. An arbitrary linear function in n dimensions has n^2 degrees of freedom. The singular value decomposition assigns $2 \times n(n-1)/2$ of these to the two orthogonal transformations R and S, with the remaining n degrees of freedom contained in the dilation Λ. The singular value decomposition appears frequently in subjects such as data analysis, where it is often used in connection with analysing non-square matrices.

4.5 Tensors and components

Many modern physics textbooks are written in the language of tensor analysis. In this approach one often works directly with the components of a vector, or linear function, in a chosen coordinate frame. The invariance of the laws under a change of frame can then be used to advantage to simplify the component equations. Since this approach is so ubiquitous it is important to establish the relationship between tensor analysis and the largely frame-free approach of the present chapter. We start by analysing Cartesian tensors, and then move onto the more general case of an arbitrary coordinate frame.

4.5.1 Cartesian tensors

The subject of Cartesian tensors arises when we restrict our frames to consist only of orthonormal vectors in Euclidean space. For these we have

$$\mathsf{e}_i \cdot \mathsf{e}_j = \delta_{ij}, \tag{4.187}$$

so there is no distinction between frames and their reciprocals. In this case we can drop all distinction between raised and lowered indices, and just work with all indices lowered. Provided both frames have the same orientation, a new frame is obtained from the $\{\mathsf{e}_k\}$ frame by a rotation,

$$\mathsf{e}_i' = R\mathsf{e}_i R^\dagger = \Lambda_{ij}\mathsf{e}_j. \tag{4.188}$$

Here R is a rotor and Λ_{ij} are the components of the rotation defined by R:

$$\Lambda_{ij} = (R\mathsf{e}_i R^\dagger)\cdot\mathsf{e}_j. \tag{4.189}$$

It follows that

$$\begin{aligned}\Lambda_{ij}\Lambda_{ik} &= (R\mathsf{e}_i R^\dagger)\cdot\mathsf{e}_j (R\mathsf{e}_i R^\dagger)\cdot\mathsf{e}_k \\ &= (R^\dagger\mathsf{e}_j R)\cdot(R^\dagger\mathsf{e}_k R) = \delta_{jk},\end{aligned} \tag{4.190}$$

and similarly

$$\Lambda_{ik}\Lambda_{jk} = \delta_{ij}. \tag{4.191}$$

A vector a has components $a_i = \mathsf{e}_i \cdot a$ and these transform under a change of frame in the obvious manner,

$$a'_i = \mathsf{e}'_i \cdot a = \Lambda_{ij} a_j. \tag{4.192}$$

It is important to realise here that it is only the components of a that change, not the underlying vector itself. The change in components is exactly cancelled by the change in the frame. Many equations in physics are invariant if the vector itself is transformed, but this is the result of an underlying symmetry in the equations, and not of the freedom to choose the coordinate system. These two concepts should not be confused!

Extending this idea, we define the components of the linear function F by

$$\mathsf{F}_{ij} = \mathsf{e}_i \cdot \mathsf{F}(\mathsf{e}_j). \tag{4.193}$$

The result of this decomposition is an $n \times n$ array of components, which can be stored and manipulated as a matrix. This definition ensures that the components of the vector $\mathsf{F}(a)$ are given by

$$\mathsf{e}_i \cdot \mathsf{F}(a) = \mathsf{e}_i \cdot \mathsf{F}(a_j \mathsf{e}_j) = \mathsf{F}_{ij} a_j, \tag{4.194}$$

which is the usual expression for a matrix acting on a column vector. Similarly, if F and G are a pair of linear functions, the components of the product function FG are given by

$$\begin{aligned}(\mathsf{FG})_{ij} &= \mathsf{FG}(\mathsf{e}_j) \cdot \mathsf{e}_i = \mathsf{G}(\mathsf{e}_j) \cdot \bar{\mathsf{F}}(\mathsf{e}_i) \\ &= \mathsf{G}(\mathsf{e}_j) \cdot \mathsf{e}_k \, \mathsf{e}_k \cdot \bar{\mathsf{F}}(\mathsf{e}_i) = \mathsf{F}_{ik} \mathsf{G}_{kj}.\end{aligned} \tag{4.195}$$

This recovers the familiar rule for multiplying matrices. If the frame is changed to a new rotated frame, the components of the tensor transform in the obvious way:

$$\mathsf{F}'_{ij} = \Lambda_{ik} \Lambda_{jl} \mathsf{F}_{kl}, \tag{4.196}$$

where the prime denotes the components in the new (primed) frame. Objects with two indices are referred to as rank-2 tensors. Rank-1 tensors are vectors, rank-3 tensors have three indices, and so on. Since rank-2 tensors appear regularly in physics they are often referred to simply as tensors. Also, it is usual to let the term tensor refer to either the component form F_{ij} or the abstract entity F.

For Cartesian tensors there are two important tensors which arise regularly in computations. These are the two *invariant* tensors. The first of these is the Kronecker δ, which transforms as

$$\delta'_{ij} = \Lambda_{ik} \Lambda_{jl} \delta_{kl} = \Lambda_{ik} \Lambda_{jk} = \delta_{ij}. \tag{4.197}$$

The components of the identity function are therefore the same in all orthonormal frames (and are those of the identity matrix in all cases). The second invariant is the alternating tensor $\epsilon_{ij\cdots k}$, where the number of indices matches the dimension of the space. This is totally antisymmetric and is defined as follows:

$$\epsilon_{ij\cdots k} = \begin{cases} 1 & i,\, j,\, \ldots,\, k = \text{even permutation of } 1, 2, \ldots, n \\ -1 & i,\, j,\, \ldots,\, k = \text{odd permutation of } 1, 2, \ldots, n \\ 0 & \text{otherwise} \end{cases} . \tag{4.198}$$

The order of a permutation is the number of pairwise swaps required to return to the original order $1, 2, \ldots, n$. If an even number of swaps is required the permutation is even, and similarly for the odd case. In three dimensions even permutations of $1, 2, 3$ coincide with cyclic orderings of the indices. The determinant of a matrix can be expressed in terms of the alternating tensor via

$$\mathsf{F}_{\alpha i}\mathsf{F}_{\beta j} \cdots \mathsf{F}_{\gamma k}\epsilon_{\alpha\beta\cdots\gamma} = \det\,(\mathsf{F})\,\epsilon_{ij\cdots k}. \tag{4.199}$$

Given this result, it is straightforward to prove the frame invariance of the alternating tensor under rotations:

$$\epsilon'_{ij\cdots k} = \Lambda_{i\alpha}\Lambda_{j\beta} \cdots \Lambda_{k\gamma}\epsilon_{\alpha\beta\cdots\gamma} = \det\,(\Lambda)\,\epsilon_{ij\cdots k}. \tag{4.200}$$

But since Λ_{ij} is a rotation matrix it has determinant $+1$, so the tensor is indeed invariant.

4.5.2 The determinant revisited

We should now establish that the definition of the determinant (4.199) agrees with our earlier definition (4.143). To prove this we first need the result that

$$\epsilon_{ij\cdots k} = \mathsf{e}_i \wedge \mathsf{e}_j \cdots \wedge \mathsf{e}_k\, I^\dagger, \tag{4.201}$$

where $I = \mathsf{e}_1\mathsf{e}_2 \cdots \mathsf{e}_n$ and the $\{\mathsf{e}_k\}$ form an orthonormal frame. The right-hand side of (4.201) is zero if any of the indices are the same, because of the antisymmetry of the outer product. If the indices form an even permutation of $1, 2, \ldots, n$ we can reorder the vectors into the order $\mathsf{e}_1\mathsf{e}_2 \cdots \mathsf{e}_n = I$, in which case the right-hand side of (4.201) returns $+1$. Similarly, any anticyclic combination of $1, 2, \ldots, n$ returns -1. Together these agree with the definition (4.198) of the alternating tensor $\epsilon_{ij\cdots k}$. We can now rearrange the left-hand side of (4.199) as follows:

$$\begin{aligned} \mathsf{F}_{\alpha i}\mathsf{F}_{\beta j} \cdots \mathsf{F}_{\gamma k}\epsilon_{\alpha\beta\cdots\gamma} &= \mathsf{F}_{\alpha i}\mathsf{F}_{\beta j} \cdots \mathsf{F}_{\gamma k}\, \mathsf{e}_\alpha \wedge \mathsf{e}_\beta \cdots \wedge \mathsf{e}_\gamma\, I^\dagger \\ &= \mathsf{F}(\mathsf{e}_i) \wedge \mathsf{F}(\mathsf{e}_j) \cdots \mathsf{F}(\mathsf{e}_k)\, I^\dagger \\ &= \det\,(\mathsf{F})\, \mathsf{e}_i \wedge \mathsf{e}_j \cdots \wedge \mathsf{e}_k\, I^\dagger \\ &= \det\,(\mathsf{F})\, \epsilon_{ij\cdots k}, \end{aligned} \tag{4.202}$$

which recovers the expected result.

We assume that most readers are familiar with the various techniques employed when computing the determinant of an $n \times n$ matrix. These can be found in most elementary textbooks on linear algebra. It is instructive to see how the same results arise in the geometric algebra treatment. We have already established that the determinant of the product of two functions is the product of the determinants, and that taking the adjoint does not change the determinant. To establish a further set of results we first introduce the (non-orthonormal) vectors $\{f_i\}$,

$$f_i \equiv \mathsf{F}(\mathsf{e}_i), \tag{4.203}$$

so that

$$\mathsf{F}_{ij} = \mathsf{e}_i \cdot f_j. \tag{4.204}$$

From equation (4.143) the determinant of F can be written

$$\det\,(\mathsf{F}) = (f_1 \wedge f_2 \wedge \cdots \wedge f_n) \cdot (\mathsf{e}_n \wedge \cdots \wedge \mathsf{e}_2 \wedge \mathsf{e}_1). \tag{4.205}$$

Expanding this product out in full recovers the standard expression for the determinant of a matrix. The first result we see is that swapping any two of the $\{f_i\}$ changes the sign of the determinant. This is the same as swapping two columns in the matrix F_{ij}. Since matrix transposition does not affect the result, the same is true for interchanging rows.

Next we single out one of the $\{\mathsf{e}_k\}$ vectors and write

$$\begin{aligned} \det\,(\mathsf{F}) &= (-1)^{j+1} (\mathsf{e}_n \wedge \cdots \check{\mathsf{e}}_j \cdots \wedge \mathsf{e}_1) \cdot \big(\mathsf{e}_j \cdot (f_1 \wedge \cdots \wedge f_n)\big) \\ &= \sum_{k=1}^{n} (-1)^{j+k}\, \mathsf{e}_j \cdot f_k \, (\mathsf{e}_n \wedge \cdots \check{\mathsf{e}}_j \cdots \wedge \mathsf{e}_1) \cdot (f_1 \wedge \cdots \check{f}_k \cdots \wedge f_n). \end{aligned} \tag{4.206}$$

The final part of each term in the sum corresponds to an $(n-1) \times (n-1)$ determinant, as can be seen by comparing with (4.205). This is equivalent to the familiar expression for the expansion of the determinant by the jth row. A further useful result is obtained from the identity

$$f_1 \wedge \cdots \wedge (f_j + \lambda f_k) \wedge \cdots \wedge f_n = f_1 \wedge \cdots \wedge f_j \wedge \cdots \wedge f_n \quad j \neq k. \tag{4.207}$$

This result means that any multiple of the kth row can be added to the jth row without changing the result. The same is true for columns. This is the key to the method of Gaussian elimination for finding a determinant. In this method the matrix is first transformed to upper (or lower) triangular form, so that the determinant is then simply the product of the entries down the leading diagonal. This is numerically a highly efficient method for calculating determinants. We can continue in this manner to give concise proofs of many of the key results for determinants. For a useful summary of these, see Turnbull (1960).

To see how these formulae also lead to the familiar expression for the inverse of a matrix, consider the decomposition:

$$
\begin{aligned}
\mathsf{F}^{-1}_{ij} &= \mathsf{e}_i \cdot \mathsf{F}^{-1}(\mathsf{e}_j) \\
&= \langle \mathsf{e}_i\, \mathsf{e}_1 \wedge \cdots \wedge \mathsf{e}_n\, \bar{\mathsf{F}}(\mathsf{e}_n \wedge \cdots \wedge \mathsf{e}_1\, \mathsf{e}_j) \rangle \det\, (\mathsf{F})^{-1} \\
&= (-1)^{i+j} \langle \mathsf{F}(\mathsf{e}_1 \wedge \cdots \check{\mathsf{e}}_i \cdots \wedge \mathsf{e}_n)\, \mathsf{e}_n \wedge \cdots \check{\mathsf{e}}_j \cdots \wedge \mathsf{e}_1 \rangle \det\, (\mathsf{F})^{-1}.
\end{aligned}
\tag{4.208}
$$

The term enclosed in angular brackets is the determinant of the $(n-1)\times(n-1)$ matrix obtained from F_{ij} by deleting the ith column and jth row. This is the definition of the i, j cofactor of F_{ij}. Equation (4.208) shows that the components of F^{-1}_{ij} are formed from the transposed matrix of cofactors, divided by the determinant $\det\, (\mathsf{F})$ — the familiar result. Similarly, all other matrix formulae have simple and often elegant counterparts in geometric algebra. Further examples of these are discussed in chapter 11.

4.5.3 General tensors

We now generalise the preceding treatment to the case of arbitrary basis sets in spaces of arbitrary (non-degenerate) signature. One reason for wanting to deal with non-orthonormal frames is that these regularly arise when working in curvilinear coordinate systems. In addition, in mixed signature spaces one has no option since it is impossible to identify a frame with its reciprocal. Suppose, then, that the vectors $\{\mathsf{e}_k\}$ constitute an arbitrary frame for n-dimensional space (of unspecified signature). The reciprocal frame is denoted $\{\mathsf{e}^k\}$ and the two frames are related by

$$\mathsf{e}^i \cdot \mathsf{e}_j = \delta^i_j. \tag{4.209}$$

Equation (4.94) for the reciprocal frame is general and still holds in mixed signature spaces.

As described in section 4.3.2, the vector a has components $(a^1, a^2, \ldots, a^n)$ in the $\{\mathsf{e}_k\}$ frame, and $(a_1, a_2, \ldots, a_n)$ in the $\{\mathsf{e}^k\}$ frame. When working with general coordinate frames we always ensure that upper and lower indices match separately on either side of an expression. Suppose we now form the inner product of two vectors a and b. We can write this as

$$a\cdot b = (a^i \mathsf{e}_i)\cdot(b_j \mathsf{e}^j) = a^i b_j\, \mathsf{e}_i \cdot \mathsf{e}^j = a^i b_j \delta^j_i = a^i b_i. \tag{4.210}$$

The general rule is that sums are only taken over pairs of indices where one is a superscript and the other a subscript. Another way to write an inner product is to introduce the *metric tensor* g_{ij}:

$$g_{ij} = \mathsf{e}_i \cdot \mathsf{e}_j. \tag{4.211}$$

In terms of its components g_{ij} is a symmetric $n \times n$ matrix. The inverse matrix is written as g^{ij} and is given by

$$g^{ij} = \mathsf{e}^i \cdot \mathsf{e}^j. \tag{4.212}$$

It is easily verified that this is the inverse of g_{ij}:

$$g^{ik} g_{kj} = \mathsf{e}^i \cdot \mathsf{e}^k \, \mathsf{e}_k \cdot \mathsf{e}_j = \mathsf{e}^i \cdot \mathsf{e}_j = \delta^i_j. \tag{4.213}$$

Employing the metric tensor we can write the inner product of two vectors in a number of equivalent forms:

$$a \cdot b = a^i b_i = a_i b^i = a^i b^j g_{ij} = a_i b_j g^{ij}. \tag{4.214}$$

Of course, all of these expressions encode the same thing and, unless there is a particular reason to introduce a frame, the index-free expression $a \cdot b$ is usually the simplest to use.

The same ideas extend to expressing the linear function F in a general non-orthonormal frame. We let F act on the frame vector e_j and find the components of the result in the reciprocal frame. The components are then given by

$$\mathsf{F}_{ij} = \mathsf{e}_i \cdot \mathsf{F}(\mathsf{e}_j). \tag{4.215}$$

Again, the set of numbers F_{ij} are referred to as the components of a rank-2 tensor and form an $n \times n$ matrix, the entries of which depend on the choice of frame. Similar expressions exist for combinations of frame vectors and reciprocal vectors, for example,

$$\mathsf{F}^{ij} = \mathsf{F}(\mathsf{e}^j) \cdot \mathsf{e}^i. \tag{4.216}$$

One use of the metric tensor is to interchange between these expressions:

$$\mathsf{F}^{ij} = \mathsf{e}^i \cdot \mathsf{F}(\mathsf{e}^j) = \mathsf{e}^i \cdot \mathsf{e}^k \, \mathsf{e}_k \cdot \mathsf{F}(\mathsf{e}_l \mathsf{e}^l \cdot \mathsf{e}^j) = g^{ik} g^{jl} \mathsf{F}_{kl}. \tag{4.217}$$

Again, we have at our disposal a variety of different ways of encoding the information in F. In terms of the abstract concept of a linear operator, the metric tensor g_{ij} is simply the identity operator expressed in a non-orthonormal frame.

If F_{ij} are the components of F in some frame then the components of $\bar{\mathsf{F}}$ are given by

$$\bar{\mathsf{F}}_{ij} = \bar{\mathsf{F}}(\mathsf{e}_j) \cdot \mathsf{e}_i = \mathsf{e}_j \cdot \mathsf{F}(\mathsf{e}_i) = \mathsf{F}_{ji}. \tag{4.218}$$

That is, viewed as a matrix, the components of $\bar{\mathsf{F}}$ are found from the components of F by matrix transposition. For mixed index tensors we have to be slightly more careful, as we now have

$$\mathsf{F}_i{}^j = \mathsf{F}(\mathsf{e}^j) \cdot \mathsf{e}_i = \mathsf{e}^j \cdot \bar{\mathsf{F}}(\mathsf{e}_i) = \bar{\mathsf{F}}^j{}_i. \tag{4.219}$$

If F is a symmetric function we have $\bar{\mathsf{F}} = \mathsf{F}$. In this case the component matrices satisfy

$$\mathsf{F}_{ij} = \mathsf{F}(\mathsf{e}_j) \cdot \mathsf{e}_i = \mathsf{F}(\mathsf{e}_i) \cdot \mathsf{e}_j = \mathsf{F}_{ji}, \tag{4.220}$$

so the components F_{ij} form a symmetric matrix. The same is true of $\mathsf{F}^{ij} = \mathsf{F}^{ji}$, but for the mixed tensor $\mathsf{F}_i{}^j$ we have $\mathsf{F}_i{}^j = \mathsf{F}^j{}_i$.

The components of the product function FG are found from the following rearrangement:

$$\begin{aligned}(\mathsf{FG})_{ij} &= \mathsf{FG}(\mathsf{e}_j)\cdot\mathsf{e}_i = \mathsf{G}(\mathsf{e}_j)\cdot\bar{\mathsf{F}}(\mathsf{e}_i) \\ &= \mathsf{G}(\mathsf{e}_j)\cdot\mathsf{e}_k\, \mathsf{e}^k\cdot\bar{\mathsf{F}}(\mathsf{e}_i) = \mathsf{F}_i{}^k\mathsf{G}_{kj}.\end{aligned} \tag{4.221}$$

Provided the correct combination of subscript and superscript indices is used, this can be viewed as a matrix product. Alternatively, one can work entirely with subscripted indices, and include suitable factors of the metric tensor,

$$(\mathsf{FG})_{ij} = \mathsf{F}_{ik}\mathsf{G}_{lj}g^{kl}. \tag{4.222}$$

Higher rank linear functions give rise to higher rank tensors. Suppose, for example, that $\phi(a_1, a_2, a_3)$ is a scalar function of three vectors, and is linear on each argument,

$$\phi(\lambda a_1 + \mu b, a_2, a_3) = \lambda\phi(a_1, a_2, a_3) + \mu\phi(b, a_2, a_3), \quad \text{etc.} \tag{4.223}$$

The components of this define a rank-3 tensor via

$$\phi_{ijk} = \phi(\mathsf{e}_i, \mathsf{e}_j, \mathsf{e}_k). \tag{4.224}$$

Using similar schemes it is a straightforward matter to set up a map between tensor equations and frame-free expressions in geometric algebra.

4.5.4 Coordinate transformations

If a second non-orthonormal frame $\{\mathsf{f}_\alpha\}$ is introduced we can relate the two frames via a transformation matrix $f_{\alpha i}$:

$$f_{\alpha i} = \mathsf{f}_\alpha\cdot\mathsf{e}_i, \quad f^{\alpha i} = \mathsf{f}^\alpha\cdot\mathsf{e}^i, \tag{4.225}$$

where Latin and Greek indices distinguish the components in one frame from the other. These matrices satisfy

$$f_{\alpha i}f^{\alpha j} = \mathsf{f}_\alpha\cdot\mathsf{e}_i\, \mathsf{f}^\alpha\cdot\mathsf{e}^j = \mathsf{e}_i\cdot\mathsf{e}^j = \delta_i^j \tag{4.226}$$

and

$$f_{\alpha i}f^{\beta i} = \mathsf{f}_\alpha\cdot\mathsf{e}_i\, \mathsf{f}^\beta\cdot\mathsf{e}^i = \mathsf{f}_\alpha\cdot\mathsf{f}^\beta = \delta_\alpha^\beta. \tag{4.227}$$

The decomposition of the vector a in terms of these frames gives

$$a = a^i\mathsf{e}_i = a^i\mathsf{f}^\alpha\mathsf{e}_i\cdot\mathsf{f}_\alpha = a^i f_{\alpha i}\mathsf{f}^\alpha. \tag{4.228}$$

If follows that the transformation law for the components is

$$a_\alpha = f_{\alpha i}a^i, \tag{4.229}$$

with similar expressions holding for the superscripted components.

These formulae extend simply to include linear functions. For example, we see that

$$\mathsf{F}_{\alpha\beta} = f_{\alpha i} f_{\beta j} \mathsf{F}^{ij}. \tag{4.230}$$

Again, similar expressions hold for superscripts and for mixtures of indices. In particular we have

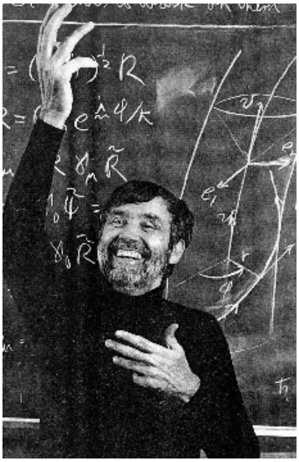
$$\mathsf{F}_\alpha{}^\beta = f_\alpha{}^i f^\beta{}_j \mathsf{F}_i{}^j. \tag{4.231}$$

Expressed in terms of matrix multiplication, this would be an equivalence transformation. Of course, the abstract frame-free function F is unaffected by any change of basis. All that changes is the particular representation of the function in the chosen coordinate system. Any set of n^2 numbers with this transformation property are called the components of a rank 2 tensor, the implication being that the underlying function is frame-independent.

In conventional accounts, the subject of tensors is often built up by taking the transformation law as fundamental. That is, a vector (rank-1 tensor) is *defined* as a set of components which transform according to equation (4.229) under a change of basis. Once one has the tools available to treat vectors and linear operations in a frame-free manner, such an approach becomes entirely unnecessary. The defining property of a tensor is that it represents a genuine geometric object (or operation) and does not depend on a choice of frame. Given this, the transformation laws (4.229) and (4.231) follow automatically. In this book the name *tensor* is applied to any frame-independent linear function, such as F. We will encounter a variety of such objects in later chapters.

4.6 Notes

The realisation that geometric algebra is a universal tool for physics was a key point in the modern development of the subject, and was first strongly promoted by David Hestenes (figure 4.5). Before his work, physicists' sole interaction with geometric algebra was through the quantum theory of spin. The Pauli and Dirac matrices form representations of Clifford algebras, a fact that was realised as soon as they were introduced. But in the 50 years since Clifford's original idea, the geometry behind his algebra had been lost as mathematicians concentrated on its algebraic properties. This discovery of the Pauli and Dirac matrices thus gave rise to two mistaken beliefs. The first was that there was something intrinsically quantum-mechanical in the non-commutative properties of the matrices. This is clearly not the case. Clifford died long before quantum theory was first formulated and was motivated entirely by classical geometry, and his algebra is today routinely employed in a range of subjects far removed from quantum theory.



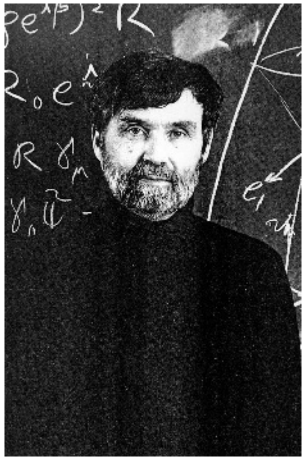


Figure 4.5 *David Hestenes.* Inventor of geometric calculus and first to draw attention to the universal nature of geometric algebra. He wrote the influential *Space-Time Algebra* in 1966, and followed this with a fully developed formalism in *Clifford Algebra to Geometric Calculus* (Hestenes & Sobczyk, 1984). This was followed by the (simpler) *New Foundations for Classical Mechanics*, first published in 1986 (second edition 1999). In a series of papers Hestenes and coworkers showed how geometric algebra could be applied in the study of classical and quantum mechanics, electrodynamics, projective and conformal geometry and Lie group theory. More recently, he has advocated the use of geometric algebra in the field of computer graphics.

The second widespread belief was that matrices were crucial to understanding the properties of Clifford algebras. This too is erroneous. The geometric algebra of a finite-dimensional vector space is an associative algebra, so always has a matrix representation. But these matrices add little, if anything, to understanding the properties of the algebra. Furthermore, an insistence on working with matrices deters one from applying geometric algebra to anything beyond the lowest dimensional spaces, because the size of the matrices increases exponentially with the dimension of the space. Working directly with the elements of the algebra imposes no such constraints, and one can easily apply the ideas to spaces of any dimension, including infinite-dimensional spaces.

Mathematicians had few such misconceptions, and Atiyah and others developed Clifford algebra as a powerful tool for geometry. Even in these developments, however, the emphasis was usually on Clifford algebra as an extra tool on top of the standard techniques for solving geometric problems. The algebra was seldom used as complete language for geometry. The picture first started to change when Hestenes recovered Clifford's original interpretation of the Pauli matrices. This led Hestenes to question whether the appearance of a Clifford algebra was telling us something about the underlying structure of quantum theory. Hestenes then went on to promote the universal nature of the algebra, which he publicised in a series of books and papers. Acceptance of this view is growing and, while not everyone is in full agreement, it is now hard to find an area of physics to which geometric algebra cannot or has not been applied without some degree of success.

4.7 Exercises

4.1 Prove that the outer product of a set of linearly dependent vectors vanishes.

4.2 In a Euclidean space, Gram–Schmidt orthogonalisation proceeds by successively replacing each vector in a set $\{a_i\}$ by one perpendicular to the preceding vectors. Prove that such a vector is given by

$$e_i = a_i - \sum_{j=1}^{i-1} \frac{a_i \cdot e_j}{e_j^2} e_j.$$

Prove that we can also write this as

$$e_i = a_i \wedge a_{i-1} \wedge \cdots \wedge a_1 (a_{i-1} \wedge \cdots \wedge a_1)^{-1}.$$

4.3 Prove that

$$(a \wedge b) \times (c \wedge d) = b \cdot c\, a \wedge d - a \cdot c\, b \wedge d + a \cdot d\, b \wedge c - b \cdot d\, a \wedge c.$$

4.4 The length of a vector in Euclidean space is defined by $|a| = \sqrt{(}a^2)$, and the angle θ between two vectors is defined by

$$\cos(\theta) = a \cdot b / (|a||b|).$$

Show that a linear transformation F which leaves lengths and angles unchanged must satisfy

$$\bar{\mathsf{F}} = \mathsf{F}^{-1}.$$

What does this imply for the determinant of F? A reflection in the (hyper)plane perpendicular to n is defined by

$$\mathsf{R}(a) = -nan,$$

where $n^2 = 1$. Show that $\bar{\mathsf{R}} = \mathsf{R}^{-1}$, and that R has determinant -1.

4.5 For the reflection in the preceding question introduce a suitable basis frame and express F in terms of a matrix F_{ij}. Verify the results for the determinant and inverse of this matrix. (Hint — align one of the basis vectors with n.)

4.6 A rotor R is defined by

$$R = \exp(-\lambda B/2).$$

By Taylor expanding in λ, prove that the operation

$$\mathsf{R}(A) = RAR^\dagger$$

preserves the grade(s) of the multivector A.

4.7 Show that the plane B is unchanged by the rotation defined by the rotor $R = \exp(B/2)$.

4.8 Analyse the properties of the matrix

$$\begin{pmatrix} 1 & 2\sinh(u) \\ 0 & 1 \end{pmatrix}.$$

To what geometric operation does this matrix correspond? Can this matrix be diagonalised, and does it have a sensible singular value decomposition?

4.9 Suppose that the linear transformation F has a complex eigenvector $e+if$ with associated eigenvector $\alpha + i\beta$. What is the effect of F on the $e\wedge f$ plane? How should one interpret the action of F in this plane?

4.10 Suppose that the vectors $\{\mathsf{e}_k\}$ form an orthonormal basis frame for n-dimensional Euclidean space. What is the effect of the transformation

$$\mathsf{T}(a) = a + \lambda a\cdot\mathsf{e}_1\,\mathsf{e}_2$$

on the rows of the matrix F_{ij} formed by decomposing F in the $\{\mathsf{e}_k\}$ frame? Use this result to prove that the determinant of a matrix is unchanged by adding a multiple of one row to another.

5

Relativity and spacetime

The geometric algebra of spacetime is called the *spacetime algebra*. Historically, the spacetime algebra was the first modern implementation of geometric algebra to gain widespread attention amongst the physics community. This is because it provides a *synthetic* framework for studying spacetime physics. There are two main approaches to the study of geometry, which can be loosely referred to as the algebraic and synthetic traditions. In the algebraic approach one works entirely with the components of a vector and manipulates these directly. Such an approach leads naturally to the subject of tensors, and places considerable emphasis on how coordinates transform under changes of frame. The synthetic approach, on the other hand, treats vectors as single, abstract entities x or a, and manipulates these directly. Geometric algebra follows in this tradition.

For much of modern physics the synthetic approach has come to dominate. The most obvious examples of this are classical mechanics and electromagnetism, both of which helped shape the development of abstract vector calculus. For these subjects, presentations typically perform all of the required calculations with the three-dimensional scalar and cross products. We have argued that geometric algebra provides extra efficiency and clarity, though it is not essential to a synthetic treatment of three-dimensional physics. But for spacetime calculations the cross product cannot be defined. Despite the obvious advantages of synthetic treatments, most relativity texts revert to a more basic, algebraic approach involving the components of 4-vectors and Lorentz-transform matrices. Such an approach has trouble encoding such basic notions as a plane in spacetime and, unsurprisingly, does a very poor job of handling the dynamics of extended bodies.

To develop a generally applicable algebra of vectors in spacetime one has little option but to use either geometric algebra, or the language of exterior forms (which is essentially a subset of geometric algebra which only employs the interior and exterior products). This is why relativistic physics still tends to dominate the literature of applications of geometric algebra. Many aspects of special relativity become clearer when viewed in the language of geometric algebra and, crucially, a wealth of new computational tools is provided which dramatically simplify relativistic problems.

5.1 An algebra for spacetime

It is not our intention in this chapter to give a fully self-contained introduction to relativity. Such an account can be found in the various books listed at the end of this chapter. In brief, a series of famous experiments conducted in the latter half of the nineteenth century showed that light did not appear to behave in quite the expected, Newtonian manner. This led Einstein to his 'second postulate', that the speed of light c is the same for all inertial (non-accelerating) observers. Combined with Einstein's 'first postulate', the principle of relativity, one is led inexorably to special relativity. The principle of relativity states simply that all inertial frames are equivalent for the purposes of physical experiment. An immediate consequence of these postulates is that the underlying geometry is no longer that of a (Euclidean) three-dimensional space, but instead the appropriate arena for physics is (Lorentzian) spacetime.

To understand why this is the case, suppose that a spherical flash of light is sent out from a source, and this event is described in two coordinate frames. We discuss the concept of a frame, as distinct from a single observer, later in this chapter. The frames are in relative motion, and their origins coincide with the location of the source at the moment the light is emitted. At this instant both frames also set their time measurements to zero. In the first frame the source is at rest and the light expands radially according to the equation

$$r = ct. \tag{5.1}$$

But the second frame must also record a radially expanding shell of light since the relative velocity of the source has no effect on the speed of light. The second frame therefore sees light expanding according to the equation

$$r' = ct'. \tag{5.2}$$

Since the two frames are in relative motion, points at a given fixed r cannot coincide with those at a fixed r'. So points reached at the same time in one frame are reached at *different* times in the second frame. But in both frames the light lies on a spherical expanding shell. So the one thing that is common to both frames is the value of

$$(ct)^2 - r^2 = (ct')^2 - (r')^2 = 0. \tag{5.3}$$

This defines the invariant interval of special relativity and is the fundamental algebraic concept we need to encode.

The preceding argument shows us that the algebra we need to construct is generated by four orthogonal vectors $\{\gamma_0, \gamma_1, \gamma_2, \gamma_3\}$ satisfying the algebraic relations

$$\gamma_0^2 = 1, \quad \gamma_0 \cdot \gamma_i = 0, \quad \gamma_i \cdot \gamma_j = -\delta_{ij}, \tag{5.4}$$

where i and j run from 1 to 3. These are summarised in relativistic notation as

$$\gamma_\mu \cdot \gamma_\nu = \eta_{\mu\nu} = \mathrm{diag}(+ - - -), \quad \mu, \nu = 0, \ldots, 3. \tag{5.5}$$

The notation $\{\gamma_\mu\}$ for a spacetime frame is a widely adopted convention in the spacetime algebra literature. The notation is borrowed from Dirac theory and we continue to employ it in this book. We have also chosen the 'particle physics' choice of signature, which has spacelike vectors with negative norm. General relativists often work with the opposite signature and swap all of the signs in $\eta_{\mu\nu}$. Both choices have their advocates and all (known) physical laws are independent of the choice of signature. Throughout we use Latin indices to denote the range 1–3 and Greek for the full spacetime range 0–3.

The $\{\gamma_\mu\}$ vectors are dimensionless, as is clear from their squares. Since we are in a space of mixed signature, we must adopt the conventions of section 4.3 and distinguish between a frame and its reciprocal. For the $\{\gamma_\mu\}$ frame the reciprocal frame vectors, $\{\gamma^\mu\}$, have $\gamma^0 = \gamma_0$ and $\gamma^i = -\gamma_i$. A general vector in the spacetime algebra can be constructed from the $\{\gamma_\mu\}$ vectors. A spacetime event, for example, is encoded in the vector x, which has coordinates x^μ in the $\{\gamma_\mu\}$ frame. Explicitly, the vector x is

$$x = x^\mu \gamma_\mu = ct\gamma_0 + x^i \gamma_i, \tag{5.6}$$

which has dimensions of distance. From this point on it will be convenient to work in units where the speed of light c is 1. Factors of c can then be inserted in any final result if the answer is required in different units. The mixed signature means that the square of a vector (a, say) is no longer necessarily positive, and instead we have

$$a^2 = aa = \epsilon |a^2|. \tag{5.7}$$

ϵ is the signature of the vector and can be ± 1 or 0. The mixed signature does not affect the validity of the axiomatic development and results of chapter 4, which made no reference to the signature.

5.1.1 The bivector algebra

There are $4 \times 3/2 = 6$ bivectors in our algebra. These fall into two classes: those that contain a timelike component (e.g. $\gamma_i \wedge \gamma_0$), and those that do not (e.g. $\gamma_i \wedge \gamma_j$). For any pair of orthogonal vectors a and b, $a \cdot b = 0$, we have

$$(a \wedge b)^2 = abab = -abba = -a^2 b^2. \tag{5.8}$$

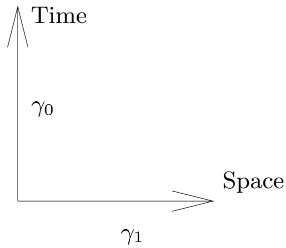


Figure 5.1 *A spacetime diagram.* Spacetime diagrams traditionally have the t axis vertical, so a suitable bivector for this plane is $\gamma_1\gamma_0$.

The two types of bivectors therefore have different signs of their squares. First, we have

$$(\gamma_i \wedge \gamma_j)^2 = -{\gamma_i}^2{\gamma_j}^2 = -1, \tag{5.9}$$

which is the familiar result for Euclidean bivectors. Each of these generates rotations in a plane. For bivectors containing a timelike component, however, we have

$$(\gamma_i \wedge \gamma_0)^2 = -{\gamma_i}^2{\gamma_0}^2 = +1. \tag{5.10}$$

Bivectors with positive square have a number of new properties. One immediate result we notice, for example, is that

$$\begin{aligned} e^{\alpha\gamma_1\gamma_0} &= 1 + \alpha\gamma_1\gamma_0 + \frac{\alpha^2}{2!} + \frac{\alpha^3}{3!}\gamma_1\gamma_0 + \cdots \\ &= \cosh(\alpha) + \sinh(\alpha)\gamma_1\gamma_0. \end{aligned} \tag{5.11}$$

This shows us that we are dealing with *hyperbolic geometry.* This will prove crucial to our treatment of Lorentz transformations. Traditionally, spacetime diagrams are drawn with the time axis vertical (see figure 5.1). For these diagrams the 'right-handed' bivector is, for example, $\gamma_1\gamma_0$. These bivectors do not generate 90° rotations, however, as we now have

$$\gamma_0 \cdot (\gamma_1\gamma_0) = -\gamma_1, \quad \gamma_1 \cdot (\gamma_1\gamma_0) = -\gamma_0. \tag{5.12}$$

5.1.2 The pseudoscalar

We define the (grade-4) pseudoscalar I by

$$I = \gamma_0\gamma_1\gamma_2\gamma_3. \tag{5.13}$$

In the literature the symbol i is often used for the pseudoscalar. We have departed from this practice to avoid confusion with the i of quantum theory. Using the latter symbol presents a potential problem because of the fact that the pseudoscalar anticommutes with vectors. The pseudoscalar defines an orientation for spacetime, and the reason for the above choice will emerge shortly. We still assume that $\{\gamma_1, \gamma_2, \gamma_3\}$ form a right-handed orthonormal set, as usual for a three-dimensional Cartesian frame. Since I is grade-4, it is equal to its own reverse:

$$\tilde{I} = \gamma_3\gamma_2\gamma_1\gamma_0 = I. \tag{5.14}$$

For relativistic applications we use the tilde ˜ to denote the reverse operation. The problem with the alternative symbol, the dagger †, is that it is usually reserved for a different role in relativistic quantum theory. The fact that $\tilde{I} = I$ makes it easy to compute the square of I :

$$I^2 = I\tilde{I} = (\gamma_0\gamma_1\gamma_2\gamma_3)(\gamma_3\gamma_2\gamma_1\gamma_0) = -1. \tag{5.15}$$

Multiplication of a bivector by I results in a multivector of grade $4 - 2 = 2$, so returns another bivector. This provides a map between bivectors with positive and negative squares, for example

$$I\gamma_1\gamma_0 = \gamma_1\gamma_0 I = \gamma_1\gamma_0\gamma_0\gamma_1\gamma_2\gamma_3 = -\gamma_2\gamma_3. \tag{5.16}$$

If we define $B_i = \gamma_i\gamma_0$ then the bivector algebra can be summarised by

$$\begin{aligned} B_i \times B_j &= \epsilon_{ijk}\, IB_k, \\ (IB_i)\times(IB_j) &= -\epsilon_{ijk}\, IB_k, \\ (IB_i)\times B_j &= -\epsilon_{ijk} B_k. \end{aligned} \tag{5.17}$$

These equations show that the pseudoscalar provides a natural complex structure for the set of bivectors. This in turn tells us that there is a complex structure hidden in the group of Lorentz transformations.

As well as the four vectors, we also have four trivectors in our algebra. The vectors and trivectors are interchanged by a duality transformation,

$$\gamma_1\gamma_2\gamma_3 = \gamma_0\gamma_0\gamma_1\gamma_2\gamma_3 = \gamma_0 I = -I\gamma_0. \tag{5.18}$$

The pseudoscalar I *anticommutes* with vectors and trivectors, as we are in a space of even dimensions. As always, I commutes with all even-grade multivectors.

5.1.3 The spacetime algebra

Combining the preceding results, we arrive at an algebra with 16 terms. The $\{\gamma_\mu\}$ define an explicit basis for this algebra as follows:

1	$\{\gamma_\mu\}$	$\{\gamma_\mu\wedge\gamma_\nu\}$	$\{I\gamma_\mu\}$	I
1 scalar	4 vectors	6 bivectors	4 trivectors	1 pseudoscalar

This is the *spacetime algebra*, $\mathcal{G}(1,3)$. The structure of this algebra tells us practically all one needs to know about (flat) spacetime and the Lorentz transformation group. A general element of the spacetime algebra can be written as

$$M = \alpha + a + B + Ib + I\beta, \tag{5.19}$$

where α and β are scalars, a and b are vectors and B is a bivector. The reverse of this element is

$$\tilde{M} = \alpha + a - B - Ib + I\beta. \tag{5.20}$$

The vector generators of the spacetime algebra satisfy

$$\gamma_\mu \gamma_\nu + \gamma_\nu \gamma_\mu = 2\eta_{\mu\nu}. \tag{5.21}$$

These are the defining relations of the Dirac matrix algebra, except for the absence of an identity matrix on the right-hand side. It follows that the Dirac matrices define a representation of the spacetime algebra. This also explains our notation of writing $\{\gamma_\mu\}$ for an orthonormal frame. But it must be remembered that the $\{\gamma_\mu\}$ are basis *vectors*, not a set of matrices in 'isospace'.

5.2 Observers, trajectories and frames

From a study of the literature on relativity one can easily form the impression that the subject is in the main concerned with transformations between frames. But it is the subject of relativistic dynamics that is of primary importance to us, and one aim of the spacetime algebra development is to minimise the use of coordinate frames. Instead, we aim to develop spacetime physics in a frame-free manner and, where necessary, then focus on the physics as seen from different observers. Developing relativistic physics in this manner has the added advantage of clarifying precisely which aspects of special relativity need modification to incorporate gravity.

5.2.1 Spacetime paths

Suppose that $x(\lambda)$ describes a curve in spacetime, where λ is some arbitrary, monotonically-increasing parameter along the curve. The tangent vector to the curve is

$$x' = \frac{dx(\lambda)}{d\lambda}. \tag{5.22}$$

Under a change of parameter from λ to τ the tangent vector becomes

$$\frac{dx}{d\tau} = \frac{d\lambda}{d\tau}\frac{dx}{d\lambda}. \tag{5.23}$$

It follows that

$$\left(\frac{dx}{d\tau}\right)^2 = \left(\frac{d\lambda}{d\tau}\right)^2 \left(\frac{dx}{d\lambda}\right)^2, \tag{5.24}$$

so the sign of $(x')^2$ is an invariant feature of the path. We assume for simplicity that this sign does not change along the path. As we are working in a space of mixed signature there are then three cases to consider.

The first possibility is that $(x')^2 > 0$, in which case the path is said to be *timelike*. Timelike trajectories are those followed by massive particles. For these paths we can define an invariant proper interval

$$\Delta\tau = \int_{\lambda_1}^{\lambda_2} \left(\frac{dx}{d\lambda} \cdot \frac{dx}{d\lambda}\right)^{1/2} d\lambda. \tag{5.25}$$

It is straightforward to check that this interval is independent of how the path is parameterised. If we consider the simplest case of a particle (or observer) at rest in the γ_0 system, its spacetime trajectory can be written as $x = t\gamma_0$. In this case it is clear that the interval defines the elapsed time in the observer's rest frame. This must be true for all possible paths, so the interval (5.25) defines the time as measured along the path. This is called the *proper time*, and is usually given the symbol τ. The proper time defines a preferred parameter along the curve with the unique property that the velocity v,

$$v = \frac{dx}{d\tau} = \dot{x}, \tag{5.26}$$

satisfies

$$v^2 = 1. \tag{5.27}$$

Throughout we use dots to denote differentiation with respect to proper time τ. The unit timelike vector v then defines the instantaneous rest frame. The definition of 'proper time' makes it clear that in relativity observers moving in relative motion measure different times.

The second case to consider is that $(x')^2 = 0$. In this case the trajectory is said to be *lightlike* or *null*. Null trajectories are followed by massless (point) particles and (in the geometric optics limit) they define possible photon paths. There is no preferred parameter along these curves, and the proper distance (or time) measured along the curve is 0. Photons do still carry an intrinsic clock, defined by their frequency, but this can tick at an arbitrary rate.

The third possibility is that $(x')^2 < 0$, in which case the trajectory is said to be *spacelike*. As with timelike paths there is a preferred (affine) parameter along the path such that $(x')^2 = -1$. In this case the parameter defines the *proper distance*. Spacelike curves cannot arise for the trajectories of (known) particles, which are constrained to move at less than (or equal to) the speed of light. Events which are separated by spacelike intervals cannot be in causal

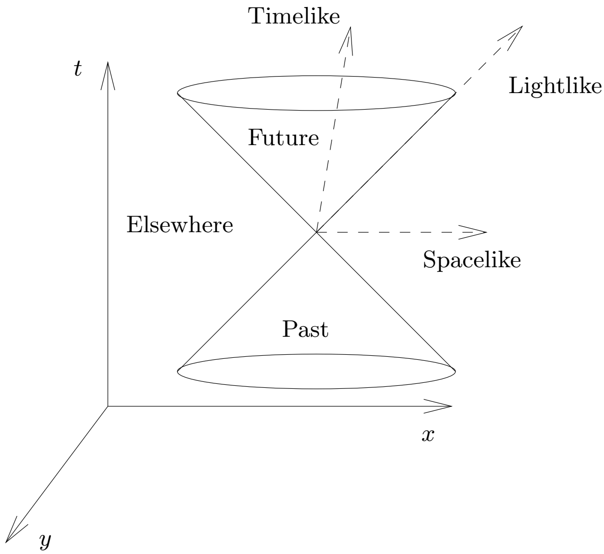


Figure 5.2 *Spacetime trajectories.* There are three different types of spacetime trajectory: timelike, lightlike and spacelike. The set of lightlike trajectories through a point separate spacetime into three regions: the past, the future and 'elsewhere'.

contact with each other and cannot exert any classical influence over each other. The three possibilities for spacetime trajectories are summarised in figure 5.2.

5.2.2 Spacetime frames

The subject of spacetime frames and coordinates dominates many discussions of the meaning of special relativity. The concept of a frame is distinct from that of an observer as it involves the notion of a coordinate lattice. We start with an inertial observer with constant velocity v. This velocity vector is then equated with the timelike vector e_0 from a spacetime frame $\{\mathsf{e}_\mu\}$. The remaining vectors e_i are chosen so that they form a right-handed set of orthonormal spacelike vectors perpendicular to $\mathsf{e}_0 = v$. The $\{\mathsf{e}_\mu\}$ then define a set of frame vectors satisfying

$$\mathsf{e}_\mu \cdot \mathsf{e}_\nu = \eta_{\mu\nu}. \tag{5.28}$$

So far these vectors are only defined at a single point on the observer's trajectory. We now assume that the vectors extend throughout all spacetime, so that any event can be given a set of spacetime coordinates

$$x^\mu = \mathsf{e}^\mu \cdot x. \tag{5.29}$$

Clearly these coordinates are a rather distinct concept from what an observer will actually measure, since the observer is constrained to remain in one place and only receives incoming photons. Frequently one sees discussions involving arrays of clocks all cleverly synchronised to read the time x^0 at each spatial location. But how such a frame is set up is not really the point. The assertion is that the coordinates as specified above are a reasonable model for the sort of distance and time measurements performed in a laboratory system using physical measuring devices. It is precisely this assertion that is challenged by general relativity, which insists that one talk entirely in terms of physically-defined coordinates, so that the x^μ defined above have no physical meaning. That said, for applications not involving gravity and for non-accelerating frames, we can safely identify the coordinates defined above with physical distances and times and will continue to do so in this chapter.

5.2.3 Relative vectors

Now suppose that we follow a timelike path with instantaneous velocity v, $v^2 = 1$. What sort of quantities do we measure? First we construct a frame of rest vectors $\{\mathsf{e}_i\}$ perpendicular to $v = \mathsf{e}_0$. We also take a point on the worldline as the spatial origin. Then a general event x can be decomposed in this frame as

$$x = t\mathsf{e}_0 + x^i\mathsf{e}_i, \tag{5.30}$$

where the time coordinate is

$$t = x\cdot\mathsf{e}_0 = x\cdot v \tag{5.31}$$

and spatial coordinates are

$$x^i = x\cdot\mathsf{e}^i. \tag{5.32}$$

Suppose now that the event is a point on the worldline of an object at rest in our frame. The three-dimensional vector to this object is

$$x^i\mathsf{e}_i = x\cdot\mathsf{e}^\mu\,\mathsf{e}_\mu - x\cdot\mathsf{e}^0\,\mathsf{e}_0 = x - x\cdot v\,v = x\wedge v\,v. \tag{5.33}$$

Wedging with v projects onto the components of the vector x in the rest frame of v. The key quantity is the spacetime bivector $x\wedge v$. We call this the *relative* vector and write

$$\boldsymbol{x} = x\wedge v. \tag{5.34}$$

With these definitions we have

$$xv = x\cdot v + x\wedge v = t + \boldsymbol{x}. \tag{5.35}$$

The invariant distance now decomposes as

$$\begin{aligned} x^2 = xvvx &= (x\cdot v + x\wedge v)(x\cdot v + v\wedge x) \\ &= (t+\boldsymbol{x})(t-\boldsymbol{x}) = t^2 - \boldsymbol{x}^2, \end{aligned} \tag{5.36}$$

recovering the invariant interval. A second observer with a different velocity performs a different split of x into time and space components. But the interval x^2 is the same for all observers as it manifestly does not depend on the choice of frame.

5.2.4 The even subalgebra

Each observer sees a set of relative vectors, which we model as spacetime bivectors. What algebraic properties do these have? To simplify matters, we take the timelike velocity vector to be γ_0 and introduce a standard frame of relative vectors

$$\boldsymbol{\sigma}_i = \gamma_i \gamma_0. \tag{5.37}$$

These define a set of spacetime bivectors representing timelike planes. (The notation is again borrowed from quantum mechanics and is commonplace in the spacetime algebra literature.) The $\{\boldsymbol{\sigma}_i\}$ satisfy

$$\begin{aligned} \boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_j &= \tfrac{1}{2}(\gamma_i\gamma_0\gamma_j\gamma_0 + \gamma_j\gamma_0\gamma_i\gamma_0) \\ &= \tfrac{1}{2}(-\gamma_i\gamma_j - \gamma_j\gamma_i) = \delta_{ij}. \end{aligned} \tag{5.38}$$

These act as vector generators for a three-dimensional algebra. This is the geometric algebra of the relative space in the rest frame defined by γ_0. Furthermore, the volume element of this algebra is

$$\boldsymbol{\sigma}_1\boldsymbol{\sigma}_2\boldsymbol{\sigma}_3 = (\gamma_1\gamma_0)(\gamma_2\gamma_0)(\gamma_3\gamma_0) = -\gamma_1\gamma_0\gamma_2\gamma_3 = I, \tag{5.39}$$

so the algebra of relative space shares the same pseudoscalar as spacetime. This was the reason for our earlier definition of I. Of course, we still have

$$\tfrac{1}{2}(\boldsymbol{\sigma}_i\boldsymbol{\sigma}_j - \boldsymbol{\sigma}_j\boldsymbol{\sigma}_i) = \epsilon_{ijk} I \boldsymbol{\sigma}_k, \tag{5.40}$$

so that both relative vectors and relative bivectors are spacetime bivectors.

The even-grade terms in the spacetime algebra define the *even subalgebra*. As we have just established, this algebra has precisely the properties of the algebra of three-dimensional (relative) space. The even subalgebra contains scalar and pseudoscalar terms, and six bivector terms. These are split into three timelike vectors and three spacelike vectors, which in turn become relative vectors and bivectors. This is called a *spacetime split*, and it is *observer-dependent*. Different velocity vectors generate different spacetime splits. Algebraically, this provides us with an extremely efficient tool for comparing physical effects in different frames.

Spacetime bivectors which are also used as relative vectors are written in bold. This conforms with our earlier usage of a bold face for vectors in three dimensions. There is a potential ambiguity here — how are we to interpret the expression $\boldsymbol{a}\wedge\boldsymbol{b}$? Our convention is that if all of the terms in an expression are bold, the dot and wedge symbols drop down to their three-dimensional meaning, otherwise they take their spacetime definition. This works pretty well in practice, though where necessary we will try to draw attention to the fact that this convention is in use.

5.2.5 Relative velocity

Suppose that an observer with constant velocity v measures the relative velocity of a particle with proper velocity $u(\tau) = \dot{x}(\tau)$, $u^2 = 1$. We have

$$uv = \frac{d}{d\tau}(x(\tau)v) = \frac{d}{d\tau}(t + \boldsymbol{x}), \tag{5.41}$$

where $t + \boldsymbol{x}$ is the description of the event x in the v frame. It follows that

$$\frac{dt}{d\tau} = u\cdot v, \quad \frac{d\boldsymbol{x}}{d\tau} = u\wedge v. \tag{5.42}$$

The relative velocity $\boldsymbol{u}$ as measured in the v frame is therefore

$$\boldsymbol{u} = \frac{d\boldsymbol{x}}{dt} = \frac{d\boldsymbol{x}}{d\tau}\frac{d\tau}{dt} = \frac{u\wedge v}{u\cdot v}. \tag{5.43}$$

This construction of the relative velocity is extremely elegant. It embodies the concept of relativity in its precise (anti)symmetry. If we interchange u and v the second observer measures precisely the same relative speed as the first, but in the opposite direction. Expressions like $u\wedge v/u\cdot v$ arise frequently in the subject of *projective geometry* (see section 10.1). The resulting bivector is homogeneous, which is to say we can rescale u and v and still recover the same result. So the choice of parameterisation of the two spacetime trajectories is irrelevant to their relative velocity. The relative velocity is determined solely by the spacetime trajectories themselves, and not by any evolution parameter.

The definition of the relative velocity ensures that the magnitude is

$$\frac{(u\wedge v)^2}{(u\cdot v)^2} = 1 - \frac{1}{(u\cdot v)^2} < 1, \tag{5.44}$$

so no two observers measure a relative velocity greater than the speed of light (which is 1 in our current choice of units). If we form the Lorentz factor γ using

$$\begin{aligned}\gamma^{-2} &= 1 - \boldsymbol{u}^2 \\ &= 1 + (u\cdot v)^{-2}[(uv - u\cdot v)(vu - v\cdot u)] = (u\cdot v)^{-2},\end{aligned} \tag{5.45}$$

we find that $\gamma = u \cdot v$. It follows that we can decompose the velocity as

$$u = uvv = (u \cdot v + u \wedge v)v = \gamma(1 + \boldsymbol{u})v, \tag{5.46}$$

which shows a neat split into a part $\gamma \boldsymbol{u} v$ in the rest space of v, and a part γv along v.

5.2.6 Momentum and wave vectors

The relativistic definitions of energy and momentum can be motivated in various ways. Perhaps the simplest is to consider photons with frequency ω and wavevector $\boldsymbol{k}$ measured in the γ_0 frame. From quantum theory, the energy and momentum are given by $\hbar\omega$ and $\hbar\boldsymbol{k}$ respectively. If we define the wavevector k by

$$k = \omega\gamma_0 + k^i\gamma_i, \tag{5.47}$$

then the energy-momentum vector for the photon is simply

$$p = \hbar k. \tag{5.48}$$

An observer with velocity v, as opposed to γ_0, measures energy and momentum given by

$$E = p \cdot v, \quad \boldsymbol{p} = p \wedge v. \tag{5.49}$$

We take this as the correct definition for massive particles as well. So a particle of rest mass m and velocity u has an energy-momentum vector $p = mu$. A spacetime split of this vector with the velocity vector v yields

$$pv = p \cdot v + p \wedge v = E + \boldsymbol{p}. \tag{5.50}$$

A significant feature of this definition is that the relative momentum is related to the velocity by

$$\boldsymbol{p} = mu \cdot v \, \boldsymbol{u} = \gamma m \boldsymbol{u}, \tag{5.51}$$

where again γ is the Lorentz factor. One sometimes sees this formula written in terms of a velocity-dependent mass $m' = \gamma m$, but we will not adopt this practice here.

From the definition of p we recover the invariant

$$m^2 = p^2 = pvvp = (E + \boldsymbol{p})(E - \boldsymbol{p}) = E^2 - \boldsymbol{p}^2. \tag{5.52}$$

Similarly, for a photon with wavevector k, $k^2 = 0$, we have

$$0 = kvvk = (\omega + \boldsymbol{k})(\omega - \boldsymbol{k}) = \omega^2 - \boldsymbol{k}^2. \tag{5.53}$$

This recovers the relation $|\boldsymbol{k}| = \omega$, which holds in all frames.

5.2.7 Proper acceleration

A final ingredient in the formulation of relativistic dynamics is the proper acceleration. A particle follows a trajectory $x(\tau)$, where τ is the proper time. The particle has velocity $v = \dot{x}$, $v^2 = 1$. The proper acceleration is simply

$$\dot{v} = \frac{dv}{d\tau}. \tag{5.54}$$

Since $v^2 = 1$, the velocity and acceleration are perpendicular

$$\frac{d}{d\tau}(v^2) = 0 = 2\dot{v}\cdot v. \tag{5.55}$$

In many physical phenomena it turns out that a more useful concept is provided by the *acceleration bivector*

$$B_v = \dot{v}\wedge v = \dot{v}v. \tag{5.56}$$

This bivector denotes the acceleration projected into the instantaneous rest frame of the particle. Typically this bivector multiplied by the rest mass is equated with a bivector encoding the forces acting on the particle. Any change in the parameter along the curve will rescale the velocity vector, so B_v can be written as

$$B_v = \frac{v'\wedge v}{(v\cdot v)^{3/2}}, \tag{5.57}$$

which is independent of the parameterisation of the trajectory.

Before applying the various preceding definitions to a range of dynamical problems, we turn to a discussion of the Lorentz transformations. This will pave the way for a powerful method for studying relativistic problems which is unique to geometric algebra.

5.3 Lorentz transformations

Lorentz transformations are usually expressed in the form of a coordinate transformation. We suppose that two inertial observers have set up 'coordinate lattices' in their own rest frames, as discussed in section 5.2.2. We denote these frames by S and S', and assume that they are set up such that their 1 and 2 axes coincide, but that S' moves at (scalar) velocity βc along the 3 axis as seen in the S frame. We denote the 0 and 3 components by t and z respectively. If the origins of the frames coincide at $t = t' = 0$, the coordinates of the same spacetime event as measured in the two frames are related by

$$t' = \gamma(t - \beta z), \quad x^{1'} = x^1, \quad x^{2'} = x^2, \quad z' = \gamma(z - \beta t), \tag{5.58}$$

where $\gamma = (1-\beta^2)^{-1/2}$ and β is the velocity in units of c ($\beta < 1$). The inverse relations are easily found to be

$$t = \gamma(t' + \beta z'), \quad x^1 = x^{1'}, \quad x^2 = x^{2'}, \quad z = \gamma(z' + \beta t'). \tag{5.59}$$

The arguments leading to these transformation laws are discussed in all introductory texts on relativity (see e.g. Rindler (1977) or French (1968)).

To get a clearer understanding of this transformation law we must first convert these relations into a transformation law for the frame vectors. The vector x has been decomposed in two frames, $\{\mathsf{e}_\mu\}$ and $\{\mathsf{e}'_\mu\}$, so that

$$x = x^\mu \mathsf{e}_\mu = x^{\mu\prime} \mathsf{e}'_\mu. \tag{5.60}$$

We then have, for example,

$$t = \mathsf{e}^0 \cdot x, \quad t' = \mathsf{e}^{0\prime} \cdot x. \tag{5.61}$$

Concentrating on the 0 and 3 components we have

$$t\mathsf{e}_0 + z\mathsf{e}_3 = t'\mathsf{e}'_0 + z'\mathsf{e}'_3, \tag{5.62}$$

and from this we derive the vector relations

$$\mathsf{e}'_0 = \gamma(\mathsf{e}_0 + \beta\mathsf{e}_3), \quad \mathsf{e}'_3 = \gamma(\mathsf{e}_3 + \beta\mathsf{e}_0). \tag{5.63}$$

These define the new frame in terms of the old. As a check the new frame vectors have the correct normalisation,

$$(\mathsf{e}'_0)^2 = \gamma^2(1-\beta^2) = 1, \quad (\mathsf{e}'_3)^2 = -1. \tag{5.64}$$

The geometry of this transformation is illustrated in figure 5.3.

We saw earlier that bivectors with positive square lead to hyperbolic geometry. This suggests that we introduce an 'angle' α with

$$\tanh(\alpha) = \beta \tag{5.65}$$

so that

$$\gamma = \bigl(1 - \tanh^2(\alpha)\bigr)^{-1/2} = \cosh(\alpha). \tag{5.66}$$

The vector e'_0 is now

$$\begin{aligned} \mathsf{e}'_0 &= \cosh(\alpha)\,\mathsf{e}_0 + \sinh(\alpha)\,\mathsf{e}_3 \\ &= \bigl(\cosh(\alpha) + \sinh(\alpha)\,\mathsf{e}_3\mathsf{e}_0\bigr)\mathsf{e}_0 \\ &= \exp(\alpha\,\mathsf{e}_3\mathsf{e}_0)\,\mathsf{e}_0, \end{aligned} \tag{5.67}$$

where we have expressed the scalar + bivector term as an exponential. Similarly, we have

$$\mathsf{e}'_3 = \cosh(\alpha)\,\mathsf{e}_3 + \sinh(\alpha)\,\mathsf{e}_0 = \exp(\alpha\,\mathsf{e}_3\mathsf{e}_0)\,\mathsf{e}_3. \tag{5.68}$$

Now recall that these are just two of four frame vectors, and the other pair

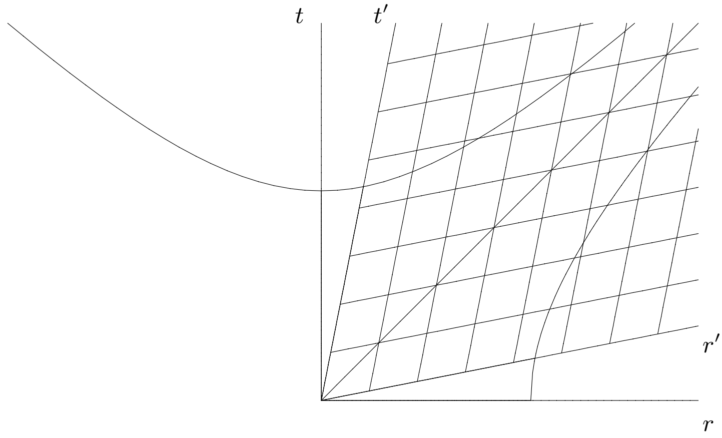


Figure 5.3 *A Lorentz transformation.* The transformation leaves the magnitude of a vector invariant. As the underlying geometry of a spacetime plane is Lorentzian, vectors of constant magnitude lie on hyperbolae, rather than circles. The transformed axes define a new coordinate grid.

are unchanged by the transformation. Since $\mathsf{e}_3\mathsf{e}_0$ anticommutes with e_0 and e_3, but commutes with e_1 and e_2, we can express the relationship between the two frames as

$$\mathsf{e}_\mu' = R\mathsf{e}_\mu\tilde{R}, \quad \mathsf{e}^{\mu\prime} = R\mathsf{e}^\mu\tilde{R}, \tag{5.69}$$

where

$$R = \mathrm{e}^{\alpha\,\mathsf{e}_3\mathsf{e}_0/2}. \tag{5.70}$$

The same rotor prescription introduced for rotations in Euclidean space also works for boosts in relativity! This is dramatically simpler than having to work with 4×4 Lorentz transform matrices.

5.3.1 Addition of velocities

As a simple example, suppose that we are in a frame with basis vectors $\{\gamma_\mu\}$. We observe two objects flying apart with 4-velocities

$$v_1 = \mathrm{e}^{\alpha_1\gamma_1\gamma_0/2}\gamma_0\mathrm{e}^{-\alpha_1\gamma_1\gamma_0/2} = \mathrm{e}^{\alpha_1\gamma_1\gamma_0}\gamma_0 \tag{5.71}$$

and

$$v_2 = \mathrm{e}^{-\alpha_2\gamma_1\gamma_0/2}\gamma_0\mathrm{e}^{\alpha_2\gamma_1\gamma_0/2} = \mathrm{e}^{-\alpha_2\gamma_1\gamma_0}\gamma_0. \tag{5.72}$$

What is the relative velocity they see for each other? We form

$$\frac{v_1 \wedge v_2}{v_1 \cdot v_2} = \frac{\left\langle \mathrm{e}^{(\alpha_1 + \alpha_2)\gamma_1\gamma_0} \right\rangle_2}{\left\langle \mathrm{e}^{(\alpha_1 + \alpha_2)\gamma_1\gamma_0} \right\rangle_0} = \frac{\sinh(\alpha_1 + \alpha_2)\gamma_1\gamma_0}{\cosh(\alpha_1 + \alpha_2)}. \tag{5.73}$$

Both observers therefore measure a relative velocity of

$$\tanh(\alpha_1 + \alpha_2) = \frac{\tanh(\alpha_1) + \tanh(\alpha_2)}{1 + \tanh(\alpha_1)\tanh(\alpha_2)}, \tag{5.74}$$

Addition of (collinear) velocities is achieved by adding hyperbolic angles, and not the velocities themselves. Replacing the tanh factors by the scalar velocities $u = c\tanh(\alpha)$ recovers the more familiar expression

$$u' = \frac{u_1 + u_2}{1 + u_1 u_2/c^2}. \tag{5.75}$$

The surprising conclusion is that addition of velocities in spacetime is really a generalized rotation in a hyperbolic space! Quite dramatically different from the Newtonian prescription of simple vector addition of the velocities.

5.3.2 Photons, Doppler shifts and aberration

For many relativistic applications involving the properties of light it is sufficient to use a simplified model of a photon as a point particle following a null trajectory. The tangent vector to the path is the wavevector k. This provides for simple formulae for Doppler shifts and aberration. Suppose that two particles follow different worldlines and that particle 1 emits a photon which is received by particle 2 (see figure 5.4). The frequency seen by particle 1 is $\omega_1 = v_1 \cdot k$, and that by particle 2 is $\omega_2 = v_2 \cdot k$. The ratio of these describes the Doppler effect, often expressed as a redshift, z:

$$1 + z = \frac{\omega_1}{\omega_2} = \frac{v_1 \cdot k}{v_2 \cdot k}. \tag{5.76}$$

This can be applied in many ways. For example, suppose that the emitter is receding in the γ_1 direction, and $v_2 = \gamma_0$. We have

$$k = \omega_2(\gamma_0 + \gamma_1), \qquad v_1 = \cosh(\alpha)\,\gamma_0 - \sinh(\alpha)\,\gamma_1, \tag{5.77}$$

so that

$$1 + z = \frac{\omega_2\big(\cosh(\alpha) + \sinh(\alpha)\big)}{\omega_2} = \mathrm{e}^{\alpha}. \tag{5.78}$$

The velocity of the emitter in the γ_0 frame is $\tanh(\alpha)$, and it is easy to check that

$$\mathrm{e}^{\alpha} = \left(\frac{1 + \tanh(\alpha)}{1 - \tanh(\alpha)}\right)^{1/2}. \tag{5.79}$$

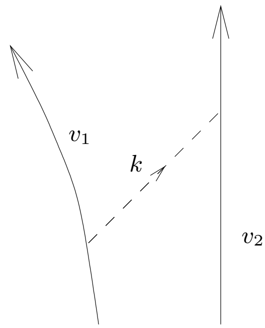


Figure 5.4 *Photon emission and absorption.* A photon is emitted by particle 1 and received by particle 2.

This formula recovers the standard expression for the relativistic Doppler effect:

$$\omega_2 = \left(\frac{1-\beta}{1+\beta}\right)^{1/2} \omega_1. \tag{5.80}$$

In its current form this formula is appropriate for a source and receiver moving away from each other at velocity βc. Had they been approaching each other the sign of β would be reversed, leading to an increased frequency at the receiver (a blueshift).

Aberration formulae can be obtained in a similar manner. Suppose that observer 1 has velocity γ_0, and that this observer receives photons at an angle θ to the 1 axis in the 12 plane. The photons are therefore on a null trajectory with tangent vector

$$n = \gamma_0 - \cos(\theta)\,\gamma_1 - \sin(\theta)\,\gamma_2, \tag{5.81}$$

and the γ_0 observer recovers the angle θ via

$$\tan(\theta) = \frac{n\cdot\gamma_2}{n\cdot\gamma_1}. \tag{5.82}$$

Suppose now that a second observer moves with velocity β relative to the first along the 1 axis. This observer's velocity is

$$v = \mathsf{e}_0 = \cosh(\alpha)\,\gamma_0 + \sinh(\alpha)\,\gamma_1 \tag{5.83}$$

and the frame vectors for this observer are

$$\mathsf{e}_1 = \cosh(\alpha)\,\gamma_1 + \sinh(\alpha)\,\gamma_0, \quad \mathsf{e}_2 = \gamma_2, \quad \mathsf{e}_3 = \gamma_3. \tag{5.84}$$

According to this observer the photons arrive at an angle

$$\tan(\theta') = \frac{n\cdot\mathsf{e}_2}{n\cdot\mathsf{e}_1} = \frac{\sin(\theta)}{\cosh(\alpha)\,\cos(\theta) + \sinh(\alpha)}. \tag{5.85}$$

A straightforward rearrangement gives

$$\cos(\theta') = \frac{\cosh(\alpha)\,\cos(\theta) + \sinh(\alpha)}{\cosh(\alpha) + \sinh(\alpha)\,\cos(\theta)} = \frac{\cos(\theta) + \beta}{1 + \beta\cos(\theta)}, \tag{5.86}$$

so observers in relative motion measure different angles to a fixed light source. This effect can be seen in observations of stars from the Earth. The Earth's orbital velocity around the sun has a β of roughly 10^{-4} so to a good approximation we have

$$\cos(\theta') \approx \cos(\theta) + \beta\sin^2(\theta). \tag{5.87}$$

The aberration angle $\phi = \theta - \theta'$ satisfies the approximate formula

$$\phi \approx \beta\sin(\theta), \tag{5.88}$$

which implies that the aberration varies over a year as θ varies through a complete cycle. This variation was first observed by James Bradley in 1727 and was explained in terms of a particle model of light. Bradley was able to use his data to give an improved estimate of the speed of light, though the full relativistic relation of (5.86) cannot be checked in this manner.

5.4 The Lorentz group

The full Lorentz group consists of the transformation group for vectors that preserves lengths and angles. These include reflections and rotations. A reflection in the hyperplane perpendicular to n is achieved by

$$a \mapsto -nan^{-1}. \tag{5.89}$$

The n^{-1} is necessary to accommodate both timelike $n^2 > 0$ and spacelike $n^2 < 0$ cases. We cannot have null n, as the inverse does not exist. A timelike n generates time-reversal transformations, whereas spacelike reflections preserve time-ordering. Pairs of either of these result in a transformation which preserves time-ordering. However, a combination of one spacelike and one timelike reflection does not preserve the time-ordering. The full Lorentz group therefore contains four sectors (table 5.1).

The structure of the Lorentz group is easily understood in the spacetime algebra. We concentrate on even numbers of reflections, which have determinant +1 and correspond to type I and type IV transformations. The remaining types are obtained from these by a single extra reflection. If we combine even numbers of reflections we arrive at a transformation of the form

$$a \mapsto \psi a \psi^{-1}, \tag{5.90}$$

where ψ is an even multivector. This expression is currently too general, as we

	Parity preserving	Space reflection
	I	II
Time order preserving	Proper orthochronous	I with space reflection
	III	IV
Time reversal	I with time reversal	I with $a \mapsto -a$

Table 5.1 *The full Lorentz group.* The group of Lorentz transformations falls into four disjoint sectors. Sectors I and IV have determinant $+1$, whereas II and III have determinant -1. Both I and II preserve time-ordering, and the proper orthochronous transformations (type I) are simply-connected to the identity.

have not ensured that the right-hand side is a vector. To see how to do this we decompose ψ into invariant terms. We first note that

$$\psi\tilde{\psi} = (\psi\tilde{\psi})^{\sim} \tag{5.91}$$

so $\psi\tilde{\psi}$ is even-grade and equal to its own reverse. It can therefore only contain a scalar and a pseudoscalar,

$$\psi\tilde{\psi} = \alpha_1 + I\alpha_2 = \rho e^{I\beta}, \tag{5.92}$$

where $\rho \neq 0$ in order for ψ^{-1} to exist. We can now define a rotor R by

$$R = \psi(\rho e^{I\beta})^{-1/2}, \tag{5.93}$$

so that

$$R\tilde{R} = \psi\tilde{\psi}(\rho e^{I\beta})^{-1} = 1, \tag{5.94}$$

as required. We now have

$$\psi = \rho^{1/2} e^{I\beta/2} R, \qquad \psi^{-1} = \rho^{-1/2} e^{-I\beta/2} \tilde{R} \tag{5.95}$$

and our general transformation becomes

$$a \mapsto e^{I\beta/2} R a e^{-I\beta/2} \tilde{R} = e^{I\beta} R a \tilde{R}. \tag{5.96}$$

The term $Ra\tilde{R}$ is necessarily a vector as it is equal to its own reverse, so we must restrict β to either 0 or π, leaving the transformation

$$a \mapsto \pm R a \tilde{R}. \tag{5.97}$$

The transformation $a \mapsto Ra\tilde{R}$ preserves causal ordering as well as parity. Transformations of this type are called ‘proper orthochronous’ transformations.

We can prove that transformations parameterised by rotors are proper orthochronous by starting with the velocity γ_0 and transforming it to $v = R\gamma_0\tilde{R}$. We require that the γ_0 component of v is positive, that is,

$$\gamma_0 \cdot v = \langle \gamma_0 R \gamma_0 \tilde{R} \rangle > 0. \tag{5.98}$$

Decomposing in the γ_0 frame we can write

$$R = \alpha + \boldsymbol{a} + I\boldsymbol{b} + I\beta \tag{5.99}$$

and we find that

$$\langle \gamma_0 R \gamma_0 \tilde{R} \rangle = \alpha^2 + \boldsymbol{a}^2 + \boldsymbol{b}^2 + \beta^2 > 0 \tag{5.100}$$

as required. Our rotor transformation law describes the group of proper orthochronous transformations, often called the *restricted Lorentz group*. These are the transformations of most physical relevance. The negative sign in equation (5.97) corresponds to $\beta = \pi$ and gives class-IV transformations.

5.4.1 Invariant decomposition and fixed points

Every rotor in spacetime can be written in terms of a bivector as

$$R = \pm e^{B/2}. \tag{5.101}$$

(The minus sign is rarely required, and does not affect the vector transformation law.) We can understand many of the features of spacetime transformations and rotors through the properties of the bivector B. The bivector B can be decomposed in a Lorentz-invariant manner by first writing

$$B^2 = \langle B^2 \rangle_0 + \langle B^2 \rangle_4 = \rho e^{I\phi}, \tag{5.102}$$

and we will assume that $\rho \neq 0$. (The case of a null bivector is treated slightly differently.) We now define

$$\hat{B} = \rho^{-1/2} e^{-I\phi/2} B, \tag{5.103}$$

so that

$$\hat{B}^2 = \rho^{-1} e^{-I\phi} B^2 = 1. \tag{5.104}$$

With this we can now write

$$B = \rho^{1/2} e^{I\phi/2} \hat{B} = \alpha \hat{B} + \beta I \hat{B}, \tag{5.105}$$

which decomposes B into a pair of bivector blades, $\alpha\hat{B}$ and $\beta I\hat{B}$. Since

$$\hat{B}(I\hat{B}) = (I\hat{B})\hat{B} = I, \tag{5.106}$$

the separate bivector blades commute. The rotor R now decomposes into

$$R = e^{\alpha\hat{B}/2} e^{\beta I\hat{B}/2} = e^{\beta I\hat{B}/2} e^{\alpha\hat{B}/2}, \tag{5.107}$$

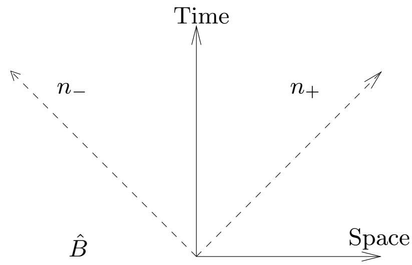


Figure 5.5 *A timelike plane.* Any timelike plane $\hat{B}$, $\hat{B}^2 = 1$, contains two null vectors n_+ and n_-. These can be normalised so that $n_+ \wedge n_- = 2\hat{B}$.

exhibiting an *invariant* split into a boost and a rotation. The boost is generated by $\hat{B}$ and the rotation by $I\hat{B}$.

For every timelike bivector $\hat{B}$, $\hat{B}^2 = 1$, we can construct a pair of null vectors $n_\pm$ satisfying

$$\hat{B} \cdot n_\pm = \pm n_\pm. \tag{5.108}$$

These are necessarily null, since

$$n_+ \cdot n_+ = (B \cdot n_+) \cdot n_+ = B \cdot (n_+ \wedge n_+) = 0, \tag{5.109}$$

with the same holding for n_-. The two null vectors can also be chosen so that

$$n_+ \wedge n_- = 2\hat{B}, \tag{5.110}$$

so that they form a null basis for the timelike plane defined by $\hat{B}$ (see figure 5.5).

The null vectors $n_\pm$ anticommute with $\hat{B}$ and therefore commute with $I\hat{B}$. The effect of the Lorentz transformation on $n_\pm$ is therefore

$$\begin{aligned} R n_\pm \tilde{R} &= \mathrm{e}^{\alpha \hat{B}/2} n_\pm \mathrm{e}^{-\alpha \hat{B}/2} \\ &= \cosh(\alpha)\, n_\pm + \sinh(\alpha)\, \hat{B} \cdot n_\pm \\ &= \mathrm{e}^{\pm\alpha} n_\pm. \end{aligned} \tag{5.111}$$

The two null directions are therefore just scaled — their direction is unchanged. It follows that every Lorentz transformation has two invariant null directions. The case where the bivector generator itself is null, $B^2 = 0$, corresponds to the special situation where these two null directions coincide.

5.4.2 The celestial sphere

One way to visualise the effect of Lorentz transformations is through their effect on the past light-cone (see figure 5.6). Each null vector on the past light-cone maps to a point on the sphere S^- — the *celestial sphere* for the observer. Suppose

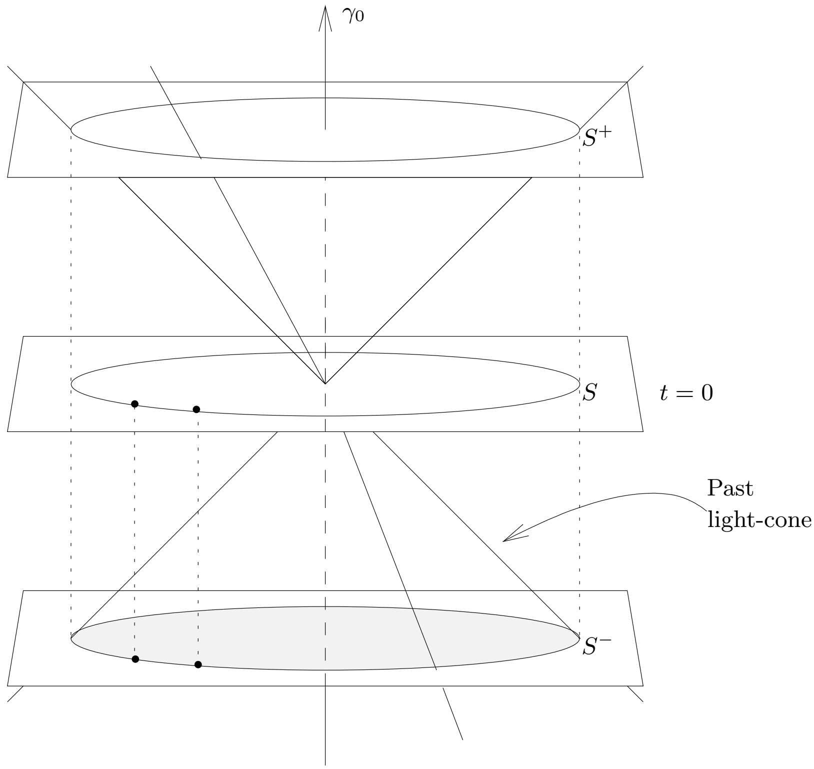


Figure 5.6 *The celestial sphere.* Each observer sees events in their past light-cone, which can be viewed as defining a sphere (shown here as a circle in a plane).

then that light is received along the null vector n, with the observer's velocity chosen to be γ_0. The relative vector in the γ_0 frame is $n\wedge\gamma_0$. This has magnitude

$$(n\wedge\gamma_0)^2 = (n\cdot\gamma_0)^2 - n^2\gamma_0^2 = (n\cdot\gamma_0)^2. \tag{5.112}$$

We therefore define the unit relative vector $\boldsymbol{n}$ by the projective formula

$$\boldsymbol{n} = \frac{n\wedge\gamma_0}{n\cdot\gamma_0}. \tag{5.113}$$

Observers passing through the same spacetime point at different velocities see different celestial spheres. If a second observer has velocity $v = R\gamma_0\tilde{R}$, the unit relative vectors in this observer's frame are formed from $n\wedge v/n\cdot v$. These can be brought to the γ_0 frame for comparison by forming

$$\boldsymbol{n}' = \tilde{R}\frac{n\wedge v}{n\cdot v}R = \frac{n'\wedge\gamma_0}{n'\cdot\gamma_0}, \tag{5.114}$$

where $n' = \tilde{R}nR$. The effects of Lorentz transformations can be visualised simply by moving around points on the celestial sphere with the map $n \mapsto \tilde{R}nR$. We know immediately, then, that two directions remain invariant and so describe the same points on the celestial spheres of two observers.

5.4.3 Relativistic visualisation

We have endeavoured to separate the concept of a single observer from that of a coordinate lattice. A clear illustration of this distinction arises when one studies how bodies appear when seen by different observers. Concentrating purely on coordinates leads directly to the conclusion that there is a measurable Lorentz contraction in the direction of motion of a body moving relative to some coordinate system. But when we consider what two different observers actually *see*, the picture is rather different.

Suppose that two observers in relative motion observe a sphere. The sphere and one of the observers are both at rest in the γ_0 system. This observer sees the edge of the sphere as a circle defined by the unit vectors

$$\boldsymbol{n} = \sin(\theta)(\cos(\phi)\,\boldsymbol{\sigma}_1 + \sin(\phi)\,\boldsymbol{\sigma}_2) + \cos(\theta)\,\boldsymbol{\sigma}_3, \quad 0 \leq \phi < 2\pi. \tag{5.115}$$

The angle θ is fixed so the sphere subtends an angle 2θ on the sky and is centred on the 3 axis (see figure 5.7). The incoming photon paths from the sphere are defined by the family of null vectors

$$n = (1 - \boldsymbol{n})\gamma_0. \tag{5.116}$$

Now suppose that a second observer has velocity $\beta = \tanh(\alpha)$ along the 1 axis, so

$$v = \cosh(\alpha)\,\gamma_0 + \sinh(\alpha)\,\gamma_1 = R\gamma_0\tilde{R}, \tag{5.117}$$

where $R = \exp(\alpha\,\gamma_1\gamma_0/2)$. To compare what these two observers see we form

$$\begin{aligned} n' = \tilde{R}nR = &\cosh(\alpha)\big(1 + \beta\sin(\theta)\,\cos(\phi)\big)\gamma_0 - \cosh(\alpha)\big(\sin(\theta)\,\cos(\phi) + \beta\big)\gamma_1 \\ &- \sin(\theta)\,\sin(\phi)\,\gamma_2 - \cos(\theta)\,\gamma_3. \end{aligned} \tag{5.118}$$

And from this the new unit relative outward vector is

$$\boldsymbol{n}' = \frac{\cosh(\alpha)\big(\sin(\theta)\,\cos(\phi) + \beta\big)\boldsymbol{\sigma}_1 + \sin(\theta)\,\sin(\phi)\,\boldsymbol{\sigma}_2 + \cos(\theta)\,\boldsymbol{\sigma}_3}{\cosh(\alpha)\big(1 + \beta\sin(\theta)\,\cos(\phi)\big)}. \tag{5.119}$$

Now consider the vector

$$\boldsymbol{c} = \boldsymbol{\sigma}_3 + \sinh(\alpha)\,\cos(\theta)\,\boldsymbol{\sigma}_1. \tag{5.120}$$

This vector satisfies

$$\boldsymbol{c}\cdot\boldsymbol{n}' = \cosh(\alpha)\,\cos(\theta), \tag{5.121}$$

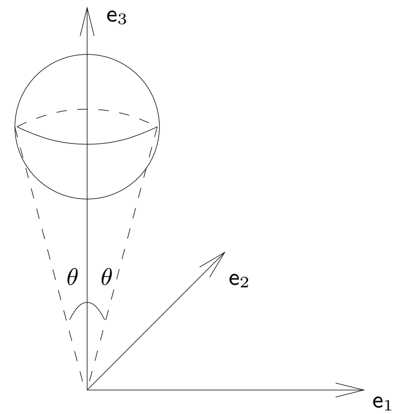


Figure 5.7 *Relativistic visualization of a sphere.* The sphere is at rest in the γ_0 frame with its centre a unit distance along the 3 axis. The sphere is simultaneously observed by two observers placed at the spatial origin. One observer is at rest in the γ_0 system, and the other is moving along the 1 axis.

which is independent of ϕ. It follows that, from the point of view of the second observer, all points on the edge of the sphere subtend the same angle to $\boldsymbol{c}$. So the vector $\boldsymbol{c}$ must lie at the centre of a circle, and the second observer still sees the edge of the sphere as circular. That is, both observers see the sphere as a sphere, and there is no observable contraction along the direction of motion. The only difference is that the moving observer sees the angular diameter of the sphere reduced from 2θ to $2\theta'$, where

$$\begin{aligned}\cos(\theta') &= \frac{\cos(\theta)\,\cosh(\alpha)}{\big(1+\sinh^2(\alpha)\,\cos^2(\theta)\big)^{1/2}},\\ \tan(\theta') &= \frac{\tan(\theta)}{\gamma}.\end{aligned} \tag{5.122}$$

More generally, moving observers see solid objects as rotated, as opposed to contracted along their direction of motion. Visualising Lorentz transformations of solid objects has now been discussed by various authors (see Rau, Weiskopf & Ruder (1998)). But the original observation that spheres remain spheres for observers in relative motion had to wait until 1959 — more than 50 years after the development of special relativity! The first authors to point out this invisibility of the Lorentz contraction were Terrell (1959) and Penrose (1959). Both authors based their studies on the fact that the Lorentz group is isomorphic to the conformal group acting on the surface of a sphere. This type of geometry is discussed in chapter 10.

5.4.4 Pure boosts and observer splits

Suppose we are travelling with velocity u and want to boost to velocity v. We seek the rotor for this which contains no additional rotational factors. We have

$$v = Lu\tilde{L} \tag{5.123}$$

with $La_\perp \tilde{L} = a_\perp$ for any vector outside the $u \wedge v$ plane. It is clear that the appropriate bivector for the rotor is $u \wedge v$, and as this anticommutes with u and v we have

$$v = Lu\tilde{L} = L^2 u \quad \Rightarrow L^2 = vu. \tag{5.124}$$

The solution to this is

$$L = \frac{1 + vu}{[2(1 + u \cdot v)]^{1/2}} = \exp\left(\frac{\alpha}{2} \frac{v \wedge u}{|v \wedge u|}\right), \tag{5.125}$$

where the angle α is defined by $\cosh(\alpha) = u \cdot v$.

Now suppose that we start in the γ_0 frame and some arbitrary rotor R takes this to $v = R\gamma_0\tilde{R}$. We know that the pure boost for this transformation is

$$L = \frac{1 + v\gamma_0}{[2(1 + v \cdot \gamma_0)]^{1/2}} = \exp\left(\frac{\alpha}{2} \frac{v \wedge \gamma_0}{|v \wedge \gamma_0|}\right), \tag{5.126}$$

where $v \cdot \gamma_0 = \cosh(\alpha)$. Now define the further rotor U by

$$U = \tilde{L}R, \qquad U\tilde{U} = \tilde{L}R\tilde{R}L = 1. \tag{5.127}$$

This satisfies

$$U\gamma_0\tilde{U} = \tilde{L}vL = \gamma_0, \tag{5.128}$$

so $U\gamma_0 = \gamma_0 U$. We must therefore have $U = \exp(I\boldsymbol{b}/2)$, where $I\boldsymbol{b}$ is a relative bivector, and U generates a pure rotation in the γ_0 frame. We now have

$$R = LU, \tag{5.129}$$

which decomposes R into a relative rotation and boost. Unlike the invariant decomposition into a boost and rotation of equation (5.107), the boost L and rotation U will not usually commute. The fact that the LU decomposition initially singled out the γ_0 vector shows that the decomposition is frame-dependent. Both the invariant split of equation (5.107) and the frame-dependent split of equation (5.129) are useful in practice.

5.5 Spacetime dynamics

Dynamics in spacetime is traditionally viewed as a hard subject. This need not be the case, however. We have now established that Lorentz transformations which preserve parity and causal structure can be described with rotors. By parameterising the motion in terms of rotors many equations are considerably simplified, and can be solved in new ways. This provides a simple understanding of the Thomas precession, as well as a new formulation of the Lorentz force law for a particle in an electromagnetic field.

5.5.1 Rotor equations and Fermi transport

A spacetime trajectory $x(\tau)$ has a future-pointing velocity vector $\dot{x} = v$. This is normalised to $v^2 = 1$ by parameterising the curve in terms of the proper time. This suggests an analogy with rigid-body dynamics. We write

$$v = R\gamma_0\tilde{R}, \tag{5.130}$$

which keeps v future-pointing and normalised. This moves all of the dynamics into the rotor $R = R(\tau)$, and this is the key idea which simplifies much of relativistic dynamics. The next quantity we need to find is the acceleration

$$\dot{v} = \frac{d}{d\tau}(R\gamma_0\tilde{R}) = \dot{R}\gamma_0\tilde{R} + R\gamma_0\dot{\tilde{R}}. \tag{5.131}$$

But just as in three dimensions, $\dot{R}\tilde{R}$ is of even grade and is equal to minus its reverse, so can only contain bivector terms. We therefore have

$$\begin{aligned}\dot{v} &= \dot{R}\tilde{R}v - v\dot{R}\tilde{R} \\ &= 2(\dot{R}\tilde{R})\cdot v.\end{aligned} \tag{5.132}$$

This equation is consistent with the fact that $v\cdot\dot{v} = 0$, which follows from $v^2 = 1$.

If we now form the acceleration bivector we obtain

$$\dot{v}v = 2(\dot{R}\tilde{R})\cdot v\, v. \tag{5.133}$$

This determines the projection of the bivector into the instantaneous rest frame defined by v. In this frame the projected bivector is purely timelike and corresponds to a pure boost. The remaining freedom in $\dot{R}\tilde{R}$ corresponds to an additional rotation in R which does not change v.

For the purposes of determining the velocity and trajectory of a particle the component of $\dot{R}\tilde{R}$ perpendicular to v is of no relevance. In some applications, however, it is useful to attach physical significance to the comoving frame vectors $\{\mathsf{e}_\mu\}$,

$$\mathsf{e}_\mu = R\gamma_\mu\tilde{R}, \tag{5.134}$$

which have $\mathsf{e}_0 = v$. The spatial set of vectors $\{\mathsf{e}_i\}$ satisfy $\mathsf{e}_i\cdot v = 0$ and span the instantaneous rest space of v. In this case, the dynamics of the e_i can be used to determine the component of $\dot{R}\tilde{R}$ which is not fixed by v alone.

The vectors $\{\mathsf{e}_i\}$ are carried along the trajectory by the rotor R. They are said to be *Fermi-transported* if their transformation from one instant to the next is

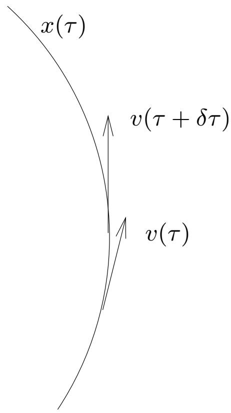


Figure 5.8 *The proper boost.* The change in velocity from τ to $\tau+\delta\tau$ should be described by a rotor solely in the $\dot{v}\wedge v$ plane.

a pure boost in the v frame. In this case the $\{\mathsf{e}_i\}$ vectors remain 'as constant as possible', subject to the constraint $\mathsf{e}_i\cdot v = 0$. For example, the direction defined by the angular momentum of an inertial guidance gyroscope (supported at its centre of mass so there are no torques) is Fermi-transported along the path of the gyroscope through spacetime.

To ensure Fermi-transport of $R\gamma_i\tilde{R}$ we need to ensure that the rotor describes pure boosts from one instant to the next (see figure 5.8). To first order in $\delta\tau$ we have

$$v(\tau+\delta\tau) = v(\tau) + \delta\tau\,\dot{v}. \tag{5.135}$$

The pure boost between $v(\tau)$ and $v(\tau+\delta\tau)$ is determined by the rotor

$$L = \frac{1+v(\tau+\delta\tau)v(\tau)}{[2(1+v(\tau+\delta\tau)\cdot v(\tau))]^{1/2}} = 1+\tfrac{1}{2}\delta\tau\,\dot{v}v, \tag{5.136}$$

to first order in $\delta\tau$. But since

$$R(\tau+\delta\tau) = R(\tau)+\delta\tau\dot{R}(\tau) = (1+\delta\tau\dot{R}\tilde{R})R(\tau), \tag{5.137}$$

the additional rotation that takes the $\{\mathsf{e}_i\}$ frame from τ to $\tau+\delta\tau$ is described by the rotor $1+\delta\tau\dot{R}\tilde{R}$. Equating this to the pure boost L of equation (5.136), we find that the correct expression to ensure Fermi-transport of the $\{\mathsf{e}_i\}$ is

$$\dot{R}\tilde{R} = \tfrac{1}{2}\dot{v}v. \tag{5.138}$$

This is as one would expect. The bivector describing the change in the rotor is simply the acceleration bivector, which is the acceleration seen in the instantaneous rest frame.

Under Fermi-transport the $\{\mathsf{e}_i\}$ frame vectors satisfy

$$\dot{\mathsf{e}}_i = 2(\dot{R}\tilde{R})\cdot\mathsf{e}_i = -\mathsf{e}_i\cdot(\dot{v}v). \tag{5.139}$$

This leads directly to the definition of the *Fermi derivative*

$$\frac{Da}{D\tau} = \dot{a} + a\cdot(\dot{v}v). \tag{5.140}$$

The Fermi derivative of a vector vanishes if the vector is Fermi-transported along the worldline. The derivative preserves both the magnitude a^2 and $a \cdot v$. The former holds because

$$\frac{d}{d\tau}(a^2) = -2a\cdot\big(a\cdot(\dot{v}\wedge v)\big) = 0. \tag{5.141}$$

Conservation of $a \cdot v$ is also straightforward to check:

$$\begin{aligned}\frac{d}{d\tau}(a\cdot v) &= -\big(a\cdot(\dot{v}v)\big)\cdot v + a\cdot\dot{v} \\ &= -a\cdot\dot{v} + a\cdot v\,\dot{v}\cdot v + a\cdot\dot{v} = 0.\end{aligned} \tag{5.142}$$

It follows that if a starts perpendicular to v it remains so. In the case where $a\cdot v = 0$ the Fermi derivative takes on the simple form

$$\frac{Da}{D\tau} = \dot{a} + a\cdot\dot{v}\,v = \dot{a} - \dot{a}\cdot v\,v = \dot{a}\wedge v\,v. \tag{5.143}$$

This is the projection of $\dot{a}$ perpendicular to v, as expected. The Fermi derivative extends simply to multivectors as follows:

$$\frac{DM}{D\tau} = \frac{dM}{d\tau} + M\times(\dot{v}v). \tag{5.144}$$

Derivatives of this type are important in gauge theories and gravity.

5.5.2 Thomas precession

As an application, consider a particle in a circular orbit (figure 5.9). The worldline is

$$x(\tau) = t(\tau)\gamma_0 + a(\cos(\omega t)\gamma_1 + \sin(\omega t)\gamma_2), \tag{5.145}$$

and the velocity is

$$v = \dot{x} = \dot{t}\big(\gamma_0 + a\omega(-\sin(\omega t)\gamma_1 + \cos(\omega t)\gamma_2)\big). \tag{5.146}$$

The relative velocity as seen in the γ_0 frame, $\boldsymbol{v} = v\wedge\gamma_0/v\cdot\gamma_0$, has magnitude $|\boldsymbol{v}| = a\omega$. We therefore introduce the hyperbolic angle α, with

$$\tanh(\alpha) = a\omega, \quad \dot{t} = \cosh(\alpha). \tag{5.147}$$

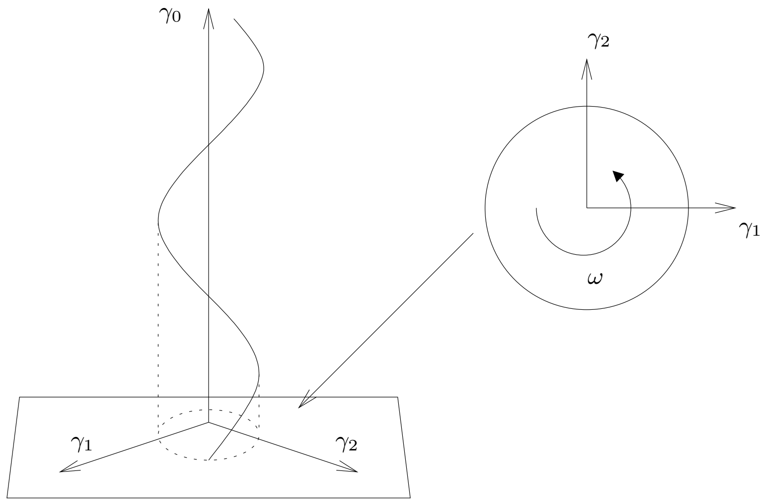


Figure 5.9 *Thomas precession.* The particle follows a helical worldline, rotating at a constant rate in the γ_0 frame.

The velocity is now

$$\begin{aligned} v &= \cosh(\alpha)\,\gamma_0 + \sinh(\alpha)\big(-\sin(\omega t)\gamma_1 + \cos(\omega t)\gamma_2\big) \\ &= \mathrm{e}^{\alpha \boldsymbol{n}/2}\gamma_0 \mathrm{e}^{-\alpha \boldsymbol{n}/2}, \end{aligned} \tag{5.148}$$

where

$$\boldsymbol{n} = -\sin(\omega t)\boldsymbol{\sigma}_1 + \cos(\omega t)\boldsymbol{\sigma}_2. \tag{5.149}$$

This form of time dependence in the rotor is inconvenient to work with. To simplify, we write

$$\boldsymbol{n} = \mathrm{e}^{-\omega t I \boldsymbol{\sigma}_3}\boldsymbol{\sigma}_2 = R_\omega \boldsymbol{\sigma}_2 \tilde{R}_\omega, \tag{5.150}$$

where $R_\omega = \exp(-\omega t I\boldsymbol{\sigma}_3/2)$. We now have

$$\mathrm{e}^{\alpha \boldsymbol{n}/2} = \exp(\alpha R_\omega \boldsymbol{\sigma}_2 \tilde{R}_\omega/2) = R_\omega R_\alpha \tilde{R}_\omega, \tag{5.151}$$

where

$$R_\alpha = \exp(\alpha \boldsymbol{\sigma}_2/2). \tag{5.152}$$

The velocity is now given by

$$v = R_\omega R_\alpha \tilde{R}_\omega \gamma_0 R_\omega \tilde{R}_\alpha \tilde{R}_\omega = R_\omega R_\alpha \gamma_0 \tilde{R}_\alpha \tilde{R}_\omega. \tag{5.153}$$

The final expression follows because R_ω commutes with γ_0.

We can now see that the rotor for the motion must have the form

$$R = R_\omega R_\alpha \Phi, \tag{5.154}$$

where Φ is a rotor that commutes with γ_0. We want R to describe Fermi transport of the $\{\mathsf{e}_i\}$, so we must have $\dot{v}v = 2\dot{R}\tilde{R}$. We begin by forming the acceleration bivector $\dot{v}v$. We can simplify this derivation by writing $v = R_\omega v_\alpha \tilde{R}_\omega$, where $v_\alpha = R_\alpha \gamma_0 \tilde{R}_\alpha$. We then find that

$$\begin{aligned}\dot{v}v &= R_\omega\big(2(\tilde{R}_\omega \dot{R}_\omega)\cdot v_\alpha\, v_\alpha\big)\tilde{R}_\omega \\ &= -\omega\,\cosh(\alpha)\,R_\omega\big((I\boldsymbol{\sigma}_3)\cdot v_\alpha\, v_\alpha\big)\tilde{R}_\omega \\ &= \omega\,\sinh(\alpha)\,\cosh(\alpha)\,R_\omega\big(-\cosh(\alpha)\,\boldsymbol{\sigma}_1 + \sinh(\alpha)\,I\boldsymbol{\sigma}_3\big)\tilde{R}_\omega. \end{aligned} \tag{5.155}$$

We also form the rotor equivalent, $2\dot{R}\tilde{R}$, which is

$$\begin{aligned}2\dot{R}\tilde{R} &= 2\dot{R}_\omega\tilde{R}_\omega + 2R_\omega R_\alpha \dot{\Phi}\tilde{\Phi}\tilde{R}_\alpha\tilde{R}_\omega \\ &= -\omega\cosh(\alpha)\,I\boldsymbol{\sigma}_3 + 2R_\omega R_\alpha\dot{\Phi}\tilde{\Phi}\tilde{R}_\alpha\tilde{R}_\omega. \end{aligned} \tag{5.156}$$

Equating the two preceding results we find that

$$\begin{aligned}2\dot{\Phi}\tilde{\Phi} &= \omega\cosh^2(\alpha)\,\tilde{R}_\alpha\big(-\sinh(\alpha)\,\boldsymbol{\sigma}_1 + \cosh(\alpha)\,I\boldsymbol{\sigma}_3\big)R_\alpha \\ &= \omega\cosh^2(\alpha)\,I\boldsymbol{\sigma}_3. \end{aligned} \tag{5.157}$$

The solution with $\Phi = 1$ at $t = 0$ is $\Phi = \exp(\omega\cosh(\alpha)tI\boldsymbol{\sigma}_3/2)$, so the full rotor is

$$R = \mathrm{e}^{-\omega t I\boldsymbol{\sigma}_3/2}\mathrm{e}^{\alpha\boldsymbol{\sigma}_2/2}\mathrm{e}^{\cosh(\alpha)\omega t I\boldsymbol{\sigma}_3/2}. \tag{5.158}$$

This form of the rotor ensures that the $\mathsf{e}_i = R\gamma_i\tilde{R}$ are Fermi transported. The fact that the 'internal' rotation rate $\omega\cosh(\alpha)$ differs from ω is due to the fact that the acceleration is formed in the instantaneous rest frame v and not the fixed γ_0 frame. This difference introduces a precession — the *Thomas precession*. We can see this effect by imagining the vector γ_1 being transported around the circle. The rotated vector is

$$\mathsf{e}_1 = R\gamma_1\tilde{R}. \tag{5.159}$$

In the low velocity limit $\cosh(\alpha) \mapsto 1$ the vector γ_1 continues to point in the γ_1 direction and the frame does not rotate, as we would expect. At larger velocities, however, the frame starts to precess. After time $t = 2\pi/\omega$, for example, the γ_1 vector is transformed to

$$\mathsf{e}_1(2\pi/\omega) = \mathrm{e}^{\alpha\boldsymbol{\sigma}_2/2}\mathrm{e}^{2\pi\cosh(\alpha)I\boldsymbol{\sigma}_3}\gamma_1\mathrm{e}^{-\alpha\boldsymbol{\sigma}_2/2}. \tag{5.160}$$

Dotting this with the initial vector $\mathsf{e}_1(0) = \gamma_1$ we see that the vector has precessed through an angle

$$\theta = 2\pi(\cosh(\alpha) - 1). \tag{5.161}$$

This shows that the effect is of order $|\boldsymbol{v}|^2/c^2$. The form of the Thomas precession justifies one of the relativistic corrections to the spin-orbit coupling in the Pauli theory of the electron.

5.5.3 The Lorentz force law

The non-relativistic form of the Lorentz force law for a particle of charge q is

$$\frac{d\boldsymbol{p}}{dt} = q(\boldsymbol{E} + \boldsymbol{v}\times\boldsymbol{B}), \tag{5.162}$$

where the $\times$ here denotes the vector cross product, and all relative vectors are expressed in some global Newtonian frame, which we will take to be the γ_0 frame. We seek a covariant relativistic version of this law. The quantity $\boldsymbol{p}$ on the left-hand side is the relative vector $p\wedge\gamma_0$. Since $dt = \gamma d\tau$, we must multiply through by $\gamma = v\cdot\gamma_0$ to convert the derivative into one with respect to proper time. The first term on the right-hand side then includes

$$\begin{aligned} v\cdot\gamma_0\, \boldsymbol{E} &= \tfrac{1}{4}\big(\boldsymbol{E}(v\gamma_0 + \gamma_0 v) + (v\gamma_0 + \gamma_0 v)\boldsymbol{E}\big) \\ &= \tfrac{1}{4}\big((\boldsymbol{E}v - v\boldsymbol{E})\gamma_0 - \gamma_0(\boldsymbol{E}v - v\boldsymbol{E})\big) \\ &= (\boldsymbol{E}\cdot v)\wedge\gamma_0. \end{aligned} \tag{5.163}$$

Recall at this point that $\boldsymbol{E}$ is a spacetime bivector built from the $\boldsymbol{\sigma}_k = \gamma_k\gamma_0$, so $\boldsymbol{E}$ *anticommutes* with γ_0.

For the magnetic term in equation (5.162) we first replace the cross product by the equivalent three-dimensional expression $(I\boldsymbol{B})\cdot\boldsymbol{v}$. Expanding out, and expressing in the full spacetime algebra, we obtain

$$\begin{aligned} \tfrac{1}{2}v\cdot\gamma_0(I\boldsymbol{B}\boldsymbol{v} - \boldsymbol{v}I\boldsymbol{B}) &= \tfrac{1}{4}\big(I\boldsymbol{B}(v\gamma_0 - \gamma_0 v) - (v\gamma_0 - \gamma_0 v)I\boldsymbol{B}\big) \\ &= \tfrac{1}{4}\big((I\boldsymbol{B}v - vI\boldsymbol{B})\gamma_0 - \gamma_0(I\boldsymbol{B}v - vI\boldsymbol{B})\big) \\ &= \big((I\boldsymbol{B})\cdot v\big)\wedge\gamma_0, \end{aligned} \tag{5.164}$$

where we use the fact that γ_0 *commutes* with $I\boldsymbol{B}$. Combining equations (5.163) and (5.164) we can now write the Lorentz force law (5.162) in the form

$$\frac{d\boldsymbol{p}}{d\tau} = \dot{p}\wedge\gamma_0 = q\big((\boldsymbol{E} + I\boldsymbol{B})\cdot v\big)\wedge\gamma_0. \tag{5.165}$$

We next define the *Faraday bivector* F by

$$F = \boldsymbol{E} + I\boldsymbol{B}. \tag{5.166}$$

This is the covariant form of the electromagnetic field strength. It unites the electric and magnetic fields into a single spacetime structure. We study this in greater detail in chapter 7. The Lorentz force law can now be written

$$\dot{p}\wedge\gamma_0 = q(F\cdot v)\wedge\gamma_0. \tag{5.167}$$

The rate of working on the particle is $q\boldsymbol{E}\cdot\boldsymbol{v}$, so

$$\frac{dp_0}{dt} = q\boldsymbol{E}\cdot\boldsymbol{v}. \tag{5.168}$$

Here, $p_0 = p\cdot\gamma_0$ is the particle's energy in the γ_0 frame. Multiplying through by $v\cdot\gamma_0$, we find

$$\dot{p}\cdot\gamma_0 = q\boldsymbol{E}\cdot(v\wedge\gamma_0) = q(F\cdot v)\cdot\gamma_0. \tag{5.169}$$

In the final step we have used $(I\boldsymbol{B})\cdot(v\wedge\gamma_0) = 0$. Adding this equation to equation (5.167), and multiplying on the right by γ_0, we find

$$\dot{p} = qF\cdot v. \tag{5.170}$$

Recalling that $p = mv$, we arrive at the relativistic form of the *Lorentz force law*,

$$m\dot{v} = qF\cdot v. \tag{5.171}$$

This is *manifestly* Lorentz covariant, because no particular frame is picked out. The acceleration bivector is

$$\dot{v}v = \frac{q}{m}F\cdot v\, v = \frac{q}{m}(F\cdot v)\wedge v = \frac{q}{m}\boldsymbol{E}_v, \tag{5.172}$$

where $\boldsymbol{E}_v$ is the relative electric field in the v frame. A charged point particle only responds to the instantaneous electric field in its frame. Algebraically, this bivector is

$$\boldsymbol{E}_v = \tfrac{1}{2}(F - vFv). \tag{5.173}$$

So $\boldsymbol{E}_v$ is the component of the bivector F which anticommutes with v.

Now suppose that we parameterise the velocity with a rotor, so that $v = R\gamma_0\tilde{R}$. We have

$$\dot{v} = 2\dot{R}\tilde{R}v = 2(\dot{R}\tilde{R})\cdot v = \frac{q}{m}F\cdot v. \tag{5.174}$$

The simplest form of the rotor equation comes from equating the projected terms:

$$\dot{R} = \frac{q}{2m}FR. \tag{5.175}$$

This is not the most general possibility as we could include an extra multiple of $F\wedge v\, v$. The rotor determined by equation (5.175) will not, in general, describe Fermi-transport of the $R\gamma_i\tilde{R}$ vectors. However, equation (5.175) is sufficient to determine the velocity of the particle, and is certainly the simplest form of rotor equation to work with. As we now demonstrate, the rotor equation (5.175) is remarkably efficient when it comes to solving the dynamical equations.

5.5.4 Constant field

Motion in a constant field is easy to solve for now. We can immediately integrate the rotor equation to give

$$R = \exp\left(\frac{q}{2m}F\tau\right)R_0. \tag{5.176}$$

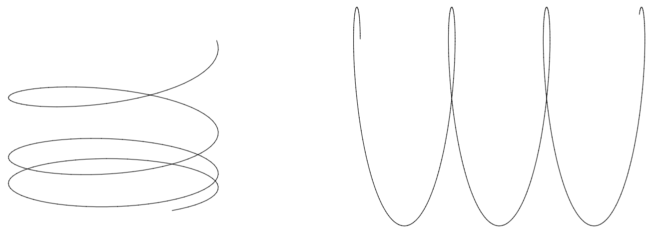


Figure 5.10 *Particle in a constant field.* The general motion is a combination of linear acceleration and circular motion. The plot on the left has $\boldsymbol{E}$ and $\boldsymbol{B}$ colinear. The plot on the right has $\boldsymbol{E}$ entirely in the $I\boldsymbol{B}$ plane, giving rise to cycloids.

To proceed and recover the trajectory we form the invariant decomposition of F. We first write

$$F^2 = \langle F^2\rangle_0 + \langle F^2\rangle_4 = \rho \mathrm{e}^{I\theta}, \tag{5.177}$$

so that we can set

$$F = \rho^{1/2}\mathrm{e}^{I\theta/2}\hat{F} = \alpha\hat{F} + I\beta\hat{F}, \tag{5.178}$$

where $\hat{F}^2 = 1$. (If F is null a slightly different procedure is followed.) We now have

$$R = \exp\left(\frac{q}{2m}\alpha\hat{F}\tau\right)\,\exp\left(\frac{q}{2m}I\beta\hat{F}\tau\right)R_0. \tag{5.179}$$

Next we decompose the initial velocity $v_0 = R_0\gamma_0\tilde{R}_0$ into components in and out of the $\hat{F}$ plane:

$$v_0 = \hat{F}^2 v_0 = \hat{F}\,\hat{F}\cdot v_0 + \hat{F}\,\hat{F}\wedge v_0 = v_{0\|} + v_{0\perp}. \tag{5.180}$$

Now $v_{0\|} = \hat{F}\,\hat{F}\cdot v_0$ anticommutes with $\hat{F}$, and $v_{0\perp}$ commutes with $\hat{F}$, so

$$\dot{x} = \exp\left(\frac{q}{m}\alpha\hat{F}\tau\right)v_{0\|} + \exp\left(\frac{q}{m}I\beta\hat{F}\tau\right)v_{0\perp}. \tag{5.181}$$

This integrates immediately to give the particle history

$$x - x_0 = \frac{\mathrm{e}^{q\alpha\hat{F}\tau/m} - 1}{q\alpha/m}\hat{F}\cdot v_0 - \frac{\mathrm{e}^{q\beta I\hat{F}\tau/m} - 1}{q\beta/m}(I\hat{F})\cdot v_0. \tag{5.182}$$

The first term gives linear acceleration and the second is periodic and drives rotational motion (see figure 5.10). One has to be slightly careful integrating the velocity equation in the case where either α or β is zero, which corresponds to perpendicular $\boldsymbol{E}$ and $\boldsymbol{B}$ fields.

5.5.5 Particle in a Coulomb field

As a further application we consider the case of a charged point particle moving in a central Coulomb field. If relativistic effects are ignored the problem reduces to the inverse-square force law described in section 3.2.1. We therefore expect that the relativistic description will add additional perturbative effects to the elliptic and hyperbolic orbits found in the inverse-square case. We assume for simplicity that the central charge has constant velocity γ_0 and is placed at the origin. The electromagnetic field is

$$F = \frac{Q\boldsymbol{x}}{4\pi\epsilon_0 r^3}, \tag{5.183}$$

where $\boldsymbol{x} = x\wedge\gamma_0$ and $r^2 = \boldsymbol{x}^2$. In this section all bold symbols denote relative vectors in the γ_0 frame. The question of how to generalise the non-relativistic definitions of centre of mass and relative separation turns out to be surprisingly complex and is not tackled here. Instead we will simply assume that the source of the Coulomb field is far heavier than the test charge so that the source's motion can be ignored.

There are two constants of motion for this force law. The first is the energy

$$E = mv\cdot\gamma_0 + \frac{qQ}{4\pi\epsilon_0 r}. \tag{5.184}$$

If the charges are opposite, qQ is negative and the potential is attractive. The force law can now be written in the γ_0 frame as

$$m\frac{d^2\boldsymbol{x}}{d\tau^2} = \frac{qQ\boldsymbol{x}}{4\pi\epsilon_0 r^3}\left(\frac{E}{m} - \frac{qQ}{4\pi\epsilon_0 mr}\right). \tag{5.185}$$

The second conserved quantity is the angular momentum, which is conserved for any central force, as is the case in equation (5.185). If we define the spacetime bivector $L = x\wedge p$ we find that

$$\dot{L} = qx\wedge(F\cdot v). \tag{5.186}$$

It follows that the trivector $L\wedge\gamma_0$ is conserved. Equivalently, we can define the relative bivector

$$I\boldsymbol{l} = L\wedge\gamma_0\,\gamma_0, \tag{5.187}$$

so that the relative vector $\boldsymbol{l}$ is conserved. This is the relative angular momentum vector and satisfies $\boldsymbol{x}\cdot\boldsymbol{l} = 0$. It follows that the test particle's motion takes place in a constant plane as seen from the source charge.

In order to integrate the rotor equation we need to find a way to express the field as a function of the particle's proper time. This is achieved by introducing an angular measure in the plane of motion. Suppose that we align the 3 axis with $\boldsymbol{l}$, so that we can write

$$\hat{\boldsymbol{x}}(\tau) = \boldsymbol{\sigma}_1 \exp\big(I\boldsymbol{\sigma}_3\theta(\tau)\big), \tag{5.188}$$

where $\hat{\boldsymbol{x}}$ is the unit relative vector $\boldsymbol{x}/r$. It follows that

$$\boldsymbol{l}^2 = m^2 r^4 \dot{\hat{\boldsymbol{x}}}^2 = m^2 r^4 \dot{\theta}^2. \tag{5.189}$$

If we set $l = |\boldsymbol{l}|$ we have $l = mr^2\dot{\theta}$, which enables us to express the Coulomb field as

$$F = \frac{Qm\dot{\theta}\boldsymbol{\sigma}_1 \exp\bigl(I\boldsymbol{\sigma}_3\theta(\tau)\bigr)}{4\pi\epsilon_0 l}. \tag{5.190}$$

If we now let

$$\kappa = \frac{qQ}{4\pi\epsilon_0 l} \tag{5.191}$$

the rotor equation takes on the simple form

$$\frac{dR}{d\theta} = \frac{\kappa}{2}\boldsymbol{\sigma}_1 \exp(I\boldsymbol{\sigma}_3\theta)R. \tag{5.192}$$

Re-expressing the differential equation in terms of θ is a standard technique for solving inverse-square problems in non-relativistic physics. But this technique fails to give a simple solution to the relativistic equation (5.185). Instead, we see that the technique gives a simple solution to the relativistic problem only if applied directly to the rotor equation.

To solve equation (5.192) we first set

$$R = \exp(-I\boldsymbol{\sigma}_3\theta/2)U. \tag{5.193}$$

It follows that

$$\frac{dU}{d\theta}\tilde{U} = \tfrac{1}{2}(\kappa\boldsymbol{\sigma}_1 + I\boldsymbol{\sigma}_3), \tag{5.194}$$

which integrates straightforwardly. The full rotor is then

$$R = \mathrm{e}^{-I\boldsymbol{\sigma}_3\theta/2}\mathrm{e}^{A\theta/2}R_0, \tag{5.195}$$

where

$$A = \kappa\boldsymbol{\sigma}_1 + I\boldsymbol{\sigma}_3. \tag{5.196}$$

The initial conditions can be chosen such that $\theta(0) = 0$, which tells us how to align the 1 axis. The rotor R_0 then specifies the initial velocity v_0. If we are not interested in transporting a frame, R_0 can be set equal to a pure boost from γ_0 to v_0.

With the rotor equation now solved, the velocity can be integrated to recover the trajectory. Clearly, different types of path are obtained for the different signs of $A^2 = \kappa^2 - 1$. The equation relating r and θ is found from the relation

$$-\frac{d}{d\theta}\left(\frac{1}{r}\right) = \frac{m}{l}\hat{\boldsymbol{x}}\cdot\dot{\boldsymbol{x}}. \tag{5.197}$$

To evaluate the right-hand side we need

$$\begin{aligned}\hat{\boldsymbol{x}}\cdot\dot{\boldsymbol{x}} &= \langle \mathrm{e}^{-I\boldsymbol{\sigma}_3\theta/2}\boldsymbol{\sigma}_1\mathrm{e}^{I\boldsymbol{\sigma}_3\theta/2}R\gamma_0\tilde{R}\gamma_0\rangle \\ &= -\langle\gamma_1\mathrm{e}^{A\theta/2}v_0\mathrm{e}^{-A\theta/2}\rangle \\ &= \langle\mathrm{e}^{-A\theta}\gamma^1 v_0\rangle. \end{aligned} \tag{5.198}$$

It follows that

$$-\frac{d}{d\theta}\left(\frac{1}{r}\right) = \frac{m}{l}\langle\mathrm{e}^{-A\theta}\gamma^1 v_0\rangle. \tag{5.199}$$

For a given l and v_0 this integrates to give the trajectory in the $\boldsymbol{Il}$ plane.

Suppose, for example, that we are interested in bound states. For these we must have $A^2 < 0$, which implies that $\kappa^2 < 1$. We write

$$|A| = (1-\kappa^2)^{1/2} \tag{5.200}$$

for the magnitude of A. To simplify the equations we will assume that $\tau = 0$ corresponds to a point on the trajectory where $\boldsymbol{v}$ is perpendicular to $\boldsymbol{x}$. In this case we have

$$v_0 = \cosh(\alpha_0)\,\gamma_0 + \sinh(\alpha_0)\,\gamma_2 \tag{5.201}$$

so that the trajectory is determined by

$$-\frac{d}{d\theta}\left(\frac{1}{r}\right) = \frac{m}{l|A|}\big(\kappa\cosh(\alpha_0)+\sinh(\alpha_0)\big)\sin(|A|\theta). \tag{5.202}$$

The magnitude of the angular momentum is given by $l = mr_0\sinh(\alpha_0)$, which can be used to write

$$m\big(\kappa\cosh(\alpha_0)+\sinh(\alpha_0)\big) = (E^2 - m^2|A|^2)^{1/2}. \tag{5.203}$$

The trajectory is then given by

$$\frac{l|A|^2}{r} = -\kappa E + (E^2 - m^2|A|^2)^{1/2}\cos(|A|\theta), \tag{5.204}$$

and since this represents a bound state, κ must be negative. The fact that the angular term goes as $\cos(|A|\theta)$ shows that this equation specifies a precessing ellipse (figure 5.11). The precession rate of the ellipse can be found simply using the technique of section 3.3.

5.5.6 The gyromagnetic moment

Particles with non-zero spin have a magnetic moment which is proportional to the spin. In non-relativistic physics we write this as $\boldsymbol{m} = \gamma\boldsymbol{s}$, where γ is the gyromagnetic ratio and $\boldsymbol{s}$ is the spin (which has units of angular momentum). The gyromagnetic ratio is usually written in the form

$$\gamma = g\frac{q}{2m}, \tag{5.205}$$

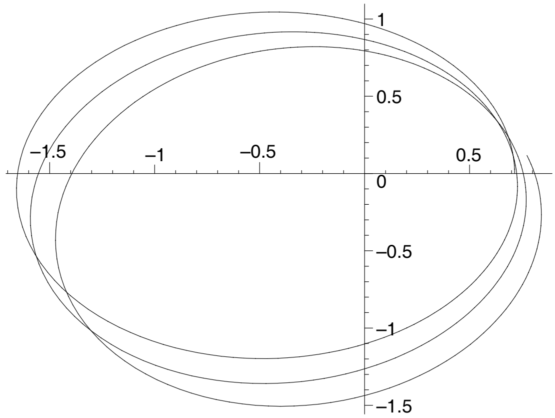


Figure 5.11 *Motion in a Coulomb field.* For bound orbits ($E < m$) the particle's motion is described by a precessing ellipse. The plot is for $|A| = 0.95$. The units are arbitrary.

where m is the particle mass, q is the charge and g is the (reduced) gyromagnetic ratio. The last is determined experimentally via the precession of the spin vector which, in classical physics, obeys

$$\dot{\boldsymbol{s}} = g\frac{q}{2m}(I\boldsymbol{B})\cdot\boldsymbol{s}. \tag{5.206}$$

We seek a relativistic extension of this equation. We start by introducing the relativistic spin vector s, which is perpendicular to the velocity v, so $s \cdot v = 0$. For a particle at rest in the γ_0 frame we have $s = \boldsymbol{s}\gamma_0$. The particle's spin will interact with the magnetic field only in the instantaneous rest frame, so we should regard equation (5.206) as referring to this frame.

Given that $\boldsymbol{s} = s\gamma_0$ we find that

$$\begin{aligned}(I\boldsymbol{B})\cdot\boldsymbol{s} &= \langle (F\wedge\gamma_0)\gamma_0 s\gamma_0\rangle_2 \\ &= (F\cdot s)\wedge\gamma_0.\end{aligned} \tag{5.207}$$

So, for a particle at rest in the γ_0 frame, equation (5.206) can be written

$$\frac{ds}{dt} = g\frac{q}{2m}(F\cdot s)\wedge\gamma_0\,\gamma_0. \tag{5.208}$$

To write down an equation which is valid for arbitrary velocity we must replace the two factors of γ_0 on the right-hand side with the velocity v. On the left-hand side we need the derivative of s which preserves $s \cdot v = 0$. This is the Fermi derivative of section 5.5.1, which tells us that the relativistic form of the spin precession equation is

$$\dot{s} + s\cdot(\dot{v}v) = g\frac{q}{2m}(F\cdot s)\wedge v\, v. \tag{5.209}$$

This equation tells us how much the spin vector rotates, relative to a Fermi-transported frame, which is physically sensible. We can eliminate the acceleration bivector $\dot{v}v$ by using the relativistic Lorentz force law to find

$$\begin{aligned}\dot{s} &= g\frac{q}{2m}(F\cdot s)\wedge v\, v - \frac{q}{m}s\cdot(F\cdot v\, v)\\ &= \frac{q}{2m}\big(g(F\cdot s)\wedge v + 2(F\cdot s)\cdot v\big)v\\ &= \frac{q}{m}F\cdot s + (g-2)\frac{q}{2m}(F\cdot s)\wedge v\, v.\end{aligned} \tag{5.210}$$

This is called the Bargmann–Michel–Telegdi equation.

For the value $g = 2$, the Bargmann–Michel–Telegdi equation reduces to

$$\dot{s} = \frac{q}{m}F\cdot s, \tag{5.211}$$

which has the same form as the Lorentz force law. In this sense, $g = 2$ is the most natural value of the gyromagnetic ratio of a point particle in relativistic physics. Ignoring quantum corrections, this is indeed found to be the value for an electron. Quantum corrections tell us that for an electron $g = 2(1+\alpha/2\pi+\cdots)$. The corrections are due to the fact that the electron is never truly isolated and constantly interacts with virtual particles from the quantum vacuum.

Given a velocity v and a spin vector s, with $v\cdot s = 0$ and s normalised to $s^2 = -1$, we can always find a rotor R such that

$$v = R\gamma_0\tilde{R}, \quad s = R\gamma_3\tilde{R}. \tag{5.212}$$

For these we have

$$\dot{v} = 2(\dot{R}\tilde{R})\cdot v, \quad \dot{s} = 2(\dot{R}\tilde{R})\cdot s. \tag{5.213}$$

For a particle with $g = 2$, this pair of equations reduces to the single rotor equation (5.175). The simple form of this equation further justifies the claim that $g = 2$ is the natural, relativistic value of the gyromagnetic ratio. This also means that once we have solved the rotor equation, we can simultaneously compute both the trajectory and the spin precession of a classical relativistic particle with $g = 2$.

5.6 Notes

There are many good introductions to special relativity. Standard references include the books by French (1968), Rindler (1977) and d'Inverno (1992). Practically all introductory books make heavy use of coordinate geometry. Geometric algebra was first systematically applied to the study of relativistic physics in the book *Space-Time Algebra* by Hestenes (1966). Since this book was published in 1966 many authors have applied spacetime algebra techniques to relativistic physics. The two most significant papers are again by Hestenes, 'Proper particle mechanics' and 'Proper dynamics of a rigid point particle' (1974a,b). These papers detail the use of rotor equations for solving problems in electrodynamics, and much of section 5.5 follows their presentation.

5.7 Exercises

5.1 Suppose that the spacetime bivector $\hat{B}$ satisfies $\hat{B}^2 = 1$. By writing $\hat{B} = \boldsymbol{a} + I\boldsymbol{b}$ in the γ_0 frame, show that we can write

$$\hat{B} = \cosh(u)\hat{\boldsymbol{a}} + \sinh(u)I\hat{\boldsymbol{b}} = \mathrm{e}^{uI\hat{\boldsymbol{b}}\hat{\boldsymbol{a}}}\hat{\boldsymbol{a}},$$

where $\hat{\boldsymbol{a}}^2 = \hat{\boldsymbol{b}}^2 = 1$. Hence explain why we can write $\hat{B} = R\boldsymbol{\sigma}_3\tilde{R}$. By considering the null vectors $\gamma_0 \pm \gamma_3$, prove that we can always find two null vectors satisfying

$$\hat{B}\cdot n_\pm = \pm n_\pm.$$

5.2 The boost L from velocity u to velocity v satisfies

$$v = Lu\tilde{L} = L^2 u,$$

with $L\tilde{L} = 1$. Prove that a solution to this equation is

$$L = \frac{1 + vu}{[2(1 + v\cdot u)]^{1/2}}.$$

Is this solution unique? Show further that this solution can be written in the form

$$L = \exp\left(\frac{\alpha}{2}\frac{v\wedge u}{|v\wedge u|}\right),$$

where $\alpha > 0$ satisfies $\cosh(\alpha) = u\cdot v$.

5.3 *Compton scattering* occurs when a photon scatters off an electron. If we ignore quantum effects this can be modelled as a relativistic collison process. The incident photon has wavelength λ_0 in the frame in which the electron is initially stationary. Show that the wavelength after scattering, λ, satisfies

$$\lambda - \lambda_0 = \frac{2\pi\hbar}{mc}\big(1 - \cos(\theta)\big),$$

where θ is the angle through which the photon scatters.

5.4 A relativistic particle has velocity $v = R\gamma_0\tilde{R}$. Show that v satisfies the Lorentz force equation $m\dot{v} = qF\cdot v$ if R satisfies

$$\dot{R} = \frac{q}{2m}FR.$$

Show that the solution to this for a constant field is

$$R = \exp(qF\tau/2m)R_0.$$

Given that F is *null*, $F^2 = 0$, show that v is given by the polynomial

$$v = v_0 + \tau\frac{q}{m}F\cdot v_0 - \tau^2\frac{q^2}{4m^2}Fv_0F.$$

Suppose now that $F = \boldsymbol{\sigma}_1 + I\boldsymbol{\sigma}_2$ and the particle is initially at rest in the γ_0 frame. Sketch the resultant motion in the $\gamma_1\gamma_3$ plane.

5.5 One way to construct the Fermi derivative of a vector a is to argue that we should 'de-boost' the vector at proper time $\tau + \delta\tau$ before comparing it with $a(\tau)$. Explain why this leads us to evaluate

$$\lim_{\delta\tau\to 0}\frac{1}{\delta\tau}\big(\tilde{L}a(\tau+\delta\tau)L - a(\tau)\big),$$

and confirm that this evaluates to $\dot{a} + a\cdot(\dot{v}v)$.

5.6 A frame is Fermi-transported along the worldline of a particle with velocity $v = R\gamma_0\tilde{R}$. The rotor R is decomposed into a rotation and boost in the γ_0 frame as $R = LU$. Show that the rotation U satisfies

$$2\dot{U}\tilde{U} = -(\tilde{L}\dot{L} + \gamma_0\tilde{L}\dot{L}\gamma_0).$$

What is the interpretation of the right-hand side in terms of the γ_0 frame?

5.7 The bivector $B = a\wedge b$ is Fermi-transported along a worldline by Fermi-transporting the two vectors a and b. Show that B remains a blade, and that the bivector satisfies

$$\frac{dB}{d\tau} + B\times(\dot{v}v) = 0.$$

5.8 A point particle with a gyromagnetic ratio $g = 2$ is in a circular orbit around a central Coulomb field. Show that in one complete orbit the spin vector rotates in the plane $A = \kappa\boldsymbol{\sigma}_1 + I\boldsymbol{\sigma}_3$ by an amount $2\pi|A|$, where

$$\kappa = \frac{qQ}{4\pi\epsilon_0 l},$$

and l is the angular momentum.

5.9 Show that the Bargmann–Michel–Telegdi equation of (5.210) for a relativistic point particle with spin vector s can be written

$$\dot{s} = \frac{q}{m}\big(F + \tfrac{1}{2}(g-2)F\wedge v\, v\big)\cdot s.$$

Given that $v = R\gamma_0\tilde{R}$ and $s = R\gamma_3\tilde{R}$, show that the rotor R satisfies the equation

$$\dot{R} = \frac{q}{2m}FR + \frac{q}{4m}(g-2)RI\boldsymbol{B}_0,$$

where

$$I\boldsymbol{B}_0 = (\tilde{R}FR)\wedge\gamma_0\, \gamma_0.$$

Assuming that the electromagnetc field F is constant, prove that $\boldsymbol{B}_0$ is also constant. Hence study the precession of s for a particle with a gyromagnetic ratio $g \neq 2$.

6

Geometric calculus

Geometric algebra provides us with an invertible product for vectors. In this chapter we investigate the new insights this provides for the subject of vector calculus. The familiar gradient, divergence and curl operations all result from the action of the vector operator, ∇. Since this operator is vector-valued, we can now form its geometric product with other multivectors. We call this the *vector derivative*. Unlike the separate divergence and curl operations, the vector derivative has the important property of being invertible. That is to say, Green's functions exist for ∇ which enable initial conditions to be propagated off a surface.

The synthesis of vector differentiation and geometric algebra described in this chapter is called '*geometric calculus*'. We will see that geometric calculus provides new insights into the subject of complex analysis and enables the concept of an analytic function to be extended to arbitrary dimensions. In three dimensions this generalisation gives rise to the angular eigenstates of the Pauli theory, and the spacetime generalisation of an analytic function defines the wavefunction for a massless spin-1/2 particle. Clearly there are many insights to be gained from a unified treatment of calculus based around the geometric product.

The early sections of this chapter discuss the vector derivative, and its associated Green's functions, in flat spaces. This way we can quickly assemble a number of results of central importance in later chapters. The generalisations to embedded surfaces and manifolds are discussed in the final section. This is a large and important subject, which has been widely discussed elsewhere. Our presentation here is kept brief, focusing on the key results which are required later in this book.

6.1 The vector derivative

The vector derivative is denoted with the symbol ∇ (or $\boldsymbol{\nabla}$ in two and three dimensions). Algebraically, this has all of the properties of a vector (grade-1) object in a geometric algebra. The operator properties of ∇ are contained in the definition that the inner product of ∇ with any vector a results in the *directional derivative* in the a direction. That is,

$$a\cdot\nabla F(x) = \lim_{\epsilon\mapsto 0} \frac{F(x+\epsilon a) - F(x)}{\epsilon}, \tag{6.1}$$

where we assume that this limit exists and is well defined. Suppose that we now define a constant coordinate frame $\{\mathsf{e}_k\}$ with reciprocal frame $\{\mathsf{e}^k\}$. Spatial coordinates are defined by $x^k = \mathsf{e}^k\cdot x$, and the summation convention is assumed except where stated otherwise. The vector derivative can be written

$$\nabla = \sum_k \mathsf{e}^k \frac{\partial}{\partial x^k} = \mathsf{e}^k \partial_k, \tag{6.2}$$

where we introduce the useful abbreviation

$$\partial_i = \frac{\partial}{\partial x^i}. \tag{6.3}$$

The frame decomposition $\nabla = \mathsf{e}^k\partial_k$ shows clearly how the the vector derivative combines the algebraic properties of a vector with the operator properties of the partial derivatives. It is a straightforward exercise to confirm that the definition of ∇ is independent of the choice of frame.

6.1.1 Scalar fields

As a first example, consider the case of a scalar field $\phi(x)$. Acting on ϕ, the vector derivative ∇ returns the *gradient*, $\nabla\phi$. This is the familiar grad operation. The result is a vector whose components in the $\{\mathsf{e}^k\}$ frame are the partial derivatives with respect to the x^k coordinates. The simplest example of a scalar field is the quantity $a\cdot x$, where a is a constant vector. We write $a\cdot x = x^j a_j$, so that the gradient becomes

$$\nabla(x\cdot a) = \mathsf{e}^i \frac{\partial x^j}{\partial x^i} a_j = \mathsf{e}^i a_j \delta^j_i. \tag{6.4}$$

But the right-hand side simply expresses the vector a in the $\{\mathsf{e}^k\}$ frame, so we are left with the frame-free result

$$\nabla(x\cdot a) = a. \tag{6.5}$$

This result is independent of both the dimensions and signature of the vector space. Many formulae for the vector derivative can be built up by combining this primitive result with the chain and product rules for differentiation. A particular application of this result is to the coordinates themselves,

$$\nabla x^k = \nabla(x \cdot \mathsf{e}^k) = \mathsf{e}^k, \tag{6.6}$$

a formula which generalises to curvilinear coordinate systems.

As a second example, consider the derivative of the scalar x^2. We first derive the result in coordinates before discussing a more elegant, frame-free derivation. We form

$$\begin{aligned}\nabla(x^2) &= \mathsf{e}^i \partial_i (x^j x^k) \mathsf{e}_j \cdot \mathsf{e}_k \\ &= \mathsf{e}^i \left(\frac{\partial x^j}{\partial x^i} x^k + \frac{\partial x^k}{\partial x^i} x^j \right) \mathsf{e}_j \cdot \mathsf{e}_k \\ &= x^k \mathsf{e}_k + x^j \mathsf{e}_j \\ &= 2x,\end{aligned} \tag{6.7}$$

which recovers the expected result. It is extremely useful to be able to perform such manipulations without reference to any coordinate frame. This requires a notation to keep track of which terms are being differentiated in a given expression. A suitable convention is to use overdots to define the scope of the vector derivative. With this notation we can write

$$\nabla(x^2) = \dot{\nabla}(\dot{x} \cdot x) + \dot{\nabla}(x \cdot \dot{x}) = 2\dot{\nabla}(\dot{x} \cdot x). \tag{6.8}$$

In the final term it is only the first factor of x which is differentiated, while the second is held constant. We can therefore apply the result of equation (6.5), which immediately gives $\nabla(x^2) = 2x$. More complex results can be built up in a similar manner.

In Euclidean spaces $\nabla\phi$ points in the direction of steepest increase of ϕ. This is illustrated in equation (6.5). To get the biggest increase in $a \cdot x$ for a given step size you must clearly move in the positive a direction, since moving in any orthogonal direction does not change the value. More generally, suppose $\nabla\phi = J$ and consider the contraction of this equation with the unit vector n,

$$n \cdot \nabla\phi = n \cdot J. \tag{6.9}$$

We seek the direction of n which maximises this value. Clearly in a Euclidean space this must be the J direction, so J points in the direction of greatest increase of ϕ. Also, setting n in the J direction shows that the magnitude of J is simply the derivative in the direction of steepest increase.

In mixed signature spaces, such as spacetime, this simple geometric picture can break down. As a simple example, consider a timelike plane defined by orthogonal basis vectors $\{\gamma_0, \gamma_1\}$, with $\gamma_0^2 = 1$ and $\gamma_1^2 = -1$. We introduce the scalar field

$$\phi = \langle x\gamma_0 x\gamma_0 \rangle = (x^0)^2 + (x^1)^2. \tag{6.10}$$

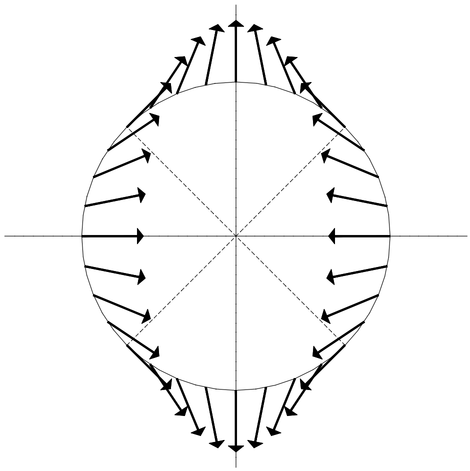


Figure 6.1 *Spacetime gradients.* The contours of the scalar field $\phi = \langle x\gamma_0 x\gamma_0\rangle$ define circles in spacetime. But the direction of the vector derivative is only in the outward normal direction along the 0 axis. Along the 1 axis the gradient points inwards, which reflects the opposite signature. Around the circle the gradient interpolates between these two extremes. At points where x is null the gradient vector is tangential to the circle.

Contours of constant ϕ are circles in the spacetime plane, so the direction of steepest increase points radially outwards. But if we form the gradient of ϕ we obtain

$$\nabla\phi = 2\dot{\nabla}\langle \dot{x}\gamma_0 x\gamma_0\rangle = 2\gamma_0 x\gamma_0. \tag{6.11}$$

Figure 6.1 shows the direction of this vector for various points on the unit circle. Clearly the vector does not point in the direction of steepest increase of ϕ. Instead, $\nabla\phi$ points in a direction 'normal' to tangent vectors in the circle. In mixed signature spaces, the 'normal' does not point in the direction our Euclidean intuition is used to. This example should be borne in mind when we consider directed integration in spaces of mixed signature. (This example may appear esoteric, but closed spacetime curves of this type are of considerable importance in some modern attempts to construct a quantum theory of gravity.)

6.1.2 Vector fields

Suppose now that we have a vector field $J(x)$. The full vector derivative ∇J contains two terms, a scalar and a bivector. The scalar term is the *divergence* of

$J(x)$. In terms of the constant frame vectors $\{\mathsf{e}_k\}$ we can write

$$\nabla \cdot J = \frac{\partial}{\partial x^k} e^k \cdot J = \frac{\partial J^k}{\partial x^k} = \partial_k J^k. \tag{6.12}$$

The divergence can also be defined in terms of the geometric product as

$$\nabla \cdot J = \tfrac{1}{2}(\nabla J + \dot{J}\dot{\nabla}). \tag{6.13}$$

The simplest example of the divergence is for the vector x itself, for which we find

$$\nabla \cdot x = \frac{\partial x^k}{\partial x^k} = n, \tag{6.14}$$

where n is the dimension of the space.

The remaining, antisymmetric, term defines the exterior derivative of the vector field. In terms of coordinates this can be written

$$\nabla \wedge J = \mathsf{e}^i \wedge (\partial_i J) = \mathsf{e}^i \wedge \mathsf{e}^j \, \partial_i J_j. \tag{6.15}$$

The components are the antisymmetrised terms in $\partial_i J_j$. In three dimensions these are the components of the curl, though $\boldsymbol{\nabla} \wedge \boldsymbol{J}$ is a bivector, rather than an (axial) vector. (In this chapter we write vectors in two and three dimensions in bold face.) The three-dimensional curl requires a duality operation to return a vector,

$$\mathrm{curl}(\boldsymbol{J}) = -I \, \boldsymbol{\nabla} \wedge \boldsymbol{J}. \tag{6.16}$$

The exterior derivative generalises the curl to arbitrary dimensions.

As an example, consider the exterior derivative of the position vector x. We find that

$$\nabla \wedge x = \mathsf{e}^i \wedge \mathsf{e}_i = \mathsf{e}^i \wedge \mathsf{e}^j \, (\mathsf{e}_i \cdot \mathsf{e}_j) = 0, \tag{6.17}$$

which follows because $\mathsf{e}^i \wedge \mathsf{e}^j$ is antisymmetric on i and j, whereas $\mathsf{e}_i \cdot \mathsf{e}_j$ is symmetric. Again, we can give an algebraic definition of the exterior derivative in terms of the geometric product as

$$\nabla \wedge J = \tfrac{1}{2}(\nabla J - \dot{J}\dot{\nabla}). \tag{6.18}$$

Equations (6.13) and (6.18) combine to give the familiar decomposition of a geometric product:

$$\nabla J = \nabla \cdot J + \nabla \wedge J. \tag{6.19}$$

So, for example, we have $\nabla x = n$.

6.1.3 Multivector fields

The preceding definitions extend simply to the case of the vector derivative acting on a multivector field. We have

$$\nabla A = \mathsf{e}^k \partial_k A, \tag{6.20}$$

and for an r-grade multivector field A_r we write

$$\nabla \cdot A_r = \langle \nabla A_r \rangle_{r-1}, \tag{6.21}$$

$$\nabla \wedge A_r = \langle \nabla A_r \rangle_{r+1}. \tag{6.22}$$

These define the interior and exterior derivatives respectively. The interior derivative is often referred to as the divergence, and the exterior derivative is sometimes called the curl. This latter name conflicts with the more familiar meaning of 'curl' in three dimensions, however, and we will avoid this name where possible.

An important result for the vector derivative is that the exterior derivative of an exterior derivative always vanishes,

$$\begin{aligned}\nabla \wedge (\nabla \wedge A) &= \mathsf{e}^i \wedge \partial_i (\mathsf{e}^j \wedge \partial_j A) \\ &= \mathsf{e}^i \wedge \mathsf{e}^j \wedge (\partial_i \partial_j A) = 0.\end{aligned} \tag{6.23}$$

This follows because $\mathsf{e}^i \wedge \mathsf{e}^j$ is antisymmetric on i, j, whereas $\partial_i \partial_j A$ is symmetric, due to the fact that partial derivatives commute. Similarly, the divergence of a divergence vanishes,

$$\nabla \cdot (\nabla \cdot A) = 0, \tag{6.24}$$

which is proved in the same way, or by using duality. (By convention, the inner product of a vector and a scalar is zero.)

Because ∇ is a vector, it does not necessarily commute with other multivectors. We therefore need to be careful in describing the scope of the operator. We use the following series of conventions to clarify the scope:

(i) In the absence of brackets, ∇ acts on the object to its immediate right.
(ii) When the ∇ is followed by brackets, the derivative acts on all of the terms in the brackets.
(iii) When the ∇ acts on a multivector to which it is not adjacent, we use overdots to describe the scope.

The 'overdot' notation was introduced in the previous section, and is invaluable when differentiating products of multivectors. For example, with this notation we can write

$$\nabla(AB) = \nabla AB + \dot{\nabla} A \dot{B}, \tag{6.25}$$

which encodes a version of the product rule. If necessary, the overdots can be replaced with partial derivatives by writing

$$\dot{\nabla} A \dot{B} = \mathsf{e}^k A\, \partial_k B. \tag{6.26}$$

Later in this chapter we also employ the overdot notation for linear functions. Suppose that $\mathsf{f}(a)$ is a position-dependent linear function. We write

$$\dot{\nabla} \dot{\mathsf{f}}(a) = \nabla \mathsf{f}(a) - \mathsf{e}^k \mathsf{f}(\partial_k a), \tag{6.27}$$

so that $\dot{\nabla}\dot{\mathsf{f}}(a)$ only differentiates the position dependence in the linear function, and not in its argument.

We can continue to build up a series of useful basic results by differentiating various multivectors that depend linearly on x. For example, consider

$$\nabla\, x \cdot A_r = \mathsf{e}^k\, \mathsf{e}_k \cdot A_r, \tag{6.28}$$

where A_r is a grade-r multivector. Using the results of section 4.3.2 we find that

$$\begin{aligned} \nabla\, x \cdot A_r &= r A_r, \\ \nabla\, x \wedge A_r &= (n-r) A_r, \\ \dot{\nabla} A_r \dot{x} &= (-1)^r (n - 2r) A_r, \end{aligned} \tag{6.29}$$

where n is the dimension of the space.

6.2 Curvilinear coordinates

So far we have only expressed the vector derivative in terms of a fixed coordinate frame (which is usually chosen to be orthonormal). In many applications, however, it is more convenient to work in a *curvilinear* coordinate system, where the frame vectors vary from point to point. A general set of coordinates consist of a set of scalar functions $\{x^i(x)\}$, $i = 1, \ldots, n$, defined over some region. In this region we can equally write $x(x^i)$, expressing the position vector x parametrically in terms of the coordinates. If one of the coordinates is varied and all of the others are held fixed we specify an associated coordinate curve. The derivatives along these curves specify a set of frame vectors by

$$\mathsf{e}_i(x) = \frac{\partial x}{\partial x^i} = \lim_{\epsilon \mapsto 0} \frac{x(x^1, \ldots, x^i + \epsilon, \ldots, x^n) - x}{\epsilon}, \tag{6.30}$$

where the ith coordinate is varied and all others are held fixed. The derivative in the e_i direction, $\mathsf{e}_i \cdot \nabla$, is found by moving a small amount along e_i. But this is precisely the same as varying the x^i coordinate with all others held fixed. We therefore have

$$\mathsf{e}_i \cdot \nabla = \frac{\partial}{\partial x^i} = \partial_i. \tag{6.31}$$

In order that the coordinate system be valid over a given region we require that throughout this region

$$\mathsf{e}_1\wedge\mathsf{e}_2\wedge\cdots\wedge\mathsf{e}_n \neq 0. \tag{6.32}$$

As this quantity can never pass through zero it follows that the frame has the same orientation throughout the valid region.

We can construct a second frame directly from the coordinate functions by defining

$$\mathsf{e}^i = \nabla x^i. \tag{6.33}$$

From their construction we see that the $\{\mathsf{e}^i\}$ vectors have vanishing exterior derivative:

$$\nabla\wedge\mathsf{e}^i = \nabla\wedge(\nabla x^i) = 0. \tag{6.34}$$

As the notation suggests, the two frames defined above are reciprocal to one another. This is straightforward to check:

$$\mathsf{e}_i\cdot\mathsf{e}^j = \mathsf{e}_i\cdot\nabla x^j = \frac{\partial x^j}{\partial x^i} = \delta^j_i. \tag{6.35}$$

This result is very useful because, when working with curvilinear coordinates, one usually has simple expressions for either $x^i(x)$ or $x(x^i)$, but rarely both. Fortunately, only one is needed to construct a set of frame vectors, and the reciprocal frame can then be constructed algebraically (see section 4.3). This construction provides a simple geometric picture for the gradient in a general space. Suppose we view the coordinate $x^1(x)$ as a scalar field. The contours of constant x^1 are a set of $(n-1)$-dimensional surfaces. The remaining coordinates $x^2, \ldots, x^n$ define a set of directions in this surface. At each point on the surface of constant x^1 the vector ∇x^1 is orthogonal to all of the directions in the surface. In Euclidean spaces this vector is necessarily orthogonal (normal) to the surface. In other spaces this construct defines what we mean by normal.

Now suppose we have a function $F(x)$ that is expressed in terms of the coordinates as $F(x^i)$. A simple application of the chain rule gives

$$\nabla F = \nabla x^i\, \partial_i F = \mathsf{e}^i\partial_i F. \tag{6.36}$$

This is consistent with the decomposition

$$\nabla = \mathsf{e}^i\frac{\partial}{\partial x^i} = \mathsf{e}^i\partial_i = \mathsf{e}^i\mathsf{e}_i\cdot\nabla, \tag{6.37}$$

which holds as the $\{\mathsf{e}_i\}$ and $\{\mathsf{e}^i\}$ are reciprocal frames.

6.2.1 Tensor analysis

A consequence of curvilinear frame vectors is that one has to be careful when working entirely in terms of coordinates, as is the case in tensor analysis. The problem is that for a vector, for example, we have $J = J^i \mathsf{e}_i$. If we just keep the coordinates J^i we lose the information about the position dependence in the coordinate frame. When formulating the derivative of J in tensor analysis we must introduce *connection coefficients* to keep track of the derivatives of the frame vectors. This can often complicate derivations.

There are two cases of the vector derivative in curvilinear coordinates that do not require connection coefficients. The first is the exterior derivative, for which we can write

$$\nabla\wedge J = \nabla\wedge(J_i \mathsf{e}^i) = (\nabla J_i)\wedge \mathsf{e}^i. \tag{6.38}$$

It follows that the exterior derivative has coordinates $\partial_i J_j - \partial_j J_i$ regardless of chosen coordinate system. The second exception is provided by the divergence of a vector. We have

$$\nabla\cdot J = \nabla\cdot(J^i \mathsf{e}_i). \tag{6.39}$$

If we define the volume factor V by

$$\mathsf{e}_1\wedge\mathsf{e}_2\wedge\cdots\wedge\mathsf{e}_n = IV, \tag{6.40}$$

where I is the unit pseudoscalar, we can write (following section 4.3)

$$\mathsf{e}_i = (-1)^{i-1}\mathsf{e}^n\wedge\mathsf{e}^{n-1}\wedge\cdots\wedge\check{\mathsf{e}}^i\wedge\cdots\wedge\mathsf{e}^1\, IV. \tag{6.41}$$

Recalling that each of the e^i vectors has vanishing exterior derivative, one can quickly establish that

$$\nabla\cdot J = \frac{1}{V}\frac{\partial}{\partial x^i}\big(VJ^i\big). \tag{6.42}$$

Similarly, the Laplacian ∇^2 can be written as

$$\nabla^2\phi = \frac{1}{V}\frac{\partial}{\partial x^i}\left(Vg^{ij}\frac{\partial\phi}{\partial x^j}\right), \tag{6.43}$$

where $g^{ij} = \mathsf{e}^i\cdot\mathsf{e}^j$.

6.2.2 Orthogonal coordinates in three dimensions

A number of the most useful coordinate systems are orthogonal systems of coordinates in three dimensions. For these systems a number of special results hold. We define a set of orthonormal vectors by first introducing the magnitudes

$$h_i = |\mathsf{e}_i| = (\mathsf{e}_i\cdot\mathsf{e}_i)^{1/2}. \tag{6.44}$$

In terms of these we can write (no sums implied)

$$\mathsf{e}_i = h_i\hat{\mathsf{e}}_i, \quad \mathsf{e}^i = \frac{1}{h_i}\hat{\mathsf{e}}_i. \tag{6.45}$$

We now use the $\{\hat{\mathsf{e}}_i\}$ as our coordinate frame and, since this frame is orthonormal, we can work entirely with lowered indices. For a vector $\boldsymbol{J}$ we have

$$\boldsymbol{J} = J_i \hat{\mathsf{e}}_i = \sum_{i=1}^{3} \frac{J_i}{h_i} \mathsf{e}_i. \tag{6.46}$$

It follows that we can write

$$\boldsymbol{\nabla} \cdot \boldsymbol{J} = \frac{1}{h_1 h_2 h_3} \left(\frac{\partial}{\partial x_1}(h_2 h_3 J_1) + \frac{\partial}{\partial x_2}(h_3 h_1 J_2) + \frac{\partial}{\partial x_3}(h_1 h_2 J_3) \right). \tag{6.47}$$

A compact formula for the Laplacian is obtained by replacing each J_i term with $1/h_i \, \partial_i \phi$,

$$\begin{aligned} \boldsymbol{\nabla}^2 \phi = &\frac{1}{h_1 h_2 h_3} \left(\frac{\partial}{\partial x_1} \left(\frac{h_2 h_3}{h_1} \frac{\partial \phi}{\partial x_1} \right) + \frac{\partial}{\partial x_2} \left(\frac{h_3 h_1}{h_2} \frac{\partial \phi}{\partial x_2} \right) \right. \\ &\left. + \frac{\partial}{\partial x_3} \left(\frac{h_1 h_2}{h_3} \frac{\partial \phi}{\partial x_3} \right) \right). \end{aligned} \tag{6.48}$$

The components of the curl can be found in a similar manner. A number of useful curvilinear coordinate systems are summarised below.

Cartesian coordinates

These are the basic starting point for all other coordinate systems. We introduce a constant, right-handed orthonormal frame $\{\boldsymbol{\sigma}_i\}$, $\boldsymbol{\sigma}_1 \boldsymbol{\sigma}_2 \boldsymbol{\sigma}_3 = I$. This notation for a Cartesian frame is borrowed from quantum theory and is very useful in practice. The coordinates in the $\{\boldsymbol{\sigma}_i\}$ frame are written, following standard notation, as (x, y, z). To avoid confusion between the scalar coordinate x and the three-dimensional position vector we write the latter as $\boldsymbol{r}$. That is,

$$\boldsymbol{r} = x\boldsymbol{\sigma}_1 + y\boldsymbol{\sigma}_2 + z\boldsymbol{\sigma}_3. \tag{6.49}$$

Since the frame vectors are orthonormal we have $h_1 = h_2 = h_3 = 1$, so the divergence and Laplacian take on their simplest forms.

Cylindrical polar coordinates

These are denoted (ρ, ϕ, z) with ρ and ϕ the standard two-dimensional polar coordinates

$$\rho = \left(x^2 + y^2\right)^{1/2}, \qquad \tan\phi = \frac{y}{x}. \tag{6.50}$$

The coordinates lie in the ranges $0 \leq r < \infty$ and $0 \leq \phi < 2\pi$. The coordinate vectors are

$$\begin{aligned} \hat{\mathsf{e}}_\rho &= \cos(\phi)\, \boldsymbol{\sigma}_1 + \sin(\phi)\, \boldsymbol{\sigma}_2, \\ \hat{\mathsf{e}}_\phi &= -\sin(\phi)\, \boldsymbol{\sigma}_1 + \cos(\phi)\, \boldsymbol{\sigma}_2, \\ \hat{\mathsf{e}}_z &= \boldsymbol{\sigma}_3. \end{aligned} \tag{6.51}$$

We have adopted the common convention of labelling the frame vectors with the associated coordinate. The magnitudes are $h_\rho = 1$, $h_\phi = \rho$ and $h_z = 1$, and the frame vectors satisfy

$$\hat{\mathsf{e}}_\rho \hat{\mathsf{e}}_\phi \hat{\mathsf{e}}_z = \boldsymbol{\sigma}_1\boldsymbol{\sigma}_2\boldsymbol{\sigma}_3 = I \tag{6.52}$$

and so form a right-handed set in the order (ρ, ϕ, z).

Spherical polar coordinates

Spherical polar coordinates arise in many problems in physics, particularly quantum mechanics and field theory. They are typically labelled (r, θ, ϕ) and are defined by

$$r = |\boldsymbol{r}| = (\boldsymbol{r}\cdot\boldsymbol{r})^{1/2}, \quad r\cos(\theta) = z, \quad \tan(\phi) = \frac{y}{x}. \tag{6.53}$$

The coordinate ranges are $0 \leq r < \infty$, $0 \leq \theta \leq \pi$ and $0 \leq \phi < 2\pi$. The ϕ coordinate is ill defined along the z axis — a reflection of the fact that it is impossible to construct a global coordinate system over the surface of a sphere. The inverse relation giving $\boldsymbol{r}(r, \theta, \phi)$ is often useful,

$$\boldsymbol{r} = r\sin(\theta)(\cos(\phi)\,\boldsymbol{\sigma}_1 + \sin(\phi)\,\boldsymbol{\sigma}_2) + r\cos(\theta)\,\boldsymbol{\sigma}_3. \tag{6.54}$$

This expression makes it a straightforward exercise to compute the orthonormal frame vectors, which are

$$\begin{aligned}
\hat{\mathsf{e}}_r &= \sin(\theta)(\cos(\phi)\,\boldsymbol{\sigma}_1 + \sin(\phi)\,\boldsymbol{\sigma}_2) + \cos(\theta)\,\boldsymbol{\sigma}_3 = r^{-1}\boldsymbol{r}, \\
\hat{\mathsf{e}}_\theta &= \cos(\theta)(\cos(\phi)\,\boldsymbol{\sigma}_1 + \sin(\phi)\,\boldsymbol{\sigma}_2) - \sin(\theta)\,\boldsymbol{\sigma}_3, \\
\hat{\mathsf{e}}_\phi &= -\sin(\phi)\,\boldsymbol{\sigma}_1 + \cos(\phi)\,\boldsymbol{\sigma}_2.
\end{aligned} \tag{6.55}$$

The associated normalisation factors are

$$h_r = 1, \quad h_\theta = r, \quad h_\phi = r\sin(\theta). \tag{6.56}$$

The orthonormal vectors satisfy $\hat{\mathsf{e}}_r\hat{\mathsf{e}}_\theta\hat{\mathsf{e}}_\phi = I$ so that $\{\hat{\mathsf{e}}_r, \hat{\mathsf{e}}_\theta, \hat{\mathsf{e}}_\phi\}$ form a right-handed orthonormal frame. This frame can be obtained from the $\{\mathsf{e}_i\}$ frame through the application of a position-dependent rotor, so that $\hat{\mathsf{e}}_r = R\boldsymbol{\sigma}_3\tilde{R}$, $\hat{\mathsf{e}}_\theta = R\boldsymbol{\sigma}_1\tilde{R}$ and $\hat{\mathsf{e}}_\phi = R\boldsymbol{\sigma}_2\tilde{R}$. The rotor is then given by

$$R = \exp(-I\boldsymbol{\sigma}_3\phi/2)\exp(-I\boldsymbol{\sigma}_2\theta/2). \tag{6.57}$$

Spheroidal coordinates

These coordinates turn out to be useful in a number of problems in gravitation and electromagnetism involving rotating sources. We introduce a vector $\boldsymbol{a}$, so that $\pm\boldsymbol{a}$ denote the foci of a family of ellipses. The distances from the foci are given by

$$r_1 = |\boldsymbol{r} + \boldsymbol{a}|, \quad r_2 = |\boldsymbol{r} - \boldsymbol{a}|. \tag{6.58}$$

From these we define the orthogonal coordinates

$$u = \tfrac{1}{2}(r_1 + r_2), \quad v = \tfrac{1}{2}(r_1 - r_2). \tag{6.59}$$

The coordinate system is completed by rotating the ellipses around the $\boldsymbol{a}$ axis. This defines an oblate spheroidal coordinate system. Prolate spheroidal coordinates are formed by starting in a plane, defining (u_1, u_2) as above, and rotating this system around the minor axis.

If we define

$$\hat{\boldsymbol{r}}_1 = \frac{\boldsymbol{r} + \boldsymbol{a}}{r_1}, \quad \hat{\boldsymbol{r}}_2 = \frac{\boldsymbol{r} - \boldsymbol{a}}{r_2}, \tag{6.60}$$

we see that

$$\mathsf{e}^u = \tfrac{1}{2}(\hat{\boldsymbol{r}}_1 + \hat{\boldsymbol{r}}_2), \quad \mathsf{e}^v = \tfrac{1}{2}(\hat{\boldsymbol{r}}_1 - \hat{\boldsymbol{r}}_2), \tag{6.61}$$

which are clearly orthogonal. The normalisation factors are found from

$$h_u^2 = \frac{u^2 - v^2}{u^2 - a^2}, \quad h_v^2 = \frac{u^2 - v^2}{a^2 - v^2}. \tag{6.62}$$

If we align $\boldsymbol{a}$ with the 3 axis and let ϕ take its spherical-polar meaning, the coordinate frame is completed with the vector $\hat{\mathsf{e}}_\phi$, and

$$h_\phi^2 = (u^2 - a^2)(a^2 - v^2). \tag{6.63}$$

The frame vectors satisfy $\hat{\mathsf{e}}_u \hat{\mathsf{e}}_\phi \hat{\mathsf{e}}_v = I$. The hyperbolic nature of the coordinate system is often best expressed by redefining the u and v coordinates as $a\cosh(w)$ and $a\cos(\vartheta)$ respectively.

6.3 Analytic functions

The vector derivative combines the algebraic properties of geometric algebra with vector calculus in a simple and natural way. In this section we show how the vector derivative can be used to extend the definition of an analytic function to arbitrary dimensions. We start by considering the vector derivative in two dimensions to establish the link with complex analysis.

6.3.1 Analytic functions in two dimensions

Suppose that $\{\mathsf{e}_1, \mathsf{e}_2\}$ define an orthonormal frame in two dimensions. This is identified with the Argand plane by singling out e_1 as the real axis. We denote coordinates by (x, y) and write the position vector as $\boldsymbol{r}$:

$$\boldsymbol{r} = x\mathsf{e}_1 + y\mathsf{e}_2. \tag{6.64}$$

With this notation the vector derivative is

$$\boldsymbol{\nabla} = \mathsf{e}_1 \frac{\partial}{\partial x} + \mathsf{e}_2 \frac{\partial}{\partial y}. \tag{6.65}$$

In section 2.3.3 we showed that complex numbers sit naturally within the geometric algebra of the plane. The pseudoscalar is the bivector $I = \mathsf{e}_1\mathsf{e}_2$, which satisfies $I^2 = -1$. Complex numbers therefore map directly onto even-grade elements in the algebra by identifying the unit imaginary i with I. The position vector $\boldsymbol{r}$ is mapped onto a complex number by pre-multiplying by the vector representing the real axis:

$$z = x + Iy = \mathsf{e}_1\boldsymbol{r}. \tag{6.66}$$

Now suppose we introduce the complex field $\psi = u + Iv$. The vector derivative applied to ψ yields

$$\boldsymbol{\nabla}\psi = \left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right)\mathsf{e}_1 + \left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right)\mathsf{e}_2. \tag{6.67}$$

The terms in brackets are precisely the ones that vanish in the Cauchy–Riemann equations. The statement that ψ is an *analytic function* (a function that satisfies the Cauchy–Riemann equations) reduces to the equation

$$\boldsymbol{\nabla}\psi = 0. \tag{6.68}$$

This is the fundamental equation which can be generalised immediately to higher dimensions. These generalisations invariably turn out to be of mathematical and physical importance, and it is is no exaggeration to say that equations of the type of equation (6.68) are amongst the most studied in physics.

To complete the link with complex analysis we recall that the complex partial derivative ∂_z is defined by the properties

$$\frac{\partial z}{\partial z} = 1, \qquad \frac{\partial z^\dagger}{\partial z} = 0 \tag{6.69}$$

with the complex conjugate satisfying

$$\frac{\partial z}{\partial z^\dagger} = 0, \qquad \frac{\partial z^\dagger}{\partial z^\dagger} = 1. \tag{6.70}$$

From these we see that

$$\frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} - I\frac{\partial}{\partial y}\right), \qquad \frac{\partial}{\partial z^\dagger} = \frac{1}{2}\left(\frac{\partial}{\partial x} + I\frac{\partial}{\partial y}\right). \tag{6.71}$$

An analytic function is one that depends on z alone. That is, we can write $\psi(x + Iy) = \psi(z)$. The function is therefore independent of $z^\dagger$, and we have

$$\frac{\partial\psi(z)}{\partial z^\dagger} = 0. \tag{6.72}$$

This summarises the content of the Cauchy–Riemann equations, though this fact is often obscured by the complex limiting argument favoured in many textbooks. Comparing the preceding forms, we see that this equation is equivalent to

$$\frac{1}{2}\left(\frac{\partial}{\partial x} + I\frac{\partial}{\partial y}\right)\psi = \tfrac{1}{2}\mathsf{e}_1\boldsymbol{\nabla}\psi = 0, \tag{6.73}$$

recovering our earlier equation.

It is instructive to see why solutions to $\boldsymbol{\nabla}\psi = 0$ can be constructed as power series in z. We first see that

$$\boldsymbol{\nabla} z = \boldsymbol{\nabla}(\mathsf{e}_1 \boldsymbol{r}) = 2\mathsf{e}_1 \cdot \boldsymbol{\nabla} \boldsymbol{r} - \mathsf{e}_1 \boldsymbol{\nabla} \boldsymbol{r} = 2\mathsf{e}_1 - 2\mathsf{e}_1 = 0. \tag{6.74}$$

This little manipulation drives most of analytic function theory! It follows immediately, for example, that

$$\boldsymbol{\nabla}(z - z_0)^n = n\boldsymbol{\nabla}(\mathsf{e}_1 \boldsymbol{r} - z_0)(z - z_0)^{n-1} = 0, \tag{6.75}$$

so a Taylor series expansion in z about z_0 automatically returns an analytic function. We will delay looking at poles until we have introduced the subject of directed integration.

6.3.2 Generalized analytic functions

There are two problems with the standard presentation of complex analytic function theory that prevent a natural generalisation to higher dimensions:

(i) Both the vector operator $\boldsymbol{\nabla}$ and the functions it operates on are mapped into the same algebra by picking out a preferred direction for the real axis. This only works in two dimensions.

(ii) The 'complex limit' argument does not generalise to higher dimensions. Indeed, one can argue that it is not wholly satisfactory in two dimensions, as it confuses the concept of a directional derivative with the concept of being independent of $z^\dagger$.

These problems are solved by keeping the derivative operator $\boldsymbol{\nabla}$ as a vector, while letting it act on general multivectors. The analytic requirement is then replaced with the equation $\boldsymbol{\nabla}\psi = 0$. Functions satisfying this equation are said to be *monogenic*. If ψ contains all grades it is clear that both the even-grade and odd-grade components must satisfy this equation independently. Without loss of generality, we can therefore assume that ψ has even grade.

We can construct monogenic functions by following the route which led to the conclusion that z is analytic in two dimensions. We recall that $\boldsymbol{\nabla}\boldsymbol{r} = 3$ and

$$\boldsymbol{\nabla}(\boldsymbol{a}\boldsymbol{r}) = -\boldsymbol{a}. \tag{6.76}$$

It follows that

$$\psi = \boldsymbol{r}\boldsymbol{a} + 3\boldsymbol{a}\boldsymbol{r} \tag{6.77}$$

is a monogenic for any constant vector $\boldsymbol{a}$. The main difference with complex analysis is that we cannot derive new monogenics simply from power series in this solution, due to the lack of commutativity. One can construct monogenic functions from series of geometric products, but a more instructive route is to classify monogenics via their angular properties.

First we assume that Ψ is a monogenic containing terms which scale uniformly with $\boldsymbol{r}$. If we introduce polar coordinates we can then write

$$\Psi(\boldsymbol{r}) = r^l \psi(\theta, \phi). \tag{6.78}$$

The function $\psi(\theta, \phi)$ then satisfies

$$l r^{l-1} \mathsf{e}_r \psi + r^l \boldsymbol{\nabla} \psi(\theta, \phi) = 0. \tag{6.79}$$

It follows that ψ satisfies the angular eigenvalue equation

$$-\boldsymbol{r} \wedge \boldsymbol{\nabla} \psi = l \psi. \tag{6.80}$$

These angular eigenstates play a key role in the Pauli and Dirac theories of the electron. Since Ψ satisfies $\boldsymbol{\nabla}\Psi = 0$, it follows that

$$\boldsymbol{\nabla}^2 \Psi = 0. \tag{6.81}$$

So each component of Ψ (in a constant basis) satisfies Laplace's equation. It follows that each component of ψ is a spherical harmonic, and hence that l is an integer. We can construct a monogenic by starting with the function $(x + yI\boldsymbol{\sigma}_3)^l$, which is the three-dimensional extension of the complex analytic function z^l. In terms of polar coordinates

$$(x + yI\boldsymbol{\sigma}_3)^l = r^l \sin^l(\theta)\, \mathrm{e}^{l\phi I \boldsymbol{\sigma}_3}, \tag{6.82}$$

which gives us our first angular monogenic function

$$\psi_l^l = \sin^l(\theta)\, \mathrm{e}^{l\phi I \boldsymbol{\sigma}_3}. \tag{6.83}$$

The remaining monogenic functions are constructed from this by acting with an operator which, in quantum terms, lowers the eigenvalue of the angular momentum around the z axis. These are discussed in more detail in section 8.4.1.

6.3.3 The spacetime vector derivative

To construct the vector derivative in spacetime suppose that we introduce the orthonormal frame $\{\gamma_\mu\}$ with associated coordinates x^μ. We can then write

$$\nabla = \gamma^\mu \frac{\partial}{\partial x^\mu} = \gamma_0 \frac{\partial}{\partial t} + \gamma^i \frac{\partial}{\partial x^i}. \tag{6.84}$$

This derivative is the key operator in all relativistic field theories, including electromagnetism and Dirac theory. If we post-multiply by γ_0 we see that

$$\nabla \gamma_0 = \partial_t + \gamma^i \gamma_0 \partial_i = \partial_t - \boldsymbol{\nabla}, \tag{6.85}$$

where $\boldsymbol{\nabla} = \boldsymbol{\sigma}_i \partial_i$ is the vector derivative in the relative space defined by the γ_0 vector. Similarly,

$$\gamma_0 \nabla = \partial_t + \boldsymbol{\nabla}. \tag{6.86}$$

These equations are consistent with

$$\nabla x = \nabla(\gamma_0 \gamma_0 x) = (\partial_t - \boldsymbol{\nabla})(t - \boldsymbol{r}) = 4, \tag{6.87}$$

where x is the spacetime position vector. The spacetime vector derivative satisfies

$$\nabla^2 = \frac{\partial^2}{\partial t^2} - \boldsymbol{\nabla}^2, \tag{6.88}$$

which is the fundamental operator describing waves travelling at the speed of light. The spacetime monogenic equation $\nabla \psi = 0$ is discussed in detail in chapters 7 and 8. We only note here that, if ψ is an even-grade element of the spacetime algebra, the monogenic equation is precisely the wave equation for a massless spin-1/2 particle.

6.3.4 Characteristic surfaces and propagation

The fact that ∇^2 can give rise to either elliptic or hyperbolic operators, depending on signature, suggests that the propagator theory for ∇ will depend strongly on the signature. This is confirmed by a simple argument which can be modified to apply to most first-order differential equations. Suppose we have a generic equation of the type

$$\nabla \psi = f(\psi, x), \tag{6.89}$$

where ψ is some multivector field, $f(\psi, x)$ is a known function and x is the position vector in an n-dimensional space. We are presented with data on some $(n-1)$-dimensional surface, and wish to propagate these initial conditions away from the surface. If surfaces exist for which this is not possible they are known as *characteristic surfaces*. Suppose that we construct a set of independent tangent vectors in the surface, $\{e_1, \ldots, e_{n-1}\}$. Knowledge of ψ on the surface enables us to calculate each of the directional derivatives $e_i \cdot \nabla \psi$, $i = 1, \ldots, n-1$. We now form the normal vector

$$n = I\, e_1 \wedge e_2 \wedge \cdots \wedge e_{n-1}, \tag{6.90}$$

where I is the pseudoscalar for the space. Pre-multiplying equation (6.89) with n we obtain

$$n \cdot \nabla \psi = -n \wedge \nabla \psi + n f(\psi, x). \tag{6.91}$$

But we have

$$n \wedge \nabla\psi = I(e_1 \wedge e_2 \wedge \cdots \wedge e_{n-1}) \cdot \nabla\psi$$
$$= I \sum_{i=1}^{n-1} (-1)^{i+1-n} (e_1 \wedge \cdots \wedge \check{e}_i \wedge \cdots \wedge e_{n-1})\, e_i \cdot \nabla\psi, \tag{6.92}$$

which is constructed entirely from known derivatives of ψ. Equation (6.91) then tells us how to propagate ψ in the n direction. The only situation in which we can fail to propagate ψ is when n still lies in the surface. This happens if n is linearly dependent on the surface tangent vectors. If this is the case we have

$$n \wedge (e_1 \wedge e_2 \wedge \cdots \wedge e_{n-1}) = 0. \tag{6.93}$$

But this implies that

$$(I^{-1} n) \wedge n = I^{-1} n \cdot n = 0. \tag{6.94}$$

We therefore only fail to propagate when $n^2 = 0$, so characteristic surfaces are always null surfaces. This possibility can only arise in mixed signature spaces, and unsurprisingly the propagators in these spaces can have quite different properties to their Euclidean counterparts.

6.4 Directed integration theory

The true power of geometric calculus begins to emerge when we study directed integration theory. This provides a very general and powerful integral theorem which enables us to construct Green's functions for the vector derivative in various spaces. These in turn can be used to generalise the many powerful results from complex function theory to arbitrary spaces.

6.4.1 Line integrals

The simplest integrals to start with are line integrals. The line integral of a multivector field $F(x)$ along a line $x(\lambda)$ is defined by

$$\int F(x) \frac{dx}{d\lambda}\, d\lambda = \int F\, dx = \lim_{n \mapsto \infty} \sum_{i=1}^{n} \bar{F}^i \Delta x^i. \tag{6.95}$$

In the final expression a set of successive points along the curve $\{x_i\}$ are introduced, with x_0 and x_n the endpoints, and

$$\Delta x^i = x_i - x_{i-1}, \quad \bar{F}^i = \tfrac{1}{2}\big(F(x_{i-1}) + F(x_i)\big). \tag{6.96}$$

If the curve is closed then $x_0 = x_n$. The result of the integral is independent of the way we choose to parameterise the curve, provided the parameterisation respects the required ordering of points along the curve. Curves that double back on themselves are handled by referring to the parameterised form $x(\lambda)$, which tells us how the curve is traversed.

The definition of the integral (6.95) looks so standard that it is easy to overlook the key new feature, which is that dx is a *vector-valued measure*, and the product $F\,dx$ is a geometric product between multivectors. This small extension to scalar integration is sufficient to bring a wealth of new features. We refer to dx, and its multivector-valued extensions, as a *directed measure*. The fact that dx is no longer a scalar means that equation (6.95) is not the most general line integral we can form. We can also consider integrals of the form

$$\int F(x)\frac{dx}{d\lambda}G(x)\,d\lambda = \int F(x)\,dx\,G(x), \tag{6.97}$$

and more generally we can consider sums of terms like these. The most general form of line integral can be written

$$\int \mathsf{L}(\partial_\lambda x; x)\,d\lambda = \int \mathsf{L}(dx), \tag{6.98}$$

where $\mathsf{L}(a) = \mathsf{L}(a;x)$ is a multivector-valued linear function of a. The position dependence in L can often be suppressed to streamline the notation.

Suppose now that the field F is replaced by the vector-valued function $v(x)$. We have

$$\int v\,dx = \int v\cdot dx + \int v\wedge dx, \tag{6.99}$$

which separates the directed integral into scalar and bivector-valued terms. If v is the unit tangent vector along the curve then the scalar integral returns the arc length. In many applications the scalar and bivector integrals are considered separately. But to take advantage of the most powerful integral theorems in geometric calculus we need to use the combined form, containing a geometric product with the directed measure.

6.4.2 Surface integrals

The natural extension of a line integral is to a directed surface integral. Suppose now that the the multivector-valued field F is defined over a two-dimensional surface embedded in some larger space. If the surface is parameterised by two coordinates $x(x^1, x^2)$ we define the directed measure by the bivector

$$dX = \frac{\partial x}{\partial x^1}\wedge\frac{\partial x}{\partial x^2}\,dx^1\,dx^2 = e_1\wedge e_2\,dx^1\,dx^2, \tag{6.100}$$

where $e_i = \partial_i x$. This measure is independent of how the surface is parameterised, provided we orient the coordinate vectors in the desired order. Sometimes more than one coordinate patch will be needed to parameterise the entire surface, but