

Physics from geometry

José B. Almeida and Franklin Potter

December 11, 2008

Copyright © José B. Almeida.

Contents

Preface	vii
1 Introduction	1
1.1 What is fundamental in physics?	1
1.2 Linear spaces	7
1.3 The scalar product	9
1.4 Complex numbers	11
1.5 Quaternions	11
1.6 The cross product	12
1.7 The outer product	13
2 Geometric algebra in two and three dimensions	17
2.1 The geometric product of vectors	17
2.2 An introduction to geometric algebra	18
2.3 Geometric algebra of the plane	20
2.4 The geometric algebra of space	23
2.5 Notation conventions	29
2.6 Reflections	33
2.7 Rotations	35
3 Axiomatic approach to geometric algebra	41
3.1 Axiomatic development	41
3.2 Reflections, rotations and boosts	54
3.3 Basis frames and components	56
3.4 Linear algebra	59
3.5 Idempotents and ideals	69
3.6 Tensors and components	71
4 Differential calculus	77
4.1 The vector derivative	77
4.2 Curvilinear coordinates	81
4.3 Analytic and monogenic functions	86
5 Five-dimensional spacetime	91

5.1	The algebra of five-dimensional spacetime	92
5.2	Matrix representation	94
5.3	The null subspace	96
5.4	The Lorentz group	97
5.5	Unitary groups	99
6	Flat space physics	109
6.1	Postulates of special relativity	109
6.2	Wave-particle duality	117
6.3	First order form of the wave equation	120
6.4	Hubble expansion	121
7	Directed integration theory	125
7.1	Line integrals	125
7.2	Surface integrals	126
7.3	n -dimensional surfaces	128
7.4	The fundamental theorem of calculus	129
7.5	The divergence and Green's theorems	134
7.6	Cauchy's integral formula	135
7.7	Green's functions in Euclidean spaces	137
7.8	Spacetime propagators	139
8	Curved manifolds	143
8.1	Geodesics	144
8.2	The pseudoscalar and projection	148
8.3	Directed integration for embedded surfaces	150
8.4	Intrinsic and extrinsic geometry	151
8.5	Coordinates and derivatives	154
8.6	Riemannian geometry	155
9	Electromagnetism	161
9.1	Vacuum equations	161
9.2	Electrodynamics	166
9.3	Electromagnetism in media	167
9.4	The field of a moving charge	168
9.5	Fermat's principle	171
9.6	Optical waveguides	174
9.7	Solitons and photons	175
10	Gravitation	177
10.1	Gravitational dynamics	177

List of Tables

5.1	Matrix representation of basis vectors	95
5.2	Alternative matrix representation of basis vectors	106

List of Tables

List of Figures

1.1	The bivector	14
1.2	The trivector	16
2.1	Vector bivector product	24
2.2	Indexing conventions	31
2.3	A reflection	33
2.4	Rotation from a to b	38
3.1	Outer product	44
5.1	Representation of $SU(4)$	102
7.1	Triangulated surface	127
7.2	Spacetime contour	141
8.1	A geodesic	144

List of Figures

Preface

There is a general consensus among physicists about what is a physical theory. The essential concept is a set of principles or axioms which are unproven statements, whose validity is sustained on the consistency of the whole theory and its ability to make correct predictions. Using standard rules of mathematics and logic it is possible to derive consequences from the set of principles, some of which are observables and can be confronted with experiment and/or observation of the physical world. In general every theory has its own application domain, that is, a set of conditions where it is capable of providing verifiable predictions.

No physical theory has yet been formulated whose application domain is universal and the search for a unified theory of physics is a strong motivation for many researchers. The goal is to establish a reduced number of principles from which one could derive a formalism applicable to physics of all scales, from particles to the cosmos, and to all times, from the origin of the Universe, through the present time, allowing predictions for the Universe's future.

This book is different from the majority of physics books because it does not pretend to formulate physics theories, although it derives formalisms applicable to physics; its essential difference is that it requires not a set of principles but rather a space, referring to a number of dimensions and a space metric,

1 Introduction

My motivation for writing this book has a lot to do with the general subject of physics unification but, as I hope to make clear within the book, my approach departs considerably from that of the Grand Unified Theories or from those approaches generally known as Theories of Everything; for the former see for instance Ross [1]. For some years I've been pursuing the idea that physics as a whole must originate from a small set of first principles using only mathematical derivations with perfectly established rules. There is a lot of unwarranted faith in this belief but the results obtained so far are so promising that my faith keeps getting stronger all the time. I have little hope that I will be the one to finally formulate physics as a unified theory but I expect to be able to pave the path that others will tread to its end. A few considerations about what is really fundamental in physics should clarify my position.

1.1 What is fundamental in physics?

Everybody makes use of constants in everyday life, quite often without even being aware of it; we use constants as a set of references in which we base our relationship and communication with other people; let me give two examples.

When we return home we expect to find it always in the same place, that is, we expect its location to be constant. Because the location does not change from day to day it is a temporal constant. It is also a constant relative to the different people who may wish to find us, because we express our house's location through a door number and we give that number to all those who want to visit us. Can we imagine the confusion if the door number were different for each visitor? Our door number is thus a constant which we use as reference, allowing us to convey the information of a constant location.

Let us examine a different example; suppose someone states that a worker's monthly pay is proportional to the time he worked during that month. That person will be implying a mathematical proportionality relation between the number that expresses the worker's pay in euros, say, and the number of work hours in the corresponding month. Implicitly he will be saying that the result of dividing pay by work time is always the same, whatever the month; that number is a proportionality constant. Everyday language is less precise than mathematical language and it is acceptable to say that two numbers are proportional even when the result of their quotient fluctuates. In this section I want to look at constants used in physics, trying to identify those that are truly fundamental, that

is, those that we have to accept without justification independently of the advancement of science.

If we use physics to solve real everyday problems, the question of which and how many constants are fundamental becomes irrelevant. For instance, we know exactly what calculations are needed in order to find the energy for heating a room and those calculations are independent of the fact that some constants may be fundamental and others just convenient. Even in relativistic or quantum mechanical problems it does not matter if the constants are fundamental or not. The question becomes meaningful only for the unification of physics, that is, when we try to place all branches of physics under the umbrella of a single theory. Someone wrote that physics will be mature when it can be condensed into a single formula to be printed on the front of a T-shirt, something like $E = mc^2$.

It is important to fully understand this objective because it does not mean that the unifying formula will spell the end of physics and certainly not the end of other sciences. I usually make an analogy to clarify the unification objective. We all make use of word processors, it is indifferent whether we work with *Microsoft Word* or *L^AT_EX*, and we associate to those programs an identity independent of the particular computer on which they are running. Computers differ from each other but they are all essentially based on electrical currents which are turned on or off; these can also be assigned to 0s or 1s of a binary algebra. Certainly we are not close to understanding how the word processor works just because we know the binary algebra or the currents flowing in the integrated circuits; this is because those programs exist on a different level, which becomes manifest if we realize that they can exist in different computers without losing their identity. In a similar way the different branches of physics, as well as the various sciences, exist on a higher level of reality, and their development can and must be carried on independently of what may become the language of physics at a fundamental level.

All physics theories are based on principles, fundamental truths that we don't question and whose validity is established when the theories' predictions agree with experimental results or observations of the world. Fundamental constants' values are also accepted without justification; they are part of a theory's essence. When several theories compete for the explanation of a given set of observations, it always happens that one of them becomes prominent because it makes better predictions or is applicable to a wider set of phenomena. Einstein's relativity dethroned Newton's dynamics because, while both theories' predictions are equally effective when low energies are involved, relativity produces far more accurate predictions for high energy experiments. One always expects any new theory to include or revert to the existing ones in those cases where the latter produce good predictions.

At present physicists have to reach out for different theories when they are dealing with cosmology and astrophysics phenomena as well as interactions between atoms or elementary particles. The three main theories, general relativity, quantum mechanics and the standard model of particle physics, coexist not always peacefully. Actually it

is only on everyday physics, what we can call human scale physics, that there are no uncertainties about which theory must be used. Here Newton's dynamics and Maxwell's electromagnetism are the rulers and one can be certain that any new theory will have to incorporate them. One should say, however, that quantum physicists are perfectly pleased with quantum mechanics and astrophysicists are equally happy with general relativity; things get more difficult when the two theories have to come together in order to build models of primordial Universe. A unified theory of physics, when it comes into existence, will have to be applicable to the whole Universe, from elementary particles to the cosmos, and to all times, from the beginning to the present time, allowing predictions for the Universe's evolution. Physicists are not only concerned with the validity of their theories but also with their elegance, a vague concept which has to do with having a small number of founding principles; this means that a theory is all the more elegant the smaller the number of its principles. It is in the scope of a search for elegance that determining which and how many fundamental constants there are finds its reason.

First of all I would like to remove from the list of fundamental constants a large set of constants relative to laws of approximate proportionality. For example, I know that the force needed for pushing a chair is approximately proportional to the velocity of the chair. Putting it differently, if I divide my force by the chair's velocity I get always approximately the same number, which I designate by the friction constant. However, everybody understands that there are so many variables in a friction problem that reducing it to a simple proportionality law can only be a rough approximation. These types of constants, constants of convenience, are typically known with low precision, so they are usually written with no more than one decimal place and are also dependent on the particular units system that is used. Those constants are the result of our inability to deeply understand the phenomenon in question or else the need to simplify a complex set of relations among all the variables involved. These constants are a practical application of Taylor's theorem, familiar to all physicists and mathematicians, and they could theoretically be eliminated by the use of exact formulas. Those constants cannot qualify as fundamental.

Some constants in physics are associated with relations of exact proportionality, unlike the preceding case; one example is the proportionality law between the mean energy of a gas' molecules and its temperature; this law involves the Boltzman constant. Let us recall that the molecules in a gas are in constant random motion, which can be measured either by the gas temperature or by the molecules' mean energy. Actually, temperature and mean energy in this case measure exactly the same thing and one wonders why one is expressed in Kelvin and the other in Joules. The main reason for this state of affairs is tradition; statistical mechanics had to be invented for us to understand that temperature measurement was indeed energy measurement and by then there were two different units for the same thing. It means that Boltzman's constant is essentially a units conversion factor. Contrary to those we called constants of convenience, it is known with high accuracy; we can however do without it, for it suffices to express temperatures

in Joules. We don't do it ordinarily because we might not be understood by a large fraction of the population, but we do it frequently when writing equations. Since they are units conversion factors, all constants like this have their value dependent on the units' systems that are used; Boltzman's constant has different values when expressed in Joules per Kelvin or in Joules per degree Fahrenheit.

In human scale physics, i. e. leaving aside the sub-atomic world and the cosmos, one usually has to deal with four fundamental constants; if we consider also those areas the number grows to over twenty. The four fundamental constants are: speed of light in vacuum, universal gravitational constant, which appears when evaluating the attraction between two masses, Planck's constant, relating a photon's frequency to its energy, and one constant chosen among electrical entities, for which we can elect the permittivity of vacuum. All these constants are known with great accuracy and their values are units' system dependent; the main question is: are they really fundamental or, just like Boltzman's constant, they are the result of parallel developments of our knowledge, which led us to define different units for things that are not essentially different?

Let us look first at the vacuum speed of light, that every one recalls being 300,000 Km/s. This constant allows us to convert a time interval into the distance light travels during that time and so, if needed, we can measure length in seconds or years. One usually writes light-seconds or light-years to stress that distance is meant but these are not different seconds or years; they are exactly the same thing. A light-second is a second but we add the light qualifier to mean we are referring to a 300,000 Km distance. With Einstein and Minkowski, distance and time measurements became undistinguishable; distance and time concepts are intimately related and don't allow a distinction in terms of units. In relativity work it is common practice to measure time and distance with the same units, eliminating the vacuum light speed constant. I am convinced that since the formulation of special relativity theory the latter is no longer a fundamental constant.

Can we do a similar thing with the other fundamental constants? As we said we have to consider four fundamental constants and significantly we have also four fundamental dimensions in any units system. The choice of these may fall, for instance, in length, time, mass and electric charge. Operating the four fundamental constants among themselves we can build standards for each of those dimensions; in doing so we make our measurements independent of standard meters or atomic clocks, because our standards are all built with constants that we consider fundamental and immutable. This procedure allows, in principle, the use of a single standard for all measures, time, length, mass, etc., since we can convert among them through the fundamental constants. The purpose of this single standard is the establishment of a universal scale factor, so that we all use the same number to express the same measurement. Measurements made in this way don't need units because they are all expressed in the same unit; they can be called non-dimensional; simultaneously all fundamental constants are eliminated because they become unity. Such units' system is usually designated by Planck or natural units but most of all I would like to stress that we can apparently do away with all the fundamental

constants.

It is not really like that. What can we say about constants like pi or the Neper number? The former pops up unavoidable every time we relate the perimeter of a circle with its diameter and the latter is similarly unavoidable although in less common situations. There is no units' system manipulation capable of changing them, quite simply because they are the result of mathematical relations and thus they are non-dimensional. These constants must not be called fundamental because their value can be obtained mathematically and does not need to be known *a priori*. There is, however, at least one non-dimensional constant in physics, which is thus units' system independent; this is the fine structure constant, which we find in the interaction between two electrons. This is truly a fundamental constant of physics, unless someone someday discovers a mathematical series with this constant as point of convergence; then its value will cease to be a postulate and this will become another number like pi.

What can we say about the *pléiade* of constants in elementary particle physics? The present state of knowledge in this area is so incipient that we can reasonably hope the fall off for several of them in the years to come, as the theory evolves. Something similar can probably happen in cosmology, where constants such as the critical density of the Universe and the cosmological constant have all the looks of temporary solutions for the insufficiencies of present science. We hope that on its coming of age physics will dispense with all the constants, exception made for the non-dimensional ones; these may eventually become mathematical, leaving physics completely devoid of fundamental constants. If and when this vision becomes real, physics will become a set of mathematical theorems and relations, derived from a small number of axioms; these will be the first principles of physics. Among the principles that will remain, one must certainly establish the number of space dimensions where physics is developed; will the four dimensions of general relativity theory be enough or will we need eleven dimensions, as super-string theorists advocate? Most scientists today think that we will need eleven dimensions but a small group of people aims to prove that five dimensions will suffice; it seems definitely clear that we will need to go beyond four dimensions.

What do physicists mean when they speak of multidimensional universes? Our senses and our instruments tell us about three dimensions to which we have grown used to. Everybody knows that three numbers are needed in order to place a point relative to the Earth: latitude, longitude and altitude; this is what we mean when we say the physical universe is three-dimensional. Einstein added time to these three dimensions raising the number to four; this dimension is notably different from the other three, at least in the way we perceive it. In fact time is different even for the relativity theory; we can never go back in time, for instance. Apparently there is no place for new dimensions because we have no direct perception of them nor do our instruments show evidence of their existence. If there are more than four dimensions, the additional ones must have distinguishing characteristics to separate them from the three physical world dimensions and the time, so that we can live perfectly unaware of them. In super-string type theories

the new dimensions are curled on themselves in such way that in the macroscopic world we live in they are impossible to detect. The concept is by no means intuitive and it is difficult to grasp but it provides motivation for the work of a large number of scientists, just because it holds the promise of one day unifying the different areas of physics. For the small community exploiting the five dimensional possibility, the new dimension is detectable only at the cosmic level. This dimension is undistinguishable from the time in ordinary laboratory experiments but separates from it at the galaxy level and in experiments involving velocities approaching the speed of light; we will be exploiting this idea further along in this book.

It is generally accepted that the unification of physics cannot be achieved without a change of paradigm; the problem is that everyone knows a change is needed but nobody can say for sure which change will work. The revolution brought by Einstein is a change of paradigm in itself because prior to that no one could conceive an upper bound to velocity; the fact that the speed of light is an absolute upper limit to all velocities means that even if it were possible for someone to travel in a train close to the speed of light and fire a bullet inside the train, also close to the speed of light, the "sum" of the two velocities would still be below that limit. Einstein was responsible for other important concepts which together made a radical change to the way we see the Universe. In a parallel development the quantum revolution initiated by Planck produced another paradigm change, this time with relevance mainly for the microscopic world. Here one says that energy cannot change continuously but just by quantum leaps. One says also that there are no certainties, just probabilities, in the quantum world. It is hard to imagine a paradigm change similar to any of those and nevertheless we will probably have to wait for a revolution of comparable caliber in order to place all physics under a single theory.

What is the purpose of a unified theory? Why are we concerned with having to change theories when we look at different problems, if the different theories are reliable within their domain? This is first of all an aesthetics question because an elegant theory of universal application will be a work of art capable of immortalizing its author. But there is more. By looking at an object from different viewpoints one gains perspective over that object; similarly, by approaching a problem in different ways one usually perceives details that are hidden in a single approach. Furthermore we are far from happy with the reliability of existing theories at the cosmic and elementary particle scales and so it is legitimate to hope that the forthcoming revolution will bring answers to some of the remaining problems.

Where do we stand then in respect of what is fundamental in Physics and in the Universe through it? There is no doubt that some concepts we use and associate to different manifestations and properties of reality are in such way related that they could be merged into a single one, removing several constants in the process. If this can one day be extended to particle physics and the cosmos is a question of faith; it is equally legitimate to believe or not believe. Because faith moves mountains, those who believe

find in it enough motivation to proceed with their efforts.

1.2 Linear spaces

Linear spaces are defined in terms of entities called *vectors* and a field, usually the real numbers, called *scalars*. An addition operation is defined for vectors with the properties:

1. Commutativity:

$$a + b = b + a. \quad (1.1)$$

2. Associativity:

$$a + (b + c) = (a + b) + c = a + b + c. \quad (1.2)$$

3. An identity element denoted 0:

$$a + 0 = a. \quad (1.3)$$

4. There is an inverse for every element denoted $-a$:

$$a + (-a) = 0. \quad (1.4)$$

Vectors can be operated on with scalars by a product operation so that given any vector a and any scalar λ , the product λa is still a vector belonging to the linear space. With λ, μ scalars and a, b vectors, we can write the following other properties:

1. Distributive:

$$\lambda(a + b) = \lambda a + \lambda b; \quad (1.5)$$

$$(\lambda + \mu)a = \lambda a + \mu a. \quad (1.6)$$

2. Associative for product:

$$(\lambda\mu)a = \lambda(\mu a). \quad (1.7)$$

3. Identity element for product, denoted 1:

If 1 is the identity element for the product of scalars, then

$$1a = a. \quad (1.8)$$

In the cases of 2 or 3-dimensional spaces it is possible to associate vectors with directed line segments, providing a direct geometric visualization of a linear space. The addition of vectors is performed graphically by the well known parallelogram construction and the product by a scalar has a graphical interpretation as dilation of a vector's length. The addition operation performed on vectors is notoriously different from the scalar addition but there is no need to adopt a different symbol for it because there is never ambiguity.

It is convenient to introduce here two definitions:

1. Two linear spaces are said to be *isomorphic* when their respective elements can be placed in a one-to-one correspondence that preserves sums and, furthermore, there is a one-to-one correspondence between scalars which preserves both sums and products.
2. When two linear spaces \mathcal{U} and \mathcal{V} share the same scalars and all the elements of \mathcal{U} are contained in \mathcal{V} , it is said that \mathcal{U} is a *subspace* of \mathcal{V} .

One usually understands the concept of space dimensionality in the simplest cases: lines are 1-dimensional, surfaces are 2-dimensional and volumes are 3-dimensional. The extension to higher dimensions can be made conceptually, although the geometric representation becomes rather difficult; actually one can only make representations of low-dimension sections of high-dimension spaces. Mathematically, however, it is possible to define the dimensionality of a linear space in such a way that it applies to any number of dimensions. Some definitions will be needed:

1. A vector b is said to be a *linear combination* of vectors a_1, \dots, a_n if n scalars can be found such that

$$b = \lambda_1 a_1 + \dots + \lambda_n a_n = \sum_{i=1}^n \lambda_i a_i. \quad (1.9)$$

2. A set of vectors $\{a_1, \dots, a_n\}$ is said to be *linearly dependent* if n scalars, not all zero, can be found such that

$$\lambda_1 a_1 + \dots + \lambda_n a_n = 0. \quad (1.10)$$

If such set of scalars cannot be found the set of vectors is said to be *linearly independent*.

3. A set of n vectors is said to *span* a linear space \mathcal{V} if all elements of \mathcal{V} can be expressed as a linear combination of the set.
4. A set of n linearly independent vectors which spans space \mathcal{V} is said to form a *basis* for \mathcal{V} and n is said to be the *dimension* of \mathcal{V} .

It is useful to think of vectors as directed segments or displacements, although linear spaces as defined are abstract mathematical entities. There are other possible geometric interpretations of vectors which can be useful at times; for instance, we will find situations when the interpretation of vectors as matrices will help us anticipate space properties which could otherwise go unnoticed.

Vector addition and multiplication by scalars are not the only operations we will need to perform with vectors as every reader most certainly knows. In 3-dimensional vector calculus the product of two vectors can be either a scalar or another vector; in the former case the operation is called scalar product and in the latter the cross product. We will next examine each of those products in order to define similar operations that can be generalized to any number of dimensions.

1.3 The scalar product

The objective of the scalar product is to allow us to deal mathematically with the concepts of distance and angle. The scalar product of two vectors a and b is a scalar that verifies the following conditions:

1. $a \cdot b = b \cdot a$;
2. $a \cdot (\lambda b) = \lambda(a \cdot b)$;
3. $a \cdot (b + c) = a \cdot b + a \cdot c$;
4. $a \cdot a > 0$, if $a \neq 0$.

The last condition will be relaxed when we study mixed signature spaces such as are found in relativity. The length of a vector can now be defined by

$$|a| = \sqrt{a \cdot a}. \quad (1.11)$$

Since there are two possible signs for the square root we adopt the convention that the positive sign is always meant when no sign is indicated. Since our linear space has been equipped with a scalar product which allowed us to define the length of a vector, it is now called a *metric space*. Notice that many scalar products can be defined in the same linear space and consequently many metric spaces coexist in the same linear space, some of which have special designations; the most common and the simplest is *Euclidean space* defined above.

Condition 4, above, defines the scalar product as positive-definite, which in turn implies the Schwarz inequality

$$|a \cdot b| \leq |a||b|. \quad (1.12)$$

The proof goes like this

$$\begin{aligned} (a + \lambda b) \cdot (a + \lambda b) &\geq 0 \quad \forall \lambda \\ \Rightarrow a \cdot a + 2\lambda a \cdot b + \lambda^2 b \cdot b &\geq 0 \\ \Rightarrow (a \cdot b)^2 &\leq (a \cdot a)(b \cdot b), \end{aligned} \quad (1.13)$$

where the last step follows from solving the quadratic inequality in λ . Equation (1.12) is recovered because all the numbers in the last inequality are positive. The *angle* θ between a and b can now be defined by

$$a \cdot b = |a||b| \cos \theta. \quad (1.14)$$

Two vectors whose scalar product is zero are said to be *orthogonal*. Among the infinite possible basis it is usually convenient to choose one whose vectors are mutually orthogonal

and have unit length; when this is done we say that we have defined an *orthonormal* basis. If the vectors $\{\sigma_1, \dots, \sigma_n\}$ denote an orthonormal basis, this fact can be summarized in the equation

$$\sigma_i \cdot \sigma_j = \delta_{ij}, \quad (1.15)$$

where δ_{ij} is the Kronecker delta defined by

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases} \quad (1.16)$$

Any vector a can now be expanded in the orthonormal basis as

$$a = \sum_{i=1}^n a_i \sigma_i. \quad (1.17)$$

We will start using *Einstein's summation convention* for which we must first introduce upper indices. From this point onwards the notation a^i is understood as meaning a upper index i and not a raised to i ; for the latter we will have to place the a in parenthesis as in $(a)^i$. This will be true even when numbers and not letters are used as upper indices; so a^2 means a upper index 2 and $(a)^2$ means a squared. Only when there is no danger of ambiguity will we relax the notation and treat raised letters or numbers as exponents. Einstein's summation convention states that any indices that appear repeated in upper and lower position must be summed over their entire span. According to the convention a vector is expanded in the orthonormal basis as

$$a = a^i \sigma_i. \quad (1.18)$$

Note that raising and lowering indices is here used only to put Einstein's summation convention into effect but we will later attribute different meanings to indices placed above and below.

The *components* $\{a^i\}$ of vector a in the basis $\{\sigma_i\}$ can be found simply by

$$a^i = \sigma_i \cdot a. \quad (1.19)$$

The scalar product of vectors $a = a^i \sigma_i$ and $b = b^j \sigma_j$ can now be evaluated easily

$$a \cdot b = (a^i \sigma_i) \cdot (b^j \sigma_j) = a^i b^j \sigma_i \cdot \sigma_j = a^i b^j \delta_{ij} = \sum_{i=1}^n a^i b^i. \quad (1.20)$$

The angle definition given in Eq. (1.14) applies only to spaces with positive-definite scalar product. Later in the book we will move into spaces where this rule does not apply and the angle definition will have to be reviewed.

1.4 Complex numbers

Complex numbers form a field and as such they are usually assigned to scalars; remember that our definition of linear space did not specify that scalars had to be real numbers. However complex numbers can also be viewed as an ordered pair of real numbers that represents a directed segment in the complex plane, as was realized by Wessel in 1797. If we take the complex $z = x + iy$, this is represented on the plane by a segment with components $\{x, y\}$ and squared length $(x)^2 + (y)^2$. This value cannot be obtained multiplying z by itself but rather by its conjugate

$$zz^* = (x + iy)(x - iy) = (x)^2 + (y)^2. \quad (1.21)$$

A product of two complex numbers $z = x + iy$ and $w = u + iv$ on the plane can be constructed as

$$zw^* = (x + iy)(u - iv) = xu + yv + i(yu - xv). \quad (1.22)$$

The real part of this product recovers the usual scalar product. In order to understand the meaning of the imaginary part we switch to polar representation

$$z = |z|e^{i\theta}, \quad w = |w|e^{i\phi}. \quad (1.23)$$

Performing the product

$$zw^* = |z||w|e^{i(\theta-\phi)}. \quad (1.24)$$

The imaginary term has magnitude $|z||w|\sin(\theta - \phi)$, where $\theta - \phi$ is the angle between the two segments. The magnitude of this term is the area of the parallelogram formed by the two segments and the sign indicates the *handedness* of the area element, that is, whether it is swept from z to w or in reverse. This suggests that we can interpret a complex number as a sum of a scalar, the real part, with an area, the imaginary part. This will later prove to be a very powerful interpretation.

1.5 Quaternions

Quaternions were invented by W. R. Hamilton as an extension of complex numbers into higher dimensions. At first sight one would expect the complex plane to be extendable to 3 dimensions by simply adding an extra imaginary axis perpendicular to the normal complex plane. Hamilton thought so but in fact this does not work because the product of the two imaginary units is a new entity which does not belong to the algebra. To cut a long story short, 3 imaginary units $\{i, j, k\}$ are needed, with the following multiplication rules:

$$(i)^2 = (j)^2 = (k)^2 = -1 \quad (1.25)$$

and

$$ij = -ji = k, \quad jk = -kj = i, \quad ki = -ik = j; \quad (1.26)$$

these can be summarized in $(i)^2 = (j)^2 = (k)^2 = ijk = -1$. With $\{a, b, c, d\}$ reals, a *quaternion* is an element of the form $q = a + bi + cj + dk$; this has a conjugate obtained by changing the signs of the imaginary terms: $q^* = a - bi - cj - dk$. It is very easy to verify that the product of a quaternion by its conjugate returns a real number, the square of its norm;

$$qq^* = (a)^2 + (b)^2 + (c)^2 + (d)^2 = |q|^2. \quad (1.27)$$

The quaternion product is *invertible*, like real and complex products, that is, for each quaternion q there is another quaternion $1/q$ such that $q(1/q) = 1$. The quaternion product is also associative but it is non-commutative, due to the multiplication rules for $\{i, j, k\}$.

Quaternions provide a very efficient way of encoding 3-dimensional rotations. If a is a *pure quaternion*, a quaternion with no real part, it can be represented by a directed segment in 3D; then if R is a unit quaternion, a new pure quaternion, representing a new directed segment, is obtained with the double-sided transformation

$$a' = RaR^*. \quad (1.28)$$

As a consequence, replacing R by i, j or k we generate rotations of π around each axis. We will later have to look closely at rotations in spaces of different dimensions but for 3-dimensional space the use of quaternions is perfectly adequate and effective.

There is only one other invertible algebra, that of *octonions*, with seven imaginary units. This algebra is non-commutative and non-associative and is seldom used.

1.6 The cross product

After our digression over complex numbers and quaternions, we return to the products of vectors to examine the cross product, which bears a strong relation to the product of pure quaternions. Suppose we want the product of two pure quaternions $a = a_1i + a_2j + a_3k$ and $b = b_1i + b_2j + b_3k$; this can be written

$$ab = - \sum_{i=1}^3 a_i b_i + c, \quad (1.29)$$

where c is a pure quaternion given by

$$c = (a_2b_3 - a_3b_2)i + (a_3b_1 - a_1b_3)j + (a_1b_2 - a_2b_1)k. \quad (1.30)$$

This is exactly the way we evaluate the cross product of two vectors expanded in an orthonormal basis.

The properties of the cross product are

1. $a \times b$ is perpendicular to the plane of a and b ,

2. $a \times b$ has magnitude $|a||b| \sin \theta$,
3. The vectors a , b , and $a \times b$ form a right-handed set.

The cross product is usually defined not with recourse to quaternions but to the product of the basis vectors. If $\{\sigma_1, \sigma_2, \sigma_3\}$ form an orthonormal basis we define

$$\sigma_1 \times \sigma_2 = \sigma_3, \quad \sigma_2 \times \sigma_3 = \sigma_1, \quad \sigma_3 \times \sigma_1 = \sigma_2; \quad (1.31)$$

more succinctly we can write

$$\sigma_i \times \sigma_j = \varepsilon_{ij}^k \sigma_k, \quad (1.32)$$

where ε_{ij}^k is an alternating tensor defined by

$$\varepsilon_{ij}^k = \begin{cases} 1 & \text{if } ijk \text{ is a cyclic permutation of } 123, \\ -1 & \text{if } ijk \text{ is an anticyclic permutation of } 123, \\ 0 & \text{otherwise.} \end{cases} \quad (1.33)$$

Expanding a and b into the $\{\sigma_i\}$ basis we have

$$\begin{aligned} a \times b &= (a^i \sigma_i) \times (b^j \sigma_j) \\ &= a^i b^j (\sigma_i \times \sigma_j) \\ &= \varepsilon_{ij}^k a^i b^j \sigma_k. \end{aligned} \quad (1.34)$$

The cross product has enormous importance in vector calculus but, as we shall see, it cannot be generalized to spaces of dimension different from 3 and needs to be replaced by a more encompassing definition.

1.7 The outer product

The cross product works well in 3 dimensions but it does not work in other spaces. In two dimensions there is no way to find a direction perpendicular to the plane of the two vectors and in higher dimensions the perpendicular direction becomes undefined. Consider for instance a space spanned by four orthonormal basis vectors $\{\sigma_1, \dots, \sigma_4\}$; any combination of vectors σ_3 and σ_4 is perpendicular to the plane of σ_1 and σ_2 .

The best generalization of the cross product was invented by the German mathematician H. G. Grassmann. He introduced the *outer product*, which we will also call the *wedge product* and write as $a \wedge b$. Grassmann realized that the outer product needed to change sign when the two vectors were interchanged. The outer product can be defined in spaces of any dimension and so we don't need to associate vectors to oriented segments; in fact we will make use of the outer product in this book in cases where the elements of

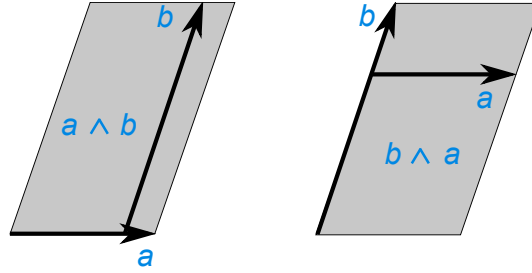


Figure 1.1: The bivector: the outer product of a and b is a directed area element of magnitude $|a||b|\sin\theta$. The orientation of the parallelogram is defined by whether the area is generated by dragging b over a (direct) or a over b (reverse).

the linear space are matrices. The outer product provides a means of encoding a plane without the notion of a vector perpendicular to it and that is why it is dimension independent. The result of the outer product is not a scalar but it is not a vector either; it is a new entity called a *bivector*. Geometrically the outer product between two directed segments a and b can be seen as the area element of magnitude $|a||b|\sin\theta$, on the plane of a and b with a plus sign if the circuit $a, b, -a, -b$ follows a clockwise (direct) direction and a negative sign if it follows an anti-clockwise (reverse) direction. Alternatively, we can look at the outer product between two vectors as the area generated by dragging one of the vectors along the other one; the sign is then defined by which vector is used as dragging basis, see Fig. 1.1.

The outer product of two vectors has the following algebraic properties:

1. The product is *antisymmetric*:

$$a \wedge b = -b \wedge a. \quad (1.35)$$

Geometrically this means that the area defined by a and b changes sign with the interchange of the two vectors; consequently, all vectors verify

$$a \wedge a = 0. \quad (1.36)$$

2. Bivectors are elements of a linear space in the same way as vectors. In two and three dimensions addition is easy to visualize but in higher dimensions it becomes more difficult because two planes need not intersect.

3. The outer product is distributive over addition:

$$a \wedge (b + c) = a \wedge b + a \wedge c. \quad (1.37)$$

The geometrical visualization of the outer product as a parallelogram may be misleading because it is a directed area of no particular shape; in fact the same result can be generated by many different vector pairs, producing different parallelograms with the same area.

In two dimensions all vectors are on the same plane and all bivectors must then exist on that plane. Suppose we have an orthonormal basis $\{\sigma_1, \sigma_2\}$ and we expand two vectors on that basis

$$a = a^1 \sigma_1 + a^2 \sigma_2, \quad b = b^1 \sigma_1 + b^2 \sigma_2. \quad (1.38)$$

The outer product $a \wedge b$ expands as

$$\begin{aligned} a \wedge b &= a^1 b^1 \sigma_1 \wedge \sigma_1 + a^1 b^2 \sigma_1 \wedge \sigma_2 + a^2 b^1 \sigma_2 \wedge \sigma_1 + a^2 b^2 \sigma_2 \wedge \sigma_2 \\ &= (a^1 b^2 - a^2 b^1) \sigma_1 \wedge \sigma_2, \end{aligned} \quad (1.39)$$

which recovers the imaginary part of Eq. (1.22). The magnitude of the product is then immediately $|a||b| \sin \theta$. The coefficient of $\sigma_1 \wedge \sigma_2$ is positive if one goes from a to b in the same angular direction as from σ_1 to σ_2 and it is negative in the opposite case.

In three dimensions a plane can be defined by the direction perpendicular to it; for that reason one can make a one-to-one correspondence between bivectors and vectors and the space of bivectors is therefore also three-dimensional. Suppose that a right-handed orthonormal basis is formed by vectors $\{\sigma_1, \sigma_2, \sigma_3\}$ and two vectors are expanded in that basis as $a = a^i \sigma_i$ and $b = b^j \sigma_j$. The bivector $a \wedge b$ can then be decomposed into a basis of bivectors as

$$\begin{aligned} a \wedge b &= (a^i \sigma_i) \wedge (b^j \sigma_j) \\ &= (a^2 b^3 - a^3 b^2) \sigma_2 \wedge \sigma_3 + (a^3 b^1 - a^1 b^3) \sigma_3 \wedge \sigma_1 \\ &\quad + (a^1 b^2 - a^2 b^1) \sigma_1 \wedge \sigma_2. \end{aligned} \quad (1.40)$$

The components in this expansion are equivalent to those in the cross product but they are components of a bivector instead of vector components. If one recalls, however, the correspondence between bivectors and vectors, the two results can be seen as equivalent.

The concept of handedness has a clear geometrical interpretation in 3-dimensional spaces but it is ambiguous in other dimensions. We are used to aligning the directions of the basis vectors $\{\sigma_1, \sigma_2, \sigma_3\}$ with the fingers of the right hand and we say that the respective frame is right-handed. Any cyclic permutation of the same vectors in their alignment with the right hand fingers is possible and the resulting frame is still right-handed. If two vectors are swapped, the alignment is no longer possible with the right hand and the frame becomes left-handed. In other dimensions we will have to establish the conventions for calling a frame right-handed and there is no agreed convention that works for all dimensions.

It is an easy task to extend the outer product to spaces of arbitrary dimension. If $a = a^i \sigma_i$ and $b = b^j \sigma_j$ are two vectors expanded in an orthonormal basis of an arbitrary

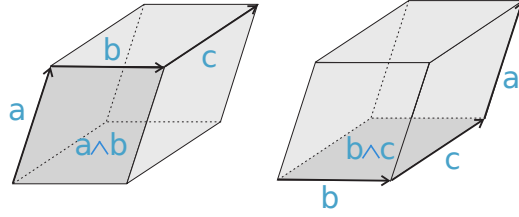


Figure 1.2: The trivector: the outer product of $a \wedge b \wedge c$ is a directed volume and is associative; the sign remains unchanged by circular permutations of the 3 vectors.

dimension space, the outer product $a \wedge b$ is given by

$$a \wedge b = a^i b^j \sigma_i \wedge \sigma_j. \quad (1.41)$$

Naturally all terms with $i = j$ will be null and each non-zero term has a matching term of opposite sign. In dimensions higher than two the product of three non-coplanar vectors becomes possible if it is established that the outer product is *associative*; it produces a new entity called a trivector. For instance, in three dimensions

$$\begin{aligned} a \wedge b \wedge c &= (a \wedge b) \wedge c = a \wedge (b \wedge c) \\ &= (a^i \sigma_i) \wedge (b^j \sigma_j) \wedge (c^k \sigma_k) = \varepsilon_{ijk} a^i b^j c^k \sigma_1 \wedge \sigma_2 \wedge \sigma_3, \end{aligned} \quad (1.42)$$

which represents a *directed volume*; see Fig, 1.2

2 Geometric algebra in two and three dimensions

In this chapter we introduce *geometric algebra* in an intuitive and informal way; the next chapter will be more formal, following the axiomatic route. The latter would normally be sufficient and it is the only one found in some reference books but we find that presenting this subject in an intuitive way may help the unfamiliar reader to become acquainted with some of the concepts. Geometric algebra was created in the nineteenth century by the English mathematician William Kingdon Clifford who joined the inner and outer products into a single *geometric product*; this has the important property of being invertible in spaces of any dimension. In the literature geometric algebra is frequently designated by *Clifford algebra* but we follow here the suggestion of recent authors that Clifford's initial designation is to be preferred.

2.1 The geometric product of vectors

In chapter 1 we studied the symmetric scalar (or inner) product, whose result is a scalar and the antisymmetric exterior (or outer) product, whose result is a bivector. We showed how these products had a geometrical interpretation in two or three dimensions and could easily be extended into any number of dimensions. Clifford's powerful idea was to add the two products in order to define the geometric product. Remember that when discussing complex numbers we showed that the product zw^* also had a scalar and an imaginary part which was later seen to have the magnitude of the outer product between the directed segments representing the two complex numbers. The geometric product between vectors a and b is thus

$$ab = a \cdot b + a \wedge b. \quad (2.1)$$

Many people find it difficult to accept that the result of the geometric product should be a sum of two different entities but our excursion over complex numbers was meant to show that the procedure is similar to the sum of real and imaginary parts in the product of two complex numbers. As we shall see, geometric algebra encompasses both complex algebra and quaternions and so the parallel with complex product is more than pedagogical.

Swapping the factors in Eq. (2.1) changes the sign of the outer product and leaves the

inner product unchanged

$$ba = b \cdot a + b \wedge a = a \cdot b - a \wedge b. \quad (2.2)$$

It then follows

$$a \cdot b = \frac{1}{2} (ab + ba) \quad (2.3)$$

and

$$a \wedge b = \frac{1}{2} (ab - ba). \quad (2.4)$$

In the axiomatic approach to geometric algebra the geometric product is introduced axiomatically, with the associative property, and the inner and outer product are then derived by means of the relations above.

The product of a vector a with another parallel vector λa is

$$a(\lambda a) = \lambda a \cdot a + \lambda a \wedge a = \lambda a \cdot a, \quad (2.5)$$

a pure scalar. When λ is equal to unity, the equation states that a^2 is a pure scalar and we can write $a^2 = |a|^2$ for the squared length of a vector. When two vectors a and b are perpendicular, their product is

$$ab = a \cdot b + a \wedge b = a \wedge b, \quad (2.6)$$

a pure bivector. Swapping the two vectors we see that

$$ba = b \cdot a + b \wedge a = -a \wedge b = -ab, \quad (2.7)$$

from which we conclude that orthogonal vectors anticommute under the geometric product. The geometric product between two vectors encodes the parallel and orthogonal contributions, placing them separately in the scalar and bivector parts.

2.2 An introduction to geometric algebra

Following Clifford's original idea we can multiply vectors successively, obtaining results which have not only scalar and bivector parts but can, in effect, have also vector, trivector, parts and higher-dimensional parts of any *grade* up to the linear space's dimension. Geometric algebra elements are called *multivectors* and are *graded* in the sense that they can be separated in terms with different *grades*. The scalar term is assigned grade-0, the vectors grade-1, the bivectors grade-2 and so on. Just as vectors define a line, bivectors a plane and trivectors a volume, terms of higher grade define hypervolumes of the corresponding dimension. The operation of selecting only the terms with grade r in a multivector is denoted $\langle \rangle_r$, so $\langle ab \rangle_2$ selects the bivector term of the geometric product

$$\langle ab \rangle_2 = a \wedge b. \quad (2.8)$$

By convention we can suppress the 0 index when referring to the scalar part

$$\langle ab \rangle_0 = \langle ab \rangle = a \cdot b. \quad (2.9)$$

The geometric product is *associative* by definition and this property allows the product of more than two vectors

$$a(bc) = (ab)c = abc, \quad (2.10)$$

which can be broken up in parts of different grades

$$a(bc) = a(\langle bc \rangle + \langle bc \rangle_2) = \langle abc \rangle_1 + \langle abc \rangle_3. \quad (2.11)$$

Arbitrary multivectors, with terms of arbitrary grades, can now be written in terms of vector products and so the product of multivectors inherits the associative property. For three general multivectors we can write

$$A(BC) = (AB)C = ABC. \quad (2.12)$$

The product of multivectors is also distributive over addition

$$A(B + C) = AB + AC. \quad (2.13)$$

We said above that the geometric product of vectors was invertible, which means we can divide by vectors. Suppose that $ab = C$, where $C = \langle ab \rangle + \langle ab \rangle_2$; we have

$$Cb = (ab)b = a(bb) = ab^2, \quad (2.14)$$

so we can make $b^{-1} = b/b^2$ and recover a from

$$a = Cb^{-1}. \quad (2.15)$$

We have thus performed a division by b , which was our goal.

As an application of the multivector product we can evaluate the square of bivector $a \wedge b$

$$\begin{aligned} (a \wedge b)(a \wedge b) &= \frac{1}{4} (ab - ba)(ab - ba) \\ &= \frac{1}{4} (abab + baba - 2a^2b^2) \\ &= \frac{1}{4} (ab + ba)^2 - a^2b^2 \\ &= (a \cdot b)^2 - a^2b^2. \end{aligned} \quad (2.16)$$

If we assume $a \cdot b = |a||b| \cos \theta$ we have the further result

$$(a \wedge b)^2 = -a^2b^2 \sin^2 \theta. \quad (2.17)$$

The magnitude of the bivector product $a \wedge b$ is therefore equal to the area of the parallelogram with sides $|a|$ and $|b|$ making an angle θ between them. Applications such as this will be commonplace along the book.

2.3 Geometric algebra of the plane

In a two-dimensional space (a plane) the geometric product becomes easy to understand, so consider that such space is spanned by the orthonormal basis $\{\sigma_1, \sigma_2\}$. The basis vectors are such that

$$(\sigma_1)^2 = (\sigma_2)^2 = 1, \quad \sigma_1 \cdot \sigma_2 = 0. \quad (2.18)$$

Because the basis vectors are orthogonal we must also consider the basis bivector

$$\sigma_1 \sigma_2 = \sigma_1 \cdot \sigma_2 + \sigma_1 \wedge \sigma_2 = \sigma_1 \wedge \sigma_2 = \sigma_{12}. \quad (2.19)$$

When referring to orthogonal basis vector products we will usually simplify the notation by using multiple indices, as in the equation above; we don't use a comma separating the two indices because we will not need to go to spatial dimensions higher than 9, so a double index will never be confused with a single double figure index. In a two-dimensional algebra there are no elements of grade higher than two, since the outer product of three vectors will necessarily return zero. The highest grade elements in an algebra are called *pseudoscalars* and their grade is equal to the underlying linear space dimension.

The full algebra is spanned by the basis set

$$\begin{array}{ccc} 1 & \{\sigma_1, \sigma_2\} & \sigma_{12} \\ 1 \text{ scalar} & 2 \text{ vectors} & 1 \text{ bivector} \end{array} \quad (2.20)$$

This algebra is denoted \mathcal{G}_2 or equivalently $\mathcal{G}_{2,0}$.¹ Any multivector can be expanded on the basis above and multivector sums and products can be calculated by application of the properties of those operations. For instance, the sum of multivectors

$$\begin{aligned} A &= a_0 + a_1 \sigma_1 + a_2 \sigma_2 + a_3 \sigma_{12}, \\ B &= b_0 + b_1 \sigma_1 + b_2 \sigma_2 + b_3 \sigma_{12}, \end{aligned} \quad (2.21)$$

produces the multivector $S = A + B$ given by

$$S = (a_0 + b_0) + (a_1 + b_1) \sigma_1 + (a_2 + b_2) \sigma_2 + (a_3 + b_3) \sigma_{12}. \quad (2.22)$$

This result for addition is exactly as one would expect; products however behave in a more surprising way. Recall from Eq. (2.19) the product of the two basis vectors; recall also that orthogonal vectors anticommute

$$\sigma_2 \sigma_1 = -\sigma_1 \sigma_2 = -\sigma_{12}. \quad (2.23)$$

¹The double index denomination will be explained when we turn our attention to mixed signature spaces. This algebra is also denoted \mathcal{Cl}_2 or $\mathcal{Cl}_{2,0}$ in those works that adopt the designation *Clifford algebra* instead of *geometric algebra*; some authors swap the indices in which case the algebra is then denoted as $\mathcal{Cl}_{0,2}$.

Now we can form products of σ_{12} with vectors, both on the left and on the right

$$\sigma_1(\sigma_{12}) = (\sigma_1\sigma_1)\sigma_2 = \sigma_2 \quad (2.24)$$

and

$$\sigma_2(\sigma_{12}) = -\sigma_2(\sigma_2\sigma_1) = -\sigma_1. \quad (2.25)$$

The effect of multiplying a vector on the right by σ_{12} is another vector rotated $\pi/2$ anti-clockwise to the first one. Similarly, multiplying on the left produces clockwise rotation

$$(\sigma_{12})\sigma_1 = -\sigma_2, \quad (\sigma_{12})\sigma_2 = \sigma_1. \quad (2.26)$$

Finally consider the product of σ_{12} by itself

$$(\sigma_{12})^2 = \sigma_1\sigma_2\sigma_1\sigma_2 = -\sigma_1\sigma_1\sigma_2\sigma_2 = -1. \quad (2.27)$$

We have found that σ_{12} is a square root of -1 by geometric considerations alone; in fact, when a vector is multiplied twice by σ_{12} it is rotated by π , that is, it is multiplied by -1 . This is very important because it gives us the possibility of interpreting geometrically an imaginary number.

We can now form the product of the two multivectors in Eq. (2.21)

$$AB = M = \mu_0 + \mu_1\sigma_1 + \mu_2\sigma_2 + \mu_3\sigma_{12}, \quad (2.28)$$

with

$$\begin{aligned} \mu_0 &= \alpha_0\beta_0 + \alpha_1\beta_1 + \alpha_2\beta_2 - \alpha_3\beta_3, \\ \mu_1 &= \alpha_0\beta_1 + \alpha_1\beta_0 + \alpha_3\beta_2 - \alpha_2\beta_3, \\ \mu_2 &= \alpha_0\beta_2 + \alpha_2\beta_0 + \alpha_1\beta_3 - \alpha_3\beta_1, \\ \mu_3 &= \alpha_0\beta_3 + \alpha_3\beta_0 + \alpha_1\beta_2 - \alpha_2\beta_1. \end{aligned} \quad (2.29)$$

We will not usually write the full product explicitly but it was important to do it here because it is the first time we find a product of multivectors. The product is defined for every pair of multivectors and its result is also a multivector; we say that the algebra is *closed* under it. We mentioned earlier that in some cases it is useful to assign the linear space elements to matrices; to see how it is done in this case, consider the assignment

$$\sigma_1 \equiv \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 \equiv \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}; \quad (2.30)$$

simultaneously replace every real scalar a by the matrix aI . It is easy to verify that these matrices verify the orthogonality relations Eq. (2.18) under the usual matrix product with the inner product evaluated as $\sigma_1 \cdot \sigma_2 = (\sigma_1\sigma_2 + \sigma_2\sigma_1)/2$.

We have already suggested that the pseudoscalar could be used to replace the complex imaginary i ; in fact, complex algebra is isomorphic to a sub-algebra of \mathcal{G}_2 , where only elements of even grade are used. If multivectors containing only scalar and bivector terms are multiplied among themselves the algebra is still closed; we therefore represent the complex number Z as

$$Z = a + bI = a + b\sigma_{12}. \quad (2.31)$$

The capital letter I will be used for the pseudoscalar of the particular algebra of interest; we will reserve the symbol i to those cases where the pseudoscalar squares to -1 and furthermore it commutes with every element of the algebra.

The usual representation of complex numbers on the Argand diagram, with the real part associated with the x coordinate and the imaginary part associated with the y coordinate, must not be confused with the graphical representation of vectors in G_2 . In the former case the complex $Z = |Z|\cos\theta$ is an instruction to rotate by θ and dilate by $|Z|$, whereas in the latter case we just represent oriented segments. Given a vector

$$x = u\sigma_1 + v\sigma_2, \quad (2.32)$$

we obtain a complex number by left multiplication by σ_1

$$\sigma_1 x = u + vI = Z. \quad (2.33)$$

This is how vectors on a plane are mapped to complex numbers.

To obtain the conjugate of Z , $Z^\dagger = u - vI$, we have to multiply by σ_1 on the right

$$Z^\dagger = u - vI = u - v\sigma_{12} = x\sigma_1. \quad (2.34)$$

Comparing Z with Z^\dagger we see that *conjugation* is equivalent to reversing the order in the product $\sigma_1 x$; we call this operation *reversion* and it is one of the possible conjugation operations allowed by geometric algebra. If we consider a second complex number $W = \sigma_1 y$, the complex product $ZW^\dagger = W^\dagger Z$ now becomes

$$W^\dagger Z = y\sigma_1\sigma_1 x = xy, \quad (2.35)$$

as expected.

In order to rotate a complex number by an angle ϕ on the Argand plane we just need to multiply by $\exp(i\phi)$

$$Z \mapsto Z' = e^{i\phi} Z; \quad (2.36)$$

the question now is how to perform a vector rotation in geometric algebra. The exponential $\exp(i\phi)$ is a complex number itself that must be written in the form of scalar plus bivector terms in G_2

$$e^{I\phi} = \cos\phi + I\sin\phi. \quad (2.37)$$

This is consistent with the series expansion of the exponential function.

Using Eq. (2.33) to transform between complex numbers and equivalent vectors

$$x' = \sigma_1 Z'. \quad (2.38)$$

Since Z' is the result of a rotation applied to Z we write

$$x' = \sigma_1 e^{i\phi} Z = \sigma_1 e^{i\phi} \sigma_1 x. \quad (2.39)$$

But

$$\begin{aligned} \sigma_1 e^{i\phi} \sigma_1 &= \sigma_1 (\cos \phi + I \sin \phi) \sigma_1 \\ &= \cos \phi - I \sin \phi = e^{-I\phi}, \end{aligned} \quad (2.40)$$

where we have used the fact that $I = \sigma_{12}$ *anticommutes* with vectors. We arrive finally at

$$x' = e^{-I\phi} x = x e^{I\phi}. \quad (2.41)$$

In order to perform a rotation by ϕ of a vector x on the plane I we have to multiply on the left by $\exp(-I\phi)$ or on the right by $\exp(I\phi)$. This procedure works only in two dimensions and we will see later how to extend rotations to spaces of arbitrary dimension.

We saw how the even subalgebra of two-dimensional geometric algebra \mathcal{G}_2 can give a geometric interpretation to the algebra of complex numbers. Once we start exploring other algebras we will see that their pseudoscalar may be more apt to adopt the role of the complex imaginary than the pseudoscalar of \mathcal{G}_2 . In any case the separation between scalar and imaginary parts acquires a geometric significance which helps extend complex analysis to spaces of any dimension.

2.4 The geometric algebra of space

The geometric algebra of three-dimensional space is the obvious generalization of \mathcal{G}_2 , the geometric algebra of the plane. We will see how it perfectly encompasses all operations of vector algebra and how it describes vectors, surfaces and planes. Spatial rotations are encoded in a very simple way, much easier to work with than matrices. It is also the perfect tool to formulate classical mechanics, although this book does not explore that area.

In order to construct the algebra of 3-dimensional space \mathcal{G}_3 we add another basis vector σ_3 to our previous basis set $\{\sigma_1, \sigma_2\}$. All three vectors are assumed to be orthonormal and anticommute between themselves. There are now three possible bivectors generated by the basis vectors

$$\{\sigma_{12}, \sigma_{23}, \sigma_{31}\}.$$

These define three independent planes in space, which is the expected number since each plane can be referred to by a normal vector. We now have also to consider the product of

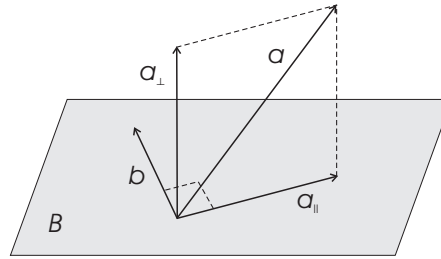


Figure 2.1: The vector a can be decomposed into the sum of a component perpendicular to the plane bivector B and a component in that plane. The bivector B can be written as the product $a_{\parallel} \wedge b$, with b normal to a_{\parallel} .

three vectors, producing a *trivector*; this has the geometric interpretation of an oriented volume (see Fig. 1.2). Just as the bivector was the unique area element of two-dimensional space, the trivector is the unique volume element of three-dimensional space. This is the highest grade element and it is unique up to scale (or volume) and handedness (or sign). According to our established conventions, this is called the pseudoscalar of the algebra.

In three dimensions the algebra is spanned by

1	$\{\sigma_i\}$	$\{\sigma_{ij}\}$	σ_{123}
1 scalar	3 vectors	3 bivectors	1 trivector

This is a graded algebra with 4 different grades and a total dimension of $2^3 = 8$.

We have now a number of new products to consider. Starting with the vector-bivector products, we recall what was found in the algebra of the plane to verify that each bivector generates a $\pi/2$ rotation of a vector on its own plane; each of the bivectors in Eq. (2.4) has the properties of the bivector studied previously for \mathcal{G}_2 . In particular

$$(\sigma_{12})^2 = (\sigma_{23})^2 = (\sigma_{31})^2 = -1 \quad (2.42)$$

and each bivector rotates by $\pi/2$ vectors on its own plane. Since vectors need not belong to the plane of a bivector, we see that products of a vector a and a bivector B , written aB , can contain both vector and trivector terms; the former accounts for the $\pi/2$ rotation of the component of a belonging to the plane of B and the latter to the volume generated by multiplying the normal component of a by B . In Fig. 2.1 we have decomposed vector a into its components in the plane of B and perpendicular to that plane

$$a = a_{\parallel} + a_{\perp}, \quad (2.43)$$

so that now we can write

$$aB = (a_{\parallel} + a_{\perp})B. \quad (2.44)$$

The bivector B can always be written as

$$B = a_{\parallel} \wedge b = a_{\parallel} b, \quad (2.45)$$

where b is a suitable vector normal to a in the B plane. This demonstrates that

$$a_{\parallel} B = a_{\parallel} (a_{\parallel} b) = a_{\parallel}^2 b \quad (2.46)$$

is a vector. This is the vector part of the product, which produces a rotation on the plane. On the other hand

$$a_{\perp} B = a_{\perp} (a_{\parallel} b) = a_{\perp} a_{\parallel} b, \quad (2.47)$$

which is the product of three orthogonal vectors and so it is a trivector. This is the trivector part of the product, which corresponds to a volume.

We want to extend the concepts of inner and outer product to the vector bivector product; in order to do this we study the product of vector a with bivector $b \wedge c$. Using Eq. (2.4) we write

$$a(b \wedge c) = \frac{1}{2} a(bc - cb) = \frac{1}{2} (abc - cba). \quad (2.48)$$

We now rearrange Eq. (2.3) as

$$ab = 2a \cdot b - ba \quad (2.49)$$

to write

$$\begin{aligned} a(b \wedge c) &= (a \cdot b)c - (a \cdot c)b - \frac{1}{2} (bac - cab) \\ &= 2(a \cdot b)c - 2(a \cdot c)b + \frac{1}{2} (bc - cb)a, \end{aligned} \quad (2.50)$$

so that

$$a(b \wedge c) - (b \wedge c)a = 2(a \cdot b)c - 2(a \cdot c)b. \quad (2.51)$$

The left hand side of this equation is the anti-symmetrized product of a vector and a bivector and the right hand side is a vector. So this operation lowers the grade of the bivector by one and we call it the inner product

$$a \cdot B = \frac{1}{2} (aB - Ba), \quad (2.52)$$

for arbitrary vector and bivector. Combining with the former equation we get the important result

$$a \cdot (b \wedge c) = (a \cdot b)c - (a \cdot c)b. \quad (2.53)$$

We can reduce the number of parenthesis by establishing the precedence of inner and outer products over the geometric product; under this convention the previous equation is simplified to

$$a \cdot (b \wedge c) = a \cdot bc - a \cdot cb. \quad (2.54)$$

The remainder part of the vector bivector product returns a trivector. Because it raises the grade it is called outer product and it is denoted with a \wedge .

$$a \wedge (b \wedge c) = \frac{1}{2} [a(b \wedge c) + (b \wedge c)a]. \quad (2.55)$$

It is easy to show that the product is associative;

$$\begin{aligned} a \wedge (b \wedge c) &= \frac{1}{2} [a(b \wedge c) + (b \wedge c)a] \\ &= \frac{1}{4} (abc - acb + bca - cba) \\ &= \frac{1}{4} [2(a \wedge b)c + bac + bca + 2c(a \wedge b) - cab - acb] \\ &= \frac{1}{2} [(a \wedge b)c + c(a \wedge b) + b(c \cdot a) - (c \cdot a)b] \\ &= (a \wedge b) \wedge c, \end{aligned} \quad (2.56)$$

so we can write the result simply as $a \wedge b \wedge c$. The outer product of vectors is therefore associative and antisymmetric on all pairs of vectors; this product was first proposed by H. G. Grassmann and is frequently known as Grassmann's exterior product. From Eq. (2.46) we see that

$$a \wedge B = a_{\perp} B = a_{\perp} \wedge B, \quad (2.57)$$

so the exterior product with a bivector selects the component perpendicular to the plane and returns a trivector (volume element). This product is symmetric in its vector and bivector arguments:

$$a \wedge B = a_{\perp} \wedge a_{\parallel} \wedge b = -a_{\parallel} \wedge a_{\perp} \wedge b = a_{\parallel} \wedge b \wedge a_{\perp} = B \wedge a. \quad (2.58)$$

The full product of vector and bivector is the sum of inner and outer product terms

$$aB = a \cdot B + a \wedge B, \quad (2.59)$$

where the inner product term is a vector and the outer product term is a trivector. In a similar manner to what was done for the vector product we can write

$$\begin{aligned} a \cdot B &= \frac{1}{2} (aB - Ba), \\ a \wedge B &= \frac{1}{2} (aB + Ba). \end{aligned} \quad (2.60)$$

Notice that the signs in these equations are opposite to those found in the vector product case. By adding the two contributions in one single product, the full product of vector and bivector becomes invertible.

In 3-dimensional algebra we have to consider the product of bivectors among themselves. We have seen that the square of a bivector produces a scalar but we must investigate the product of bivectors on orthogonal planes; for instance the product

$$\sigma_{12}\sigma_{23} = \sigma_1\sigma_2\sigma_2\sigma_3 = \sigma_{13}, \quad (2.61)$$

results in a third bivector. In a similar way

$$\sigma_{23}\sigma_{12} = \sigma_2\sigma_3\sigma_1\sigma_2 = \sigma_3\sigma_2\sigma_2\sigma_1 = \sigma_{31} = -\sigma_{13}, \quad (2.62)$$

so the product of orthogonal bivectors is antisymmetric. The product of two bivectors is a scalar if they are on the same plane and another bivector if they are orthogonal; in the general case it has both parts.

If we label the basis bivectors as

$$B_1 = \sigma_{23}, \quad B_2 = \sigma_{31}, \quad B_3 = \sigma_{12}, \quad (2.63)$$

their product satisfies

$$B_i B_j = -\delta_{ij} - \varepsilon_{ijk} B_k. \quad (2.64)$$

This is very similar to the product of vectors, in that we find a scalar and a bivector term, the former being symmetric and the latter antisymmetric. This only happens in 3 dimensions, for in higher dimensions the symmetric product can have a fourvector term; we will eventually denote the scalar part with a dot and the fourvector part with a wedge. The anti-symmetrized product is always a bivector and bivectors form a closed algebra under this product.

We recall that the basis bivectors satisfy

$$(B_1)^2 = (B_2)^2 = (B_3)^2 = -1 \quad (2.65)$$

and

$$B_i B_j = -B_j B_i, \quad i \neq j. \quad (2.66)$$

These properties are reminiscent of quaternions but we cannot make a direct identification between basis bivectors $\{B_i\}$ and the generators of quaternion algebra $\{i, j, k\}$ because Hamilton specified that $ijk = -1$ and we have

$$B_1 B_2 B_3 = \sigma_2 \sigma_3 \sigma_3 \sigma_1 \sigma_1 \sigma_2 = 1. \quad (2.67)$$

In order to set up an isomorphism we have to change one of the signs; for instance

$$i \leftrightarrow B_1, \quad j \leftrightarrow -B_2, \quad k \leftrightarrow B_3. \quad (2.68)$$

We learn from this that proper quaternions are a *left-handed* set of bivectors, whereas Hamilton tried to interpret $\{i, j, k\}$ as a *right-handed* set of vectors.

Given three vectors a , b and c , the trivector $a \wedge b \wedge c$ is formed by sweeping the bivector $a \wedge b$ along vector c (see Fig. 1.2). Graphically this produces a parallelepiped but the shape is really meaningless, just as with the product of bivectors. The trivector represents a volume and its orientations. The figure helps interpret the algebraic properties of the trivector. The same oriented volume is obtained sweeping $a \wedge b$ along c or $b \wedge c$ along a ; mathematically this is written as $(a \wedge b) \wedge c = a \wedge (b \wedge c)$, that is, the product is associative. If any two vectors are swapped, the orientation by which the volume is swept is changed and under two successive swaps the trivector returns to the original result, so

$$a \wedge b \wedge c = c \wedge a \wedge b = b \wedge c \wedge a. \quad (2.69)$$

This can also be seen in Fig. 1.2.

The unit right-handed pseudoscalar for the algebra of space is $\sigma_{123} = \sigma_1 \sigma_2 \sigma_3$. We can check that it squares to -1 by

$$(\sigma_{123})^2 = \sigma_1 \sigma_2 \sigma_3 \sigma_1 \sigma_2 \sigma_3 = (\sigma_1)^2 (\sigma_{23})^2 = -1. \quad (2.70)$$

It also commutes with any vector; see for instance the product

$$\sigma_{123} \sigma_1 = \sigma_1 \sigma_2 \sigma_3 \sigma_1 = \sigma_{23}, \quad (2.71)$$

the same result that is obtained with

$$\sigma_1 \sigma_{123} = \sigma_1 \sigma_1 \sigma_2 \sigma_3 = \sigma_{23}. \quad (2.72)$$

Similar behaviour would be obtained with the remaining basis vectors and hence for all vectors. Because it commutes with all vectors, the unit pseudoscalar also commutes with all elements of the algebra; since it squares to -1 we denote it as

$$i = \sigma_{123}. \quad (2.73)$$

This is therefore a new candidate to replace the complex imaginary unit and it is, in effect, the best choice in some applications. Given any set of three vectors we must have

$$a \wedge b \wedge c = \alpha i, \quad (2.74)$$

where α is a scalar. From what we said previously it is obvious that $|\alpha|$ is the volume of the parallelepiped with sides defined by a , b and c . The sign of α indicates whether the set $\{a, b, c\}$ is right-handed or left-handed.

By looking at Eqs. (2.71) and (2.72) we can say the the product of the pseudoscalar by a vector returns a bivector whose plane is normal to the vector. We can express each of the basis bivectors as the product of the pseudoscalar by the *dual* vector:

$$\sigma_{12} = i \sigma_3, \quad \sigma_{23} = i \sigma_1, \quad \sigma_{31} = i \sigma_2. \quad (2.75)$$

The operation that consists in obtaining one algebra element from the product of the pseudoscalar by another algebra element is called *Hodge duality*.

Finally consider the dual of the outer product of two vectors

$$i(\sigma_1 \wedge \sigma_2) = \sigma_1 \sigma_2 \sigma_3 \sigma_1 \sigma_2 = (\sigma_{12})^2 \sigma_3 = -\sigma_3. \quad (2.76)$$

The result is a vector perpendicular to the plane of the two vectors with its orientation given by the left-hand rule; it is the opposed sign to the result of a cross product. So we see that the cross product has an equivalent in 3-dimensional geometric algebra as

$$a \times b = -ia \wedge b. \quad (2.77)$$

The cross product is specific to 3 dimensions and is unnecessary in this book, so we will reserve the \times symbol to another product to be defined later on. The result of a cross product is indeed a bivector, which is represented by its 3-dimensional dual vector; this is the source of many difficulties which we shall avoid by replacing it with the more general wedge product.

It is useful to make a small digression over the Pauli algebra, to see how it is indeed perfectly encompassed by the geometric algebra of space. The full geometric product of two basis vectors is written

$$\sigma_i \sigma_j = \sigma_i \cdot \sigma_j + \sigma_i \wedge \sigma_j = \delta_{ij} + i\varepsilon_{ijk} \sigma_k. \quad (2.78)$$

Notice how the notation we chose for the basis vectors leads us directly to the Pauli algebra of quantum mechanics. If we choose the Pauli matrices for elements of our linear space, the following assignments become natural

$$\sigma_1 \equiv \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 \equiv \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 \equiv \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (2.79)$$

These matrices satisfy

$$\sigma_i \sigma_j = \delta_{ij} \mathbb{I} + i\varepsilon_{ijk} \sigma_k, \quad (2.80)$$

where \mathbb{I} is the 2×2 identity matrix and i can be seen both as the complex imaginary unit or the product of the three Pauli matrices; naturally all scalars have to be multiplied by \mathbb{I} to become matrices. There is perfect isomorphism between Pauli algebra and \mathcal{G}_3 , so we don't need to choose different symbols for denoting matrices; there is really no need to include the identity matrix symbol in the matrix equation and we can jump between matrix and directed segment representation when needed.

2.5 Notation conventions

In this sections we will review the conventions that have been made in previous paragraphs and establish new ones to be used in the remainder of the book. These conventions

are meant to simplify expressions and to make equations easier to interpret both in geometric and physical terms.

We will usually denote the vectors of an orthonormed basis as $\{\sigma_i\}$; later on we will have to consider non-orthonormed frames whose vectors will then be denoted as $\{g_i\}$. The basis multivectors resulting from the product of orthonormed basis vectors will be denoted with multi-index, as in $\sigma_1\sigma_2\sigma_3 = \sigma_{123}$; this convention does not apply to non-orthonormed frames, as we shall see in due time. The unit right-handed pseudoscalar in any algebra can be replaced by the symbol I , but when it squares to -1 and simultaneously commutes with all vectors it is then preferably written as i ; for instance, in \mathcal{G}_2 , $\sigma_{12} = I$ and in \mathcal{G}_3 , $\sigma_{123} = i$.

The number of brackets is greatly reduced by establishing that *inner and outer products take precedence over the geometric product* and also that *inner products are performed before outer products*. For example, in \mathcal{G}_3 we are allowed to write

$$i(a \wedge b) = ia \wedge b \quad (2.81)$$

or

$$(a \cdot b)c = a \cdot bc. \quad (2.82)$$

The latter rule establishes that

$$(a \cdot b)c \wedge d = a \cdot bc \wedge d. \quad (2.83)$$

In a general multivector the terms of grade r are selected with angle brackets $\langle \rangle_r$; for instance

$$\langle ab \rangle_2 = a \wedge b. \quad (2.84)$$

The special case of selecting the scalar part (grade zero) is denoted simply with $\langle \rangle$. The scalar part of any pair of multivectors is symmetric

$$\langle AB \rangle = \langle BA \rangle. \quad (2.85)$$

As a result, the scalar part satisfies the cyclic reordering property

$$\langle AB \dots C \rangle = \langle BC \dots A \rangle. \quad (2.86)$$

We will frequently use the operation of *reversion*, which reverses the order of all vector and multivector products, it is denoted with a tilde \tilde{A} . For orthonormed basis vector products, reversion changes the sign of orders

$$(2, 3, 6, 7, 10, 11, \dots);$$

for instance

$$\begin{aligned} \tilde{\sigma}_{12} &= \sigma_2\sigma_1 = -\sigma_{12}, \\ \tilde{\sigma}_1 &= \sigma_1, \\ \tilde{i} &= \sigma_3\sigma_2\sigma_1 = -i. \end{aligned} \quad (2.87)$$

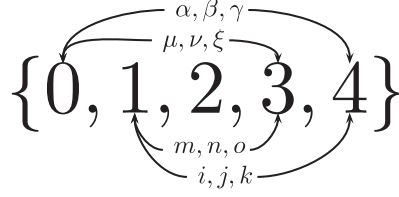


Figure 2.2: Indices in the range $\{0, 4\}$ will be denoted with Greek letters α, β, γ . Indices in the range $\{0, 3\}$ will also receive Greek letters but chosen from μ, ν, ξ . For indices in the range $\{1, 4\}$ we will use Latin letters i, j, k and finally for indices in the range $\{1, 3\}$ we will use also Latin letters chosen from m, n, o .

Another useful operation is *conjugation*, also known as *Hermitian conjugation*, denoted with a dagger A^\dagger . This operation is equivalent to the operation with the same designation performed over matrices, if an isomorphism is made between geometric algebra and some matrix algebra. So, conjugation performed in space algebra \mathcal{G}_3 is equivalent to Hermitian conjugation performed in Pauli algebra with the matrix assignment of Eq. (2.79). Without recourse to matrix representation, the conjugation consists on reversing the products after changing the sign of basis vectors with negative norm. In the algebras we have studied so far conjugation and reversion are equivalent but that is not usually the case in mixed signature algebras. A general multivector in \mathcal{G}_3 can be written

$$M = \alpha + a + B + \beta i, \quad (2.88)$$

where α and β are scalars, a is a vector and B a bivector. The reverse and conjugate of M are identical

$$\widetilde{M} = M^\dagger = \alpha + a - B - \beta i. \quad (2.89)$$

A multivector is said to be *homogeneous* if it contains terms of a single grade. The designation *inner* product is reserved to the lowest grade part in the product of homogeneous multivectors; the inner product of two homogeneous multivectors of the same grade always returns a scalar. The terms *exterior* and *outer* product both designate the highest grade part of the product of two homogeneous multivectors. Colloquially the designations *dot product* and *wedge product* can sometimes replace the proper inner and outer product.

Later in the book we will deal with 5-dimensional space but we will also be interested in two of its 4-dimensional and one 3-dimensional subspaces; ideally our choice of indices should clearly identify their ranges in order to avoid the need to specify the latter in every equation. The diagram in Fig. 2.2 shows the index naming convention used in this book; Einstein's summation convention will be adopted as well as the compact notation

for partial derivatives $\partial_\alpha = \partial/\partial x^\alpha$.

Derivatives for both flat and curved space will be defined later on but their indexing conventions can be explained here. For derivatives applied in subspaces of 5-dimensional spacetime we use a pre-index convention, where the chosen character indicates the subspace to which the operator applies; for instance in 4D spacetime

$${}^\mu\nabla\psi = \sigma^\mu\partial_\mu\psi, \quad {}^\mu\mathbf{D}\psi = g^\mu\partial_\mu\psi; \quad (2.90)$$

and in 3D space

$${}^m\nabla\psi = \sigma^m\partial_m\psi, \quad {}^m\mathbf{D}\psi = g^m\partial_m\psi; \quad (2.91)$$

finally in 4D space

$${}^i\nabla\psi = \sigma^i\partial_i\psi, \quad {}^i\mathbf{D}\psi = g^iu\partial_i\psi. \quad (2.92)$$

Mapping between geometry and physics is facilitated if one chooses to work always with non-dimensional quantities, which can be obtained with a suitable choice of standards for the fundamental units. In this book all problems of dimensional homogeneity are avoided through the use of *normalizing factors* listed below for all units, defined with recourse to the *fundamental constants*: $\hbar \rightarrow$ Planck constant divided by 2π , $G \rightarrow$ gravitational constant, $c \rightarrow$ speed of light and $\varepsilon_0 \rightarrow$ vacuum permittivity.

Dimension	Standard	Value
Length	$\sqrt{G\hbar/c^3}$	$6.4473 \times 10^{-36} \text{ m.}$
Time	$\sqrt{G\hbar/c^5}$	$2.15059 \times 10^{-44} \text{ s.}$
Mass	$\sqrt{\hbar c/G}$	$8.68356 \times 10^{-9} \text{ Kg.}$
Permittivity	ε_0	$8.85419 \times 10^{-12} \text{ Asm}^{-1}\text{V}^{-1}.$

This normalization defines a system of *non-dimensional units* (natural units) with important consequences, namely: 1) All the fundamental constants, \hbar , G , c , ε_0 , become unity; 2) a particle's Compton frequency, defined by $\nu = mc^2/\hbar$, becomes equal to the particle's mass; 3) the frequent term $GM/(c^2r)$ is simplified to M/r . The non-dimensional constants such as the fine structure constant α remain unchanged.

In the non-dimensional units system the value of the proton charge can be derived from the relation

$$\alpha = \frac{e^2}{4\pi\varepsilon_0}. \quad (2.93)$$

Using the accepted value $\alpha \approx 7.29735 \times 10^{-3}$ we obtain the proton charge $e \approx 0.302822$. Vacuum permeability can be derived from the fact that the vacuum speed of light provides a relation between permittivity and permeability: $c = 1/\sqrt{\varepsilon_0\mu_0}$; as a result vacuum permeability is also unity in the non-dimensional units' system.

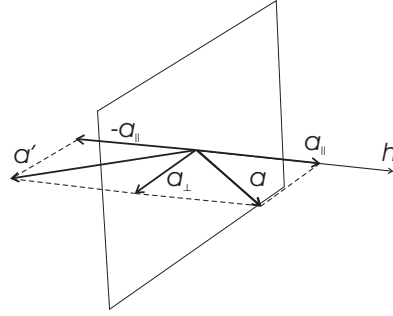


Figure 2.3: The vector a is reflected in the hyperplane perpendicular to h , which is valid in arbitrary dimensions. The result a' is formed by reversing the sign of a_{\parallel} , the component of a in the direction of h .

2.6 Reflections

One of the best ways to show the power of geometric algebra is to formulate reflections and rotations; the simplicity of the equations when compared to alternative matrix methods should convince everybody that this is the more appropriate formulation. Let us consider an arbitrary vector a and a unit vector h ($h^2 = 1$). This can be achieved simply by

$$\begin{aligned} a &= h^2 a \\ &= h(h \cdot a + h \wedge a) \\ &= a_{\parallel} + a_{\perp}. \end{aligned} \tag{2.94}$$

The parallel component $a_{\parallel} = a \cdot h h$ is easily recognized as the projection of a onto the direction of h and the remainder must be the perpendicular component $a_{\perp} = h h \wedge a$. The confirmation that a_{\perp} is perpendicular to h can be made by

$$h \cdot a_{\perp} = \langle h h (h \wedge a) \rangle = \langle h \wedge a \rangle = 0. \tag{2.95}$$

In this example we made use of the grade selector to simplify derivations.

The result of reflecting a on the plane orthogonal to h is shown in Fig. 2.3 and can be written

$$\begin{aligned} a' &= a_{\perp} - a_{\parallel} = h h \wedge a - a \cdot h h \\ &= -h \cdot a h - h \wedge a h \\ &= -h a h. \end{aligned} \tag{2.96}$$

A formula as compact as this could not be written without the use of geometric algebra and is valid in spaces of any dimension.

A check can be made that lengths and angles are preserved by reflection of two vectors:

$$(-hah) \cdot (-hbh) = \langle (-hah)(-hbh) \rangle = \langle habh \rangle = \langle abhh \rangle = a \cdot b. \quad (2.97)$$

In the derivation we made use of the cycle reordering property of the scalar part, defined in Eq. (2.86).

Complex conjugation performs a reflection on the real axis, when complex numbers are represented on the Argand plane. If $x = \mu\sigma_1 + \nu\sigma_2$ is a point on the σ_{12} plane, it is associated with a complex number by the relation (2.33) that we reproduce here

$$Z = \sigma_1 x; \quad (2.98)$$

on the other hand we obtain x from Z by

$$x = \sigma_1 Z. \quad (2.99)$$

Reflecting x on the σ_{31} plane is performed by the operation

$$x' = -\sigma_2 x \sigma_2 = -\sigma_2 \sigma_{12} x \sigma_1 \sigma_{12} = \sigma_1 x \sigma_1 = \sigma_1 Z^\dagger. \quad (2.100)$$

So, if we decide to represent complex numbers on the plane we know how to handle complex conjugation in geometric algebra. If, on the other hand we opt for a complex representation in \mathcal{G}_3 as scalar plus pseudoscalar parts, the conjugation is performed in the usual way and corresponds to both reversion and Hermitian conjugation.

The reflection of bivector $B = a \wedge b$ on the plane perpendicular to h involves reflecting both vectors on this plane; the result is

$$B' = (-hah) \wedge (-hbh). \quad (2.101)$$

The simplification follows

$$\begin{aligned} (-hah) \wedge (-hbh) &= \frac{1}{2} (hahhbh - hbhah) \\ &= \frac{1}{2} h(ab - ba)h \\ &= hBh. \end{aligned} \quad (2.102)$$

This formula shows that a bivector reflects with a formula similar to that of vectors, *except for a change of sign*. In space algebra bivectors can be represented by their vector duals but the transformation law for the bivectors must be applied to their duals. This is the reason why 3-dimensional vectors are frequently classified in axial and polar vectors. This can be a source of great confusion and is specific to 3-dimensional space; we will not have to be concerned with such distinction with our adopted notation.

Next we need to look at the reflection of a trivector $a \wedge b \wedge c$. Reflecting each vector individually we write

$$\begin{aligned} (-hah) \wedge (-hbh) \wedge (-hch) &= \langle (-hah)(-hbh)(-hch) \rangle_3 \\ &= -\langle habch \rangle_3. \end{aligned} \quad (2.103)$$

The product abc can only contain vector and trivector terms, but the former is excluded by the grade-3 selector, so we are left with

$$(-hah) \wedge (-hbh) \wedge (-hch) = -\langle ha \wedge b \wedge ch \rangle_3. \quad (2.104)$$

We have already seen that trivectors commute with all vectors, so a further simplification is possible

$$(-hah) \wedge (-hbh) \wedge (-hch) = -a \wedge b \wedge c. \quad (2.105)$$

The reflection of a trivector on a plane simply flips its sign; this means that if the vectors in a trivector form a right-handed set, after reflection they form a left-handed set and vice-versa.

2.7 Rotations

Suppose we want to rotate a vector $a = a^m \sigma_m$, around the σ_3 direction, by an angle 2ϕ in the forward direction. We know from trigonometry that the basis vectors must be transformed according to the rules

$$\sigma_1 \mapsto \cos(2\phi)\sigma_1 + \sin(2\phi)\sigma_2, \quad (2.106)$$

$$\sigma_2 \mapsto -\sin(2\phi)\sigma_1 + \cos(2\phi)\sigma_2, \quad (2.107)$$

$$\sigma_3 \mapsto \sigma_3, \quad (2.108)$$

so vector a is transformed as

$$\begin{aligned} a \mapsto & [a^1 \cos(2\phi) - a^2 \sin(2\phi)] \sigma_1 \\ & + [a^1 \sin(2\phi) + a^2 \cos(2\phi)] \sigma_2 + a^3 \sigma_3. \end{aligned} \quad (2.109)$$

This operation is conveniently written as

$$a \mapsto e^{-\sigma_{12}\phi} a e^{\sigma_{12}\phi}, \quad (2.110)$$

as we shall demonstrate.

First of all let us consider the exponential of multivector terms in two cases; if u is such that $u^2 = -1$ and ϕ is a scalar

$$\begin{aligned}
 e^{u\phi} &= 1 + u\phi - \frac{\phi^2}{2!} - u\frac{\phi^3}{3!} + \frac{\phi^4}{4!} + \dots \\
 &= 1 - \frac{\phi^2}{2!} + \frac{\phi^4}{4!} - \dots \{= \cos \phi\} + \\
 &\quad + u\phi - u\frac{\phi^3}{3!} + \dots \{= u \sin \phi\} \\
 &= \cos \phi + u \sin \phi.
 \end{aligned} \tag{2.111}$$

Conversely if h is such that $h^2 = 1$

$$\begin{aligned}
 e^{h\phi} &= 1 + h\phi + \frac{\phi^2}{2!} + h\frac{\phi^3}{3!} + \frac{\phi^4}{4!} + \dots \\
 &= 1 + \frac{\phi^2}{2!} + \frac{\phi^4}{4!} + \dots \{= \cosh \phi\} + \\
 &\quad + h\phi + h\frac{\phi^3}{3!} + \dots \{= h \sinh \phi\} \\
 &= \cosh \phi + h \sinh \phi.
 \end{aligned}$$

We are now in position to apply the transformation in Eq. (2.110) to each of the basis vectors individually; we see that

$$\begin{aligned}
 e^{-\sigma_{12}\phi} \sigma_1 e^{\sigma_{12}\phi} &= (\cos \phi - \sigma_{12} \sin \phi) \sigma_1 (\cos \phi + \sigma_{12} \sin \phi) \\
 &= (\sigma_1 \cos \phi + \sigma_2 \sin \phi) (\cos \phi + \sigma_{12} \sin \phi) \\
 &= \sigma_1 (\cos^2 \phi - \sin^2 \phi) + 2\sigma_2 \sin \phi \cos \phi \\
 &= \sigma_1 \cos(2\phi) + \sigma_2 \sin(2\phi).
 \end{aligned} \tag{2.112}$$

A similar procedure with σ_2 would show that it is also rotated by 2ϕ in the forward direction, while σ_3 is left unchanged by the transformation. We conclude that Eq. (2.110) effectively rotates a as desired. Note that although we had specified a rotation around an axis aligned with σ_3 , this vector does not intervene in the final expression; it is the bivector defining a plane normal to the axis that plays the decisive role. The possibility of specifying an axis for a rotation is specific to 3 dimensions but the transformation rule of Eq. (2.110) is easily extrapolated to arbitrary dimensions because it specifies the plane where the rotation takes place. The multivector $R = \exp(\sigma_{12}\phi)$ is called a *rotor*. Rotors are generally defined as multivectors with scalar and bivector parts such that $R\tilde{R} = 1$. In Euclidean 3-dimensional space it is indifferent to write R^\dagger or \tilde{R} .

If we use B to represent a bivector and define its norm by $|B| = (B\tilde{B})^{1/2}$, a general 3-dimensional rotation is performed by the rotor

$$R = e^{-B/2} = \cos\left(\frac{|B|}{2}\right) - \frac{B}{|B|} \sin\left(\frac{|B|}{2}\right) \tag{2.113}$$

and the rotation transformation is encoded in the rule

$$a \mapsto Ra\tilde{R}. \quad (2.114)$$

The rotation angle is $|B|$ and the rotation plane is defined by B ; the same applies to spaces of arbitrary dimension. Given any two unit vectors n and m , making an angle ϕ between them, a rotor can be constructed multiplying one by the other

$$R = nm = n \cdot m + n \wedge m = \cos \phi - \frac{m \wedge n}{\sin \phi} \sin \phi. \quad (2.115)$$

We define the unitary bivector

$$\hat{B} = \frac{m \wedge n}{\sin \phi}, \quad (2.116)$$

so that the rotor becomes

$$R = e^{-\hat{B}\phi}. \quad (2.117)$$

The rotation transformation in Eq. (2.114) can be written with the new rotor form as

$$a \mapsto nmam, \quad (2.118)$$

showing that a rotation can be obtained by two successive reflections in many different combinations.

It is easy to prove that the rotation transformation in Eq. (2.114) preserves angles and lengths, as one expects from a rotation. Suppose that $a' = Ra\tilde{R}$ and $b' = Rb\tilde{R}$; then

$$\begin{aligned} a' \cdot b' &= \frac{1}{2} (Ra\tilde{R}Rb\tilde{R} + Rb\tilde{R}Ra\tilde{R}) \\ &= \frac{1}{2} R(ab + ba)\tilde{R} \\ &= a \cdot b R\tilde{R} \\ &= a \cdot b. \end{aligned} \quad (2.119)$$

We can also see that the inverse transformation is given by

$$a = \tilde{R}a'R. \quad (2.120)$$

The proof is given below

$$\tilde{R}a'R = \tilde{R}Ra\tilde{R} = a. \quad (2.121)$$

The rotor written as a sum of scalar and bivector parts has no particular geometric significance; in order to understand the geometric meaning of a rotor one must write it in exponential form. When this form is used it becomes very easy to understand that the bivector in the exponent defines both the plane and angle of rotation.

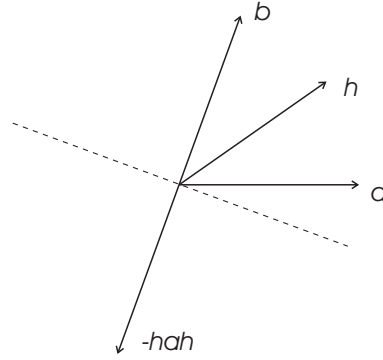


Figure 2.4: The vector a is rotated onto b by first reflecting in the plane perpendicular to h and then in the plane perpendicular to b . All the vectors have unit length.

We may need to find the rotor responsible for rotating a unit vector a into another unit vector b , leaving unchanged the vectors perpendicular to the plane $a \wedge b$. In order to do this we have to reflect a in a plane perpendicular to a vector h that bisects the angle between a and b ; this produces a vector opposite to b , which must then be further reflected in a plane perpendicular to b (see Fig. 2.4). The vector h can be found by

$$h = \frac{a + b}{|a + b|}. \quad (2.122)$$

The first reflection produces

$$-b = -hah. \quad (2.123)$$

This has to be combined with the reflection induced by b , so we have to consider the rotor

$$R = bh = \frac{1 + ba}{|b + a|} = \frac{1 + ba}{\sqrt{2(1 + b \cdot a)}}, \quad (2.124)$$

which represents a rotation in the $a \wedge b$ plane. Multiplying on the right by a we see that

$$Ra = \frac{a + b}{\sqrt{2(1 + b \cdot a)}} = a \frac{1 + ab}{\sqrt{2(1 + b \cdot a)}} = a\tilde{R}. \quad (2.125)$$

We can then write

$$Ra\tilde{R} = R^2a = a(\tilde{R})^2. \quad (2.126)$$

This is always possible for vectors in the plane of rotation and explains why in this case it is possible to perform single-sided rotations, as one usually does with complex numbers on the Argand plane. Returning to the polar form $R = \exp(-\hat{B}\theta/2)$, where \hat{B} represents the $a \wedge b$ plane, we see that

$$R^2 = e^{-\hat{B}\theta}, \quad (2.127)$$

so, the rotation of a onto b can be performed by

$$b = e^{-\hat{B}\theta} a = a e^{\hat{B}\theta}. \quad (2.128)$$

This formula does not generalize to higher dimensions and one should always use the double-sided rotation formula with half-angles in the exponent. If c is a vector perpendicular to the $a \wedge b$ plane

$$cR = c \frac{1 + ba}{\sqrt{2(1 + b \cdot a)}} = \frac{1 + ba}{\sqrt{2(1 + b \cdot a)}} c = cR, \quad (2.129)$$

so that

$$Rc\tilde{R} = cR\tilde{R} = c; \quad (2.130)$$

and the vector remains unchanged. The correct interpretation of the double-sided rotation law means that unit bivectors such as

$$\sigma_{12} = e^{\sigma_{12}\pi/2} \quad (2.131)$$

generate rotations of π and not of $\pi/2$. Bivectors square to -1 but because they act double sidedly, the rotor -1 is the identity operation. More generally, both R and $-R$ generate the same rotation, so there is a two-to-one map between rotors and rotations; this is known as *double-cover* representation of the rotation group.

We will now check how bivectors are affected by the rotation of their vectors; suppose that the two vectors forming bivector $B = a \wedge b$ are both rotated, as in

$$\begin{aligned} B' = a' \wedge b' &= \frac{1}{2} (Ra\tilde{R}Rb\tilde{R} - Rb\tilde{R}Ra\tilde{R}) \\ &= \frac{1}{2} R(ab - ba)\tilde{R} \\ &= Ra \wedge b\tilde{R} \\ &= RB\tilde{R}. \end{aligned} \quad (2.132)$$

Bivectors are rotated using exactly the same formula as vectors; the same could be concluded for all multivectors, as we shall see in the next chapter. In three dimensions we need only check that it is true for the pseudoscalar, which is very easy because i commutes with everything and should remain unchanged after rotation

$$Ri\tilde{R} = R\tilde{R}i = i. \quad (2.133)$$

We now have a method of rotating all geometric objects in three dimensions; this will later be extended to spaces of arbitrary dimension.

It may come as a surprise that rotors follow a different transformation law that sets them apart from geometric objects. To see how this works we consider a vector a which is first rotated to b by rotor R_1 and then to vector c by rotor R_2 . It is then

$$b = R_1 a \tilde{R}_1 \quad (2.134)$$

and

$$c = R_2 b \tilde{R}_2 = R_2 R_1 a \tilde{R}_1 \tilde{R}_2 = R_2 R_1 a (\widetilde{R_2 R_1}). \quad (2.135)$$

So we can define

$$R = R_2 R_1, \quad (2.136)$$

to obtain the transformation law

$$c = R a \tilde{R}. \quad (2.137)$$

Equation (2.136) establishes that *rotor composition* is achieved by a geometric product. Rotors form a group under the composition rule, because the product of two rotors is still a rotor. The rotor composition rule establishes also that rotors must be transformed by single-sided rule, as opposed to geometric objects, so they stand on their own as *operators*.

We can now give a very natural explanation to something that is usually thought to be specific to quantum mechanics: the transformation rule for spinors. Suppose that in Eq. (2.135) we make $R_2 = \exp(-B\pi)$, so that vector b is taken on a 2π excursion back to itself. The final rotor R is

$$R = e^{-B\pi} R_1 = -R_1. \quad (2.138)$$

The rotor has changed sign under a 2π rotation, just as spinors do. The sign change in a rotor is interpreted geometrically as a change in the direction of the rotation. To illustrate this, suppose we want to rotate σ_1 into σ_2 ; we have the option of going $\pi/2$ in a clockwise direction or $3\pi/2$ in an anti-clockwise direction. Either rotation is achieved by the rotor

$$R(\theta) = e^{-\sigma_{12}\theta/2}. \quad (2.139)$$

The rotor for the rotation in the clockwise direction is then

$$R(\pi/2) = \frac{1}{\sqrt{2}} (1 - \sigma_{12}). \quad (2.140)$$

The alternative in the anti-clockwise direction is associated with the rotor

$$R(-3\pi/2) = -\frac{1}{\sqrt{2}} (1 - \sigma_{12}) = -R(\pi/2). \quad (2.141)$$

Both rotors define the same rotation but their sign gives an indication about the handedness of the rotation.

3 Axiomatic approach to geometric algebra

In the previous chapter we looked at geometric algebra in two and three dimensions using an intuitive approach, in order to make the introduction of this subject as easy and gradual as possible to unfamiliar readers. We will now devote this chapter to a formal presentation, which will be also more general; we will address algebras of arbitrary dimension and signature, establishing the foundation for the applications that will follow in the forthcoming chapters. Naturally there will be overlap between the subjects covered in the present chapter and those of the previous one, but that is an inevitable consequence of looking at the particular before addressing the general; it should help clarify the meaning and use of several concepts.

3.1 Axiomatic development

From the previous chapter we should have now an intuitive feel for the elements of geometric algebra, the multivectors, and some of the properties of their product. We will here define a set of axioms and conventions that will enable us to manipulate them efficiently. We will follow the method of Hestenes and Sobczyk [2], raising the geometric product to primary status in the algebra and deriving inner and outer product properties from those of the geometric product; this approach is opposite to what we used in the intuitive approach but it is the most sensible to use for an axiomatic development.

We will consider a linear space over the field of reals. Vectors can be operated under the geometric product, which is governed by three main axioms.

1. The geometric product is associative:

$$a(bc) = (ab)c = abc. \quad (3.1)$$

2. The geometric product is distributive over addition:

$$a(b + c) = ab + ac. \quad (3.2)$$

3. The square of any vector is a real scalar: $a^2 \in \mathbb{R}$.

The last axiom forces the square of any vector to be real but not necessarily positive; this is crucial to accommodate mixed signature algebras, as we shall see. If we were to consider a linear space over the complex field, vectors with complex square would

be allowed; we do not do it in general because the algebras that would result would be isomorphic to some real algebras. It may be useful sometimes to call to the attention the fact that a given formulation can have an alternative in a complex algebra. One case where such dualism exists is with the complexified algebra of spacetime, which is isomorphic to the algebra of 5-dimensional spacetime; we will later take some advantage of that isomorphism. Nothing is said about the commutative properties of the geometric product and indeed it is always possible to represent multivectors of any geometric algebra by suitably chosen matrices, multiplied according to the rules of the usual matrix product, which is neither commutative or anti-commutative.

The complete algebra is generated by successive multiplication of vectors, thus producing multivectors with terms of every possible grade. We will usually write vectors in lower case and multivectors in upper case, with a few exceptions when referring to physical entities which have long been represented in a particular manner. 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. Any multivector can be written as a sum of geometric products of vectors. They can be multiplied using the geometric product and inherit properties 1. and 2. above. So, we can write for multivectors A , B and C

$$(AB)C = A(BC) = ABC \quad (3.3)$$

and

$$A(B + C) = AB + AC. \quad (3.4)$$

If we form the square of vector $a + b$ we know that the result must be a scalar, so

$$(a + b)^2 = a^2 + ab + ba + b^2. \quad (3.5)$$

Bringing the symmetrized product to the left hand side

$$ab + ba = (a + b)^2 - a^2 - b^2; \quad (3.6)$$

and so it must be a scalar. This fact is used to define the *inner* product

$$a \cdot b = \frac{1}{2} (ab + ba). \quad (3.7)$$

Since the geometric product need not be symmetric, the remaining part must be anti-symmetric and is defined as the *outer* product

$$a \wedge b = \frac{1}{2} (ab - ba) \quad (3.8)$$

and returns a bivector. The full geometric product is then the sum

$$ab = a \cdot b + a \wedge b. \quad (3.9)$$

The path we followed here is opposite to what we used in the previous chapter; we defined both inner and outer products from the decomposition of the geometric product instead of defining the former from the sum of the latter. We feel that the present approach is more elegant, because it involves a smaller number of axioms, although the alternative may appear more natural, since it uses more familiar concepts.

Our definition of the outer product of two vectors asserted that it returns a new entity called bivector (a grade-2 multivector). This is the first step in defining a grade operation for the entire algebra. We start by defining the outer product for an arbitrary number of vectors. The outer product of vectors a_1, \dots, a_r is denoted by $a_1 \wedge a_2 \wedge \dots \wedge a_r$ and is defined by the totally antisymmetrized sum of all geometric products

$$a_1 \wedge a_2 \wedge \dots \wedge a_r = \frac{1}{r!} \sum (-1)^\varepsilon a_{k_1} a_{k_2} \dots a_{k_r}. \quad (3.10)$$

The sum includes all the permutations of the k_1, k_2, \dots, k_r indices and ε is $+1$ for an even permutation of $1, 2, \dots, r$ and -1 for an odd permutation. For two vectors

$$a_1 \wedge a_2 = \frac{1}{2} (a_1 a_2 - a_2 a_1), \quad (3.11)$$

as established above.

The outer product is null if any two vectors are the same and so it also vanishes when one vector is a linear combination of the remaining vectors; the outer product therefore records the *dimensionality* of the object formed by a set of vectors. This is what we call the *grade* of that object, so we define the outer product of r linearly independent vectors as having grade r . Any multivector that can be written as an outer product of a set of vectors is called a *blade* and any multivector can be written as a sum of blades; these blades have definite grades and define the grades of the multivector.

We don't usually use Eq. (3.10) when studying blades, for every blade can more conveniently be written as a geometric product of orthogonal anticommuting vectors; the anticommutation of orthogonal vectors takes care of the antisymmetry of the outer product. In Euclidean space, given any two vectors a and b , we form the orthogonal to a by

$$b' = b - \lambda a. \quad (3.12)$$

We then see that

$$a \wedge (b - \lambda a) = a \wedge b - \lambda a \wedge a = a \wedge b. \quad (3.13)$$

The result is the same irrespective of the value of λ (see Fig. 3.1). The bivector encodes an oriented area; interchanging b and b' changes neither the magnitude or the orientation, so returns the same bivector. We now form the inner product

$$a \cdot b' = a \cdot (b - \lambda a) = a \cdot b - \lambda a^2. \quad (3.14)$$

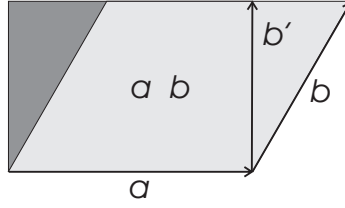


Figure 3.1: The outer product $a \wedge b$ is independent of the shape of the parallelogram formed by the two vectors; the two vectors b and b' generate the same outer product.

So we can set $\lambda = a \cdot b / a^2$; then we have $a \cdot b' = 0$ and it is

$$a \wedge b = a \wedge b' = ab'. \quad (3.15)$$

Proceeding in the same manner one is always able to construct any blade by a geometric product of suitably chosen orthogonal vectors. We have seen that

$$\begin{aligned} b' &= b - a^{-1}a \cdot b \\ &= a^{-1}(ab - a \cdot b) \\ &= a^{-1}(a \wedge b). \end{aligned} \quad (3.16)$$

This shows that $ab' = a \wedge b$ and is easily extended to higher grades.

The process works fine for Euclidean spaces but it cannot be applied in spaces of mixed signature, such as Minkowski spacetime, because in some cases it is $a^2 = 0$ and so the inverse of a does not exist. A procedure that works in every case is needed and is relatively straightforward. We start by defining the symmetric matrix

$$M_{ij} = a_i \cdot a_j. \quad (3.17)$$

This symmetric matrix can be diagonalized with an orthogonal matrix R_i^j ,

$$R_i^k M_{kl} R_j^{T\ l} = R_i^k R_j^l M_{kl} = \Lambda_{ij}. \quad (3.18)$$

where $R_i^{T\ j}$ is the transpose of R_i^j and Λ_{ij} is diagonal. We then define the new vectors

$$\sigma_i = R_i^j a_j. \quad (3.19)$$

The inner product of any two of these vectors satisfies

$$\begin{aligned} \sigma_i \cdot \sigma_j &= (R_i^k a_k) \cdot (R_j^l a_l) \\ &= R_i^k R_j^l M_{kl} \\ &= \Lambda_{ij}. \end{aligned} \quad (3.20)$$

The vectors $\sigma_1, \dots, \sigma_r$ are thus orthogonal and as such they anticommute. Their geometric product is totally antisymmetric and we can write

$$\begin{aligned}\sigma_1 \sigma_2 \cdots \sigma_r &= \sigma_1 \wedge \sigma_2 \wedge \cdots \wedge \sigma_r \\ &= (R_1^i a_i) \wedge \cdots \wedge (R_r^k a_k) \\ &= \det(R_i^j) a_1 \wedge a_2 \wedge \cdots \wedge a_r.\end{aligned}\tag{3.21}$$

The determinant appears due to the total antisymmetry of the expression and will be explained in Sec. 3.6. But R_i^j is an orthogonal matrix and so it must have determinant equal to ± 1 ; if it happens to be negative we can swap two of the σ_i to change the sign, so it is always possible to find a set of vectors such that

$$a_1 \wedge a_2 \wedge \cdots \wedge a_r = \sigma_1 \sigma_2 \cdots \sigma_r.\tag{3.22}$$

This result will be used several times in this chapter.

For a given space an orthonormal frame $\{\sigma_i\}$, $i = 1, \dots, n$ provides a natural way to represent vectors and indeed the entire algebra. We denote this algebra by \mathcal{G}_{pq} , $p+q = n$, where p is the number of basis vectors with positive square and q is the number of basis vectors with negative square. In this chapter most of the results will depend only on the total dimension of the space and not on the particular signature, so we will denote by \mathcal{G}_n all the algebras of dimension n . There are degenerate cases where some basis vectors have null square but we can ignore those for the time being. Multiplying the basis vectors among themselves we build up the basis for the whole algebra as

$$1, \quad \sigma_{ij} \ (i < j), \quad \sigma_{ijk} \ (i < j < k), \quad \dots\tag{3.23}$$

Each product in the set is totally antisymmetric because the basis vectors anticommute. The product of r distinct basis vectors is by definition a grade- r multivector. The basis (3.23) therefore provides a basis for all grade- r subspaces of \mathcal{G}_n ; we denote these subspaces as \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 ; these are given by the binomial coefficients, so

$$\dim(\mathcal{G}_n^r) = \binom{n}{r};\tag{3.24}$$

the total dimension of the algebra is thus

$$\dim(\mathcal{G}_n) = \sum_{r=0}^n \binom{n}{r} = 2^n.\tag{3.25}$$

For example, we have already seen that in two dimensions the algebra contains terms of grades 0, 1, 2, with each subspace having dimension 1, 2, 1, respectively, and the total

dimension of the algebra being 4. Similarly, in three dimensions there are grades 0, 1, 2, 3 and the respective subspace dimensions are 1, 2, 2, 1; the total dimension is 8.

Having established the grade operation for an algebra \mathcal{G}_n , an arbitrary multivector can be decomposed into a sum of pure grade terms

$$A = \langle A \rangle_0 + \langle A \rangle_1 + \cdots + \langle A \rangle_n = \sum_{r=0}^n \langle A \rangle_r. \quad (3.26)$$

The operator $\langle A \rangle_r$ returns the grade- r terms in the argument. Multivectors containing terms of only one grade are called *homogeneous* and their notation is sometimes simplified to A_r , that is,

$$\langle A_r \rangle_r = A_r. \quad (3.27)$$

We will only use this notation when there is no danger of confusion with frame indices, like in $\{\sigma_i\}$. The grade-0 terms in \mathcal{G}_n are real scalars that commute with every element in the algebra. We use the simplified notation

$$\langle A \rangle = \langle A \rangle_0 \quad (3.28)$$

for the operation that consists in taking the scalar part of A .

One might have thought that an homogeneous multivector would always be a blade, but this must not be so in dimensions higher than three. Suppose that $\{\sigma_i\}$ form an orthonormal basis for the Euclidean algebra \mathcal{G}_4 . There are six independent basis bivectors in the algebra, from which we can build terms like

$$B = \alpha \sigma_{12} + \beta \sigma_{34}, \quad (3.29)$$

where α and β are scalars. B is a pure bivector, so it is homogeneous, but there is no way to obtain B from the outer product of two vectors and so it is not a blade. The reason why this is not possible is that the planes $\sigma_1 \wedge \sigma_2$ and $\sigma_3 \wedge \sigma_4$ do not share a common vector.

We have already established that for vectors

$$ab = 2a \cdot b - ba; \quad (3.30)$$

this formula will be used several times below. Consider the case of a geometric product of vectors; we can write successively

$$\begin{aligned} aa_1 a_2 \cdots a_r &= 2a \cdot a_1 a_2 \cdots a_r - a_1 a a_2 \cdots a_r \\ &= 2a \cdot a_1 a_2 \cdots a_r - 2a \cdot a_2 a_1 a_3 \cdots a_r + a_1 a_2 a a_3 \cdots a_r \\ &= 2 \sum_{k=1}^r (-1)^k a \cdot a_k a_1 a_2 \cdots \check{a}_k \cdots a_r + (-1)^r a_1 a_2 \cdots a_r a, \end{aligned} \quad (3.31)$$

where the check on \check{a}_k means that a_k is missing from the product. Notice that we follow the precedence convention of the dot product over the geometric product.

If the vectors a_1, \dots, a_r are replaced by a set of anticommuting vectors $\sigma_1, \dots, \sigma_r$, we find that

$$\frac{1}{2} [a\sigma_1\sigma_2 \cdots \sigma_r - (-1)^r \sigma_1\sigma_2 \cdots \sigma_r a] = \sum_{k=1}^r (-1)^{k+1} a \cdot \sigma_k \sigma_1\sigma_2 \cdots \check{\sigma}_k \cdots \sigma_r. \quad (3.32)$$

The right-hand side contains a sum of terms of grade $r-1$. The left-hand side is a sum of two terms where a product of r anticommuting vectors is multiplied by a on both sides. Since any grade- r multivector can be written as a sum of terms formed from the product of anticommuting vectors, we apply this result to an homogeneous multivector to define the inner product between a vector and a grade- r multivector

$$a \cdot A_r = \frac{1}{2} [aA_r - (-1)^r A_r a]. \quad (3.33)$$

The inner product between a vector and a grade- r multivector returns a grade- $(r-1)$ multivector.

The remaining part of the product $a \langle A \rangle_r$ is called the exterior or outer product. In the case where $r = 1$ we have the product of two vectors; the remaining part is the antisymmetric product, and so it is a bivector. We will prove that in general

$$\frac{1}{2} [a(a_1 \wedge a_2 \wedge \cdots \wedge a_r) + (-1)^r (a_1 \wedge a_2 \wedge \cdots \wedge a_r)a] = a \wedge a_1 \wedge a_2 \wedge \cdots \wedge a_r. \quad (3.34)$$

First we use Eq. (3.22) to write the blade as a product of anticommuting vectors

$$\frac{1}{2} [a\sigma_1\sigma_2 \cdots \sigma_r + (-1)^r \sigma_1\sigma_2 \cdots \sigma_r a] = a \wedge \sigma_{12\dots r}. \quad (3.35)$$

For $r = 1$ the right-hand side only defines the bivector $a \wedge \sigma_1$. For $r > 1$ we proceed with

$$a \wedge \sigma_{12\dots r} = \frac{1}{r+1} a\sigma_{12\dots r} + \frac{1}{r+1} \sum_{k=1}^r (-1)^k \sigma_k a \wedge \sigma_{1\dots \check{k} \dots r}. \quad (3.36)$$

This result is established writing out all the terms in the antisymmetric product and collecting those that start with the same vector. We now assume that Eq. (3.35) applies

to an $r - 1$ blade and expand the term inside the sum as follows:

$$\begin{aligned}
 & \sum_{k=1}^r (-1)^k \sigma_k a \wedge \sigma_{1 \dots \check{k} \dots r} \\
 &= \frac{1}{2} \sum_{k=1}^r (-1)^k \sigma_k [a \sigma_{1 \dots \check{k} \dots r} + (-1)^{r-1} \sigma_{1 \dots \check{k} \dots r} a] \\
 &= \frac{1}{2} \sum_{k=1}^r (-1)^k \sigma_k a \sigma_{1 \dots \check{k} \dots r} + \frac{r}{2} (-1)^r \sigma_{1 \dots r} a \\
 &= \sum_{k=1}^r (-1)^k (\sigma_k \cdot a) \sigma_{1 \dots \check{k} \dots r} + \frac{r}{2} [a \sigma_{1 \dots r} + (-1)^r \sigma_{1 \dots r} a] \\
 &= \frac{r-1}{2} a \sigma_{1 \dots r} + \frac{r+1}{2} (-1)^r \sigma_{1 \dots r} a.
 \end{aligned} \tag{3.37}$$

This derivation makes use of Eq. (3.32). Substituting in Eq. (3.36) provides the proof of Eq. (3.35) for a grade- r blade, assuming it is true for grade $r - 1$; since it had been proven for grade-1 in Eq. (3.34) it is then proven for all blades and for all multivectors.

The outer product definition is extended by

$$a \wedge A_r = \frac{1}{2} [a A_r + (-1)^r A_r a]. \tag{3.38}$$

And the geometric product between a vector and a grade- r multivector is

$$a A_r = a \cdot A_r + a \wedge A_r, \tag{3.39}$$

which is a very natural extension of the geometric product between two vectors. Equation (3.38) is an example of how the geometric product can help simplify some calculations; while on the left-hand side we have a totally antisymmetrized product of $r + 1$ vectors, on the right-hand side we have only two products. The product $a A_r$ has $r - 1$ and $r + 1$ grades, that is

$$a A_r = \langle a A_r \rangle_{r-1} + \langle a A_r \rangle_{r+1}, \tag{3.40}$$

where

$$\langle a A_r \rangle_{r-1} = a \cdot A_r, \quad \langle a A_r \rangle_{r+1} = a \wedge A_r. \tag{3.41}$$

The inner product by a vector reduces the grade by 1 and the outer product increases the grade by the same amount.

A homogeneous multivector can be written as a sum of blades, each of which can be written as a geometric product of anticommuting vectors. Therefore, the product of two homogeneous multivectors can be decomposed as

$$A_r B_s = \langle A_r B_s \rangle_{|r-s|} + \langle A_r B_s \rangle_{|r-s|+2} + \langle A_r B_s \rangle_{r+s}. \tag{3.42}$$

The inner and outer products are reserved for the lower and higher grade parts, respectively

$$\begin{aligned} A_r \cdot B_s &= \langle A_r B_s \rangle_{|r-s|}, \\ A_r \wedge B_s &= \langle A_r B_s \rangle_{r+s}. \end{aligned} \quad (3.43)$$

These definitions are compatible with the earlier ones for products involving vectors (grade-1 multivectors). We can check that the outer product remains associative by

$$(A_r \wedge B_s) \wedge C_t = \langle A_r B_s \rangle_{r+s} \wedge C_t = \langle (A_r B_s) C_t \rangle_{r+s+t}. \quad (3.44)$$

Since the geometric product is associative so must be the outer product

$$\langle (A_r B_s) C_t \rangle_{r+s+t} = \langle A_r B_s C_t \rangle_{r+s+t} = A_r \wedge B_s \wedge C_t. \quad (3.45)$$

Equation (3.32) can be generalized to a useful formula; we start by writing

$$a \cdot (a_1 \wedge a_2 \wedge \cdots \wedge a_r) = a \cdot \langle a_1 a_2 \cdots a_r \rangle_r, \quad (3.46)$$

where a_1, a_2, \dots, a_r are a general set of vectors. The geometric product $a_1 a_2 \cdots a_r$ can only contain terms of grades $r, r-2, \dots$, so

$$\begin{aligned} \frac{1}{2} [a a_1 a_2 \cdots a_r - (-1)^r a_1 a_2 \cdots a_r a] \\ = a \cdot \langle a_1 a_2 \cdots a_r \rangle_r + a \cdot \langle a_1 a_2 \cdots a_r \rangle_{r-2} + \cdots \end{aligned} \quad (3.47)$$

What we are after is the $r-1$ grade part, so

$$a \cdot (a_1 \wedge a_2 \wedge \cdots \wedge a_r) = \frac{1}{2} \langle a a_1 a_2 \cdots a_r - (-1)^r a_1 a_2 \cdots a_r a \rangle_{r-1}. \quad (3.48)$$

We can apply Eq. (3.31) inside the grade projection operator to get

$$\begin{aligned} a \cdot (a_1 \wedge a_2 \wedge \cdots \wedge a_r) &= \sum_{k=1}^r (-1)^{k+1} a \cdot a_k \langle a_1 \cdots \check{a}_k \cdots a_r \rangle_{r-1} \\ &= \sum_{k=1}^r a \cdot a_k a_1 \wedge \cdots \wedge \check{a}_k \wedge \cdots \wedge a_r. \end{aligned} \quad (3.49)$$

The importance of this formula can be seen in the simplest cases

$$\begin{aligned} a \cdot (a_1 \wedge a_2) &= a \cdot a_1 a_2 - a \cdot a_2 a_1, \\ a \cdot (a_1 \wedge a_2 \wedge a_3) &= a \cdot a_1 a_2 \wedge a_3 - a \cdot a_2 a_1 \wedge a_3 + a \cdot a_3 a_1 \wedge a_2. \end{aligned} \quad (3.50)$$

The first case had already been found in the geometric algebra of space where it was written in Eq. (2.53).

We are now in position to look at the first of the two conjugation operations that were defined in chapter 2. The *reversion* operation reverses the order of vectors in every product, so

$$\widetilde{ab \cdots c} = c \cdots ba. \quad (3.51)$$

For each blade in a multivector, reversion can be performed by a series of swaps, each swap resulting in a sign change. The first vector has to swap past $r - 1$ vectors, the second past $r - 2$ vectors and so forth. The final result is

$$\tilde{A}_r = (-1)^{r(r-1)/2} A_r. \quad (3.52)$$

The reversing operation changes the sign of orders 2, 3, 6, 7, ..., alternating with a period of 4 orders. Considering only the scalar part of two grade- r multivectors, we see that

$$\langle A_r B_r \rangle = \langle \widetilde{A_r B_r} \rangle = \langle \tilde{B}_r \tilde{A}_r \rangle = (-1)^{r(r-1)} \langle B_r A_r \rangle = \langle B_r A_r \rangle, \quad (3.53)$$

so, for general A and B ,

$$\langle AB \rangle = \langle BA \rangle. \quad (3.54)$$

For an arbitrary number of multivectors

$$\langle AB \cdots C \rangle = \langle CA \cdots B \rangle. \quad (3.55)$$

There is a cyclic ordering property in the scalar part of a product, which is very useful in the manipulation of expressions. The product (3.54) is called *scalar product* and is sometimes denoted $A * B$. So, we write

$$A * B = \langle AB \rangle. \quad (3.56)$$

The last product we will consider is the *commutator product*, which has considerable importance. This is denoted by \times , the same symbol we used before for the cross product of vectors. We are no longer interested in this latter product and so there is no cause for confusion. The definition of the commutator product is

$$A \times B = \frac{1}{2} (AB - BA). \quad (3.57)$$

This differs from the commutator of two operators usually used in quantum mechanics by the factor $1/2$. The commutator product satisfies the *Jacobi identity*

$$A \times (B \times C) + B \times (C \times A) + C \times (A \times B) = 0, \quad (3.58)$$

which is easily verified by expanding all the products.

The commutator product of two vectors coincides with the outer product and the commutator product between a vector a and a bivector B coincides with the inner product, that is

$$B \times a = \frac{1}{2} (Ba - aB) = B \cdot a, \quad (3.59)$$

and returns another vector. Now consider the product of a bivector and a blade formed by anticommuting vectors

$$\begin{aligned} B\sigma_{12\dots r} &= 2(B \times \sigma_1)\sigma_{2\dots r} + \sigma_1 B\sigma_{2\dots r} \\ &= 2(B \times \sigma_1)\sigma_{2\dots r} + \dots + 2\sigma_1 \dots (B \times \sigma_r) + \sigma_{12\dots r} B. \end{aligned} \quad (3.60)$$

It follows that

$$B \times \sigma_{12\dots r} = \sum_{i=1}^r \sigma_1 \dots (B \cdot \sigma_i) \dots \sigma_r. \quad (3.61)$$

The right-hand side contains terms of 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

$$\begin{aligned} \widetilde{(B \times A_r)} &= \frac{1}{2} (\widetilde{BA_r - A_r B}) \\ &= \frac{1}{2} (-\tilde{A_r} B + B \tilde{A_r}) \\ &= (-1)^{r(r-1)/2} B \times A_r. \end{aligned} \quad (3.62)$$

We conclude that $B \times A_r$ has the same reversion properties of A_r . We know, however, that multivectors of grades r and $r-2$ always behave differently under reversion. The commutator product of Eq. (3.61) must then result in a homogeneous multivector of grade r ; there can be no $r-2$ terms. Since this has been proven for any grade- r basis element, it must be true for any multivector. That is

$$B \times A_r = \langle B \times A_r \rangle_r. \quad (3.63)$$

The commutator product with a bivector is therefore a grade preserving operation. Consequently the commutator product of two bivectors must result in a third bivector, which is relevant for the incorporation of Lie groups into geometric algebra.

A similar argument shows that the symmetric product with a bivector must raise or lower the grade by 2. In summary this is

$$BA_r = \langle BA_r \rangle_{r-2} + \langle BA_r \rangle_r + \langle BA_r \rangle_{r+2}, \quad (3.64)$$

where

$$\frac{1}{2} (BA_r - A_r B) = B \times A_r \quad (3.65)$$

and

$$\frac{1}{2}(BA_r + A_rB) = B \cdot A_r + B \wedge A_r. \quad (3.66)$$

In the equations above it is assumed that $r > 1$.

The outer product of n vectors defines a grade- n blade; for a given linear space the highest grade element is unique, up to magnitude and sign. The outer product of n vectors is therefore a multiple of the *unit pseudoscalar* for \mathcal{G}_n . This is denoted by I , but if it happens to square to -1 and to commute with all vectors it can also be denoted by i . Because of its definition the unit pseudoscalar verifies

$$|I^2| = 1. \quad (3.67)$$

The sign of I^2 will depend on the particular space being considered. Later in the book we will be concentrating on 5-dimensional spacetime algebra $\mathcal{G}_{4,1}$ and this is one case where it becomes convenient to denote the unit pseudoscalar by i ; another such case is the algebra of space $\mathcal{G}_{3,0}$.

The unit pseudoscalar defines an *orientation*. For any ordered set of n vectors, their outer product will have either the sign of I or the opposite sign. Those with the same sign are said to have a positive orientation while those with opposite sign have a negative orientation. In 3 dimensions it is common practice to make the positive orientation coincide with a right-handed one but in higher dimensions there is a choice; one chooses one of the orientations as positive and adheres to it. Swapping two vectors in an outer product always changes the orientation of the set.

The product of the grade- n pseudoscalar I with a grade- r multivector A_r produces a grade- $(n-r)$ multivector. This operation is called *duality*; when performed on a grade- r blade it returns the *orthogonal complement* of the blade, that is, the blade from the space of vectors not contained in the original blade. Every blade acts as pseudoscalar for the space of vectors spanned by its generating vectors. So, when working in 3 dimensions, σ_{12} can always be treated as pseudoscalar for the algebra of the plane.

In spaces of odd dimension I commutes with all vectors and so it commutes with every element in the algebra. In spaces of even dimension I anticommutes with vectors and all odd-grade multivectors. I will always commute with even-grade multivectors. This can be summarized in the formula

$$IA_r = (-1)^{r(n-1)} A_r I. \quad (3.68)$$

The unit pseudoscalar can be used to interchange inner and outer products; for vector

products it is

$$\begin{aligned}
 a \cdot (A_r I) &= \frac{1}{2} [a A_r I - (-1)^{n-r} A_r I a] \\
 &= \frac{1}{2} [a A_r I - (-1)^{n-r} (-1)^{n-1} A_r a I] \\
 &= \frac{1}{2} [a A_r + (-1)^r A_r a] I \\
 &= a \wedge A_r I.
 \end{aligned} \tag{3.69}$$

For general products of homogeneous multivectors we have

$$\begin{aligned}
 A_r \cdot (B_s I) &= \langle A_r B_s \rangle_{|r-(n-s)|} \\
 &= \langle A_r B_s I \rangle_{n-(r+s)} \\
 &= \langle A_r B_s \rangle_{r+s} I \\
 &= A_r \wedge B_s I.
 \end{aligned} \tag{3.70}$$

This sort of manipulation is frequently very useful in derivations. In those cases where I can be replaced by i it is sometimes useful to think of the algebra as a complex structure, by associating grade- r and grade- $(n-r)$ blades. When this is done a *complex scalar* is a two-part multivector of the type $\alpha + i\beta$, with α and β scalars. Similarly, a *complex vector* is a two-part multivector of the type $a + ib$, with a and b vectors.

The second form of conjugation, known as *Hermitian conjugation* or simply *conjugation*, differs from reversion in spaces with mixed signature. Conjugation involves changing the sign to vectors belonging to the negative subspace and then performing reversion. As an example, Minkowski spacetime can be associated with the algebra \mathcal{G}_{13} , with one basis vector in the positive subspace and three basis vectors in the negative subspace. If we denote the basis vectors by $\{\sigma_\mu\}$, according to our indexing conventions, the conjugates are as follows

$$\begin{aligned}
 \sigma_0^\dagger &= \sigma_0, \\
 \sigma_m^\dagger &= -\sigma_m.
 \end{aligned} \tag{3.71}$$

So, a general vector $a = a^\mu \sigma_\mu$ will form the conjugate

$$a^\dagger = (a^0 \sigma_0 + a^m \sigma_m) = a^0 \sigma_0 - a^m \sigma_m. \tag{3.72}$$

A multivector H is called *Hermitian* if it is equal to its conjugate, that is

$$H = H^\dagger. \tag{3.73}$$

When an isomorphism is considered between a geometric algebra and an adequate matrix algebra, the conjugation operation performed in geometric algebra becomes equivalent to the Hermitian conjugation performed in the matrix algebra.

3.2 Reflections, rotations and boosts

Reflections were studied in 3 dimensions in chapter 2 and they were found to be governed by the formula

$$a \mapsto a' = -hah, \quad (3.74)$$

with h a unit vector perpendicular to the plane of reflection. This formula can be extended to an arbitrary number of Euclidean dimensions. Provided that $h^2 = 1$, we see that h and the parallel component a_{\parallel} of a are transformed as

$$h \mapsto -h h h = -h, \quad (3.75)$$

while the perpendicular component a_{\perp} is left unchanged

$$a_{\perp} \mapsto -h a_{\perp} h = a_{\perp} h h = a_{\perp}. \quad (3.76)$$

So, the transformation of any vector changes the sign of the parallel component and leaves unchanged the perpendicular component, which is exactly what we expect from a reflection on an hyperplane perpendicular to h .

Things are different in spaces with negative or mixed signature. Suppose u is a unit vector belonging to a negative signature space or to the negative subspace of a mixed signature space. We then have $u^2 = -1$ and it transforms under reflection as

$$u \mapsto -u u u = u. \quad (3.77)$$

The perpendicular component of some vector a transforms as

$$a_{\perp} = -u a_{\perp} u = a_{\perp} u^2 = -a_{\perp}. \quad (3.78)$$

That is, the perpendicular component of a vector changes sign under reflection, while the parallel component remains unchanged. We will have to pay attention to this unexpected behaviour when working in spaces of mixed signature.

Returning to Euclidean spaces, two successive reflections in hyperplanes perpendicular to h_1 and h_2 define a rotation in the $h_1 \wedge h_2$ hyperplane. The rotation is generated by the rotor

$$R = h_1 h_2 = e^{-\hat{B}\theta/2}, \quad (3.79)$$

where

$$\cos(\theta/2) = h_1 \cdot h_2, \quad \hat{B} = \frac{h_1 \wedge h_2}{\sin(\theta/2)}. \quad (3.80)$$

The rotation formula is naturally

$$a \mapsto a' = R a \tilde{R}. \quad (3.81)$$

The combination of two rotations results in a third rotation and so rotations form a group; the same must be true for rotors. If R_1 and R_2 are rotors, the combined rotations transform a as

$$a \mapsto R_2(R_1 \tilde{R}_1) \tilde{R}_2 = R_2 R_1 a \tilde{R}_1 \tilde{R}_2. \quad (3.82)$$

We therefore define the rotor

$$R = R_2 R_1, \quad (3.83)$$

so that the combined rotation can be described by $Ra\tilde{R}$, as usual. The product R of m rotors is still a rotor, and in general it will consist of the geometric product of an even number of unit vectors,

$$R = h_1 h_2 \cdots h_{m-1} h_m. \quad (3.84)$$

We can use this as a definition for a rotor. The reversed rotor is then

$$\tilde{R} = h_m h_{m-1} \cdots h_2 h_1. \quad (3.85)$$

The result of the map $a \mapsto Ra\tilde{R}$ returns a vector for any vector a , since

$$Ra\tilde{R} = h_1 h_2 \cdots [h_{m-1} (h_m a h_m) h_{m-1}] \cdots h_2 h_1 \quad (3.86)$$

and we see that each time the vector is sandwiched with a unit vector on either side it returns a new vector, so the whole process must also return a vector.

We can see immediately that

$$R\tilde{R} = h_1 h_2 \cdots h_{m-1} h_m h_m h_{m-1} \cdots h_2 h_1 = 1 = \tilde{R}R. \quad (3.87)$$

In Euclidean or negative signature spaces this condition is always verified but in mixed signature spaces the product of an even number of unit vectors can return -1 ; such a product does not define a rotor and we will consider such cases at the end of this section. Since the product of rotors is a new rotor, the rotors form a *group*, similar but not identical to the rotation matrices group, due to the already mentioned two-to-one mapping between rotations and rotors. We will come back to this point several times along the book.

Every rotor can be written as the exponential of a bivector

$$R = e^{-B/2}. \quad (3.88)$$

The bivector B defines the hyperplane or hyperplanes where the rotation takes place, the angles and the *direction* of the rotation. The rotation formula for vector a is

$$a \mapsto e^{-B/2} a e^{B/2}. \quad (3.89)$$

This result has been proven for 3 dimensions in Eq. (2.113) and following; the proof for general Euclidean spaces is identical.

In mixed signature spaces the bivector part of a rotor can be formed by the product of vectors belonging to subspaces with opposite signature. If h and u are unit vectors belonging to positive and negative subspaces of a mixed signature space, we have $h^2 = 1$ and $u^2 = -1$; note that the two vectors are necessarily orthogonal because they belong to disjunct subspaces, so that $u \wedge h = uh$. For $R = \alpha + \beta uh$, with α and β scalars, to be a rotor, we must have

$$(\alpha + \beta uh)(\alpha + \beta hu) = \alpha^2 - \beta^2 = 1. \quad (3.90)$$

So, we can set

$$\cosh \theta/2 = \alpha, \quad \sinh(\theta/2) = \beta \quad (3.91)$$

and the rotor becomes

$$R = e^{uh\theta/2}. \quad (3.92)$$

This is still in the form of Eq. (3.88) and rotors in mixed signature spaces form a group, in the same way as they do in Euclidean spaces. The rotor group of spaces with just one dimension in one of their subspaces is usually known as the *Lorentz group*, particularly in the case when the global dimension is 4. Rotations performed by rotors that can be written as exponentials of positive-square grade-2 blades are sometimes called *boosts* and the corresponding rotors can also be called *boosters*.

If every vector in a blade undergoes the same rotation we are dealing with a rotation of an area, a volume or even a higher order hypervolume. The r -blade A_r can be written

$$A_r = a_1 \wedge \cdots \wedge a_r = \frac{1}{r!} \sum (-1)^\varepsilon a_{k_1} a_{k_2} \cdots a_{k_r}, \quad (3.93)$$

where the sum runs over all the permutations. When every rotor in a geometric product is rotated, the result is

$$\begin{aligned} (Ra_1 \tilde{R})(Ra_2 \tilde{R}) \cdots (Ra_r \tilde{R}) &= Ra_1 \tilde{R} Ra_2 \tilde{R} \cdots Ra_r \tilde{R} \\ &= Ra_1 a_2 \cdots a_r \tilde{R}. \end{aligned} \quad (3.94)$$

This must be true for every term in the right-hand side of Eq. (3.93), so the transformation law for blade A_r is

$$A_r \mapsto RA_r \tilde{R}. \quad (3.95)$$

Blades transform with the same law as vectors and the same applies to general multivectors; every geometric object in the algebra is transformed with the same rule.

3.3 Basis frames and components

The basis for the vectors in a geometric algebra can be formed by any set of linearly independent vectors in number equal to the dimension of the algebra. Such sets are

usually referred to as *frames*. The basis for the whole algebra is built from the basis vectors by successive use of the outer product. We use the symbols g_1, \dots, g_n or $\{g_\kappa\}$ to denote a frame for n -dimensional space. We reserve the symbols $\{\sigma_n\}$ for the particular cases where the basis vectors are orthogonal, and so anticommuting.

The volume element for the $\{g_\kappa\}$ frame is defined by

$$E_n = g_1 \wedge g_2 \wedge \dots \wedge g_n. \quad (3.96)$$

The grade- n multivector E_n is a pseudoscalar, multiple of the unit pseudoscalar for the space spanned by $\{g_\kappa\}$. We are certain that $E_n \neq 0$ because all the frame vectors are linearly independent. Associated with any frame $\{g_\kappa\}$ there is a reciprocal frame $\{g^\kappa\}$ such that

$$g^\kappa \cdot g_\lambda = \delta^\kappa_\lambda, \quad \forall \kappa, \lambda = 1 \dots n. \quad (3.97)$$

The Kronecker delta δ^κ_λ has value +1 for $\kappa = \lambda$ and is zero otherwise. In order to determine $\{g^\kappa\}$ we first form the symmetric tensor

$$g_{\kappa\lambda} = g_\kappa \cdot g_\lambda. \quad (3.98)$$

Note that the double index is here used to refer to a scalar resulting from the inner product of two frame vectors, while the convention for orthogonal basis establishes that $\sigma_{\kappa\lambda}$ is the geometric product of two frame vectors. The symmetric tensor $g_{\kappa\lambda}$ is called the *metric tensor*. Next we define $g^{\kappa\lambda}$ as the inverse of $g_{\kappa\lambda}$, that is

$$g^{\kappa\alpha} g_{\alpha\lambda} = \delta^\kappa_\lambda. \quad (3.99)$$

Using the definition (3.98) we get

$$(g^{\kappa\alpha}) g_\alpha \cdot g_\lambda = \delta^\kappa_\lambda. \quad (3.100)$$

Now, comparing with Eq. (3.97) we can write

$$g^{\kappa\alpha} g_\alpha = g^\kappa. \quad (3.101)$$

It would be easy to verify that the following formula is also true

$$g_{\kappa\alpha} g^\alpha = g_\kappa. \quad (3.102)$$

The basis vectors $\{g_\kappa\}$ are a linearly independent set and so are the reciprocal frame vectors $\{g^\kappa\}$; any vector a can thus be written uniquely in terms of either set as

$$a = a^\kappa g_\kappa = a_\kappa g^\kappa. \quad (3.103)$$

The set of scalars (a^1, a^2, \dots, a_n) are the *components* of vector a in the frame $\{g_\kappa\}$ and naturally the set of scalars (a_1, a_2, \dots, a_n) are the components of the same vector in the frame $\{g^\kappa\}$. To find the components we form the inner products

$$a \cdot g^\kappa = a^\lambda g_\lambda \cdot g^\kappa = a^\lambda \delta^\kappa_\lambda = a^\kappa \quad (3.104)$$

and

$$a \cdot g_\kappa = a_\lambda g^\lambda \cdot g_\kappa = a_\lambda \delta_\kappa^\lambda = a_\kappa. \quad (3.105)$$

These formulae explain the rationale of the indexing scheme: for basis vectors we use lower indices for the main frame and upper indices for the reciprocal frame; conversely vectors' components get upper indices when referred to the main frame and lower indices when referred to the reciprocal frame.

Combining Eqs. (3.104) and (3.105) we can write

$$a \cdot g_\kappa g^\kappa = a \cdot g^\kappa g_\kappa = a, \quad (3.106)$$

which must be true for any vector in the $\{g_\kappa\}$ space. This result can also be applied to general multivectors; for a bivector $a \wedge b$ it is

$$\begin{aligned} g_\kappa g^\kappa \cdot (a \wedge b) &= \frac{1}{2} g_\kappa g^\kappa \cdot (ab - ba) \\ &= g_\kappa g^\kappa \cdot ab - g_\kappa g^\kappa \cdot ba \\ &= ab - ba \\ &= 2a \wedge b. \end{aligned} \quad (3.107)$$

For an arbitrary grade- r multivector it is then

$$g_\kappa g^\kappa \cdot A_r = r A_r. \quad (3.108)$$

But we know that $g_\kappa g^\kappa = n$, so we can also write

$$g_\kappa g^\kappa \wedge A_r = g_\kappa (g^\kappa A_r - g^\kappa \cdot A_r) = (n - r) A_r. \quad (3.109)$$

Since $g^\kappa A_r = (-1)^r A_r g^\kappa$ we can write

$$g_\kappa A_r g^\kappa = (-1)^r (g_\kappa g^\kappa \wedge A_r + g_\kappa g^\kappa \cdot A_r). \quad (3.110)$$

Then, using Eqs. (3.109) and (3.109),

$$g_\kappa A_r g^\kappa = (-1)^r (n - 2r) A_r. \quad (3.111)$$

The $\{g_k\}$ basis is easily extended to form a basis for the whole algebra generated by the basis vectors. Any multivector A can be decomposed into a set of components through

$$A_{\iota, \kappa, \dots, \lambda} = (g_\lambda \wedge \dots \wedge g_\kappa \wedge g_\iota) \cdot A \quad (3.112)$$

and

$$A = \sum_{\iota < \kappa < \lambda} A_{\iota \kappa \dots \lambda} g^\iota \wedge g^\kappa \wedge \dots \wedge g^\lambda. \quad (3.113)$$

The components $A_{\iota,\kappa,\dots,\lambda}$ are totally antisymmetric on all indices and A is usually called an *antisymmetric tensor*.

It will always be possible to find an orthonormed frame $\{\sigma_\kappa\}$ spanning the same space as $\{g_\kappa\}$, so any vector a can be expressed in terms of components in both frames. We then get

$$a = \sigma_\lambda \sigma^\lambda \cdot a = \sigma_\lambda \sigma^\lambda \cdot g_\kappa g^\kappa \cdot a = g_\kappa g^\kappa \cdot a, \quad (3.114)$$

because

$$g_\kappa = \sigma_\lambda \sigma^\lambda \cdot g_\kappa. \quad (3.115)$$

We define the transformation matrix

$$f^\lambda{}_\kappa = \sigma^\lambda \cdot g_\kappa \quad (3.116)$$

and replace above to get

$$g_\kappa = \sigma_\lambda f^\lambda{}_\kappa. \quad (3.117)$$

Later in this chapter we will extend the transformation matrix to general cases of coordinate changes.

The orthonormed frame $\{\sigma_\kappa\}$ has an associated reciprocal frame $\{\sigma^\kappa\}$ such that

$$\sigma^\kappa \cdot \sigma_\lambda = \delta^\kappa{}_\lambda, \quad \forall \kappa, \lambda = 1 \dots n. \quad (3.118)$$

Because the frame vectors are orthonormed, $\sigma^\kappa = \sigma_\kappa$ for all basis vectors belonging to a positive signature subspace and $\sigma^\kappa = -\sigma_\kappa$ for all basis vectors belonging to a negative signature subspace. The non-orthonormed reciprocal frame vectors can be expressed as a linear combination of the orthonormed reciprocal frame ones, so we have

$$g^\kappa = f^\kappa{}_\lambda \sigma^\lambda. \quad (3.119)$$

Equation (3.97) implies that $f^\kappa{}_\lambda$ is the inverse of $f^\lambda{}_\kappa$; this can be expressed as

$$f^\alpha{}_\lambda f^\lambda{}_\kappa = \delta^\alpha{}_\kappa. \quad (3.120)$$

In a forthcoming chapter we will define another matrix, called *refractive index*, which must not be confused with the present transformation matrix.

3.4 Linear algebra

Linear algebra is the study of *linear functions*. We will use sans-serif symbols to denote linear functions, in order to make it easier to distinguish these from their multivector arguments. The theory developed in this section applies to spaces of arbitrary dimension and signature and we will assume the reader has some knowledge of linear transformations in matrix formalism. Suppose that we are discussing a transformation F which maps

vectors linearly into the same space, that is, F takes a vector in a given space to produce another vector linearly in the same space. The linearity is expressed by the relation

$$F(\lambda a + \mu b) = \lambda F(a) + \mu F(b), \quad (3.121)$$

with λ and μ scalars and a and b vectors. The action of F on a vector a can be seen as an instruction to rotate and dilate the vector to a new one. Clearly this operation is independent of coordinates, although it can also be expressed in terms of transformations of the vector coordinates; the former approach is notoriously simpler. We have already seen an example of such a transformation when we discussed rotations; in fact a rotation can be written as a function:

$$R(a) = Ra\tilde{R}, \quad (3.122)$$

where R is a rotor. The linearity of this function can be demonstrated easily.

Any function defined for vectors has a natural extension to multivectors, because one just needs to apply individually to each vector in every blade. This will extend the linear function to the whole of geometric algebra. Starting with bivectors, we can let function F act on both vectors of bivector $a \wedge b$, resulting in the bivector $F(a) \wedge F(b)$. This will be our definition of the action of F on a bivector blade

$$F(a \wedge b) = F(a) \wedge F(b). \quad (3.123)$$

The function transforms blades into blades and the linearity of F determines that it is also linear when acting on bivector blades

$$F(a \wedge b + c \wedge d) = F(a \wedge b) + F(c \wedge d). \quad (3.124)$$

The process can be extended naturally to blades of any grade, so

$$F(a \wedge b \wedge \cdots \wedge c) = F(a) \wedge F(b) \wedge \cdots \wedge F(c). \quad (3.125)$$

Linearity now ensures that F can be applied to multivectors of any sort and that it is also linear over multivectors

$$F(\lambda A + \mu B) = \lambda F(A) + \mu F(B). \quad (3.126)$$

Grades are preserved:

$$F(A_r) = \langle F(A_r) \rangle_r, \quad (3.127)$$

with A_r a grade- r multivector. Rotations have already given us an example of this as Eq. (3.95) allows us to write

$$R(A) = RA\tilde{R} \quad (3.128)$$

for any multivector A .

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

$$(FG)(a) = F[G(a)] = FG(a). \quad (3.129)$$

The final expression is a simplification to enable the reduction of brackets. The linearity of FG comes as a consequence of the original functions' linearity

$$FG(\lambda a + \mu b) = F[\lambda G(a) + \mu G(b)] = \lambda FG(a) + \mu FG(b). \quad (3.130)$$

Next we check that the product of functions can be extended to a blade; this is shown by

$$\begin{aligned} FG(a \wedge b \wedge \cdots \wedge c) &= F[G(a)] \wedge F[G(b)] \wedge \cdots \wedge F[G(c)] \\ &= F[G(a) \wedge G(b) \wedge \cdots \wedge G(c)] \\ &= F[G(a \wedge b \wedge \cdots \wedge c)]. \end{aligned} \quad (3.131)$$

The product of linear functions produces a new linear function, so we are allowed to make $H = FG$ such that

$$H(A) = FG(A), \quad (3.132)$$

since we have already defined the meaning of the right-hand side.

For any linear function F , we define its *transpose* \bar{F} , by

$$a \cdot \bar{F}(b) = F(a) \cdot b, \quad (3.133)$$

for any two vectors a and b . If F is a mapping from one linear space to another, then the transpose maps from the second space back to the first. In terms of an arbitrary frame $\{g_\kappa\}$ we have

$$g_\kappa \cdot \bar{F}(a) = a \cdot F(g_\kappa), \quad (3.134)$$

so we can construct the transpose by

$$\bar{\bar{F}}(a) = g^\kappa g_\kappa \cdot \bar{F}(a) = g^\kappa a \cdot F(g_\kappa). \quad (3.135)$$

Taking the transpose of the transpose of a function one obtains the original function, as expected. This is verified by

$$\bar{\bar{F}}(a) = g^\kappa a \cdot \bar{F}(g_\kappa) = g^\kappa g_\kappa \cdot F(a) = F(a). \quad (3.136)$$

The transpose of a product of two functions is found by

$$\begin{aligned} \overline{FG}(a) &= g^\kappa a \cdot FG(g_\kappa) = \bar{F}(a) \cdot G(g_\kappa) g^\kappa \\ &= \bar{G} \bar{F}(a) \cdot g_\kappa g^\kappa = \bar{G} \bar{F}(a). \end{aligned} \quad (3.137)$$

The transpose of a product of functions reverses the order in which the transposed functions are applied. A function is called *symmetric* when it equals its transpose, $\bar{F} = F$. Two particular examples of symmetric functions are provided by $F\bar{F}$ and $\bar{F}F$, as we verify next.

$$\overline{F\bar{F}} = \bar{\bar{F}F} = F\bar{F} \quad (3.138)$$

and similarly for $\bar{F}F$.

Being a linear function, the transpose can be extended to multivectors in the familiar way. Considering the case of a blade

$$\bar{F}(a \wedge b \wedge \cdots \wedge c) = \bar{F}(a) \wedge \bar{F}(b) \wedge \cdots \wedge \bar{F}(c). \quad (3.139)$$

If we take two bivectors $a_1 \wedge a_2$ and $b_1 \wedge b_2$, we see that

$$\begin{aligned} (a_1 \wedge a_2) \cdot F(b_1 \wedge b_2) &= a_1 \cdot F(b_2) a_2 \cdot F(b_1) - a_1 \cdot F(b_1) a_2 \cdot F(b_2) \\ &= \bar{F}(a_1) \cdot b_2 \bar{F}(a_2) \cdot b_1 - \bar{F}(a_1) \cdot b_1 \bar{F}(a_2) \cdot b_2 \\ &= \bar{F}(a_1 \wedge a_2) \cdot (b_1 \wedge b_2). \end{aligned} \quad (3.140)$$

This result can then be written for any two bivectors as

$$B_1 \cdot \bar{F}(B_2) = F(B_1) \cdot B_2. \quad (3.141)$$

It can also be extended to general multivectors

$$\langle A \bar{F}(B) \rangle = \langle F(A) B \rangle. \quad (3.142)$$

An even more general formula can be written; for that purpose consider the following

$$\begin{aligned} F(a \wedge b) \cdot c &= F(a) F(b) \cdot c - F(b) F(a) \cdot c \\ &= F[ab \cdot \bar{F}(c) - ba \cdot \bar{F}(c)] \\ &= F[(a \wedge b) \cdot \bar{F}(c)]. \end{aligned} \quad (3.143)$$

In this derivation we used the result from Eqs. (3.50). The previous formula can then be made more general:

$$\begin{aligned} A_r \cdot \bar{F}(B_s) &= \bar{F}[F(A_r) \cdot B_s] & r \leq s, \\ F(A_r) \cdot B_s &= F[A_r \cdot \bar{F}(B_s)] & r \geq s. \end{aligned} \quad (3.144)$$

If $r = s$ this equation reduces to Eq. (3.142).

Linear functions are grade preserving, so every linear function must produce a pseudoscalar when its argument is the unit pseudoscalar I . Since all the pseudoscalars in the algebra can be obtained multiplying the unit pseudoscalar by a real number, we can write

$$F(I) = \det(F)I, \quad (3.145)$$

where $\det(F)$ is a real number called the *determinant* of F . We see immediately that the determinant is the volume scale factor for the operation of the linear function. Acting on the unit hypercube the linear function returns a directed hypervolume FI .

As an example of the power of geometric algebra, consider the product of two linear functions F and G .

$$\det(FG)I = FG(I) = \det(G)F(I) = \det(F)\det(G)I. \quad (3.146)$$

We thus demonstrated that the determinant of the product of two functions equals the product of their determinants. We can just as easily show that the determinant of the transpose equals that of the original function;

$$\det(F) = \langle F(I)I^{-1} \rangle = \langle I\bar{F}(I^{-1}) \rangle = \det(\bar{F}). \quad (3.147)$$

Example 3.1

Consider the linear function

$$F(a) = a + \alpha a \cdot f_1 f_2, \quad (3.148)$$

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

We start by forming the action on a bivector blade

$$\begin{aligned} F(a \wedge b) &= (a + \alpha a \cdot f_1 f_2) \wedge (b + \alpha b \cdot f_1 f_2) \\ &= a \wedge b + \alpha(b \cdot f_1 a - a \cdot f_1 b) \wedge f_2 \\ &= a \wedge b + \alpha[(a \wedge b) \cdot f_1] \wedge f_2. \end{aligned} \quad (3.149)$$

We then extend to a general multivector

$$F(A) = A + \alpha(A \cdot f_1) \wedge f_2. \quad (3.150)$$

The determinant is now calculated by applying to the unit pseudoscalar

$$\begin{aligned} F(I) &= I + \alpha(I \cdot f_1) \wedge f_2 \\ &= I + \alpha f_1 \cdot f_2 I, \end{aligned} \quad (3.151)$$

so we have

$$\det(F) = 1 + \alpha f_1 \cdot f_2. \quad (3.152)$$

Next we derive a formula for the inverse of a linear function. Considering a multivector B and the unit pseudoscalar I , using Eqs. (3.144), we have

$$\det(F)IB = F(I)B = F[I\bar{F}(B)]. \quad (3.153)$$

The inner product with a pseudoscalar was replaced with a geometric product, since there are no other grades present. Replacing IB by A we find

$$\det(F)A = F[I\bar{F}(I^{-1}A)] \quad (3.154)$$

and similarly for the transpose. We can then write

$$\begin{aligned} F^{-1}(A) &= I\bar{F}(I^{-1}A)\det(F)^{-1}, \\ \bar{F}^{-1}(A) &= IF(I^{-1}A)\det(F)^{-1}. \end{aligned} \quad (3.155)$$

The derivation of these formulae shows once more the power of geometric algebra when compared to matrix methods.

Example 3.2

Find the inverse of the function defined in Eq. (3.148).

From Eq. (3.150) we get

$$\begin{aligned} \langle A_r F(B_r) \rangle &= \langle A_r B_r \rangle + \alpha \langle A_r (B_r \cdot f_1) \wedge f_2 \rangle \\ &= \langle A_r B_r \rangle + \alpha \langle f_2 \cdot A_r B_r f_1 \rangle, \end{aligned} \quad (3.156)$$

hence the transpose is

$$\bar{F}(A) = A + \alpha f_1 \wedge (f_2 \cdot A). \quad (3.157)$$

The inverse then follows;

$$\begin{aligned} F^{-1}(A) &= I \{ I^{-1}A + \alpha f_1 \wedge [f_2 \cdot (I^{-1}A)] \} (1 + \alpha f_1 \cdot f_2)^{-1} \\ &= [A + \alpha f_1 \cdot (f_2 \wedge A)] (1 + \alpha f_1 \cdot f_2)^{-1} \\ &= A - \frac{\alpha}{1 + \alpha f_1 \cdot f_2} f_2 \wedge (f_1 \cdot A). \end{aligned} \quad (3.158)$$

Example 3.3

Find the inverse of the rotation

$$R(a) = Ra\tilde{R}, \quad (3.159)$$

where R is a rotor.

The action of R on a general multivector has already been established;

$$R(A) = RAR\tilde{R}, \quad \bar{R}(A) = \tilde{R}AR. \quad (3.160)$$

Therefore the determinant verifies

$$\det(R)I = RI\tilde{R} = IR\tilde{R} = I, \quad (3.161)$$

so $\det(R) = 1$. The inverse is given by

$$R^{-1}(A) = I\tilde{R}I^{-1}AR = \tilde{R}AR = \bar{R}(A). \quad (3.162)$$

The inverse rotation is the same as the transpose as we expected. Transformations whose inverse equals the transpose are called *orthogonal transformations*.

The concept of linear function can be extended to that of *complex linear function*, keeping the definition (3.121) but allowing a and b to be complex vectors and λ and μ to be complex scalars. Complex linear functions naturally preserve the complex grade of their arguments and can have complex determinant, but otherwise they behave like linear functions. These functions become necessary when one wants to find the geometric algebra counterparts of *eigenvectors* and *eigenvalues* defined in matrix theory. A complex linear function F has a complex eigenvector g if

$$F(g) = \lambda g. \quad (3.163)$$

The complex scalar λ is the associated eigenvalue. It follows that

$$\det(F - \lambda I) = 0, \quad (3.164)$$

with I the identity function. This equation produces a grade- n equation whose solutions are the eigenvalues of F . The process for determining eigenvectors and eigenvalues is well developed in matrix formalism and as such implemented in software packages; it will generally be easier to switch to a matrix representation of the algebra when solving this kind of problem. So, in \mathcal{G}_{30} one would choose the Pauli matrices' representation of Eq. (2.78); in \mathcal{G}_{41} the representation would be based on Dirac matrices, as we shall see later on.

Although our definition of eigenvectors and eigenvalues was made only for complex spaces, spaces whose geometric algebra has a complex structure, it can also be applied to general spaces. The idea then is to complexify the algebra by providing a unit imaginary that commutes with vectors and so adding one extra dimension to the original space. The complexified algebra allows the definition of a complex linear function which can have some real eigenvectors and eigenvalues; these will then belong to the original non-complexified space.

Symmetric functions in Euclidean space can be expressed in the *canonical* form via the spectral decomposition. If g_i and g_j are eigenvectors of a function with eigenvalues λ_i and λ_j , we have

$$g_i \cdot F(g_j) = \lambda_j g_i \cdot g_j. \quad (3.165)$$

By the definition (3.133), if the function is symmetric, this also equals

$$\bar{F}(g_i) \cdot g_j = F(g_i) \cdot g_j = \lambda_i g_i \cdot g_j. \quad (3.166)$$

It then follows

$$(\lambda_i - \lambda_j) g_i \cdot g_j = 0, \quad (3.167)$$

so eigenvectors of a symmetric function with distinct eigenvalues are orthogonal and we can always choose them to be orthonormal; consequently we adopt the notation σ_i to replace g_i .

If we admit the existence of complex eigenvectors and eigenvalues we also find that

$$\sigma^\dagger \cdot F(\sigma) = \lambda \sigma^\dagger \cdot \sigma = F(\sigma^\dagger) \cdot \sigma = \lambda^\dagger \sigma^\dagger \cdot \sigma. \quad (3.168)$$

So, any symmetric function verifies also

$$(\lambda - \lambda^\dagger) \sigma^\dagger \cdot \sigma = 0. \quad (3.169)$$

In case $\sigma^\dagger \cdot \sigma \neq 0$ we can conclude that the eigenvalue, and hence the eigenvector, is real. In Euclidean space there are no null vectors and the inequality is always verified; every symmetric function in an n -dimensional space has then a spectral decomposition

$$F(a) = \lambda_1 P_1(a) + \lambda_2 P_2(a) + \cdots + \lambda_m P_m(a). \quad (3.170)$$

Where $\lambda_1 < \lambda_2 < \cdots < \lambda_m$ are the m distinct eigenvalues ($m \leq n$) and the P_i are the projections onto each of the invariant subspaces defined by the eigenvectors. For the case of a projection onto a one-dimensional subspace we have (no sums implied)

$$P_i(a) = a \cdot \sigma^i \sigma_i. \quad (3.171)$$

The eigenvectors form an orthonormal frame, which is the natural frame in which to study the linear function. When two eigenvalues are the same it is always possible to choose the eigenvectors so that they remain orthogonal. In non-Euclidean spaces there are null vectors and one can have $\sigma^* \cdot \sigma = 0$; for this reason it may not be possible to find an orthonormal frame.

One can define *eigenblades* as a natural extension of the eigenvector concept. We define an eigenblade A_r as a grade- r blade satisfying

$$F(A_r) = \lambda A_r, \quad (3.172)$$

where λ is real. One immediate example is the unit pseudoscalar, for which $\lambda = \det(F)$. More generally, each eigenblade determines an invariant subspace of the transformation. As an example consider a function that performs a $\pi/2$ rotation and dilation, defined by

$$F(\sigma_1) = \lambda \sigma_2, \quad F(\sigma_2) = -\lambda \sigma_1. \quad (3.173)$$

One can easily check that $\sigma_1 \pm i\sigma_2$ are complex eigenvectors with complex eigenvalues $\mp i\lambda$. If this is applied in the 3-dimensional algebra \mathcal{G}_3 , the complex eigenvectors become $\sigma_1 \pm \sigma_{31}$ and the corresponding complex eigenvalues are $\mp \lambda\sigma_{123}$. One can however use the identity

$$F(\sigma_{12}) = \lambda^2 \sigma_{12}, \quad (3.174)$$

to identify the plane σ_{12} as an eigenbivector of F . Strictly in two dimensions, the latter is the only possible answer to the question of eigenelements of F , but in three dimensions both solutions are equally valid and each of them has a different geometrical interpretation.

Antisymmetric functions have $\bar{F}(A) = -F(A)$, hence

$$a \cdot F(a) = \bar{F}(a) \cdot a = -F(a) \cdot a = 0. \quad (3.175)$$

For the study of antisymmetric functions one considers the bivector

$$F = \frac{1}{2} g^i \wedge F(g_i), \quad (3.176)$$

where the $\{g_i\}$ are an arbitrary frame for the space acted on by F . The bivector F is an invariant since it is independent of the choice of frame. The bivector F actually encodes the action of F , since

$$\begin{aligned} 2a \cdot F &= a \cdot [g^i \wedge F(g_i)] \\ &= a \cdot g^i F(g_i) - g^i a \cdot F(g_i) \\ &= F(a \cdot g^i g_i) + g^i g_i \cdot F(a) \\ &= 2F(a). \end{aligned} \quad (3.177)$$

The action of an antisymmetric function can therefore be reduced to the contraction with the *characteristic bivector* F .

$$F(a) = a \cdot F. \quad (3.178)$$

An antisymmetric function can be reduced to its simplest form if one splits F into a set of commuting unit blades:

$$F = \lambda_1 \hat{F}_1 + \lambda_2 \hat{F}_2 + \cdots + \lambda_k \hat{F}_k. \quad (3.179)$$

This decomposition is always possible in Euclidean space. Each component of F is an eigenblade of F and determines an invariant subspace. In non-Euclidean spaces, such decomposition may not be possible.

Linear functions without symmetry can be given a canonical form by means of the singular value decomposition. Suppose a linear function F acting on n -dimensional Euclidean space; suppose also that the determinant of F is non-zero. We form a symmetric function D by

$$D(a) = \bar{F}F(a). \quad (3.180)$$

Because it is symmetric, function D has n orthogonal eigenvectors, with real eigenvalues; these are positive because

$$\bar{F}F(g) = \lambda g \Rightarrow F(g) \cdot F(g) = \lambda g^2. \quad (3.181)$$

The eigenvalue λ is positive because in Euclidean space the square of a vector is a positive scalar. There are no null eigenvalues since we restricted the discussion to the cases where $\det(F) \neq 0$. It follows that

$$D(a) = \sum_{i=1}^n \lambda_i a \cdot \sigma^i \sigma_i, \quad (3.182)$$

where the $\{\sigma_i\}$ are an orthonormal frame of eigenvectors. As we said before, degenerate eigenvalues can be dealt with by picking up a set of orthonormal vectors in the invariant subspace.

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

$$D^{1/2} = \sum_{i=1}^n \lambda_i^{1/2} a \cdot \sigma^i \sigma_i, \quad (3.183)$$

which is invertible

$$D^{-1/2} = \sum_{i=1}^n \lambda_i^{-1/2} a \cdot \sigma^i \sigma_i, \quad (3.184)$$

We now set

$$S = FD^{-1/2}. \quad (3.185)$$

This verifies

$$\bar{S}S = D^{-1/2}\bar{F}FD^{-1/2} = D^{-1/2}DD^{-1/2} = I, \quad (3.186)$$

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

$$F = SD^{-1/2}. \quad (3.187)$$

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

When an $n \times n$ matrix is used for the representation of F in some frame, it is usual to include a further rotation R to align this arbitrary frame with the frame provided by the eigenvectors; in this case one writes

$$F = S\Lambda^{1/2}R, \quad (3.188)$$

where Λ is a diagonal matrix. When this is done the dilation appears sandwiched between two rotations and the matrix is said to be in the singular value decomposition. Of the n^2 degrees of freedom present in F , the two rotations are responsible for $2 \times n(n-1)/2$ and the dilation Λ provides the remaining n .

3.5 Idempotents and ideals

Some important functions are not grade preserving and don't follow the rules of linear functions. Probably the most significant case is the left (right) product by a general multivector. We have already seen in Eq. (2.136) how a rotor is transformed in a rotation by left multiplication by another rotor. Operators are generally transformed by single-sided transformation and thus the grades in the composite operator are generally different from those in the original one. For functions which result from the single-sided product by a multivector the definitions of eigenelements given above cannot be applied; in this section we look at how the concept must be adapted to those functions.

If we take any non-scalar unitary Hermitian element h of the algebra, that is $h^2 = 1$, two *idempotents* can be defined by the relations

$$f_+ = \frac{1}{2}(1 + h), \quad f_- = \frac{1}{2}(1 - h). \quad (3.189)$$

These are called idempotents because any power of either of them reproduces itself; for instance $(f_+)^2 = f_+$. The two idempotents defined with recourse to the same unitary Hermitian element are said to be *orthogonal idempotents* because their product is zero. These two idempotents are also complementary because they add up to unity. Given any multivector M it is then possible to separate it into the sum of two multivectors by right (left) multiplication by the two complementary idempotents, as in

$$M = M(f_+ + f_-) = M_+ + M_-. \quad (3.190)$$

The product of a multivector on the right (left) by an idempotent produces what is known as a left (right) *ideal*. A left (right) ideal remains a left (right) ideal when it is multiplied on the left (right) by any multivector.

Example 3.4

Using the Pauli matrix representation of \mathcal{G}_3 from Eq. (2.79) one sees immediately that two idempotents can be formed by letting $h = \sigma_3$, resulting in

$$f_+ = \frac{1}{2}(1 + \sigma_3) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad (3.191)$$

$$f_- = \frac{1}{2}(1 - \sigma_3) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \quad (3.192)$$

A general multivector is represented in matrix form by a general 2×2 matrix; when this is multiplied on the right by f_+ the result is a matrix with zeros in the second column and when the same matrix is multiplied on the right by f_- the result is a matrix with zeros in the first column.

It is important to note, however, that the choice of σ_3 was arbitrary and that a formally equivalent result would have been obtained had we chosen any unitary vector. The matrix representation of the left ideals resulting from a different choice of idempotents would not show us an obvious result but geometrically one choice of idempotents is as good as any other.

The commuting unitary Hermitian elements $\{h_i\}$ such that $h_m \neq h_i h_j$, with $m \neq i \neq j$, are called *primitive*. If the algebra allows the choice of up to k primitive commuting unitary Hermitian elements $\{h_1, h_2, \dots, h_k\}$, we can define 2^k mutually annihilating idempotents by the formula

$$f = \frac{1}{2} (1 \pm h_1) \frac{1}{2} (1 \pm h_2) \cdots \frac{1}{2} (1 \pm h_k). \quad (3.193)$$

These idempotents are called *primitive idempotents* because they cannot be obtained by addition of two idempotents. In fact the addition of any two idempotents is still an idempotent and if we include 0 and 1 also as idempotents, they form an additive group of order 2^{2^k} . The number of commuting unitary Hermitian elements that can be found in an algebra is given by a complex formula that can be found in Lounesto [3]; for us it will be sufficient to know that \mathcal{G}_3 allows just one such element, while $\mathcal{G}_{1,3}$ and $\mathcal{G}_{4,1}$ allow two elements each.

Ideals defined with recourse to primitive idempotents are called *minimal ideals*; when we switch to matrix representation it may be possible, by suitable choice of idempotents, to assign minimal left (right) ideals to matrices with just one non-zero column (row). We use minimal ideals to define *eigenideals*. If M is a multivector, ψ is a minimal left ideal and λ is a scalar, we say that ψ is an eigenideal and λ is an eigenvalue of M if

$$M\psi = \lambda\psi. \quad (3.194)$$

This multivector equation translates into a set of 2^k equations that are solved using the methods of matrix algebra; the eigenvalues determined in this way are not necessarily scalars. Since this is a transposition from matrix algebra, the type of eigenvalues that is determined depends on the particular matrix algebra that is isomorphic to the geometric algebra where the eigenvalue equation is written. So, for instance, geometric algebra $\mathcal{G}_{1,3}$ is isomorphic to the real algebra of 4×4 matrices but the eigenvalues of such matrices can have complex values that lie outside the algebra. On the other hand, geometric algebra $\mathcal{G}_{4,1}$ is isomorphic to the complex algebra of 4×4 matrices and this ensure that all the eigenvalues found belong to the algebra, although they may be scalars, pseudo-scalars or a combination of both.

In quantum mechanics one usually refers to eigenideals as eigenvectors because they can be written as column matrices; the exposition above shows that this confusion of terms can be misleading and we will avoid it by restricting the use of eigenvectors to the case of linear functions.

Eigenideals can be used to define the trace and determinant of a multivector. If $\{f_i\}$ is a complete set of primitive idempotents, that is, a set of primitive idempotents adding to unity, we find the trace of a multivector M by first solving the equation

$$M\psi = \left(\sum_{i=1}^{2^k} \lambda_i f_i \right) \psi, \quad (3.195)$$

where λ_i are eigenvalues of M and ψ is the sum of the corresponding eigenideals. The trace of M is then calculated by

$$\text{tr}(M) = \sum_{i=1}^{2^k} \lambda_i. \quad (3.196)$$

Having found the trace, the determinant is easily evaluated by the formula

$$\det(M) = e^{i \text{tr}(M)}. \quad (3.197)$$

Under the definitions above, both trace and determinant may evaluate to non-scalars and the same considerations made for eigenvalues apply here.

3.6 Tensors and components

Traditionally, tensor calculus is done with recourse to a large number of indices but geometric algebra gives us a way to simplify expressions and gives a geometric interpretation to otherwise abstract concepts. Tensors with indices are the components of multivectors in some given frame but in essence a multivector is an entity independent of frames. A grade-1 multivector is a vector, which does not change because we choose to express it in one frame or another; the same happens with areas (grade-2 multivectors), volumes (grade-3 multivectors) and so forth. We start off by analysing Cartesian tensors, the designation given to tensors in a frame of orthonormal vectors in Euclidean space. For these we have

$$\sigma_\kappa \cdot \sigma_\lambda = \delta_{\kappa\lambda}, \quad (3.198)$$

so there is no distinction between frame vectors and their reciprocals. The use of upper and lower indices is, in this case, just a convenience to make the sum convention work. For frames with the same orientation, a new frame can be obtained from the $\{\sigma_\kappa\}$ by a rotation

$$\sigma'_\kappa = R \sigma_\kappa \tilde{R} = \Lambda_\kappa^\lambda \sigma_\lambda. \quad (3.199)$$

Here R is a rotor and Λ_κ^λ are the components of the rotation in the $\{\sigma^\kappa\}$ frame:

$$\Lambda_\kappa^\lambda = R \sigma_\kappa \tilde{R} \cdot \sigma^\lambda. \quad (3.200)$$

In spite of the equivalence between frame and reciprocal frame vectors, the equations for the reciprocal frame must be consistent in terms of index placement. The corresponding equations are

$$\sigma'^{\kappa} = R\sigma^{\kappa}\tilde{R} = \Lambda^{\kappa}_{\lambda}\sigma^{\lambda} \quad (3.201)$$

and

$$\Lambda^{\kappa}_{\lambda} = R\sigma^{\kappa}\tilde{R}\cdot\sigma_{\lambda}. \quad (3.202)$$

It follows that

$$\begin{aligned} \Lambda^{\kappa}_{\lambda}\Lambda^{\lambda}_{\iota} &= (R\sigma^{\kappa}\tilde{R})\cdot\sigma^{\lambda}(R\sigma^{\lambda}\tilde{R})\cdot\sigma_{\iota} \\ &= (\tilde{R}\sigma^{\lambda}R)\cdot(\tilde{R}\sigma_{\iota}R) = \delta^{\lambda}_{\iota}; \end{aligned} \quad (3.203)$$

and similarly

$$\Lambda^{\iota}_{\kappa}\Lambda^{\lambda}_{\iota} = \delta^{\lambda}_{\kappa}. \quad (3.204)$$

A vector a has components $a^{\kappa} = \sigma^{\kappa}\cdot a$, which transform in the obvious way

$$a'^{\kappa} = \sigma'^{\kappa}\cdot a = \Lambda^{\kappa}_{\lambda}a^{\lambda}. \quad (3.205)$$

The vector remains unchanged because it is frame independent; only the coordinates change when we choose to expand the vector in a different frame. The invariance of a under a coordinate transformation has no physical significance, contrary to what would happen if some equation involving vector a would remain unchanged under a transformation of the vector itself. This would express an underlying symmetry and would have real physical significance.

Components can be defined for a general linear function F in a similar way

$$F^{\kappa}_{\lambda} = \sigma^{\kappa}\cdot F(\sigma_{\lambda}). \quad (3.206)$$

This results in an $n \times n$ array, which can naturally be treated as a matrix. This definition ensures that the components of vector $F(a)$ are given by

$$\sigma^{\kappa}\cdot F(a) = \sigma^{\kappa}\cdot F(a^{\lambda}\sigma_{\lambda}) = F^{\kappa}_{\lambda}a^{\lambda}, \quad (3.207)$$

which is the usual matrix expression for the product of a matrix by a vector. In a similar fashion, if F and G are two linear functions, we obtain for the product components

$$\begin{aligned} (FG)^{\kappa}_{\lambda} &= FG(\sigma_{\lambda})\cdot\sigma^{\kappa} = G(\sigma_{\lambda})\cdot\bar{F}(\sigma^{\kappa}) \\ &= G(\sigma_{\lambda})\cdot\sigma^{\iota}\sigma_{\iota}\cdot\bar{F}(\sigma^{\kappa}) = F^{\kappa}_{\iota}G^{\iota}_{\lambda}. \end{aligned} \quad (3.208)$$

This is the familiar formula for multiplying matrices. If the frame is rotated into a new frame, components are transformed in an obvious way

$$F'^{\kappa}_{\lambda} = F(\sigma'_{\lambda}) = \sigma'^{\kappa}\cdot F(\sigma'_{\lambda}) = \Lambda^{\kappa}_{\iota}\Lambda^{\eta}_{\lambda}F^{\iota}_{\eta}, \quad (3.209)$$

where the prime refers to the rotated frame. Objects with two indices are referred to as rank-2 tensors and vectors are also known as rank-1 tensors; the rule for coordinate transformations of higher rank tensors can be induced in an obvious way from the previous ones.

For Cartesian tensors there are two important tensors whose components are invariant under coordinate transformations. The first of these is the Kronecker δ , which transforms as

$$\delta'^{\kappa}_{\lambda} = \Lambda^{\kappa}_{\iota} \Lambda_{\lambda}^{\eta} \delta^{\iota}_{\eta} = \Lambda^{\kappa}_{\iota} \Lambda_{\lambda}^{\iota} = \delta^{\kappa}_{\lambda}. \quad (3.210)$$

The Kronecker δ represents the identity function, whose components are independent of the chosen frame. The second coordinate invariant tensor is the alternating tensor $\varepsilon_{ij\dots k}$, where the number of indices matches the dimension of space. This is totally antisymmetric and is defined as follows

$$\varepsilon_{ij\dots k} = \begin{cases} 1 & i, j, \dots, k = \text{even permutation of } 1, 2, \dots, n \\ -1 & i, j, \dots, k = \text{odd permutation of } 1, 2, \dots, n \\ 0 & \text{otherwise} \end{cases} \quad (3.211)$$

A permutation is said odd (even) if an odd (even) number of swaps is required in order to obtain that permutation from the original order $1, 2, \dots, n$. The determinant of a matrix can be expressed in terms of the alternating tensor as

$$F^{\alpha}_i F^{\beta}_j \dots F^{\gamma}_k \varepsilon_{\alpha\beta\dots\gamma} = \det(F) \varepsilon_{ij\dots k}. \quad (3.212)$$

A rotation of the alternating tensor is given by

$$\varepsilon'_{ij\dots k} = \Lambda_i^{\alpha} \Lambda_j^{\beta} \dots \Lambda_k^{\gamma} \varepsilon_{\alpha\beta\dots\gamma} = \det(\Lambda) \varepsilon_{ij\dots k}. \quad (3.213)$$

But since λ is a rotation, it has unit determinant, so the alternating tensor is coordinate invariant.

In Eq. (3.212) we established a definition for the determinant which is different from the one given in Eq. (3.145). We now show the agreement between the two definitions. If $\{\sigma_{\iota}\}$ form an orthonormal frame we have

$$\varepsilon_{\iota\kappa\dots\lambda} = \sigma_{\iota} \wedge \sigma_{\kappa} \wedge \dots \wedge \sigma_{\lambda} I^{\dagger}. \quad (3.214)$$

The right-hand side of the equation is zero if any two indices are the same. If the indices form an even permutation we can reorder them into $\sigma_{12\dots n} = I$ and the product returns $+1$. In the case of an odd permutation the right-hand side product returns -1 . We can now rearrange the left-hand side of Eq. (3.212) as follows

$$\begin{aligned} F^{\alpha}_i F^{\beta}_j \dots F^{\gamma}_k \varepsilon_{\alpha\beta\dots\gamma} &= F^{\alpha}_i F^{\beta}_j \dots F^{\gamma}_k \sigma_{\alpha} \wedge \sigma_{\beta} \wedge \dots \wedge \sigma_{\gamma} I^{\dagger} \\ &= F(\sigma_{\alpha}) \wedge F(\sigma_{\beta}) \wedge \dots \wedge F(\sigma_{\gamma}) I^{\dagger} \\ &= \det(F) \sigma_{\alpha} \wedge \sigma_{\beta} \wedge \dots \wedge \sigma_{\gamma} I^{\dagger} \\ &= \det(F) \varepsilon_{ij\dots k}, \end{aligned} \quad (3.215)$$

as expected.

We now generalize the previous treatment to tensors defined in arbitrary basis sets, for spaces with arbitrary, non-degenerate, signature. We are now going to examine the cases where the basis vectors are non-orthonormal, allowing them to vary from point to point. We also allow negative square basis vectors, to account for mixed signature spaces; we will only leave out cases where some basis vectors have zero square. Suppose that $\{g_\kappa\}$ constitute an arbitrary frame for an n -dimensional space; the reciprocal frame is formed by the $\{g^\kappa\}$, verifying Eq. (3.97). Equations (3.101) and (3.102) are valid in mixed signature spaces.

Suppose now that we have two vectors $a = a^\kappa g_\kappa = a_\kappa g^\kappa$ and $b = b^\kappa g_\kappa = b_\kappa g^\kappa$. Their inner product can be written as

$$a \cdot b = (a^\iota g_\iota) \cdot (b_\kappa g^\kappa) = a^\iota b_\kappa g_\iota \cdot g^\kappa = a^\iota b_\kappa \delta_\iota^\kappa = a^\iota b_\iota. \quad (3.216)$$

Now, if we don't use the reciprocal frame the inner product is written

$$a \cdot b = (a^\iota g_\iota) \cdot (b^\kappa g_\kappa) = a^\iota b^\kappa g_{\iota\kappa}, \quad (3.217)$$

where $g_{\iota\kappa}$ is the metric tensor defined by

$$g_{\iota\kappa} = g_\iota \cdot g_\kappa. \quad (3.218)$$

We had already encountered the metric tensor in Eq. (3.98) but it is recalled here because we are going to discuss tensors in general. The inner product of two vectors can then be written in a number of different ways

$$a \cdot b = a^\iota b_\iota = a_\iota b^\iota = a^\iota b^\kappa g_{\iota\kappa} = a_\iota b_\kappa g^{\iota\kappa}. \quad (3.219)$$

All these formulas express exactly the same thing; all but the first one are frame dependent and so this is the one we prefer whenever possible.

Expressing a linear function in terms of components goes in the same way. The components in the reciprocal frame are given by

$$F_{\iota\kappa} = g_\iota \cdot F(g_\kappa). \quad (3.220)$$

The set of numbers $F_{\iota\kappa}$ can be arranged in an $n \times n$ matrix and is referred to as the components of a rank-2 tensor. The linear function is frame independent, although its components will vary with the choice of frame. Instead of working with the frame vectors, we can work with the reciprocal frame as in

$$F^{\iota\kappa} = F(g^\kappa) \cdot g^\iota. \quad (3.221)$$

The metric tensor can be used to change from one set of components to another:

$$F^{\iota\kappa} = g^\iota \cdot F(g^\kappa) = g^{\iota\lambda} g_\lambda \cdot F(g_\mu g^{\mu\kappa}) = g^{\iota\lambda} g^{\kappa\mu} F_{\lambda\mu}. \quad (3.222)$$

Again, the linear function is the same, whatever the set of components that is used to express it.

The relation between a linear function's components $F_{\iota\kappa}$ and those of its transpose are given by

$$\bar{F}_{\iota\kappa} = \bar{F}(g_\kappa) \cdot g_\iota = g_\kappa \cdot F(g_\iota) = F_{\kappa\iota}. \quad (3.223)$$

In matrix terms this has an obvious interpretation as the exchange between rows and columns. For mixed index tensors we must exercise some care, as we now have

$$F_\iota{}^\kappa = F(g^\kappa) \cdot g_\iota = g^\kappa \cdot \bar{F}(g_\iota) = \bar{F}_\iota{}^\kappa. \quad (3.224)$$

If F is a symmetric function we have $\bar{F} = F$ and the components satisfy

$$F_{\iota\kappa} = F(g_\kappa) \cdot g_\iota = F(g_\iota) \cdot g_\kappa = F_{\kappa\iota}, \quad (3.225)$$

that is the components $F_{\iota\kappa}$ form a symmetric matrix. The same is true for $F^{\iota\kappa} = F^{\kappa\iota}$ but for the mixed components we have $F_\iota{}^\kappa = F^\kappa{}_\iota$.

If we are dealing with a product of linear functions FG , the components can be found by

$$\begin{aligned} (FG)_{\iota\kappa} &= FG(g_\kappa) \cdot g_\iota = G(g_\kappa) \cdot \bar{F}(g_\iota) \\ &= G(g_j) \cdot g_\lambda g^\lambda \cdot \bar{F}(g_\iota) = F_\iota{}^\lambda G_{\lambda\kappa}. \end{aligned} \quad (3.226)$$

In matrix terms this can be viewed as a matrix product. Alternatively we can use the metric tensor to work only with subscripted indices:

$$(FG)_{\iota\kappa} = F_{\iota\lambda} G_{\mu\kappa} g^{\lambda\mu}. \quad (3.227)$$

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

$$\phi(\lambda a_1 + \mu b, a_2, a_3) = \lambda \phi(a_1, a_2, a_3) + \mu \phi(b, a_2, a_3) \quad (3.228)$$

and similarly for the other two arguments. The components of this function define a rank-3 tensor via

$$\phi_{\iota\kappa\lambda} = \phi(g_\iota, g_\kappa, g_\lambda). \quad (3.229)$$

This shows how we can write frame-independent expressions equivalent to tensor equations.

In Eq. (3.116) we defined a matrix to convert between an arbitrary frame and an underlying orthonormal frame. In general, if a second frame $\{f_\kappa\}$ is introduced, we can relate the two frames via a transformation matrix

$$f_{\iota\kappa} = f_\iota \cdot g_\kappa, \quad f^{\iota\kappa} = f^\iota \cdot g^\kappa. \quad (3.230)$$

The components of the transformation matrix satisfy

$$f_{\iota\kappa} f^{\iota\lambda} = f_{\iota} \cdot g_{\kappa} f^{\iota} \cdot g^{\lambda} = g_{\kappa} \cdot g^{\lambda} = \delta_{\kappa}^{\lambda} \quad (3.231)$$

and

$$f_{\iota\kappa} f^{\lambda\kappa} = f_{\iota} \cdot g_{\kappa} f^{\lambda} \cdot g^{\kappa} = f_{\iota} \cdot f^{\lambda} = \delta_{\iota}^{\lambda}. \quad (3.232)$$

We can naturally define mixed index matrices by

$$f^{\iota}_{\kappa} = f^{\iota} \cdot g_{\kappa}, \quad f_{\iota}^{\kappa} = f_{\iota} \cdot g^{\kappa}. \quad (3.233)$$

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

$$a = a^{\iota} g_{\iota} = a^{\iota} f_{\kappa} g_{\iota} \cdot f^{\kappa} = a^{\iota} f^{\kappa}_{\iota} f_{\kappa}. \quad (3.234)$$

It follows that the transformation law for the components is

$$a'^{\kappa} = f^{\kappa}_{\iota} a^{\iota}, \quad (3.235)$$

where the prime indicates components in the alternate frame. A similar expression holds for reciprocal frame components.

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

$$F'_{\iota\kappa} = f_{\iota}^{\lambda} f_{\kappa}^{\mu} F_{\lambda\mu}. \quad (3.236)$$

If there are mixed indices, similar expressions exist; for instance

$$F'^{\kappa}_{\iota} = f_{\iota}^{\lambda} f^{\kappa}_{\mu} F_{\lambda}^{\mu}. \quad (3.237)$$

In tensor calculus one usually defines tensors as entities that transform according to the previous laws under coordinate transformations but in geometric algebra formalism one sees that the underlying entities (vectors or linear functions) are really invariant and only their components change with the particular frame that is chosen.

4 Differential calculus

We call *geometric calculus* to the compound subject of vector differentiation and geometric algebra. In standard vector calculus one uses gradient, divergence and curl but we will see how geometric calculus can combine and generalize all those operations under the action of a single *vector derivative* denoted ∇ . Extending the definition for the curl of a vector to spaces of arbitrary dimension becomes just as simple as it was to extend the cross product. The vector derivative is actually a vector and can be operated via the geometric product with all multivectors. A further advantage of the vector derivative, when compared to separate gradient, divergence and curl, is that it is invertible. As a consequence Green's functions exist for ∇ , which enable initial conditions to be propagated off a surface. We will generalize the concept of analytical functions to spaces of arbitrary dimension and curvature, paving the ground for forthcoming chapters where these will be used extensively.

4.1 The vector derivative

If we consider a constant frame $\{g_\iota\}$ spanning an n -dimensional space, we represent each point by its position vector $x = g_\iota x^\iota$. The *vector derivative* ∇ with respect to x is defined by

$$\nabla = \sum_{\iota=1}^n g^\iota \frac{\partial}{\partial x^\iota} = g^\iota \partial_\iota, \quad (4.1)$$

where we use the convenient notation ∂_ι for partial derivatives. If we make the inner product of ∇ and an arbitrary vector a , we obtain the *directional derivative* in the a direction

$$a \cdot \nabla F(x) = \lim_{\varepsilon \rightarrow 0} \frac{F(x + \varepsilon a) - F(x)}{\varepsilon}, \quad (4.2)$$

where we assume that this limit exists and is well defined; the function $F(x)$ can be any multivector valued function of position. The definition of ∇ is frame independent, just as any other vector, and this becomes clear if we use Eq. (4.2) as the defining equation.

We can start applying ∇ by considering a scalar function $\phi(x)$. Acting on ϕ , the vector derivative returns its gradient, entirely equivalent to the familiar grad operator. The vector's $\nabla\phi$ components in the $\{g^\iota\}$ frame are the partial derivatives with respect to the x^ι coordinates. In Euclidean spaces $\nabla\phi$ always points in the direction of steepest increase of ϕ , but this may not be the case in mixed signature spaces. A simple example

of a scalar function is $a \cdot x$, where a is a constant vector. We write $a \cdot x = a_\kappa x^\kappa$, so that the gradient becomes

$$\nabla(a \cdot x) = g^\iota \partial_\iota x^\kappa a_\kappa = g^\iota a_\kappa \delta_\iota^\kappa = g^\kappa a_\kappa = a. \quad (4.3)$$

This result is valid whatever the dimension and signature of space and can be used to build other useful results; for instance

$$\nabla x^\iota = \nabla(x \cdot g_\iota) = g_\iota, \quad (4.4)$$

a formula that we will meet again when discussing curvature.

Consider now the gradient of $(x)^2$. Working with coordinates we have

$$\begin{aligned} \nabla(x)^2 &= g^\iota \partial_\iota (x^\kappa x^\lambda) g_\kappa \cdot g_\lambda \\ &= g^\iota (\partial_\iota x^\kappa x^\lambda + \partial_\iota x^\lambda x^\kappa) g_\kappa \cdot g_\lambda \\ &= x^\kappa g_\kappa + x^\lambda g_\lambda \\ &= 2x. \end{aligned} \quad (4.5)$$

This is the result that we expected but we made use of vector components, which should desirably be avoided. We need a convention to specify the argument of the vector derivative when this is not adjacent on its right; for that purpose we use over arrows. Under this convention we write

$$\nabla(x)^2 = \overrightarrow{\nabla}(\overleftarrow{x} \cdot x) + \overrightarrow{\nabla}(x \cdot \overleftarrow{x}) = 2\overrightarrow{\nabla}(\overleftarrow{x} \cdot x). \quad (4.6)$$

In the final term it is only the first factor that is derived, while the second factor is held constant, so we can apply Eq. (4.3) to obtain immediately $\nabla(x)^2 = 2x$.

In Euclidean spaces $\nabla\phi$ always points in the direction of steepest increase of ϕ . Equation (4.3) illustrates this because clearly, in order to get the steepest increase of $a \cdot x$ one must move in the direction of a . In the case of a general scalar function, suppose that $\nabla\phi = J$ and consider the contraction of this with the unit vector n

$$n \cdot (\nabla\phi) = \nabla \cdot J. \quad (4.7)$$

We want the direction of n which maximizes the result. In Euclidean space this must be obviously in the direction of J , so J points in the direction of steepest increase. When n is given the direction of J , we see that the magnitude of J is the derivative in the direction of greatest increase.

In mixed signature spaces everything becomes more complicated. As an example consider the plane defined by the orthonormal vectors $\{\sigma_0, \sigma_1\}$, with $(\sigma_0)^2 = -1$ and $(\sigma_1)^2 = 1$. We introduce the scalar function

$$\phi = \langle x \sigma_1 x \sigma_1 \rangle = (x^0)^2 + (x^1)^2. \quad (4.8)$$

Imposing ϕ to remain constant we obtain a circle centered on the origin, so the direction of steepest increase points radially outwards. The gradient, however, does not point in this direction, as we see by

$$\nabla\phi = 2\overrightarrow{\nabla}\langle\overleftarrow{x}\sigma_1x\sigma_1\rangle = 2\sigma_1x\sigma_1. \quad (4.9)$$

If x is a point in the x^1 axis, $x = x^1\sigma_1$, the gradient becomes $2\sigma_1x\sigma_1 = 2x^1\sigma_1$, clearly radial in an outward direction, but if the point lies on the x^0 axis, $x = x^0\sigma_0$, we obtain $\sigma_1x\sigma_1 = -2x^0\sigma_0$, pointing inwards. For other points the gradient will not be radially directed.

If we consider now a vector valued function $J(x)$, the full vector derivative contains two terms, a scalar and a bivector. The scalar term is the *divergence* of $J(x)$. In terms of constant frame vectors $\{g_\iota\}$ we have

$$\nabla \cdot J = \partial_\iota(g^\iota \cdot J) = \partial_\iota J^\iota. \quad (4.10)$$

For example, the divergence of x

$$\nabla \cdot x = \partial_\iota x^\iota = n, \quad (4.11)$$

where n is the dimension of space.

The bivector term of the derivative produces what is known as *exterior derivative*. In terms of coordinates we can write

$$\nabla \wedge J = g^\iota \wedge (\partial_\iota J) = g^\iota \wedge g_\kappa \partial_\iota J^\kappa. \quad (4.12)$$

The components are antisymmetrized in ι and κ . In three dimensions these are the components of the curl, although $\nabla \wedge J$ is a bivector and $\text{curl}(J)$ is a vector. The three-dimensional curl requires a duality operation by means of the pseudovector:

$$\text{curl}(J) = -i\nabla \wedge J. \quad (4.13)$$

This works only in three dimensions but the exterior derivative is defined for an arbitrary number of dimensions.

Consider, for instance, the exterior derivative of the position vector

$$\nabla \wedge x = g^\iota \wedge g_\iota = g^\iota \wedge g^\kappa g_{\iota\kappa} = 0, \quad (4.14)$$

because $g^\iota \wedge g^\kappa$ is antisymmetric in ι and κ while $g_{\iota\kappa}$ is symmetric. The two terms of the vector derivative can be combined in the usual form of the geometric product of two vectors

$$\nabla J = \nabla \cdot J + \nabla \wedge J. \quad (4.15)$$

So we have $\nabla x = n$.

The vector derivative can be operated with any multivector, just as any other vector, extending the definition to the whole multivector field. It is

$$\nabla A = g^\iota \partial_\iota A; \quad (4.16)$$

in particular, for an r -grade multivector field A_r we have

$$\nabla \cdot A_r = \langle \nabla A_r \rangle_{r-1}, \quad (4.17)$$

$$\nabla \wedge A_r = \langle \nabla A_r \rangle_{r+1}. \quad (4.18)$$

These define the interior and exterior derivatives, respectively. The interior derivative is often called divergence.

We can verify the important result that the double exterior derivative always vanishes

$$\begin{aligned} \nabla \wedge (\nabla \wedge A) &= g^\iota \wedge \partial_\iota (g^\kappa \wedge \partial_\kappa A) \\ &= g^\iota \wedge g^\kappa \wedge (\partial_\iota \partial_\kappa A) = 0. \end{aligned} \quad (4.19)$$

This is true because $g^\iota \wedge g^\kappa$ is antisymmetric in ι and κ , whereas $\partial_\iota \partial_\kappa A$ is symmetric. The double divergence also vanishes; in order to prove this we set $B = AI$ and use the duality equation (3.70), as follows

$$\begin{aligned} \nabla \cdot (\nabla \cdot A) &= \nabla \cdot [\nabla \cdot (BI)] \\ &= \nabla \cdot (\nabla \wedge BI) \\ &= \nabla \wedge \nabla \wedge BI = 0. \end{aligned} \quad (4.20)$$

The vector derivative ∇ does not necessarily commute with all multivectors, so we need a series of notational conventions when writing expressions with this operator; these are:

- If there are no brackets, ∇ acts on the object immediately to its right.
- When ∇ is followed by brackets, the derivative is applied to the full contents of the bracket.
- When ∇ acts on a multivector to which it is not adjacent or which is adjacent on the left, we use arrows above the operator and its argument.

For example, using this notation we can write

$$\nabla(AB) = \nabla AB + \overrightarrow{\nabla} A \overleftarrow{B}, \quad (4.21)$$

which is just the rule for derivation of a product. Later in this chapter we will also use the arrow notation with linear functions. If $\mathbf{f}(a)$ is a position dependent linear function, we write

$$\overrightarrow{\nabla} \overleftarrow{\mathbf{f}}(a) = \nabla \mathbf{f}(a) - g^\iota \mathbf{f}(\partial_\iota a), \quad (4.22)$$

so that $\overrightarrow{\nabla} \overleftarrow{f}(a)$ only differentiates the point dependence in f and not in its argument.

There are a number of important results that we can build for multivectors depending linearly on x . For example,

$$\nabla x \cdot A_r = g_\iota g_\iota \cdot A_r, \quad (4.23)$$

with A_r a grade- r multivector. Using the results from Eqs. (3.108) and following we see that

$$\begin{aligned} \nabla x \cdot A_r &= r A_r, \\ \nabla x \wedge A_r &= (n - r) A_r, \\ \overrightarrow{\nabla} A_r \overleftarrow{x} &= (-1)^r (n - 2r) A_r, \end{aligned} \quad (4.24)$$

where n is the dimension of space.

4.2 Curvilinear coordinates

So far we have assumed that the frame vectors were fixed and formulated the vector derivative in terms of such fixed frames. Sometimes it is very convenient to allow the frame to be point dependent, namely when the problem under study has some symmetry, which recommends the use of specific coordinates. In those cases we define *curvilinear coordinates*. A general set of coordinates consists of a set of scalar functions $\{x^\iota(x)\}$, $\iota = 1, \dots, n$, defined over some region. In this region the functions can be inverted and we can write $x(x^\iota)$, expressing the position vector x parametrically in terms of the coordinates. If one of the coordinates is varied while all the others remain constant, the point describes a coordinate curve. The derivatives along these curves specify a set of frame vectors by

$$g_\iota(x) = \partial_\iota x = \lim_{\varepsilon \rightarrow 0} \frac{x(x^1, \dots, x^\iota + \varepsilon, \dots, x^n) - x}{\varepsilon}, \quad (4.25)$$

where coordinate x^ι is varied while the others remain fixed. The derivative in the g_ι direction, $g_\iota \cdot \nabla$ is found by moving slightly along g_ι , which is precisely the same as varying the x^ι coordinate with the others fixed. Therefore it is

$$g_\iota \cdot \nabla = \partial_\iota. \quad (4.26)$$

The coordinate system is valid over a given region on the condition that

$$g_1 \wedge g_2 \wedge \dots \wedge g_n \neq 0. \quad (4.27)$$

Since this quantity is never zero it follows that the frame keeps the same orientation over a given region.

A second frame can be constructed directly from the coordinate functions by

$$g^t = \nabla x^t. \quad (4.28)$$

This implies that the $\{g^t\}$ frame vectors have vanishing exterior derivative:

$$\nabla \wedge g^t = \nabla \wedge (\nabla x^t) = 0. \quad (4.29)$$

The notation we have chosen suggests that the two frames are reciprocal to one another. This can be verified

$$g_t \cdot g^\kappa = g_t \cdot \nabla x^\kappa = \frac{\partial x^\kappa}{\partial x^t} = \delta_t^\kappa. \quad (4.30)$$

When working in curvilinear coordinates, placing the indices above or below is no longer just a question of making the summation rule work; this was also true for mixed signature spaces but here it acquires an increased relevance. Also, when taking derivatives we cannot just think of the coordinate variation but also must take into account the variation of frame vectors themselves. Suppose we have $J = J^t g_t$. If we just derived the components J^t we would lose the variation due to the fact that the g_t are not constant. The variation of frame vectors provides the *connection coefficients* of tensor analysis.

There are two cases where the connection coefficients of curvilinear coordinates can be ignored. The first is the exterior derivative, for which we can write

$$\nabla \wedge J = \nabla \wedge (J_t g^t) = (\nabla J_t) \wedge g^t. \quad (4.31)$$

The exterior derivative has coordinates $\partial_t J_\kappa - \partial_\kappa J_t$, whatever the chosen coordinates. The second case is the divergence of a vector. We have

$$\nabla \cdot J = \nabla \cdot (J^t g_t). \quad (4.32)$$

It is convenient to define the volume factor V by

$$g_1 \wedge g_2 \wedge \cdots \wedge g_n = IV, \quad (4.33)$$

where I is the unit pseudoscalar. It is then clear that the frame vector g_t can be obtained by

$$g_t = (-1)^{t-1} g^n \wedge g^{n-1} \wedge \cdots \check{g}^t \wedge \cdots \wedge g^1 IV, \quad (4.34)$$

with the check over g^t meaning that this vector is excluded from the outer product. Recalling that each of the g^t has null exterior derivative, it is straightforward to write

$$\nabla \cdot J = \frac{1}{V} \partial_t (V J^t). \quad (4.35)$$

A consequence of this result is that the Laplacian ∇^2 of a scalar field is also connection independent, as we see from

$$\nabla^2 \phi = \frac{1}{V} \partial_t (V g^{t\kappa} \partial_\kappa \phi), \quad (4.36)$$

where $g^{\iota\kappa} = g^\iota \cdot g^\kappa$.

We will now examine some of the most frequently used orthogonal curvilinear coordinates. We define a set of orthonormal vectors by taking an orthogonal frame $\{g_\iota\}$ and defining the magnitudes

$$h_\iota = \begin{cases} (g_\iota \cdot g_\iota)^{1/2}, & \text{for } g_\iota \cdot g_\iota > 0, \\ (-g_\iota \cdot g_\iota)^{1/2}, & \text{for } g_\iota \cdot g_\iota < 0. \end{cases} \quad (4.37)$$

The magnitudes are positive numbers, no matter whether the respective frame vectors have positive or negative square. We don't consider here degenerate cases, where the frame vectors might have zero magnitude. In terms of the magnitudes we can write (no sums implied)

$$g_\iota = h_\iota \sigma_\iota, \quad g^\iota = \frac{1}{h_\iota} \sigma^\iota. \quad (4.38)$$

We can then use the $\{\sigma_\iota\}$ frame, which is orthonormed with vectors of positive or negative square. For a vector J we have

$$J = J^\iota \sigma_\iota = \sum_{\iota=1}^n \frac{J_\iota}{h_\iota} g^\iota. \quad (4.39)$$

The position vector x is therefore

$$x = x^\iota g_\iota = \sum_{\iota=1}^n x^\iota h_\iota \sigma_\iota. \quad (4.40)$$

Cartesian coordinates

These are the basic coordinates for all other coordinate systems; we will have to consider the special cases of 3-dimensional Euclidean space and 5-dimensional Minkowski spacetime, with the more traditional 4-dimensional spacetime of special relativity being derived from the latter. The notation conventions set out in Sec. 2.5 will now become important in order to identify the space or subspace where we want an equation or an expression to be valid. In 3-dimensional Euclidean space we will work with the right handed frame $\{\sigma_m\}$, $\sigma_{123} = i$; similarly, in 5-dimensional Minkowski space we will choose the frame $\{\sigma_\alpha\}$, with $(\sigma_0)^2 = -1$, $(\sigma_i)^2 = 1$ and $\sigma_{01234} = i$. It follows that 3D Euclidean space is a subspace of 5D Minkowski spacetime. The coordinates will frequently be denoted (t, x, y, z, τ) , instead of the conventional x^α notation. The case of special relativity need not be dealt with specifically since we map bivectors of 5D Minkowski spacetime to vectors of 4D Minkowski spacetime by the relation

$$\sigma_{\mu 4} \mapsto \gamma_\mu. \quad (4.41)$$

With this definition we see immediately that we have $(\gamma_0)^2 = 1$, $(\gamma_m)^2 = -1$, producing the usual $(+ - - -)$ signature.

Cylindrical polar coordinates

We are here dealing with 3D Euclidean space; the coordinates are denoted (ρ, ϕ, z) , where ρ and ϕ are the standard 2-dimensional polar coordinates

$$\rho = (x^2 + y^2)^{1/2}, \quad \tan \phi = \frac{y}{x}. \quad (4.42)$$

The coordinates lie in the ranges $0 \leq \rho < \infty$ and $0 \leq \phi < 2\pi$. The frame vectors are

$$\begin{aligned} \sigma_\rho &= \cos(\phi)\sigma_1 + \sin(\phi)\sigma_2, \\ \sigma_\phi &= -\sin(\phi)\sigma_1 + \cos(\phi)\sigma_2, \\ \sigma_z &= \sigma_3. \end{aligned} \quad (4.43)$$

Here we have used the convention that the frame vector indices indicate the coordinate to which they are associated. The frame vectors verify

$$\sigma_\rho \sigma_\phi \sigma_z = \sigma_{123} = \mathbf{i}. \quad (4.44)$$

The position vector x is given by

$$x = \rho\sigma_\rho + z\sigma_z. \quad (4.45)$$

If a point is moved along each coordinate in turn, we get the following relations

$$\begin{aligned} \partial_\rho x &= \sigma_\rho, \\ \partial_\phi x &= \rho\partial_\phi\sigma_\rho = \rho\sigma_\phi, \\ \partial_z x &= \sigma_z, \end{aligned} \quad (4.46)$$

from which we determine the magnitudes $h_\rho = 1$, $h_\phi = \rho$, $h_z = 1$.

Spherical polar coordinates

Spherical polar coordinates arise very frequently in physics and in engineering; they are labelled (r, θ, ϕ) and are defined by

$$r = (x^2 + y^2 + z^2)^{1/2}, \quad \cos(\theta) = \frac{z}{r}, \quad \tan(\phi) = \frac{y}{x}. \quad (4.47)$$

The coordinate ranges are $0 \leq r < \infty$, $0 \leq \theta < \pi$ and $0 \leq \phi < 2\pi$. Over the z axis the coordinate ϕ is undefined. The position vector x has components in the Euclidean frame

$$x = r \sin(\theta) \cos(\phi)\sigma_1 + r \sin(\theta) \sin(\phi)\sigma_2 + \cos(\theta)\sigma_3. \quad (4.48)$$

From this expression one can compute the orthonormal frame vectors, which are

$$\begin{aligned}\sigma_r &= \sin(\theta) \cos(\phi) \sigma_1 + \sin(\theta) \sin(\phi) \sigma_2 + \cos(\theta) \sigma_3, \\ \sigma_\theta &= \cos(\theta) \cos(\phi) \sigma_1 + \cos(\theta) \sin(\phi) \sigma_2 - \sin(\theta) \sigma_3, \\ \sigma_\phi &= -\sin(\phi) \sigma_1 + \cos(\phi) \sigma_2.\end{aligned}\tag{4.49}$$

In terms of these, the position vector is quite simply

$$x = r \sigma_r.\tag{4.50}$$

Variations along each of the coordinate lines produce

$$\begin{aligned}\partial_r x &= \sigma_r, \\ \partial_\theta x &= r \partial_\theta \sigma_r = r \sigma_\theta, \\ \partial_\phi x &= r \partial_\phi \sigma_r = r \sin(\theta) \sigma_\phi.\end{aligned}\tag{4.51}$$

From these relations we derive the magnitude factors $h_r = 1$, $h_\theta = r$ and $h_\phi = r \sin(\theta)$.

It is sometimes useful to have the full list of connection coefficients; these are the components of partial derivatives of the frame vectors; the latter can easily be worked out as being

$$\begin{aligned}\partial_r \sigma_r &= 0, & \partial_\theta \sigma_r &= \sigma_\theta, & \partial_\phi \sigma_r &= \sin \theta \sigma_\phi, \\ \partial_r \sigma_\theta &= 0, & \partial_\theta \sigma_\theta &= -\sigma_r, & \partial_\phi \sigma_\theta &= \cos \theta \sigma_\phi, \\ \partial_r \sigma_\phi &= 0, & \partial_\theta \sigma_\phi &= 0, & \partial_\phi \sigma_\phi &= -\sin \theta \sigma_r - \cos \theta \sigma_\theta.\end{aligned}\tag{4.52}$$

The orthonormal frame vectors satisfy $\sigma_r \sigma_\theta \sigma_\phi = \mathbf{i}$, so that $\{\sigma_r, \sigma_\theta, \sigma_\phi\}$ form a right-handed frame. This frame can be obtained from the $\{\sigma_m\}$ frame by application of a rotor, so that $\sigma_r = R \sigma_3 \tilde{R}$, $\sigma_\theta = R \sigma_1 \tilde{R}$ and $\sigma_\phi = R \sigma_2 \tilde{R}$. The rotor is given by

$$R = e^{-i\sigma_3\phi/2} e^{-i\sigma_2\theta/2}.\tag{4.53}$$

Hyperspherical coordinates

When discussing cosmology we will find the need to use hyperspherical coordinates in 4-dimensional Euclidean space, seen as a subspace of 5D Minkowski spacetime. Those coordinates are a natural extension of the 3-dimensional case; they are denoted by $(\rho, \theta, \phi, \tau)$; coordinates ρ , θ and ϕ are angles and coordinate τ is a distance. The frame vectors can be expressed in the orthonormal frame $\{\sigma_i\}$ by the following definitions

$$\begin{aligned}\sigma_\tau &= \sin(\rho) \sin(\theta) \cos(\phi) \sigma_1 + \sin(\rho) \sin(\theta) \sin(\phi) \sigma_2 + \sin(\rho) \cos(\theta) \sigma_3 + \cos(\rho) \sigma_4, \\ \sigma_\rho &= \cos(\rho) \sin(\theta) \cos(\phi) \sigma_1 + \cos(\rho) \sin(\theta) \sin(\phi) \sigma_2 + \cos(\rho) \cos(\theta) \sigma_3 - \sin(\rho) \sigma_4, \\ \sigma_\theta &= \cos(\theta) \cos(\phi) \sigma_1 + \cos(\theta) \sin(\phi) \sigma_2 - \sin(\theta) \sigma_3, \\ \sigma_\phi &= -\sin(\phi) \sigma_1 + \cos(\phi) \sigma_2.\end{aligned}\tag{4.54}$$

The position vector is simply

$$x = \tau \sigma_\tau. \quad (4.55)$$

We now write the variations along coordinate lines, as we did before;

$$\begin{aligned} \partial_\tau x &= \sigma_\tau, \\ \partial_\rho x &= \tau \sigma_\rho, \\ \partial_\theta x &= \tau \sin(\rho) \sigma_\theta, \\ \partial_\phi x &= \tau \sin(\rho) \sin(\theta) \sigma_\phi. \end{aligned} \quad (4.56)$$

These relations are a natural extension of the 3-dimensional ones.

The partial derivatives of the frame vectors, producing the connection coefficients, are summarized in the list

$$\begin{aligned} \partial_\rho \sigma_\tau &= \sigma_\rho, & \partial_\theta \sigma_\tau &= \sin \rho \sigma_\theta, & \partial_\phi \sigma_\tau &= \sin \theta \sin \rho \sigma_\phi, \\ \partial_\rho \sigma_\rho &= -\sigma_\tau, & \partial_\theta \sigma_\rho &= \cos \rho \sigma_\theta, & \partial_\phi \sigma_\rho &= \sin \theta \cos \rho \sigma_\phi, \\ \partial_\rho \sigma_\theta &= 0, & \partial_\theta \sigma_\theta &= -\sin \rho \sigma_\tau - \cos \rho \sigma_\rho, & \partial_\phi \sigma_\theta &= \cos \theta \sigma_\phi, \\ \partial_\rho \sigma_\phi &= 0, & \partial_\theta \sigma_\phi &= 0, & \partial_\phi \sigma_\phi &= -\sin \rho \sin \theta \sigma_\tau \\ & & & & & -\cos \rho \sin \theta \sigma_\rho - \cos \theta \sigma_\theta. \end{aligned} \quad (4.57)$$

All the derivatives with respect to τ are zero and were not listed.

4.3 Analytic and monogenic functions

In this section we show that the vector derivative provides a natural way of extending the analytic function concept beyond the field of complex numbers.

Suppose we have a complex function $\psi(z) = u + iv$, with a complex argument $z = x + iy$. This function is analytic if it satisfies the Cauchy-Riemann equations

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}. \quad (4.58)$$

These equations can be applied directly in algebras where the unit pseudoscalar I is imaginary and commutes with all vectors, such as the algebras \mathcal{G}_{30} and \mathcal{G}_{41} , but they have no direct equivalent in other algebras. We are looking for a concept which generalizes analytic functions to algebras of arbitrary dimension and signature; it is then natural to start with two dimensions and the Argand plane.

We use $\{\sigma_1, \sigma_2\}$ to define an orthonormal frame in two dimensions. The Argand plane derives from this by setting σ_1 as the real axis. The coordinates are denoted (x, y) and the position vector is written

$$r = x\sigma_1 + y\sigma_2. \quad (4.59)$$

With this notation the vector derivative becomes

$$\nabla = \sigma_1 \partial_x + \sigma_2 \partial_y; \quad (4.60)$$

the coordinates are used as indices in the partial derivatives under the convention used before. We saw in Sec. 2.3 that the position vector is mapped onto a complex number by pre-multiplying with the vector representing the real axis:

$$z = x + Iy = \sigma_1 r. \quad (4.61)$$

Applying the vector derivative to the function ψ we get

$$\nabla \psi = (\partial_x u - \partial_y v) \sigma_1 + (\partial_x v + \partial_y u) \sigma_2. \quad (4.62)$$

The statement that ψ is analytic is now equivalent to the equation

$$\nabla \psi = 0. \quad (4.63)$$

This equation can be immediately generalized to any algebra. We call functions verifying condition (4.63) *monogenic functions* and these are equivalent to analytic functions in the particular case of two dimensions.

It is instructive to deepen the link with complex analysis by looking at some other relations. The complex partial derivative ∂_z is defined by the properties

$$\frac{\partial z}{\partial z} = 1, \quad \frac{\partial z^\dagger}{\partial z} = 0 \quad (4.64)$$

and similarly for the complex conjugate

$$\frac{\partial z}{\partial z^\dagger} = 0, \quad \frac{\partial z^\dagger}{\partial z^\dagger} = 1. \quad (4.65)$$

We can now write the complex partial derivatives in terms of coordinate ones as

$$\frac{\partial}{\partial z} = \frac{1}{2} (\partial_x - I \partial_y), \quad \frac{\partial}{\partial z^\dagger} = \frac{1}{2} (\partial_x + I \partial_y). \quad (4.66)$$

An analytic function depends only on z , that is, it is independent of z^\dagger . We then have

$$\frac{\partial \psi(z)}{\partial z^\dagger} = 0. \quad (4.67)$$

Using the relations above, this can be written

$$\frac{1}{2} (\partial_x + I \partial_y) \psi = \frac{1}{2} \sigma_1 \nabla \psi = 0, \quad (4.68)$$

recovering the previous equation.

We can easily show that a Taylor series expansion in z about z_0 automatically returns an analytic function. We first see that

$$\nabla z = \nabla(\sigma_1 r) = 2\sigma_1 \cdot r - \sigma_1 \cdot r = 2\sigma_1 - 2\sigma_1 = 0. \quad (4.69)$$

From this we form immediately

$$\nabla(z - z_0)^n = n\nabla(\sigma_1 r - z_0)(z - z_0)^{n-1} = 0, \quad (4.70)$$

thus proving the statement.

In order to analyse monogenic functions in 3D we let the vector derivative ∇ operate on general multivectors, imposing the condition $\nabla\psi = 0$. If ψ contains all grades it is straight forward to show that even-grade and odd-grade components must satisfy the condition independently; we can then concentrate on even-grade solutions.

In order to construct monogenic functions in 3D we recall that $\nabla r = 3$ and

$$\nabla(ar) = -a \quad (4.71)$$

for any constant vector a . It follows that

$$\psi = ra + 3ar \quad (4.72)$$

is a monogenic function. The power series route used in two dimensions does not work here due to the lack of commutativity, but we can still construct monogenic functions via their angular properties.

We assume that Ψ is a monogenic function that depends on the radial coordinate r by a power law. Using polar coordinates we can then write

$$\Psi = r^l \psi(\theta, \phi). \quad (4.73)$$

The function $\psi(\theta, \phi)$ must verify

$$lr^{l-1}\sigma_r\psi + r^l\nabla\psi = 0. \quad (4.74)$$

An eigenvalue equation can then be written for ψ :

$$-r\sigma_r\wedge\nabla\psi = l\psi. \quad (4.75)$$

These angular eigenstates play a key role in quantum mechanics. Since Ψ is monogenic it must have zero Laplacian, that is

$$\nabla^2\Psi = 0. \quad (4.76)$$

In a constant basis each component of Ψ satisfies Laplace's equation. It follows that each component of ψ is a spherical harmonic and hence that l is an integer. We can start with the function $(x + i\sigma_3 y)^l$, which we know is monogenic from the two-dimensional analysis. If we return to polar coordinates

$$(x + i\sigma_3 y)^l = r^l \sin^l(\theta) e^{l\phi i\sigma_3}, \quad (4.77)$$

which gives us our first family of monogenic functions

$$\psi_l^l = \sin^l(\theta) e^{l\phi i\sigma_3}. \quad (4.78)$$

The remaining monogenic functions will be constructed from this one when we study the hydrogen atom.

The algebra of 5-dimensional spacetime \mathcal{G}_{41} is seldom used in physics textbooks but it happens to take a key role in this book. We construct the vector derivative for this algebra by introducing an orthonormal frame $\{\sigma_\alpha\}$, with signature $(- + + +)$ and associated coordinates x^α . The vector derivative is then

$$\nabla = \sigma^\alpha \partial_\alpha. \quad (4.79)$$

It will sometimes be convenient to replace the notation for coordinate x^0 by t ; when this is done the vector derivative can also be written

$$\nabla = \sigma^0 \partial_t + \sigma^\mu \partial_\mu = \sigma^0 \partial_t + {}^\mu \nabla. \quad (4.80)$$

The index preceding the derivative operator indicates that the latter is applied in the subspace spanned by the basis $\{\sigma_\mu\}$, $\mu = 1 \cdots 4$. Written in this form, the vector derivative will find application in the Dirac theory. If we form the Laplacian we see that

$$\nabla^2 = -\frac{\partial^2}{\partial t^2} + {}^\mu \nabla^2, \quad (4.81)$$

which is the fundamental operator describing waves travelling at unit speed in 4-dimensional space. The monogenic equation $\nabla\psi = 0$ will be discussed at length in forthcoming chapters.

5 Five-dimensional spacetime

We stated in the introduction to this book that our objective was to derive the main equations of physics from simple geometric postulates, these remaining unchanged for the different areas of physics. Thus we want to find those geometric principles that lead to the equations of General Relativity, Electromagnetism, Quantum Mechanics and hopefully also Particle Physics. The reader is warned that this ultimate goal has not been achieved, however we intend to show that a great deal of unification is possible under a geometrical approach to physics. We believe that with work and time the full unification of physics under geometry will be possible and we hope to open paths leading in that direction. Can we imagine a way in which physics cannot be unified via a geometrical approach? Yes, particularly if the unification requires some new area of mathematics that is not included in the three great realms of mathematics: geometry, algebra and analysis. The example of number theory in today's mathematics might be a hint that there's much more mathematics to be discovered, for number theory has numerous unsolved problems and open-ended challenges that probably lead to new areas of mathematics. There could be some topic in number theory – undiscovered so far – that is closely related to the real foundation of physics. Hopefully, geometric algebra will be able to include this or any foundation at its own level not far removed from the intrinsic origin.

The fundamental principles of any theory are postulates and one should not try to justify their adoption other than by the consistency of predictions made by the theory with observations in the Universe, both observations of phenomena occurring naturally and laboratory experiments. Many readers will be skeptical about the adoption of 5-dimensional space without a postulate about the compactness of the fifth dimension to make sure that it remains undetected at the macroscopic scale; this was the approach brought by Klein to what is now known as Kaluza-Klein theory.[4] Most people will be quite happy to work with complex numbers in 4-dimensional spacetime but will find unnatural to work with real numbers in 5-dimensional spacetime, yet the applicable algebras are isomorphic. When there is isomorphism between two algebras one has the choice of working with the one that best suits him, however, the geometric interpretation and the intuition one gains for the problems at hand is strongly influenced by the choice that is made. We have already seen that the algebra of 3-dimensional space is isomorphic to the algebra of 2×2 matrices, but while the former is strongly related to geometry the latter is not. For instance, it would have been extremely difficult to introduce the vector derivative and monogenic functions if we had worked solely with matrices. The choice between 4- or 5-dimensional spacetime is not unlike the choice between matrix or

geometric algebras; basically, the scalar imaginary of 4 dimensions upsets the connection with geometry, while it can be assigned to the unitary volume element in 5 dimensions.

We will take 5-dimensional spacetime as the geometry we want to explore, for the simple reason that it provides many relations that can be interpreted physically. At the outset we will not seek any correspondence between the dimensions of 5-dimensional spacetime and those used in physics; we will let that relationship appear naturally from the interpretation of geometric equations as physically meaning ones. The question arises, why should we start with five dimensions and one particular signature and why should we choose a configuration very seldom found in physics books. The answer is that the author did not start this process axiomatically but rather by induction; what today is submitted as first principle is the result of many trials and errors, rejecting the options that led nowhere and keeping the most fruitful and simple ones. It may well happen that today's option will tomorrow need to be replaced by a different one; this is an endless process where we get more educated all the time. We believe that the level of unification achieved by the present geometric approach to physics amply justifies the time and effort the reader will be asked to invest into understanding this proposal.

5.1 The algebra of five-dimensional spacetime

In physics the term spacetime usually designates a space with mixed signature, where only one dimension has signature opposite to all the others. The most common spacetime in physics is 4-dimensional and it can appear with the two signatures $(-+++)$ or its opposite $(+---)$. No physical significance is attributed to the choice of either signature, this being made by mere convenience, but the latter seems to have more followers than the former. The algebra associated that signature is denoted $\mathcal{G}_{1,3}$ and is spanned by the orthonormed frame $\{\gamma_\mu\}$, verifying the relation

$$\gamma_\mu \gamma_\nu + \gamma_\nu \gamma_\mu = \text{diag}(2, -2, -2, -2)_{\mu\nu}. \quad (5.1)$$

The unit pseudoscalar of the algebra, $I = \gamma_{0123}$, squares to -1 but it does not commute with vectors, as such it cannot be assigned to the complex imaginary. We can, therefore, complexify the algebra with the introduction of the complex imaginary i , thus adding one extra dimension to 4-dimensional spacetime. Five-dimensional spacetime then results from the assignment:

$$\sigma_\mu \equiv -iI\gamma_\mu, \quad \sigma_4 \equiv -iI. \quad (5.2)$$

Five-dimensional spacetime is not indifferent to the signature choice; the two main consequences of the signature option are: signature $(-++++)$ is associated with \mathcal{G}_{41} algebra, which is a complex algebra with an imaginary pseudoscalar and is isomorphic to the complex algebra of 4×4 matrices; it is said to be a *simple algebra*. Signature $(+----)$, on the other hand, is associated with \mathcal{G}_{14} , which does not provide an imaginary

pseudoscalar and is not a simple algebra. In this book the designation 5-dimensional spacetime stands for 5-dimensional space with signature $(-+++)$, associated with geometric algebra \mathcal{G}_{41} .

The algebra we need to construct is spanned by five orthonormed vectors $\{\sigma_\alpha\}$ satisfying the algebraic relations

$$(\sigma_0)^2 = -1, \quad \sigma_0 \cdot \sigma_i = 0, \quad \sigma_i \cdot \sigma_j = \delta_{ij}, \quad (5.3)$$

where the indices follow the indexing conventions of Sec. 2.5. These relations can be summarized in the relation

$$\sigma_\alpha \sigma_\beta + \sigma_\beta \sigma_\alpha = 2\eta_{\alpha\beta} = \text{diag}(-2, 2, 2, 2, 2). \quad (5.4)$$

Associated to the main frame we must consider a reciprocal frame $\{\sigma^\alpha\}$, such that $\sigma^\alpha \sigma_\beta = \delta^\alpha_\beta$; the reciprocal frame vectors then verify

$$\sigma^0 = -\sigma_0, \quad \sigma^i = \sigma_i. \quad (5.5)$$

A general vector a can be expanded in its components as

$$a = a^\alpha \sigma_\alpha. \quad (5.6)$$

Since we are in mixed signature space, the square of a vector can be positive, negative or null; the latter happens when

$$(a^0)^2 = \sum_{i=1}^4 (a^i)^2. \quad (5.7)$$

The subspace of nilpotent vectors is denominated *null subspace* and is particularly significant in physics, as we shall see.

Our algebra has 10 bivectors, $\sigma_{\alpha\beta}$, which fall into two categories: those containing σ_0 (e.g. σ_{0i}) and those that do not contain σ_0 (e.g. σ_{ij}). The two categories have different sign to their squares; those containing σ_0 produce

$$(\sigma_{0i})^2 = \sigma_0 \sigma_i \sigma_0 \sigma_i = -\sigma_0 \sigma_i \sigma_i \sigma_0 = -(\sigma_0)^2 = 1, \quad (5.8)$$

while for the second category it is

$$(\sigma_{ij})^2 = \sigma_i \sigma_j \sigma_i \sigma_j = -\sigma_i \sigma_j \sigma_j \sigma_i = -(\sigma_i)^2 = -1. \quad (5.9)$$

As discussed in Sec. 3.2, bivectors with negative square generate pure rotations and bivectors with positive square generate boosts; specifically we have

$$\begin{aligned} e^{\alpha\sigma_{0i}} &= \cosh(\alpha) + \sinh(\alpha)\sigma_{0i}, \\ e^{\alpha\sigma_{ij}} &= \cos(\alpha) + \sin(\alpha)\sigma_{ij}. \end{aligned} \quad (5.10)$$

This will prove crucial to our treatment of Lorentz transformations.

The unit pseudoscalar of \mathcal{G}_{41} algebra is

$$i = \sigma_{01234}. \quad (5.11)$$

We use the complex imaginary notation i because the pseudoscalar squares to -1 and commutes with every element of the algebra. From what we said above it becomes clear that this algebra has a basis with 32 elements, as follows:

$$\begin{array}{cccccc} 1 & \{\sigma_\alpha\} & \{\sigma_{\alpha\beta}\} & \{i\sigma_{\alpha\beta}\} & \{i\sigma_\alpha\} & i \\ 1 \text{ scalar} & 5 \text{ vectors} & 10 \text{ bivectors} & 10 \text{ trivectors} & 5 \text{ fourvectors} & 1 \text{ pseudoscalar} \end{array}$$

Almost everything we need to know about 5-dimensional flat spacetime and Lorentz transformations is encoded in the structure of this algebra. A general element of the algebra can be expanded into

$$M = \alpha + a + B + iC + ib + i\beta, \quad (5.12)$$

with α and β scalars, a and b vectors and B and C bivectors. The reverse of this element is given by

$$\widetilde{M} = \alpha + a - B - iC + ib + i\beta. \quad (5.13)$$

We can expand even further, in terms of components, as

$$\begin{aligned} M = & \alpha + a^0\sigma_0 + a^i\sigma_i + B^{0i}\sigma_{0i} + B^{ij}\sigma_{ij} \\ & + iC^{0i}\sigma_{0i} + iC^{ij}\sigma_{ij} + ib^0\sigma_0 + ib^i\sigma_i + i\beta, \end{aligned} \quad (5.14)$$

which allows us to write the conjugate

$$\begin{aligned} M^\dagger = & \alpha - a^0\sigma_0 + a^i\sigma_i + B^{0i}\sigma_{0i} - B^{ij}\sigma_{ij} \\ & - iC^{0i}\sigma_{0i} + iC^{ij}\sigma_{ij} + ib^0\sigma_0 - ib^i\sigma_i - i\beta. \end{aligned} \quad (5.15)$$

The modulus of M is obtained, as usual, by $|M| = (MM^\dagger)^{1/2}$.

5.2 Matrix representation

Every expression and equation that can be written in geometric algebra has an equivalent matrix formulation that one can arrive at by a suitable choice of an isomorphism between the particular geometric algebra under consideration and an appropriate matrix algebra. This statement can be demonstrated, but that demonstration is not our concern. Our choice is to work with geometric algebra whenever possible, since that is the way to bring out the geometric meaning of the expressions. The use of matrices is justified in two circumstances: when there is a need to demonstrate the equivalence between expressions

we arrive at and those found in the literature, and when dealing with symmetry groups. Quantum mechanics is overwhelmingly treated in the literature with a matrix approach and this situation implies that we appeal to a matrix isomorphism so that readers can compare the formulation found in this book with the standard one. Symmetry groups are the realm of matrices; this is a case where matrix formulation is actually clearer than the equivalent geometric algebra alternative, but we can have the best of both worlds by keeping the two formulations in parallel.

In the case of G_{41} algebra the appropriate isomorphism is with 4×4 matrix algebra over the complex field. There are many possible choices for the specific representation of geometric algebra elements but we will adopt the one in Table 5.1, because it is consistent with Dirac-Pauli representation commonly found in the quantum mechanics literature. We use the same symbols to represent geometric algebra elements and their matrix

Table 5.1: Matrix representation of basis vectors

$$\begin{aligned} \sigma_0 &\equiv \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, & \sigma_1 &\equiv \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{pmatrix}, & \sigma_2 &\equiv \begin{pmatrix} 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & i \\ 0 & 0 & -i & 0 \end{pmatrix}, \\ \sigma_3 &\equiv \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, & \sigma_4 &\equiv \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}. \end{aligned}$$

equivalents, since the context will usually clarify which ones are meant; actually working with either representation is formally equivalent and so there is no harm in switching between representations if that contributes to clarify the meaning of expressions. The only adaptation that is needed when converting a geometric expression to matrices is the need to include an identity matrix next to scalar and pseudoscalar elements. For instance, the scalar α of geometric algebra has a matrix equivalent αI , with I the 4-dimensional identity matrix. We will usually suppress the identity matrix, so that matrix and geometric algebra expressions become fully equivalent.

The equivalence to the Dirac-Pauli representation is easily demonstrated by defining the bivector elements

$$\gamma_\mu = \sigma_{4\mu}. \quad (5.16)$$

It will then be found that the γ_μ bivectors verify the relation (5.1), which is the defining relation of Dirac algebra.

Some concepts and definitions can be directly imported from matrix algebra into the algebra of 5D spacetime. For instance, we say that an element of the algebra M is *Hermitian* if it is the same as its conjugate, $M = M^\dagger$; we also say that M is unitary if it has unit modulus, $MM^\dagger = 1$. A useful theorem from matrix algebra allows us to say that if H is Hermitian, then

$$U = e^{iH} \quad (5.17)$$

is unitary. The demonstration using geometric algebra is straightforward.

5.3 The null subspace

Several derivations in the remainder of this book will bring us to the null subspace of 5D spacetime. Having some knowledge of physics, the reader certainly expects that somehow the original 5 dimensions are to be reduced by one, in order to allow a proper assignment to the entities of observable physical world. The null subspace of 5D spacetime is actually 4-dimensional, but it is a space with special characteristics; all tangent vectors have null length. In order to understand what this means, we increase the dimensionality in steps, starting with the null subspace in two dimensions. The null subspace contains every point on the plane but restricts the number of allowed paths through each point to two perpendicular straight lines. Assuming that $\{\sigma_0, \sigma_1\}$ are the orthogonal vectors of the plane, the allowed paths are those that bisect the angle between σ_0 and σ_1 . The path length is always null by definition but we can still characterize any displacement using either coordinate as parameter. The null condition implies that

$$\frac{dx^1}{dx^0} = 1, \quad (5.18)$$

so we can say that for elementary null displacements it is $dx^0 = dx^1$.

One dimension higher, the null subspace limits the allowed paths through each point to straight lines on the surface of a cone with apex at that point. This can be seen if we allow paths in 3 dimensions to be parameterized by an arbitrary parameter λ ; assuming an orthonormal basis $\{\sigma_0, \sigma_1, \sigma_2\}$, these paths have tangent vector given by

$$\frac{dx}{d\lambda} = \sigma_0 \frac{dx^0}{d\lambda} + \sigma_1 \frac{dx^1}{d\lambda} + \sigma_2 \frac{dx^2}{d\lambda}. \quad (5.19)$$

If the tangent vector is nilpotent we get

$$\left(\frac{dx^0}{d\lambda}\right)^2 - \left(\frac{dx^1}{d\lambda}\right)^2 - \left(\frac{dx^2}{d\lambda}\right)^2 = 0. \quad (5.20)$$

But since λ is an arbitrary parameter we are allowed to replace it by x^0 to obtain

$$\left(\frac{dx^1}{dx^0}\right)^2 + \left(\frac{dx^2}{dx^0}\right)^2 = 1, \quad (5.21)$$

which we rearrange as

$$(dx^0)^2 = (dx^1)^2 + (dx^1)^2. \quad (5.22)$$

This tells us that in null elementary displacements the variation of coordinate x^0 equals the radial displacement $[(dx^1)^2 + (dx^1)^2]^{1/2}$, that is, the displacement takes place along the generatrix of a $\pi/2$ cone. Now note that we could, for instance, have chosen x^2 for parameter, which would lead us to the result

$$(dx^2)^2 = (dx^0)^2 - (dx^1)^2. \quad (5.23)$$

In 5-dimensional spacetime the alternative choices for the parameter will give us different perspectives over physical phenomena.

Extending the reasoning to 4 dimensions becomes straightforward; assuming the orthonormal basis $\{\sigma_\mu\}$ we obtain for an elementary null displacement

$$(dx^0)^2 = \sum_{m=1}^3 (dx^m)^2. \quad (5.24)$$

The right-hand side represents the squared radius of a sphere centered on the point where we want the paths to pass. This radius is matched by the variation of coordinate x^0 . Obviously in 5 dimensions the process leads to

$$(dx^0)^2 = \sum_{i=1}^4 (dx^i)^2 \quad (5.25)$$

and we say that the right-hand side is the radius of a 4-dimensional hypersphere. The possibility of choosing any coordinate as parameter is especially interesting when we opt for x^4 because we get a familiar relation from special relativity; it is

$$(dx^4)^2 = (dx^0)^2 - \sum_{m=1}^3 (dx^m)^2. \quad (5.26)$$

Later on we will return to the two equations above, discussing them in more detail.

5.4 The Lorentz group

The Lorentz group of transformations is known from Minkowski spacetime. In geometric algebra this means that the analysis must be performed in $\mathcal{G}_{1,3}$ or in $\mathcal{G}_{3,1}$; surprisingly similar conclusions can be drawn in either of those algebras, in spite of there being no isomorphism between them. From the standpoint of 5-dimensional spacetime and $\mathcal{G}_{4,1}$ algebra, the easiest approach, the one we shall adopt, consists on working on the subspace spanned by the 4 vectors $\{\sigma_\mu\}$, as if σ_4 did not exist. Obviously, Lorentz transformations

are then a subgroup of transformations defined for the whole 5-dimensional spacetime. We don't have a clear idea of the possible physical significance of this extended group, whereas the interpretation of the Lorentz group defined in Minkowski spacetime is well understood.

The Lorentz group consists of the transformation group for 4-dimensional vectors that preserves length and angles. These include reflections and rotations. A reflection in the hyperplane normal to n is achieved by

$$a \mapsto -nan^{-1}. \quad (5.27)$$

This works for both $n^2 > 0$ and $n^2 < 0$, but a nilpotent normal is not allowed because n^{-1} does not exist. The reflections with $n^2 > 0$ preserve the ordering with respect to coordinate x^0 ; this is usually referred to in the literature as time ordering. Pairs of either type of reflection are also x^0 ordering preserving, but mixed pairs do not preserve x^0 ordering. We concentrate on even numbers of reflections which have determinant $+1$. If we combine even numbers of reflections we arrive at a transformation of the form

$$a \mapsto \psi a \psi^{-1}, \quad (5.28)$$

where ψ is a special kind of even multivector. We note that $\psi = n_1 \cdots n_n$ is the product of an even number of vectors, so that

$$\psi \tilde{\psi} = n_1 \cdots n_n n_n \cdots n_1 = \rho, \quad (5.29)$$

with $\rho \neq 0$ a scalar. We can now define a rotor R by

$$R = \psi \rho^{-1/2}, \quad (5.30)$$

so that

$$R \tilde{R} = \psi \tilde{\psi} \rho^{-1} = 1, \quad (5.31)$$

as required for a rotor. We then have

$$\psi = \rho^{1/2} R, \quad \psi^{-1} = \rho^{-1/2} \tilde{R}, \quad (5.32)$$

and our general transformation becomes

$$a \mapsto Ra\tilde{R}. \quad (5.33)$$

Every rotor in 4-dimensional algebra can be written in terms of bivectors as

$$R = e^{B/2}, \quad (5.34)$$

with B a grade-2 multivector.

Since we are essentially working in 4-dimensional algebra it is useful to define the unit pseudoscalar of this algebra as $I = \sigma_{01234}$. We then find that the square of bivector B can have only scalar and pseudoscalar terms and can, therefore, be written as

$$B^2 = \langle B^2 \rangle_0 + \langle B^2 \rangle_4 = \rho e^{I\phi}, \quad (5.35)$$

and we assume that the bivector is not nilpotent, so, $\rho \neq 0$. A new unitary bivector can be defined by

$$\hat{B} = \rho^{-1/2} e^{-I\phi/2} B. \quad (5.36)$$

Indeed we have

$$\hat{B}^2 = \rho^{-1} e^{-I\phi} B^2 = 1. \quad (5.37)$$

The unitary bivector is instrumental in decomposing bivector B into two components; it is

$$B = \rho^{1/2} e^{I\phi/2} \hat{B} = \alpha \hat{B} + \beta I \hat{B}. \quad (5.38)$$

The two terms commute, which allows us to decompose rotor R into

$$R = e^{\alpha \hat{B}/2} e^{\beta I \hat{B}/2}. \quad (5.39)$$

This is a very convenient split into a boost, generated by \hat{B} and a rotation, generated by $I\hat{B}$.

The extension of the Lorentz group into 5-dimensional spacetime involves not only the consideration of bivectors involving σ_4 but also four vectors, because the product of two rotations induced by bivectors can generate a fourvector rotor. This extended group is not discussed further in this book, which does not mean it lacks physical relevance; it may indeed happen that some equations of physics may be related to the extra symmetry that it entails.

5.5 Unitary groups

We have already discussed the Lorentz group, which is particularly important for the understanding of special relativity; in this section we concentrate on the groups that are associated with *gauge theories*. We don't need to be concerned with what is meant by the designation gauge theories because we are here just discussing geometry; later on we will make a physical interpretation of the symmetries that geometry presents us with.

If U and V are two unitary multivectors their product is still unitary, since

$$(UV)(UV)^\dagger = UVV^\dagger U^\dagger = 1; \quad (5.40)$$

hence unitary multivectors form a group. Because the geometric algebra \mathcal{G}_{41} is isomorphic to the complex algebra of 4×4 matrices, this group is known as the *unitary group* $U(4)$.

If we consider only the multivectors with unit determinant we form a subgroup of $U(4)$, known as the *special unitary group* $SU(4)$. These elements form a group since

$$\det(UV) = \det(U)\det(V) = 1. \quad (5.41)$$

Unitary multivectors can be written in the general form

$$U = e^{iH}, \quad (5.42)$$

where H is an Hermitian multivector. Hermitian multivectors in the geometric algebra \mathcal{G}_{41} have 16 degrees of freedom and can be expanded in terms of the unitary Hermitian basis elements as

$$H = a + b^i \sigma_i + c^i \sigma_{0i} + d^{jk} i \sigma_{jk} + e i \sigma_0, \quad \text{with } j < k. \quad (5.43)$$

Inserting into Eq. (3.197), we see that the trace of the Hermitian multivector is responsible for a phase factor $\exp[i \operatorname{tr}(H)]$, which belongs to group $U(1)$. If we want elements of $SU(4)$ group, we have to use traceless Hermitian multivectors in the exponent, so, elements of group $SU(4)$ have 15 degrees of freedom. The Hermitian basis elements used for the expansion of Hermitian traceless matrices are called *generators* of group $SU(4)$. Any set of 15 independent unitary non-scalar Hermitian multivectors can be used as generators of $SU(4)$. In the literature it is usual to choose 15 special traceless unitary Hermitian matrices as generators, but this has obviously no geometric significance, since the transposition to the geometric algebra varies with the particular assignment that is made between basis elements and matrices.

Unitary multivectors take a particularly simple form when the Hermitian multivector in the exponent is such that $H^2 = \alpha^2$, with α scalar. When this happens we can write

$$U = e^{i\alpha h} = \cos(\alpha) + i h \sin(\alpha), \quad (5.44)$$

with h a unitary Hermitian multivector. In order to understand what is needed for this condition to be met, we choose two commuting Hermitian basis elements h_1 and h_2 . Then there is always a third basis element h_3 which commutes with the other two and verifies

$$h_3 = h_1 h_2, \quad (h_3)^2 = 1. \quad (5.45)$$

The demonstration that the three elements commute is straightforward. The commuting Hermitian basis elements can be arranged in groups of three but no more than that.[3] For instance, we can set $h_1 = \sigma_3$, $h_2 = \sigma_{04}$ resulting in $h_3 = -\sigma_{034}$; alternatively we could set $h_1 = -\sigma_{034}$, $h_2 = \sigma_{1234}$, resulting in $h_3 = \sigma_{012}$. We can even generalize the process and take any two commuting unitary Hermitian multivectors in the algebra and multiply them to obtain the third one. Then we ask how many different unitary

Hermitian multivectors can be formed by linear combinations of $\{h_1, h_2, h_3\}$ and the unit scalar; the answer is given by the solutions of the equation

$$(a^0 + a^1 h_1 + a^2 h_2 + a^3 h_3)^2 = 1. \quad (5.46)$$

This results in the following set of simultaneous scalar equations

$$\begin{aligned} (a^0)^2 + (a^1)^2 + (a^2)^2 + (a^3)^2 &= 1, \\ a^0 a^1 + a^2 a^3 &= 0, \\ a^0 a^2 + a^1 a^3 &= 0, \\ a^0 a^3 + a^1 a^2 &= 0. \end{aligned} \quad (5.47)$$

The first equation tells us that the solutions lie on the surface of a 3-sphere on a space spanned by $\{1, h_1, h_2, h_3\}$; with the other three equations the number of points on the 3-sphere becomes limited to 16, as follows:

- Singular solutions: $1, h_1, h_2, h_3,$
- o**: $(1 + h_1 - h_2 + h_3)/2,$
- m**: $(1 - h_1 + h_2 + h_3)/2,$
- i**: $(1 + h_1 + h_2 - h_3)/2,$
- a**: $(1 - h_1 - h_2 - h_3)/2,$

and their symmetrics. In Fig. 5.1 we represented graphically the 3-dimensional projection of the solutions with $a^0 = \pm 1/2$; the representation shows that these solutions lie on the vertices of a cube, each of the sets resulting from the sign choice for a^0 forming a tetrahedron. The designations for the solutions lying on the vertices are the initials of the characters in the play *The three sisters* by Anton Chekov: Olga, Masha, Irina and Andrey; the three female characters are related to the solutions with just one minus sign in their expression, while the male character is related to the solution with three minus signs.

The 16 solutions listed above don't make the complete set of unitary Hermitian multivectors. Suppose that we attribute $h_1 = \sigma_3$, $h_2 = \sigma_{04}$ and $h_3 = -\sigma_{034}$; we are then making a choice to build these unitary Hermitian elements with recourse to the orthogonal vectors σ_3 and σ_4 but any set of two orthogonal vectors selected from the Euclidean 4-space spanned by $\{\sigma_i\}$ would be equally effective. Furthermore, we can associate the trivector element h_3 with another trivector h_4 , such that $h_3 h_4 = \sigma_{1234}$ to get an alternative set of three commuting unitary Hermitian elements. This will generate another set of 16 solutions, not completely independent of the first one, since the solutions ± 1 and $\pm h_3$ are shared between the two sets.

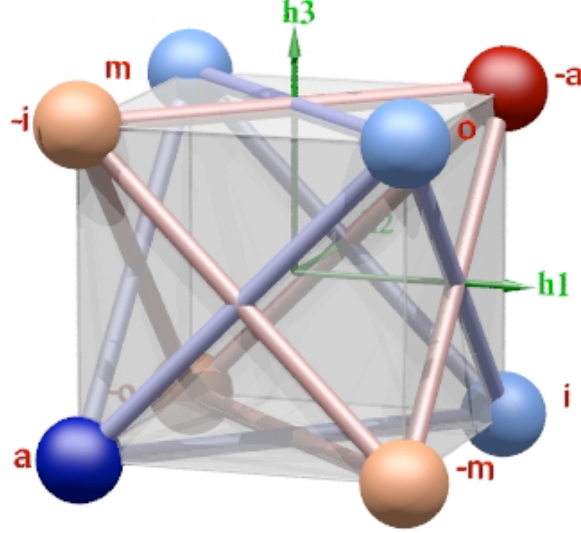


Figure 5.1: There is a total of 16 unitary Hermitian multivectors that can be built with linear combinations of three commuting unitary Hermitian multivectors $\{h_1, h_2, h_3\}$ and the unit scalar. The coefficients of the linear combinations for the solutions with the scalar coefficient equal to $\pm 1/2$ are represented graphically, highlighting their symmetry.

Recalling the discussion in Sec. 3.5 we note that if we take any unitary Hermitian multivector h from the algebra, two idempotents can be formed by

$$f = \frac{1}{2} (1 \pm h). \quad (5.48)$$

Conversely, if f is idempotent then $h = 2f - 1$ is unitary Hermitian; the proof is as follows

$$(2f - 1)^2 = 4f^2 - 4f + 1 = 1. \quad (5.49)$$

By taking two commuting unitary Hermitian multivectors, h_1 and h_2 , we can then form

four orthogonal idempotents by

$$\begin{aligned}
 f_1 &= \frac{1}{4} (1 + h_1)(1 + h_2) = \frac{1}{4} (1 + h_1 + h_2 + h_3), \\
 f_2 &= \frac{1}{4} (1 + h_1)(1 - h_2) = \frac{1}{4} (1 + h_1 - h_2 - h_3), \\
 f_3 &= \frac{1}{4} (1 - h_1)(1 - h_2) = \frac{1}{4} (1 - h_1 - h_2 + h_3), \\
 f_4 &= \frac{1}{4} (1 - h_1)(1 + h_2) = \frac{1}{4} (1 - h_1 + h_2 - h_3).
 \end{aligned} \tag{5.50}$$

These idempotents are mutually orthogonal because the product of any two of them returns zero. The four idempotents can be added among themselves, resulting in new idempotents, so that we have, with zero, a total of 16 idempotents, arranged in five categories:

- 1 null element,
- 4 single term elements called *primitive idempotents*,
- 6 double term elements,
- 4 triple term elements and
- 1 unit element.

Idempotents are also *projectors* under the transformation

$$M \rightarrow fMf. \tag{5.51}$$

Analysing the different classes of idempotents in the previous list, we verify that each of the four f_i primitive idempotents selects 1/8 of the original space, reducing the dimensionality from 5 to 2, the double term elements reduce the dimensionality to 3 and the triple term elements reduce the dimensionality to 4.

Example 5.1

Using the matrix representation of table 5.1 we let $h_1 = \sigma_3$ and $h_2 = \sigma_{04}$. The four primitive idempotents then have matrix representations consisting of a single non-zero element, one unit element on the diagonal. Each double term idempotent has two unit elements and each triple term idempotent has three diagonal unit elements. It follows that the projection induced by Eq. (5.51) selects only the matrix elements that lie on the same lines and columns of the non-zero diagonal elements.

Take for instance

$$f_2 + f_3 = \frac{1}{2} (1 - \sigma_{04}), \tag{5.52}$$

which has just the first and second elements of the diagonal non-zero. After the projection, an arbitrary multivector will be represented by a matrix of the form

$$\begin{pmatrix} \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

There are 8 degrees of freedom, corresponding to matrices with 4 complex elements, which shows that this is indeed a representation of a 3-dimensional space. The projection of arbitrary multivectors forms a group whose identity element is $f_2 + f_3$, the idempotent used for the projection.

The projection of the Hermitian basis vectors $\sigma_1, \sigma_2, \sigma_3$ produces the following representations

$$\begin{aligned} (f_2 + f_3)\sigma_1(f_2 + f_3) &= \frac{1}{2}\sigma_1(1 - \sigma_{04}) \equiv \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \\ (f_2 + f_3)\sigma_2(f_2 + f_3) &= \frac{1}{2}\sigma_2(1 - \sigma_{04}) \equiv \begin{pmatrix} 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \\ (f_2 + f_3)\sigma_3(f_2 + f_3) &= \frac{1}{2}\sigma_3(1 - \sigma_{04}) \equiv \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \end{aligned} \quad (5.53)$$

The remaining Hermitian basis elements either project to zero, the projector itself, or reproduce the previous ones. The top left 2×2 corner of the matrices reproduces Pauli matrices, which are the generators of $SU(2)$ group, the group of unitary matrices with dimension 2 and unit determinant. The generators of $SU(2)$ are actually given by

$$\lambda_m = (f_2 + f_3)\sigma_m(f_2 + f_3). \quad (5.54)$$

These are Hermitian and verify the commutation relation

$$\lambda_m \times \lambda_n = i\lambda_o, \quad (5.55)$$

where (m, n, o) is a cyclic permutation of $(1, 2, 3)$ and the symbol \times is the commutator product defined in Eq. (3.57).

Repeating the exercise with the idempotent $f_2 + f_3 + f_4$ we would find that only the 4th row/column of the matrices would be made zero, producing a representation for 4-dimensional space. There is one unexpected problem, however, because three of the diagonal elements produced are not traceless, namely

$$\begin{aligned}(f_1 + f_2 + f_3)\sigma_3(f_1 + f_2 + f_3) &= -\frac{1}{4}(1 - 3\sigma_3 + \sigma_{04} - \sigma_{034}), \\(f_1 + f_2 + f_3)\sigma_{04}(f_1 + f_2 + f_3) &= -\frac{1}{4}(1 + \sigma_3 - 3\sigma_{04} - \sigma_{034}), \\(f_1 + f_2 + f_3)\sigma_{034}(f_1 + f_2 + f_3) &= \frac{1}{4}(1 + \sigma_3 + \sigma_{04} + 3\sigma_{034}).\end{aligned}\tag{5.56}$$

This is a sign that the matrix algebra isomorphic to 4-dimensional spacetime is not the complex algebra of 3×3 matrices but rather the algebra of 2×2 matrices over the field of quaternions. In order to obtain a suitable set of generators for the special unitary group $SU(3)$ one needs to replace the diagonal elements obtained by projection with traceless linear combinations of them.

In general terms, the projection of Hermitian basis elements induced by double term idempotents produces generators of $SU(2)$, the diagonal element being obtained by the projection of either of the unitary Hermitian terms $\{h_1, h_2, h_3\}$ that go into the idempotents. Triple term idempotents project Hermitian basis elements into $SU(3)$ generators, under the condition that the diagonal elements be combined to produce traceless ones.

We can bring together everything we said above about unitary groups to try and find generators for the mysterious $U(1) \times SU(2) \times SU(3)$ group characteristic of the standard model of particle physics. It has frequently been argued that, since this group has 12 degrees of freedom, a space with 11 dimensions over the field of reals is needed to accommodate it. This is indeed one of the arguments in favour of string theories; we will show next that such reasoning is incorrect. Because 5-dimensional geometric algebras have 32 dimensions, there are sufficient degrees of freedom in the algebra to generate this group. Counting degrees of freedom is not a sufficient argument in favour of our proposition, as such we will now propose a set of generators for the standard model gauge group.

Elements of this group can be represented by ordered sets such as $(e^{ia}, e^{ib^m\lambda_m}, e^{ic^i\varepsilon_i})$, where a, b^m and c^i are real coefficients, λ_m are $SU(2)$ generators, ε_i are $SU(3)$ generators and the indices take values between 1 and 3 and between 1 and 8, for m and i , respectively. We start with the definition of $SU(2)$ generators by

$$\lambda_m = \frac{1}{2}(\sigma_m + \sigma_{0m4}).\tag{5.57}$$

Next we define $SU(3)$ generators by the following list

$$\begin{aligned}
 \varepsilon_1 &= \frac{1}{2}(\sigma_2 + \sigma_{01}), \\
 \varepsilon_2 &= \frac{1}{2}(-\sigma_1 + \sigma_{02}), \\
 \varepsilon_3 &= \frac{1}{2}(-\sigma_{034} + \sigma_{1234}), \\
 \varepsilon_4 &= \frac{1}{2}(-\sigma_{014} + \sigma_{023}), \\
 \varepsilon_5 &= \frac{1}{2}(-\sigma_{013} - \sigma_{024}), \\
 \varepsilon_6 &= \frac{1}{2}(-\sigma_4 - \sigma_{03}), \\
 \varepsilon_7 &= \frac{1}{2}(-\sigma_3 + \sigma_{04}), \\
 \varepsilon_8 &= \frac{1}{2\sqrt{3}}(2\sigma_{012} - \sigma_{034} - \sigma_{1234}).
 \end{aligned} \tag{5.58}$$

These verify the commutation relations

$$\varepsilon_i \times \varepsilon_j = i\delta_{ij}^k \varepsilon_k, \tag{5.59}$$

where the δ_{ij}^k are the structure constants. The δ_{ij}^k are odd in the interchange of any pair of indices, and the non-vanishing ones are given by the permutations of $\delta_{12}^3 = 1$, $\delta_{14}^7 = \delta_{24}^6 = \delta_{25}^7 = \delta_{34}^5 = \delta_{51}^6 = \delta_{63}^7 = 1/2$, $\delta_{45}^8 = \delta_{67}^8 = \sqrt{3}/2$.

One alternative way to verify the correctness of the generators above we can find a matrix assignment for the basis vectors, such that Gell-Mann matrices are reproduced in the upper-left 3×3 corner; this is shown in Table 5.2. We must ascertain that groups

Table 5.2: Alternative matrix representation of basis vectors

$$\begin{aligned}
 \sigma_0 &\equiv \begin{pmatrix} -i & 0 & 0 & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & -i & 0 \\ 0 & 0 & 0 & i \end{pmatrix}, & \sigma_1 &\equiv \begin{pmatrix} 0 & i & 0 & 0 \\ -i & 0 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \end{pmatrix}, & \sigma_2 &\equiv \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{pmatrix}, \\
 \sigma_3 &\equiv \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix}, & \sigma_4 &\equiv \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.
 \end{aligned}$$

$SU(2)$ and $SU(3)$ generated by $\{\lambda_m\}$ and $\{\varepsilon_i\}$, respectively, don't overlap. This is easily done, because all linear combinations of $\{\lambda_m\}$ are made of vectors in σ_{123} space and trivectors obtained from the former by a product with σ_{40} . In order to obtain linear combinations of vectors in σ_{123} from the $SU(3)$ generator list, we would choose $\{\varepsilon_1, \varepsilon_2, \varepsilon_6\}$ but then the trivector component would not be reproduced. We are then confident that two symmetry groups are distinct, and we are allowed to pair them in order to produce the product of the two groups.

It is very elucidating to relate the three sisters, Olga, Marsha and Irina, from Fig. 5.1, to the generators of unitary groups; this exercise will become useful in the discussion of elementary particles of next chapter. In order to understand the relation to $SU(2)$ we must take only two sisters; accordingly we set $h_1 = \sigma_{034}$, $h_2 = \sigma_{04}$, $h_3 = -\sigma_3$ and select the two younger sisters, Masha and Irina, which share h_2 , and note, from Eq. (5.57) that their respective unitary Hermitian multivectors can be written as

$$\begin{aligned} h_o &= \frac{1}{2}(1 - \lambda_1 + \sigma_{034}), \\ h_i &= \frac{1}{2}(1 + \lambda_1 + \sigma_{034}). \end{aligned} \quad (5.60)$$

The two elements don't belong to an $SU(2)$ group, because they are not traceless, however they can be transformed into each other by $SU(2)$ generators with the double-sided transformation

$$h' = (\lambda_2 \pm i\lambda_3)h(\lambda_2 \mp i\lambda_3). \quad (5.61)$$

So, we say that this pair of sisters exhibits the same symmetry as the elements of an $SU(2)$ group. Naturally we could have selected another pair of sisters and draw a similar conclusion.

Turning our attention now to group $SU(3)$, we set $h_1 = \sigma_{034}$, $h_2 = \sigma_{012}$, $h_3 = -\sigma_{1234}$. Obviously there is still an $SU(2)$ symmetry between any pair of sisters, but we are interested in the relationship between the three sisters and group $SU(3)$. The three Hermitian multivectors can then be re-written as

$$\begin{aligned} h_o &= \frac{1}{2} - \frac{2}{\sqrt{3}}\varepsilon_8 + \frac{1}{6}(\sigma_{012} + \sigma_{034} + \sigma_{1234}), \\ h_m &= \frac{1}{2} + \frac{1}{\sqrt{3}}\varepsilon_8 + \varepsilon_3 + \frac{1}{6}(\sigma_{012} + \sigma_{034} + \sigma_{1234}), \\ h_i &= \frac{1}{2} + \frac{1}{\sqrt{3}}\varepsilon_8 - \varepsilon_3 + \frac{1}{6}(\sigma_{012} + \sigma_{034} + \sigma_{1234}). \end{aligned} \quad (5.62)$$

6 Flat space physics

At this point we have a sufficient mathematical background in order to derive a number of equations with physical relevance; in doing so we will understand that the different areas of physics, most notably special relativity and quantum mechanics, are indeed united under the umbrella of geometry. The last sections in the previous chapter have already pointed the way ahead but the subject deserves a full chapter and the present one is entirely devoted to physics.

6.1 Postulates of special relativity

Special relativity is usually introduced in textbooks with a set of postulates from which one can derive the laws of dynamics, excluding gravitation, which is the realm of general relativity. This procedure is perfectly legitimate and it has produced excellent texts, but we will here diverge from the standard approach, mainly because we wish to find a simple set of geometric principles that allows the derivation of all the foundational equations of physics.

Our point of departure is the algebra of 5-dimensional spacetime, where we pay special attention to monogenic functions, that is functions verifying the relation

$$\nabla\psi = 0. \quad (6.1)$$

Since monogenic functions have null Laplacian, we can try solutions for the second order equation

$$\nabla^2\psi = 0. \quad (6.2)$$

Expanding the Laplacian we see that

$$\sum_{i=1}^4 \partial_{ii}\psi = \partial_{00}\psi, \quad (6.3)$$

for which we immediately propose a solution of the type

$$\psi = \psi_0 e^{ip_\alpha x^\alpha} = \psi_0 e^{ipx}, \quad (6.4)$$

where $p = p_\alpha \sigma^\alpha$ and $x = \sigma_\beta x^\beta$. Notice that we could also have proposed different solutions of the general form

$$\psi = \psi_0 e^{ihp_\alpha x^\alpha} = \psi_0 e^{ihpx}, \quad (6.5)$$

with h a unitary Hermitian element, that is, an element such that $h^2 = 1$. When we are concerned only with trajectories and not with the quantum nature of particles, choice between the different alternative solutions is irrelevant and it is easier to just set h to unity.

For Eq. (6.3) to be verified p must be a nilpotent vector, that is

$$p^2 = p \cdot p = 0. \quad (6.6)$$

Since the Laplacian is a scalar operator, the second order operation imposes no condition on ψ_0 , however, if we insert solution (6.4) into Eq. (6.1) the following must be true

$$p\psi_0 = 0. \quad (6.7)$$

So, ψ_0 must include the factor p and thus be nilpotent as well.

The function ψ has all the characteristics of a plane wave but it is called an hyperplane wave because it exists in higher-dimensional space. Imposing the condition that ψ remains constant is equivalent to the same condition being imposed on the exponent; consequently

$$p \cdot x = \text{constant}. \quad (6.8)$$

Differentiating both sides of the previous equation, this results in

$$p \cdot dx = 0. \quad (6.9)$$

In this equation we have the inner product between the nilpotent vector p and the displacement vector dx ; the latter can be expanded as

$$dx = \sigma_\alpha dx^\alpha. \quad (6.10)$$

For Eq. (6.9) to be verified there are two possibilities: either vectors p and dx are perpendicular to each other or they are collinear, consequently they are both nilpotent.

Considering displacements dx perpendicular to p we describe a 3-dimensional wavefront that can be called an hyperplane. Suppose, for instance, $p = \sigma^0 + \sigma^4$. All displacements in the directions $\sigma^1, \sigma^2, \sigma^3$ or any arbitrary combination of those will be perpendicular to the p vector, so the wavefront coincides with the 3-dimensional space spanned by σ_{123} . The same would be true for any orientation of the p vector. On the other hand dx can be nilpotent; this is written as

$$(dx)^2 = -(dx^0)^2 + \sum_{i=1}^4 (dx^i)^2 = 0. \quad (6.11)$$

It is now a question of simple manipulation to obtain Eqs. (5.25) and (5.26). A physical interpretation of those equations becomes possible if we assign coordinate x^0 to time and

coordinates x^m to Cartesian coordinates in 3-dimensional space; we will therefore replace x^0 by t , while keeping the indices of coordinates x^m . For coordinate x^4 we will use the Greek letter τ so that it stands out from the others. The equations mentioned above then become:

$$(dt)^2 = \sum_{m=1}^3 (dx^m)^2 + (d\tau)^2; \quad (6.12)$$

$$(d\tau)^2 = (dt)^2 - \sum_{m=1}^3 (dx^m)^2. \quad (6.13)$$

The two equations above carry exactly the same information, although their physical interpretation may be quite different. They both apply to the null subspace of 5-dimensional spacetime, but whereas in Eq. (6.12) it is suggested that t is used as parameter for displacements, in Eq. (6.13) the suggestion is that τ should take that role. The reason why this happens is that we cannot use displacement length for parameter, because displacements have null length; we are forced to choose a different parameter and all coordinates stand as good candidates. There is a good geometrical reason for not opting for any of the three x^m coordinates but we must leave it for later on.

If we take Eq. (6.12) and divide both sides by $(dt)^2$, we obtain

$$1 = \mathbf{v}^2 + \dot{\tau}^2, \quad (6.14)$$

where we suppressed the parenthesis and $\mathbf{v}^2 = \sum (\dot{x}^m)^2$ is the squared modulus of 3-dimensional vector \mathbf{v} that we associate with physical velocity. The physical velocity has a modulus that can never be greater than unity, because this would imply $\dot{\tau}^2 < 0$, which is not allowed. This is, of course, if we express t in terms of length but expressing t in terms of time units we would find physical velocity limited by the speed of light in vacuum. Equation (6.14) suggests the consideration of a unit length 4-component vector $v^m \sigma_m + \dot{\tau} \sigma_4$; this vector can be interpreted as a particle's 4-dimensional velocity, which has always the same length but makes different angles with 3-dimensional space. In common terms one could say that every particle travels at the speed of light in 4-dimensional space and the conventional velocity is the immediately detectable 3D component of the particle's velocity. The term $\dot{\tau}^2$ on the right hand side of Eq. (6.14) is related to the Lorentz factor γ from special relativity by $\gamma^2 = 1/\dot{\tau}^2$ and we will understand why it must never be negative further on.

Going over to Eq. (6.13) and dividing both sides by $(d\tau)^2$ we get

$$1 = \left(\frac{dt}{d\tau} \right)^2 - \sum_{m=1}^3 \left(\frac{dx^m}{d\tau} \right)^2. \quad (6.15)$$

The right hand side represents the modulus of a bivector known in special relativity as 4-velocity and treated as a vector in Minkowski spacetime. To understand how this

happens, we define the 4-velocity bivector

$$\nu = \frac{dt}{d\tau} \sigma_{40} + \frac{dx^m}{d\tau} \sigma_{4m}. \quad (6.16)$$

Using the definitions from Eq. (5.16) we can also write

$$\nu = \frac{dx^\mu}{dx^4} \gamma_\mu. \quad (6.17)$$

The mapping made between $\sigma_{4\mu}$ bivectors of 5-dimensional spacetime and γ_μ vectors of Minkowski spacetime allows us to understand concepts of special relativity in our 5-dimensional spacetime setting.

A physical interpretation of vector p can be sought with recourse to Eq. (6.9). Expanding this equation and dividing by dt we get

$$p_0 + p_m \dot{x}^m + p_4 \dot{\tau} = 0. \quad (6.18)$$

This tells us that p must be collinear with a vector $\sigma^0 + \sum_{m=1}^3 \sigma^m \dot{x}^m + \sigma^4 \dot{\tau}$ but it poses no constraint on a multiplying factor, since this will not alter the fact that p must remain a nilpotent. The best choice for this multiplying factor arises from wave-particle duality considerations, which we will develop in the next section. We can anticipate that this multiplying factor is $m/\dot{\tau}$ and the p vector can then be expanded as

$$\begin{aligned} p &= \frac{m}{\dot{\tau}} (\sigma^0 + \sum_{m=1}^3 \sigma^m \dot{x}^m + \sigma^4 \dot{\tau}) \\ &= \sigma_0 E + \mathbf{p} + \sigma^4 m, \end{aligned} \quad (6.19)$$

with E the total energy, \mathbf{p} the 3-dimensional momentum and m the mass. Because p is nilpotent, this option leads immediately to the relativistic relation

$$E^2 = \mathbf{p}^2 + m^2. \quad (6.20)$$

Massless particles, such as photons, have $m = 0$ which implies also null $\dot{\tau}$. This is consistent with the interpretation of proper time remaining constant for photons.

The first principle of special relativity is that of invariance under a Lorentz transformation. As we have seen in Sec. 5.4 the full Lorentz group in 5-dimensional spacetime is larger than the one considered in special relativity, the extension being driven by bivectors containing σ_4 . We will designate by *special Lorentz group* the group of Lorentz transformations restricted to those allowed by special relativity, generated by 4D bivectors $\sigma_{\mu\nu}$. A special Lorentz transformation may comprise a boost and a pure rotation, generated, respectively, by the time and space components of its bivector generator, as defined in Eq. (5.39). A Lorentz transformation is applied to the frame vectors but that must not affect the displacement dx , so its expansion in terms of components gets altered.

Example 6.1

Suppose that we have two vectors $a = \sigma_0 a^0 + \sigma_3 a^3$ and $b = \sigma^0 b_0 + \sigma^3 b_3$, which are to remain invariant under a transformation generated by bivector $B = \theta/2\sigma_{03}$; the vectors obviously remain the same if we multiply on the left and on the right by unity, so we get for vector a

$$\begin{aligned} a &= \sigma_0 a^0 + \sigma_3 a^3 \\ &= e^B e^{-B} (\sigma_0 a^0 + \sigma_3 a^3) e^B e^{-B}. \end{aligned} \quad (6.21)$$

Now we note that

$$e^{-B} \sigma_0 = \sigma_0 e^B, \quad e^{-B} \sigma_3 = \sigma_3 e^B. \quad (6.22)$$

Inserting above we get

$$\begin{aligned} a &= e^B \sigma_0 e^B a^0 e^B e^{-B} + e^B \sigma_3 e^B a^3 e^B e^{-B} \\ &= e^B \sigma_0 e^{-B} e^B a^0 e^B + e^B \sigma_3 e^{-B} e^B a^3 e^B \\ &= (\sigma_0 \cosh \theta + \sigma_3 \sinh \theta) (\cosh \theta + \sigma_{03} \sinh \theta) a^0 \\ &\quad + (\sigma_3 \cosh \theta + \sigma_0 \sinh \theta) (\cosh \theta + \sigma_{03} \sinh \theta) a^3 \\ &= (\sigma_0 \cosh \theta + \sigma_3 \sinh \theta) (\cosh \theta a^0 - \sinh \theta a^3) \\ &\quad + (\sigma_3 \cosh \theta + \sigma_0 \sinh \theta) (\cosh \theta a^3 - \sinh \theta a^0). \end{aligned} \quad (6.23)$$

We now define the transformed basis vectors

$$\begin{aligned} \sigma'_0 &= \sigma_0 \cosh \theta + \sigma_3 \sinh \theta; \\ \sigma'_3 &= \sigma_3 \cosh \theta + \sigma_0 \sinh \theta; \end{aligned} \quad (6.24)$$

and the transformed expansion coefficients

$$\begin{aligned} a'^0 &= \cosh \theta a^0 - \sinh \theta a^3; \\ a'^3 &= \cosh \theta a^3 - \sinh \theta a^0; \end{aligned} \quad (6.25)$$

in order to write the expansion of vector a as

$$a = \sigma'_0 a'^0 + \sigma'_3 a'^3. \quad (6.26)$$

For the transformation of vector b we impose that each term in $a \cdot b = a^0 b_0 + a^3 b_3$ must remain invariant, so the coordinates of b must incur a transformation which opposes that of a 's coordinates. It is then

$$\begin{aligned} b'_0 &= \cosh \theta b_0 + \sinh \theta b_3; \\ b'_3 &= \cosh \theta b_3 + \sinh \theta b_0. \end{aligned} \quad (6.27)$$

The reciprocal frame vectors must also be transformed opposite to the direct frame ones:

$$\begin{aligned}\sigma'^0 &= \sigma^0 \cosh \theta - \sigma^3 \sinh \theta; \\ \sigma'^3 &= \sigma^3 \cosh \theta - \sigma^0 \sinh \theta.\end{aligned}\tag{6.28}$$

Vectors such as σ_α , represented with lower indices, are said to be *covariant* and are called *covectors*; on the other hand, vectors such as σ^α , represented with upper indices, are said to be *contravariant* and are called *contravectors*.

With this example we show that what is transformed is not vector a or b but only their expansion in terms of the transformed frames.

We will now verify the consequences of applying the Lorentz invariance principle to the displacement vector dx given by Eq. (6.10), that is, we will impose that the displacement remains unchanged when the basis vectors on the right hand side undergo a special Lorentz transformation. Accordingly we write

$$\sigma_\alpha dx^\alpha = e^{B_M} \sigma_\alpha e^{-B_M} dx'^\alpha, \tag{6.29}$$

where x'^α designates the transformed coordinates. It becomes immediately obvious that basis vector σ_4 remains unchanged and so must the associated coordinate x^4 . If this is carried to Eq. (6.13) we get the invariance of the relativistic interval under a special Lorentz transformation. In special relativity there are only four coordinates and one uses Eq. (6.13) to define the invariant interval; in our approach coordinate x^4 remains unchanged under special Lorentz transformations, providing compatibility between the two approaches, as long as only special Lorentz transformations are allowed. This restriction is rather unnatural from a geometric point of view; although we cannot offer a plausible explanation for it, we believe it must carry some deep physical meaning. We turn our attention now to the invariance vector $p = p_\alpha \sigma^\alpha$. Naturally things are very similar to vector dx and we write

$$p_\alpha \sigma^\alpha = p'_\alpha e^{-B_M} \sigma^\alpha e^{B_M}. \tag{6.30}$$

The second postulate of special relativity, known as the constancy of speed of light in vacuum is a direct consequence of what we said before. Since, for photons, we have $dx = \sigma_0 dt + \sigma_m dx^m$ and the displacement must remain null, we conclude that

$$(dt)^2 = \sum_{m=1}^3 (dx^m)^2. \tag{6.31}$$

Dividing both sides by $(dt)^2$ we get the expected result

$$1 = v^2. \tag{6.32}$$

This result is invariant under a special Lorentz transformation, as expected.

Example 6.2

Let us examine the case of a particle with mass m , moving along σ_3 with velocity v . We want to replace the frame $\{\sigma_i\}$ with a frame where the particle is at rest. The p vector is

$$p = m \left(-\frac{1}{\sqrt{1-v^2}} \sigma^0 + \frac{v}{\sqrt{1-v^2}} \sigma^3 + \sigma^4 \right). \quad (6.33)$$

In the new frame we want to have

$$p = -m\sigma'^0 + m\sigma^4. \quad (6.34)$$

Equating the right hand sides of the two equations we see that

$$\sigma'^0 = \frac{1}{\sqrt{1-v^2}} (\sigma^0 - v\sigma^3). \quad (6.35)$$

Applying the corresponding transformation to the components of p we get

$$\begin{aligned} p &= \left(\frac{1}{\sqrt{1-v^2}} p_0 + \frac{v}{\sqrt{1-v^2}} p_3 \right) \sigma'^0 \\ &\quad + \left(\frac{1}{\sqrt{1-v^2}} p_3 + \frac{v}{\sqrt{1-v^2}} p_0 \right) \sigma'^3 + p_4 \sigma^4 \\ &= m \left[\left(\frac{v^2 - 1}{1 - v^2} \right) \sigma'^0 + \left(\frac{v - v}{1 - v^2} \right) \sigma'^3 + \sigma^4 \right]; \end{aligned} \quad (6.36)$$

and we recover Eq. (6.34).

The analysis can alternatively be made with the displacement vector; this is

$$dx = \sigma_0 dt + \sigma_3 dx^3 + \sigma_4 dx^4. \quad (6.37)$$

Replacing dx^3 with $v dt$ it becomes

$$dx = (\sigma_0 + v\sigma_3) dt + \sigma_4 dx^4. \quad (6.38)$$

In the particle's frame there must be no displacement along σ'_3 , so the displacement vector is

$$dx = \sigma'_0 dt' + \sigma_4 dx^4. \quad (6.39)$$

Equating the right hand sides of the two last equations we conclude that

$$\sigma'_0 = \frac{1}{\sqrt{1-v^2}} (\sigma_0 + v\sigma_3); \quad (6.40)$$

and accordingly

$$\sigma'_3 = \frac{1}{\sqrt{1-v^2}} (\sigma_3 + v\sigma_0). \quad (6.41)$$

The coordinates become transformed as

$$dt' = \frac{1}{\sqrt{1-v^2}} (dt - v dx^3) = \sqrt{1-v^2} dt; \quad (6.42)$$

$$dx'^3 = \frac{1}{\sqrt{1-v^2}} (dx^3 - v dt) = 0. \quad (6.43)$$

The transformation is the same whichever route we choose.

In general, if we have the representation of a vector in a given frame and wish to represent the same vector in a frame moving with velocity \mathbf{v} with respect to the original one, we form the bivector

$$B = \frac{\text{atanh}|\mathbf{v}|}{2|\mathbf{v}|} \sigma_0 \mathbf{v} \quad (6.44)$$

and use this as generator for the Lorentz transformation. With this general procedure we can work out the addition law for velocities, as we do in the next example

Example 6.3

Consider a moving particle with 3-dimensional velocity $\mathbf{u} = \sigma_1 u^1 + \sigma_2 u^2 + \sigma_3 u^3$ in the laboratory frame and find its velocity components in a frame moving with velocity v along σ^3 .

We look at the transformation for the p vector components; the p vector is

$$p = \frac{m}{\sqrt{(\mathbf{u})^2}} \left[-\sigma^0 + \sum_{m=1}^3 \sigma^m u^m + \sqrt{(\mathbf{u})^2} \sigma^4 \right]. \quad (6.45)$$

The transformed components are:

$$\begin{aligned} p'_0 &= \frac{1}{\sqrt{1-v^2}} p_0 + \frac{v}{\sqrt{1-v^2}} p_3 = \frac{1 - u^3 v}{\sqrt{[1-v^2][1-(\mathbf{u})^2]}}, \\ p'_1 &= p_1 = \frac{u^1}{\sqrt{1-\mathbf{u}^2}}, \\ p'_2 &= p_2 = \frac{u^2}{\sqrt{1-\mathbf{u}^2}}, \\ p'_3 &= \frac{1}{\sqrt{1-v^2}} p_3 + \frac{v}{\sqrt{1-v^2}} p_0 = \frac{u^3 - v}{\sqrt{[1-v^2][1-(\mathbf{u})^2]}}, \\ p'_4 &= p_4 = m. \end{aligned} \quad (6.46)$$

Only the momentum along σ^3 and total energy are transformed, however, all components of 3D velocity get transformed in the process. In order to get the velocity components we

need only to divide the 3 p_m components by the energy p_0 ; the result is

$$\begin{aligned} u'^1 &= \frac{u^1 \sqrt{1-v^2}}{1-u^3 v}, \\ u'^2 &= \frac{u^2 \sqrt{1-v^2}}{1-u^3 v}, \\ u'^3 &= \frac{u^3 - v}{1-u^3 v}. \end{aligned} \quad (6.47)$$

One would probably expect the velocity components normal to the frame velocity to stay unaltered, but since the time coordinate changes it produces a corresponding change in the velocity, even in the plane normal to the frame displacement where the momentum does not change.

6.2 Wave-particle duality

It is now time to return to the monogenic condition upon which we based all our deductions for the kinematics of particles. The solutions for the monogenic condition are hyperplane waves given by Eq. (6.4), which we can now write in a slightly different form:

$$\psi = \psi_0 e^{i p x} = \psi_0 e^{i(p_m x^m + m \tau \pm E t)}. \quad (6.48)$$

The \pm sign was inserted in the exponent to leave it clear that both positive and negative values for energy are allowed; if E is taken as a positive quantity, the minus sign applies to a wave propagating forward in time and the plus sign to a backward propagating wave. A particle is then associated with a 4-dimensional wave propagating at unit velocity with wave vector $k = p_m \sigma^m + m \sigma^4$ and angular frequency E . The particle's 4-dimensional wavelength is naturally

$$\lambda = \frac{1}{\sqrt{\mathbf{p}^2 + m^2}} = \frac{1}{E}. \quad (6.49)$$

This wavelength is associated to two other wavelengths: the first one is known as *Compton wavelength* and is just the inverse of mass in non-dimensional units; the second one is known as *de Broglie wavelength*, given by

$$\lambda_B = \frac{1}{\sqrt{\mathbf{p}^2}}. \quad (6.50)$$

The interaction between two particles involves the consideration of the p vectors for both particles, before and after the interaction. If we designate the p vectors for the two particles by $(p)_1$ and $(p)_2$, respectively, a simple elastic collision can be modelled by $(p)_1 + (p)_2 = (p')_1 + (p')_2$; expanding in terms of components we get

$$\begin{aligned} &[(E)_1 + (E)_2] \sigma^0 + (\mathbf{p})_1 + (\mathbf{p})_2 + [(m)_1 + (m)_2] \sigma^4 \\ &= [(E')_1 + (E')_2] \sigma^0 + (\mathbf{p}')_1 + (\mathbf{p}')_2 + [(m)_1 + (m)_2] \sigma^4. \end{aligned} \quad (6.51)$$

From this we derive separate conservation equations along σ^0 and σ^m , known respectively as energy and momentum conservation equations..

Example 6.4

The Compton effect, for instance, can also be modelled in a similar fashion, but now we have the interaction of a photon with a stationary particle; the conservation equation that must be written is now

$$\begin{aligned} (E_p + E_e)\sigma^0 + \mathbf{p}_p + m_e\sigma^4 \\ = (E'_p + E'_e)\sigma^0 + \mathbf{p}'_p + \mathbf{p}'_e + m_e\sigma^4. \end{aligned} \quad (6.52)$$

This splits into energy and momentum conservation equations as

$$E_p + E_e = E'_p + E'_e, \quad (6.53)$$

$$\mathbf{p}_p = \mathbf{p}'_p + \mathbf{p}'_e. \quad (6.54)$$

The energy equation can be rewritten as

$$E_p + m_e = E'_p + \sqrt{(\mathbf{p}'_e)^2 + (m_e)^2}. \quad (6.55)$$

This we solve for \mathbf{p}'_e to get

$$(\mathbf{p}'_e)^2 = (E_p + m_e - E'_p)^2 - (m_e)^2. \quad (6.56)$$

From the momentum conservation equation we take

$$\begin{aligned} (\mathbf{p}'_e)^2 &= (\mathbf{p}_p - \mathbf{p}'_p)^2 \cdot (\mathbf{p}_p - \mathbf{p}'_p) \\ &= (\mathbf{p}_p)^2 + (\mathbf{p}'_p)^2 - 2\mathbf{p}_p \cdot \mathbf{p}'_p \\ &= (\mathbf{p}_p)^2 + (\mathbf{p}'_p)^2 - 2|\mathbf{p}_p||\mathbf{p}'_p|\cos\theta \end{aligned} \quad (6.57)$$

Since $E_p = |\mathbf{p}_p|$ and $E'_p = |\mathbf{p}'_p|$ we can also write

$$(\mathbf{p}'_e)^2 = (E_p)^2 + (E'_p)^2 - 2E_pE'_p\cos\theta. \quad (6.58)$$

Now we combine Eqs. (6.56) and (6.58) to get

$$\begin{aligned} (E_p)^2 + (E'_p)^2 - 2E_pE'_p\cos\theta &= (E_p)^2 + (m_e)^2 + (E'_p)^2 \\ &\quad - 2E_pE'_p + 2(E_p - E'_p)m_e - (m_e)^2. \end{aligned} \quad (6.59)$$

After cancelling out the common terms this becomes

$$-2E_pE'_p\cos\theta = -2E_pE'_p + 2(E_p - E'_p)m_e. \quad (6.60)$$

A simple manipulation produces

$$\frac{1}{E'_p} - \frac{1}{E_p} = \frac{1 - \cos \theta}{m_e} \quad (6.61)$$

or in terms of wavelength

$$\lambda' - \lambda = \frac{1 - \cos \theta}{m_e}. \quad (6.62)$$

Before leaving this section we must say a word about the uncertainty principle which states that the distributions of momentum and position for a given particle are inversely proportional. A perfectly localized particle must have a distribution of momentum from $-\infty$ to $+\infty$ and conversely a perfectly stationary particle cannot be localized. If we look at the solution for monogenic functions given by Eq. (6.4), it tells us precisely the particle's momentum, which we extract from $p = E\sigma^0 + \mathbf{p} + m\sigma^4$. For a stationary particle the momentum is precisely zero. This solution, however, does not tell us anything about the particle's localization. In order to have localization we must have a superposition of solutions in much the same way as in optics we need a superposition of plane waves in order to get a beam of light. Fourier optics tells us that the spread in the angles of the individual plane waves is inversely proportional to the beam width [5, 6] and the same happens here with 4 dimensions instead of 3. But the p component along σ^4 is always numerically equivalent to the mass, so we are left with the 3 momentum components and energy to build the wavepacket that localizes the particle in position and time.

The localization of a particle is introduced in Eq. (6.4) by letting ψ be a general function of the four x^μ coordinates and preserving the harmonic dependence on τ provided by mass; this can be achieved with the superposition of hyperplane waves with a distribution of p -vectors, in a process similar to the production of a beam of light by superposition of plane waves. This type of analysis involves the evaluation of the Fourier transform for ψ , which we write as

$$\Psi(p_\mu, \tau) = \frac{1}{(2\pi)^2} \iiint_{-\infty}^{+\infty} \psi(x^\mu, \tau) e^{i(p_\mu x^\mu)} dx^\mu. \quad (6.63)$$

This verifies all the theorems of Fourier transforms, namely the *similarity theorem*, which states that if $\Psi(p_\mu, \tau)$ is the Fourier transform of $\psi(x^\mu, \tau)$, then the Fourier transform of $\psi(a_\mu x^\mu, \tau)$, with a_μ constants, is given by

$$\mathcal{F}\{\psi(a_\mu x^\mu, \tau)\} = \frac{1}{|\prod a_\mu|} \Psi\left(\frac{p_\mu}{a_\mu}, \tau\right), \quad (6.64)$$

that is, a stretch of the coordinates x^μ results in a compression in the momentum/energy domain, plus a change in the overall amplitude of momentum/energy. The association of a

particle with a wave packet is not entirely satisfactory because a wave packet necessarily diverges; this would imply that a particle would be less localized with the passage of time. We don't have an explanation for the fact that a particle's wave packet seems not to diverge, nevertheless we may invoke the Fourier analysis to gain some understanding of the uncertainty principle.

6.3 First order form of the wave equation

The monogenic condition (6.1) produces a first order differential equation, which we turned into a second order equation (6.2) because the latter can be solved as if it were scalar. Although this procedure is acceptable for modelling kynematics it ignores some features of the solutions which become important when we consider particles' properties such as spin. The strategy now will be to pick up the solution for the second order equation (6.4) and check what further restrictions the monogenic condition imposes on it; inserting this solution in Eq. (6.1) we get

$$p\psi_0 e^{ipx} = 0. \quad (6.65)$$

Since we know already that p must be a nilpotent vector in order to verify the second order equation, we are led to conclude that ψ_0 must itself include the factor p ; this is the only way the previous equation can be verified. For the moment we can let $\psi_0 = p$, since we have no obvious reason to include more factors.

Suppose now that we take three commuting Hermitian elements of the algebra $\{h_1, h_2, h_3\}$ and define a set of primitive idempotents as in Eq. (5.50). Since the four primitive idempotents add to unity we can take a monogenic function $\psi = p \exp(ip \cdot x)$ and produce a four-fold split into left ideals by right multiplication of ψ_0 with the sum of the four idempotents. We then write

$$\psi_i = \psi_0 f_i e^{ipx}, \quad (6.66)$$

such that

$$\psi = \sum_{i=1}^4 \psi_i. \quad (6.67)$$

If ψ is monogenic, so must be all the ψ_i , because the product by idempotents is made on the right of ψ_0 while the vector derivative acts on its left.

The product by an idempotent produces a curious effect on the wavefunction. Start by noticing that if f is idempotent, then $h = 2f - 1$ is unitary Hermitian, as demonstrated in Eq. (5.49). Now we multiply an exponential with pseudoscalar exponent by

an idempotent to get

$$\begin{aligned}
 f e^{i\alpha} &= f(\cos \alpha + i \sin \alpha) \\
 &= f[\cos \alpha + i(2f - 1) \sin \alpha]f \\
 &= f(\cos \alpha + i h \sin \alpha) \\
 &= f e^{ih\alpha}.
 \end{aligned} \tag{6.68}$$

One case of special interest for us happens when $f = (1 \pm \sigma_{034})/2$, for which we have $h = \pm \sigma_{034}$; applying to the previous equation we get

$$\frac{1 \pm \sigma_{034}}{2} e^{i\alpha} = \frac{1 \pm \sigma_{034}}{2} e^{\pm \sigma_{12}\alpha}. \tag{6.69}$$

So we notice that the choice of idempotent originates an operator that produces rotations on the σ_{12} plane, whose direction depends on the sign within the idempotent.

Let us return to Eq. (6.65) with the knowledge that we have gained with the discussion above. For the equation to be verified in all circumstances we must have $p\psi_0 = 0$; we expand p and write

$$(-\sigma^0 E + p_m \sigma^m + m \sigma^4) \psi_0 = 0. \tag{6.70}$$

It is useful to rewrite the equation as an eigenvalue equation, which we can easily do by left multiplication by σ^0 and rearrangement of the terms; it is then

$$(p_m \sigma^{m0} + m \sigma^{40}) \psi_0 = E \psi_0. \tag{6.71}$$

We could certainly resort to the matrix formulation in order to solve the equation, but it is more instructive and more general to find the solutions without leaving the geometric algebra approach. We know already that ψ_0 must include the factor p but we are at liberty to multiply on the right by an idempotent of our choice; we can then choose the set of four primitive idempotents generated by $h_1 = \sigma_{034}$ and $h_2 = \sigma_{04}$. The former of these Hermitian elements produces rotations on the σ_{12} plane and can be assigned to spin, but the latter produces a kind of 3-dimensional rotation on the $i\sigma_{04} = \sigma_{123}$ volume; this may possibly be related to isospin. Our option for a particular type of eigenideals produced solutions that can be described as having left or right spin and isospin; these solutions are effective in modelling leptons. Other possible choices for the idempotents are applicable to the various particles.

6.4 Hubble expansion

In Sec. 6.1 we verified that Eq. (6.2) accepted hyperplane wave solutions from which we could derive a null displacement condition expressed by Eq. (6.10); we can then conclude that hyperplane waves have displacement paths on the light hypercone of 5D.

This is easier to understand if we downgrade to 3-dimensional spacetime; in 3 dimensions geometrical representation becomes possible, so it is usually useful to start with 3-dimensional analysis and progress gradually to higher dimensions. A null displacement in 3-dimensional spacetime is specified by $(dx^0)^2 - (dx^1)^2 - (dx^2)^2 = 0$. All null displacements take place on the surface of a cone with apex at the current position and this cone is usually called the light cone, because light travels along null displacement paths. If we want to restrict our attention to null displacements we don't need 3 coordinates because we know beforehand that we are restricted to a 2-dimensional surface. Since the cone's axis is parallel to x^0 we can specify any displacement on its surface by a radial displacement (distance to the axis) and a polar angle variation. We have just switched to cylindrical polar coordinates because they are the most convenient to express the symmetry implied by null displacements. If we denote the radial and angular coordinates centered at the current position by τ and ρ , respectively, the null displacement condition becomes $(dx^0)^2 - (d\tau)^2 - \tau^2(d\rho)^2 = 0$.

Physical spacetime is 4-dimensional and so null displacements are performed on a 3-dimensional surface to which we can call a light hypercone; in special relativity one usually refers to the light cone, in spite of the fact that this is a 3-dimensional hypersurface, but in this book it is convenient to make the distinction clear. Again, if we are interested only in null displacements, 4 coordinates are one too many and the equations will gain clarity if we adopt spherical polar coordinates. One dimension above, the situation specified by Eq. (6.10) is again one of null displacements but now in 5-dimensional spacetime. One sees immediately that all displacements are now on the surface of a 4-dimensional light hypercone and that the appropriate coordinates to bring out the implied symmetry are hyperspherical polar coordinates, comprising one radial distance and 3 polar angles; we denote those coordinates by $\{\tau, \rho, \theta, \phi\}$, respectively. In hyperspherical coordinates a null displacement is expressed by

$$-(dt)^2 + (d\tau)^2 + \tau^2[(d\rho)^2 + \sin^2 \rho(d\theta)^2 + \sin^2 \rho \sin^2 \theta(d\phi)^2] = 0. \quad (6.72)$$

Here we see that the association of coordinate τ with proper time implies that the 3 coordinates associated with physical space are turned into distances measured on the surface of a 3-sphere so, when we use Cartesian coordinates for physical space we are actually referring to points on the hyperplane tangent to the hypersphere and not to points of constant τ . We can make an analogy to displacements on the Earth's surface, which can be approximated with displacements on a plane locally tangent to Earth, as long as displacements remain small. What we are saying here is that Cartesian coordinates can only be used on a local scale but not for cosmological distances.

One case of special interest considers displacements with a common origin. If displacements start at the apex of the light hypercone, they can only follow generatrices characterized by $d\rho = d\theta = d\phi = 0$ and have a very simple null condition given by

$$(dt)^2 = (d\tau)^2. \quad (6.73)$$

For convenience we will set $\tau = t$, that is, the coordinate's origin is placed at the hypercone's apex. Inserting this into the equation above and denoting derivatives with respect to t by a dot, we have

$$\dot{\tau} = 1, \quad (6.74)$$

for displacements along a generatrix.

All generatrices diverge from the hypercone's apex; naturally the distance between two generatrices increases with the radial coordinate τ . If we wish to compute distances to a particular point, measured over the hyperspherical directrix, we can set the origin for coordinate ρ at that point and introduce a distance coordinate

$$r = \tau\rho. \quad (6.75)$$

A simple manipulation allows us to write

$$\dot{r} = \dot{\tau}\rho + \tau\dot{\rho} = \dot{\tau}\rho = \frac{\dot{\tau}}{\tau} r. \quad (6.76)$$

But we have seen above that $\dot{\tau}$ is unity, so the previous equation states that two generatrices are coming apart with a velocity \dot{r} proportional to their distance r ; this is exactly what the Hubble relation says about galaxies and we are led to define the Hubble parameter by

$$H = \frac{1}{\tau}. \quad (6.77)$$

This is an extremely important conclusion. A physical interpretation of our geometric argument allows us to say that the wavefunctions of cosmical objects, such as galaxies or even galaxy clusters, are monogenic functions of 5-dimensional spacetime, diverging from a common origin, the big bang. As a consequence the Universe has an overall hyperspherical symmetry and we are wrong when we choose Cartesian coordinates for its description. By choosing the appropriate hyperspherical polar coordinates we detect immediately that distances between cosmical objects must grow at a rate which is exactly proportional to those distances. The Hubble relation thus acquires a purely geometrical explanation and results from the consideration of an empty Universe; there is no need for a critical mass density to justify this flat rate expansion.

In this model cosmological objects such as galaxy clusters are still objects, evolving along generatrices of the light hypercone. This contradicts the equivalence principle, in the sense that we now have an absolute definition of motion; our privileged frame, the frame of absolute stillness, is provided by all the galaxy clusters in spite of their apparent relative motion. Still objects are characterized by the relation $\dot{\tau} = 1$, while objects in motion relative to the privileged frame are characterized by $\dot{\tau} < 1$. From Eq. (6.72) we derive

$$1 - \tau^2[\dot{\rho}^2 + \sin^2(\rho)\dot{\theta}^2 + \sin^2(\rho)\sin^2(\theta)\dot{\phi}^2] - \dot{\tau}^2 = 0. \quad (6.78)$$

A physical interpretation of this equation becomes easier if we make the replacement $\tau\rho \rightarrow r$:

$$1 - \left[1 + \left(\frac{r}{\tau}\right)^2\right] \dot{\tau}^2 - \dot{r}^2 + 2\frac{r}{\tau} \dot{\tau}\dot{r} - r^2[\dot{\theta}^2 + \sin^2(\theta)\dot{\phi}^2] = 0. \quad (6.79)$$

What we usually designate by velocity is the quantity v verifying the relation

$$v^2 = \dot{r}^2 + r^2[\dot{\theta}^2 + \sin^2(\theta)\dot{\phi}^2]. \quad (6.80)$$

Replacing above we get

$$1 - v^2 - \left[1 + \left(\frac{r}{\tau}\right)^2\right] \dot{\tau}^2 + 2\frac{r}{\tau} \dot{\tau}\dot{r} = 0. \quad (6.81)$$

In laboratory experiments $r \ll \tau$, the equation simplifies to $\dot{\tau}^2 + v^2 = 1$ and we are taken back to special relativity.

Equation (6.81) expresses the compatibility between relativistic physics, which applies in laboratory experiments including those of high energy physics, and the physics of cosmology, which we must start to consider when the distances to our observational point become comparable to the size of the Universe. The orbits of planets involve distances which are still small compared to the Universe and we expect relativistic dynamics to work right for them, but what about the dynamics of galaxies? Since it involves the consideration of space curvature we shall have a look at this subject in a later chapter.

7 Directed integration theory

Directed integration is one area where geometric algebra provides a degree of simplification and unification that one should not neglect.

7.1 Line integrals

We can start with line integrals, the simplest cases to examine. The line integral of a multivector field $F(x)$ along a line $x(\lambda)$ is defined by

$$\int F(x) \frac{dx}{d\lambda} d\lambda = \int F dx = \lim_{n \rightarrow \infty} \sum_{i=1}^n \text{av}(F^i) \Delta x^i. \quad (7.1)$$

The final expression introduces a series of points along the curve, with x^0 and x^n the endpoints, and

$$\Delta x^i = x^i - x^{i-1}, \quad \text{av}(F^i) = \frac{1}{2} [F(x^{i-1}) + F(x^i)]. \quad (7.2)$$

If the curve is closed, then x^i and x^n coincide. The result of the integral is independent of the way we parameterize the curve, providing the ordering of the points along the curve is respected.

It is easy to overlook the importance of Eq. (7.1), since it looks very standard. One should note, however, that dx is a vector valued quantity and the product $F dx$ is a geometric product. The scope of the definition is then wider than one could have anticipated. This definition can be generalized to

$$\int F(x) \frac{dx}{d\lambda} G(x) d\lambda = \int F(x) dx G(x), \quad (7.3)$$

and we can also consider sums of terms like this. The most general form of the integral is

$$\int \mathbf{L}(\partial_\lambda x; x) d\lambda = \int \mathbf{L}(dx), \quad (7.4)$$

where $\mathbf{L}(a) = \mathbf{L}(a; x)$ is a multivector valued linear function of a . The position dependence of \mathbf{L} can be suppressed to simplify the notation.

If F is replaced by the vector valued function $v(x)$ we have

$$\int v \, dx = \int v \cdot dx + \int v \wedge dx, \quad (7.5)$$

which separates the line integral into scalar and bivector valued terms. If v is the unit tangent vector along the curve then the scalar part returns the arc length and the bivector part encodes the curve bending.

7.2 Surface integrals

Generalizing line integrals naturally takes us to surface integrals. We consider a multivector valued function F defined over a two-dimensional surface embedded in some larger space. The surface can be parameterized by two coordinates $x(x^1, x^2)$ and we define a directed measure by the bivector

$$dX = \partial_1 x \wedge \partial_2 x \, dx^1 dx^2 = g_1 \wedge g_2 \, dx^1 dx^2. \quad (7.6)$$

This measure is independent of the choice of coordinates, as long as the ordering of the frame vectors is maintained; the wrong ordering would change the sign of the measure. In some cases one may need to resort to more than one coordinate path to parameterize the surface but the measure will still be defined everywhere. A directed surface integral then takes the form

$$\int F \, dX = \int F \, g_1 \wedge g_2 \, dx^1 dx^2, \quad (7.7)$$

or a sum of terms like this if more than one coordinate patch is needed. As in the case of the line integral, the integrand can multiply the measure from left, right or both and a sum of such terms is possible.

As an example of a surface integral in three dimensions, consider a closed surface with unit outward normal n . Let F be a bivector valued function $\phi n i^{-1}$, with ϕ a scalar field. The surface integral then becomes

$$\oint \phi n i^{-1} dX = \oint \phi |dS|. \quad (7.8)$$

We have here replaced the scalar $n i^{-1} dX$ with the measure of the surface $|dS|$. A common convention for the directed measure gives $n dX$ the orientation of i . As a second example we make $F = 1$, so that the integral is

$$\oint dX = 0, \quad (7.9)$$

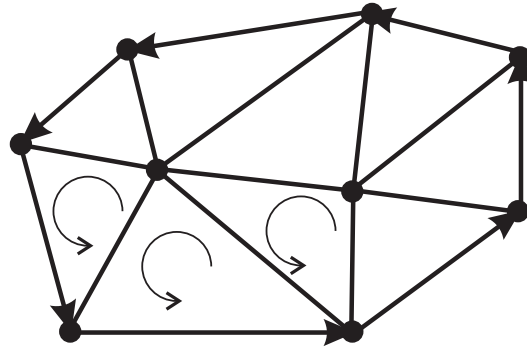


Figure 7.1: A triangulated surface is built by choosing a number of points on the surface; each set of three adjacent points defines a triangle or a *simplex*. Adding more points one gets a closer fit to the surface. All the simplices are given the same orientation.

which is true for any closed surface; for open surfaces the integral depends on the boundary.

A surface integral is the limit of a sum over a *triangulated surface* (see Fig. 7.1). A set of points is chosen over the surface and triangles or *simplices* are defined by sets of three adjacent points. The number of points is then increased until, in the limit, one obtains the original surface. Each simplex has an attached orientation such that adjacent simplices have the common edge traversed in reverse directions. In this way the orientation for the entire surface is built and the boundary is all traversed in the same direction. For some surfaces it is impossible to define a consistent orientation; one example is the Möbius strip. For such surfaces it is impossible to define a directed integral, so we restrict our presentation to *orientable surfaces*.

We now take one simplex defined by three points x_0, x_1, x_2 , with orientation defined by $x_0 \rightarrow x_1 \rightarrow x_2 \rightarrow x_0$. Two vectors are defined by

$$g_1 = x_1 - x_0, \quad g_2 = x_2 - x_0. \quad (7.10)$$

The surface measure for the simplex is then defined by

$$\Delta X = \frac{1}{2} g_1 \wedge g_2 = \frac{1}{2} (x_1 \wedge x_2 + x_2 \wedge x_0 + x_0 \wedge x_1). \quad (7.11)$$

The orientation of the surface measure is defined by the border and the area is equal to that of the simplex; an even permutation of the vertices leaves ΔX invariant. The integral can now be defined in terms of simplices as follows

$$\int F dX = \lim_{n \rightarrow \infty} \sum_{k=1}^n \text{av}(F^k) \Delta X^k. \quad (7.12)$$

The sum runs over all the simplices making up the surface and for each simplex $\text{av}(F)$ is the average value of F over the simplex. The way the sum is taken to the limit does not affect the final value of the integral for well-behaved integrals.

7.3 n -dimensional surfaces

The simplex mechanism allows us to extend easily the directed integral definition to surfaces of arbitrary dimension. The number of vertices in a simplex is one unit greater than the surface dimension. For an n -dimensional surface we consider one simplex with vertices x_0, \dots, x_n , whose order specifies the orientation. For this simplex we define the vectors

$$g_\iota = x_\iota - x_0, \quad \iota = 1, \dots, n, \quad (7.13)$$

and the measure for the simplex is

$$\Delta X = \frac{1}{n!} g_1 \wedge \dots \wedge g_n. \quad (7.14)$$

A point in the simplex can be described in terms of coordinates $\lambda_1, \dots, \lambda_n$ by writing

$$x = x_0 + \lambda^\iota g_\iota. \quad (7.15)$$

Each coordinate can take values between 0 and 1 and the coordinates also satisfy

$$\sum_{\iota=1}^n \lambda^\iota \leq 1. \quad (7.16)$$

Consider now a multivector field $F(x)$ defined over the entire surface and denote by $F_\iota = F(x_\iota)$ the value of the field at each vertex. A new function is then defined to interpolate the F_ι over the simplex. This can be

$$f(x) = F_0 + \lambda^\iota (F_\iota - F_0). \quad (7.17)$$

As the number of points increases and the simplices grow smaller, $f(x)$ becomes an even closer approximation of $F(x)$ and the triangulated surface approaches the original surface.

The directed integral of $F(x)$ over the surface is now approximated by the sum of f integrated over each simplex. The integral over the simplex uses λ^ι as coordinates, so that

$$dX = g_1 \wedge \dots \wedge g_n d\lambda^1 \dots d\lambda^n. \quad (7.18)$$

It is then straightforward to establish that

$$\int_{\text{simplex}} dX = \Delta X \quad (7.19)$$

and

$$\int_{\text{simplex}} \lambda^\iota dX = \frac{1}{n+1} \Delta X, \quad \forall \lambda^\iota. \quad (7.20)$$

The integral of $f(x)$ over a single simplex then evaluates to

$$\int_{\text{simplex}} f dX = \frac{1}{n+1} \left(\sum_{\iota=0}^n F_\iota \right) \Delta X. \quad (7.21)$$

The function is thus replaced by its average value over a simplex, which we denote by $\text{av}(F)$. Summing over all the simplices and taking the limit we can write

$$\int F dX = \lim_{n \rightarrow \infty} \sum_{\kappa=1}^n \text{av}(F^\kappa) \Delta X^\kappa. \quad (7.22)$$

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

$$\int \mathbf{L}(dX) = \lim_{n \rightarrow \infty} \sum_{\kappa=1}^n \text{av}(\mathbf{L}^k)(\Delta X^\kappa), \quad (7.23)$$

with $\text{av}(\mathbf{L}^k)$ the average value of \mathbf{L} over each simplex.

7.4 The fundamental theorem of calculus

Most readers will be familiar with several integral theorems, such as divergence and Stokes' and Cauchy integral formula for complex analysis. For the majority those will be separate theorems, each with its own application domain. Geometric algebra provides a means of unifying all those theorems under just one single formulation. In this section we will sketch the proof of this unified theorem. Those more interested in the applications than in the proof of the theorem can jump directly to the next section.

We start by introducing a notation aimed at simplifying the definition of the border operator. The k -dimension simplex (x_0, x_1, \dots, x_k) is represented by

$$(x)_{(k)} = (x_0, x_1, \dots, x_k). \quad (7.24)$$

The order of the points specifies the simplex orientation, the swap of two adjacent points changing the sign of the simplex. The border operator for a simplex is denoted \mathbf{B} and is defined by

$$\mathbf{B}(x)_{(k)} = \sum_{\iota=0}^k (-1)^\iota (x_0, \dots, \check{x}_\iota, \dots, x_k) = \sum_{\iota=0}^k (-1)^\iota (\check{x}_\iota)_{(k-1)}, \quad (7.25)$$

where the check denotes that the vertex x_ℓ is absent from the $(k-1)$ -simplex. The border operator returns the poly-hyper-line made up by simplices of dimension $k-1$. For example, for $k=1$,

$$B(x_0, x_1) = (x_1) - (x_0), \quad (7.26)$$

the border operator returns the endpoints of a line segment. The boundary of a boundary naturally vanishes:

$$BB(x)_{(k)} = 0. \quad (7.27)$$

We don't prove this but the result should be obvious.

The contents of a simplex can be a directed length, a directed area or a directed hyper-area, depending on the simplex dimension. We define the directed contents of a simplex by

$$\Delta(x)_{(k)} = \frac{1}{k!} (x_1 - x_0) \wedge (x_2 - x_0) \wedge \cdots \wedge (x_k - x_0). \quad (7.28)$$

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

$$\int_{(x)_{(k)}} dX = \Delta(x)_{(k)} = \Delta X. \quad (7.29)$$

The directed contents of a border is null:

$$\Delta[B(x)_{(k)}] = 0. \quad (7.30)$$

As an example consider the a two-dimensional simplex consisting of three points. The border is

$$B(x_0, x_1, x_2) = (x_1, x_2) - (x_0, x_2) + (x_0, x_1); \quad (7.31)$$

so, the directed contents of the border is

$$\Delta[B(x_0, x_1, x_2)] = (x_2 - x_1) - (x_2 - x_0) + (x_1 - x_0) = 0. \quad (7.32)$$

It would be easy to build up from the previous result to the case of an arbitrary simplex. These results are all that is needed to establish that the directed integral over the border of a simplex is zero:

$$\oint_{B(x)_{(k)}} dL = \sum_{\ell=0}^k (-1)^\ell \int_{(\check{x}_\ell)_{(k-1)}} dX = \Delta[B(x)_{(k-1)}] = 0. \quad (7.33)$$

A general surface is built up by a set of simplices, defined in such way that any common boundary, the directed hyper-lines shared by two simplices, are equal and opposite. Consequently, the line integrals over two simplices cancel out over their common border.

The line integral over the boundary of the surface can therefore be replaced by the sum of the line integrals over each simplex in the set. If the border is closed we establish that

$$\oint dL = \lim_{n \rightarrow \infty} \sum_{\kappa=1}^n \oint dL^\kappa = 0. \quad (7.34)$$

The sum runs over each simplex in the hyper-line, with κ labelling the simplex. Everything that was said above applies to hyper-lines bounding a surface that can be filled with a connected set of simplices. So, as well as being oriented, the surface must be simply connected.

Now recall Eq. (7.20) and consider a constant vector b . If we define $b_\iota \cdot g^\iota$ we see that

$$b_\iota \lambda^\iota = b \cdot (x - x_0), \quad (7.35)$$

which is valid for all vectors x in the simplex of interest. Multiplying Eq. (7.20) by b_ι and summing over all ι we get

$$\int_{(x)_{(k)}} b \cdot (x - x_0) dX = \frac{1}{k+1} \sum_{\iota=1}^k b \cdot g^\iota \Delta X. \quad (7.36)$$

Rearranging the equation we obtain

$$\begin{aligned} \int b \cdot x dX &= \frac{1}{k+1} \left[\sum_{\iota=1}^k b \cdot (x_\iota - x_0) + (k+1)b \cdot x_0 \right] \Delta X \\ &= b \cdot \text{av}(x) \Delta X, \end{aligned} \quad (7.37)$$

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

$$\text{av}(x) = \frac{1}{k+1} \sum_{\iota=1}^k x_\iota. \quad (7.38)$$

Now suppose we have a k -simplex, specified by $k+1$ points (x_0, x_1, \dots, x_k) and we form the directed line integral of $b \cdot x$; we get

$$\oint_{B(x)_{(k)}} b \cdot dL = \frac{1}{k+1} \sum_{\iota=1}^k (-1)^\iota b \cdot (x_0 + \dots \check{x}_\iota \dots + x_n) \Delta(\check{x}_\iota)_{(k-1)}. \quad (7.39)$$

The right hand side sum can be evaluated with the help of the result

$$\sum_{\iota=1}^k (-1)^\iota b \cdot (x_0 + \dots \check{x}_\iota \dots + x_n) \Delta(\check{x}_\iota)_{(k-1)} = \frac{1}{k!} b \cdot (g_1 \wedge \dots \wedge g_n). \quad (7.40)$$

This can be proven algebraically and we leave it for the reader. We have now established the simple result

$$\oint_{B(x)_{(k)}} b \cdot x dL = b \cdot (\Delta X). \quad (7.41)$$

where $\Delta X = \Delta[(x)_{(k)}]$. The order and orientation in this equation are relevant. The simplex $(x)_{(k)}$ is oriented, and the order of points specifies how its boundary is traversed. With dL the oriented element over each boundary, and ΔX the surface element for the simplex, we find that the correct expression for the line integral is $b \cdot (\Delta X)$.

Now we can apply these results to the interpolated function $f(x)$ of Eq. (7.17). We assume that we are working in a n -dimensional space and consider a simplex with points (x_0, x_1, \dots, x_n) . The simplex is chosen so that its surface is non-zero, so the n vectors $g_\iota = x_\iota - x_0$ define a frame. We therefore write

$$g_\iota = x_\iota - x_0, \quad (7.42)$$

and introduce the reciprocal frame $\{g^\iota\}$. For these vectors it is

$$g^\iota \cdot (x - x_0) = \lambda^\iota. \quad (7.43)$$

It follows that the line integral of $f(x)$ over the boundary is given by

$$\begin{aligned} \oint_{B(x)_{(k)}} f(x) dL &= \sum_{\iota=1}^n (F_\iota - F_0) \oint g^\iota \cdot (x - x_0) dL \\ &= \sum_{\iota=1}^n (F_\iota - F_0) g^\iota \cdot (\Delta X). \end{aligned} \quad (7.44)$$

Taking into account the directional derivatives of $f(x)$ we see that

$$\frac{\partial f(x)}{\partial \lambda^\iota} = F_\iota - F_0. \quad (7.45)$$

The result for the line integral can then be written

$$\begin{aligned} \oint_{B(x)_{(k)}} f(x) dL &= \sum_{\iota=1}^n (F_\iota - F_0) g^\iota \cdot (\Delta X) \\ &= \sum_{\iota=1}^n \frac{\partial f}{\partial \lambda^\iota} g^\iota \cdot (\Delta X) = \overrightarrow{f} \cdot \overleftarrow{\nabla} \cdot (\Delta X). \end{aligned} \quad (7.46)$$

We used the definition $\nabla = g^\iota \partial_\iota$, that results from the λ^ι being a set of coordinates.

The result of the last equation can now be added over all the simplices. The interpolated function $f(x)$ takes on the same value over the common boundary of two adjacent

simplices so, after summation, the result is just the integral over the surface boundary. Therefore we arrive at

$$\oint f(x) dL = \sum_{\kappa} \vec{f} \overleftarrow{\nabla} \cdot (\Delta X), \quad (7.47)$$

with the sum running over all the simplices. In the limit, more points are added and the simplices shrink in size towards zero, allowing us to write the first form of the fundamental theorem,

$$\oint_{BS} dL = \int_S \vec{F} \overleftarrow{\nabla} dX. \quad (7.48)$$

The replacement of f by F is valid in the limit, when the interpolated function coincides with its original. The inner product was replaced with a geometric product because, in the limit, ∇ lies entirely on the space defined by the pseudoscalar measure dX .

We are now familiar with the generalization procedure and apply these results to the fundamental theorem, as formulated above, to get

$$\oint_{BS} dLG = \int_S \vec{\nabla} dX \overleftarrow{G}. \quad (7.49)$$

Since ∇ is a vector, it may or may not commute with dX , depending on the dimension of space. A further generalization is still possible, allowing for different terms with the function on the left or right of the dL . We introduce a linear function $L(A_{n-1}) = L(A_{n-1}; x)$. This function takes a multivector of grade $n-1$ as its argument and returns a general multivector. L is also position-dependent, and its interpolation over a simplex is defined by

$$L(A) = L(A; x_0) + \sum_{\iota=1}^n \lambda^{\iota} [L(A; x_{\iota}) - L(A; x_0)]. \quad (7.50)$$

Because $L(A)$ is linear, sums and integrals can be moved inside the argument, so we can write

$$\begin{aligned} \oint L(dL) &= L\left(\oint dL; x_0\right) + \sum_{\iota=1}^n L\left(\oint \lambda^{\iota} dL; x_{\iota}\right) - \sum_{\iota=1}^n L\left(\oint \lambda^{\iota} dL; x_0\right) \\ &= \sum_{\iota=1}^n L(g^{\iota} \Delta X; x_{\iota}) - L(g^{\iota} \Delta X; x_0) \\ &= \vec{L}(\overleftarrow{\nabla} \Delta X). \end{aligned} \quad (7.51)$$

There is no position dependence in the final term because the derivative is constant over the simplex. Building up a chain of simplices and taking the limit we get the general result

$$\oint_{BS} L(dL) = \int_S \vec{L}(\overleftarrow{\nabla} dX). \quad (7.52)$$

This holds for any linear function $L(A_{n-1})$ integrated over a closed region of n -dimensional flat space. We recall that the notions of line and surface integrals are here taken in a very general sense; for instance, in 3 dimensions the line integral would become a surface integral and the surface integral would be called a volume integral. This is still not the most general statement of the fundamental theorem because we require that space is flat.

7.5 The divergence and Green's theorems

For the first application of the fundamental theorem we let $L(A)$ be the scalar valued function

$$L(A) = \langle JAI^{-1} \rangle. \quad (7.53)$$

Here J is a vector, and I is the unit pseudoscalar for the n -dimensional space. The argument A is a multivector of grade $n - 1$. Equation (7.52) gives

$$\int_S \langle \vec{J} \overleftarrow{\nabla} dX I^{-1} \rangle = \int_S \nabla \cdot J |dX| = \oint_{BS} \langle J dL I^{-1} \rangle, \quad (7.54)$$

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

$$n|dL| = dL I^{-1}, \quad (7.55)$$

where $|dL|$ is the scalar measure over the boundary. With this definition we make sure that, in Euclidean spaces, $n dL$ has the orientation defined by I , and that n points outwards. With this definition we arrive at

$$\int_S \nabla \cdot J |dX| = \oint_{BS} n \cdot J |dL|. \quad (7.56)$$

This may seem an unusual form of the divergence theorem but its transposition to 3-dimensions and common vector analysis notation makes it immediately recognizable:

$$\int_V \text{div} J dV = \oint_S n \cdot J dS. \quad (7.57)$$

Now go back to the fundamental theorem in the form of Eq. (7.49) and let G be the vector J in two-dimensional Euclidean space. The theorem becomes

$$\oint_{BS} dL J = \int_S \vec{\nabla} dX \overleftarrow{J} = - \int_S \nabla J dX, \quad (7.58)$$

where we used the fact that dX is a pseudoscalar and so it anticommutes with vectors in two dimensions. Introducing Cartesian coordinates, we have $dX = I dx dy$, so

$$\oint_{BS} dL J = - \int_S \nabla J I dx dy. \quad (7.59)$$

If we let $J = P\sigma_1 + Q\sigma_2$ and take the scalar part of both sides, we prove Green's theorem in the plane

$$\oint Pdx + Qdy = \int (\partial_x Q - \partial_y P) dxdy. \quad (7.60)$$

The line integral is taken around the perimeter of the area, in a positive sense, as specified by $I = \sigma_{12}$.

7.6 Cauchy's integral formula

Cauchy's integral theorem of complex variable theory can be seen as a special case of the fundamental theorem. We let ψ be an even-grade multivector, therefore it commutes with dX in two dimensions. So, we can write

$$\int \nabla \psi dX = \oint dL\psi = \oint \frac{\partial \vec{r}}{\partial \lambda} \psi d\lambda. \quad (7.61)$$

Here λ is a parameter along the closed curve. Now we recall from Sec. 4.3 that a complex number z is formed by $z = \sigma_1 \vec{r}$. The special notation for vectors is used in this section to distinguish from complex numbers. Therefore we have

$$\oint \psi dz = \int \sigma_1 \nabla \psi dX, \quad (7.62)$$

where the left hand side has become a complex line integral. We know that ψ is analytic if $\nabla \psi = 0$, so we have just proved that the line integral of an analytic function over a closed curve always vanishes.

For a contour C enclosing a point a , Cauchy's integral formula states that

$$f(a) = \frac{1}{2\pi i} \oint_C \frac{f(z)}{z-a} dz, \quad (7.63)$$

where f is an analytic function; the result is independent of contour shape. To understand this independence consider a second contour enclosing C ; the difference between the two contour integrals is a contour integral on a region not enclosing a . In such region $f(z)/(z-a)$ is analytic, so the difference has zero contribution.

We concentrate now on the properties of Cauchy's kernel $1/(z-a)$. We start with

$$\frac{1}{z-a} = \frac{(z-a)^\dagger}{|z-a|^2} = \frac{\vec{r}-\vec{a}}{(\vec{r}-\vec{a})^2} \sigma_1. \quad (7.64)$$

with $\vec{a} = a\sigma_1$, the vector corresponding to the complex number a . The important factor here is $(\vec{r}-\vec{a})/(\vec{r}-\vec{a})^2$, which can be written as

$$\vec{r}-\vec{a}(\vec{r}-\vec{a})^2 = \nabla \ln |\vec{r}-\vec{a}|. \quad (7.65)$$

It happens that $\ln |\vec{r} - \vec{a}|$ is Green's function for the Laplacian operator in two dimensions,

$$\nabla^2 \ln |\vec{r} - \vec{a}| = 2\pi\delta(\vec{r} - \vec{a}). \quad (7.66)$$

We then conclude that the vector part of Cauchy's kernel must verify

$$\nabla \frac{\vec{r} - \vec{a}}{(\vec{r} - \vec{a})^2} = 2\pi\delta(\vec{r} - \vec{a}). \quad (7.67)$$

The Cauchy kernel is the Green's function for the two-dimensional vector derivative.

The Cauchy integral formula now follows from Eq. (7.62),

$$\begin{aligned} \oint \frac{f(z)}{z-a} dz &= \sigma_1 \int \nabla \left[\frac{\vec{r} - \vec{a}}{(\vec{r} - \vec{a})^2} \sigma_1 f(x) \right] dX \\ &= \sigma_1 \int \left[2\pi\delta(x - \vec{a}) \sigma_1 f(z) + \nabla f(z) \frac{\vec{r} - \vec{a}}{(\vec{r} - \vec{a})^2} \sigma_1 \right] I |dX| \\ &= 2\pi I f(a). \end{aligned} \quad (7.68)$$

Establishing the theorem in terms of geometric algebra ensures that we are now able to proceed to its generalization. We have already proved a theorem in two dimensions, applying to non-analytic functions. For these we can write

$$2\pi I f(a) = \oint \frac{f}{z-a} dz - 2 \int \frac{\partial f}{\partial z^\dagger} \frac{1}{z-a} I |dX|, \quad (7.69)$$

using relations from Sec. 4.3.

In complex analysis one is frequently faced with the series expansion of a function. Particularly for an analytic function $f(z)$, apart from a pole of order n at $z = a$, the function has a Laurent series given by

$$f(z) = \frac{a_{-n}}{(z-a)^n} \cdots \frac{a_{-1}}{z-a} + \sum_{\iota=0}^{\infty} a_{\iota} (z-a)^{\iota}. \quad (7.70)$$

The residue theorem then states that for this function

$$\oint_C f(z) = 2\pi i a_{-1}. \quad (7.71)$$

The residue term in a Laurent expansion is a weighted Green's function; the theorem just recovers the weight.

Suppose now that we start with a function $f(x)$ defined on the real axis, which we want to propagate into the upper half-plane, subject to the boundary condition that f falls to zero as $|z| \rightarrow \infty$. The Cauchy formula tells us we should propagate according to the formula

$$f(a) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{f(x)}{x-a} dx. \quad (7.72)$$

If we now form the Fourier transform of the initial function $f(x)$,

$$f(x) = \int_{-\infty}^{\infty} \frac{dk}{2\pi} F(k) e^{ikx}. \quad (7.73)$$

Replacing above we have

$$f(a) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{dk}{2\pi} F(k) \int_{-\infty}^{\infty} \frac{e^{ikx}}{x-a} dx. \quad (7.74)$$

We can now close the x integral in the upper half-plane for positive k . For negative k there is no residue term because a lies in the upper half-plane. The Cauchy integral returns

$$f(a) = \int_0^{\infty} \frac{dk}{2\pi} F(k) e^{ika}. \quad (7.75)$$

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

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

$$f(\theta) = \sum_{-\infty}^{\infty} f_n e^{in\theta}. \quad (7.76)$$

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

$$f(z) = \sum_{n=1}^{\infty} f_{-n} z^{-n}, \quad |z| > 1. \quad (7.77)$$

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

7.7 Green's functions in Euclidean spaces

Our excursion through complex variable theory and its relation to two-dimensional space was merely a step towards a generalized theory. An analytic function is a 2-dimensional

monogenic function, so we now look at multivector functions satisfying $\nabla\psi = 0$. For simplicity we will only work with even-grade multivectors. The fundamental theorem states that

$$\oint_{BS} \psi dL = \int \nabla\psi dX = 0, \quad (7.78)$$

where we have used the fact that ψ commutes with the pseudoscalar measure dX . For any monogenic function, the directed integral over a closed boundary (hyper-line) must vanish.

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

$$G(x; y) = \frac{1}{S_n} \frac{x - y}{|x - y|^n}, \quad (7.79)$$

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

$$\nabla G(x; y) = \nabla \cdot G(x; y) = \delta(x - y). \quad (7.80)$$

When applying the fundamental theorem we must allow for the lack of commutativity between G and ψ . We write

$$\begin{aligned} \oint_{BS} G dL &= \int_S (\vec{G} \overleftarrow{\nabla} \psi + G \nabla \psi) dX \\ &= \int_S \vec{G} \overleftarrow{\nabla} \psi dX. \end{aligned} \quad (7.81)$$

The last relation is justified by ψ being monogenic. Letting G be the Green's function of Eq. (7.79), we find that the Cauchy's theorem in n dimensions can be written as

$$\psi(y) = \frac{1}{S_n} \oint_{BS} \frac{x - y}{|x - y|^n} dL \psi(x). \quad (7.82)$$

This provides a relation between the value of a monogenic function at a point, with the value of a directed integral over a region surrounding the point.

Equation (7.79) provides a generalization of Liouville's theorem for monogenic functions in Euclidean spaces. We first need a definition for the modulus function:

$$|M| = \langle MM^\dagger \rangle^{1/2}, \quad (7.83)$$

which is a well defined positive-definite function for all multivectors M in any geometric algebra. The modulus function satisfies Schwarz inequality in the form

$$|A + B| \leq |A| + |B|, \quad (7.84)$$

as one can easily demonstrate. If we let a denote a unit vector and ∇_y denote the derivative with respect to vector y we find that

$$a \cdot \nabla_y \psi(y) = \frac{1}{IS_n} \oint_{BS} \frac{a(x-y)^2 + na \cdot (x-y)(x-y)}{|x-y|^2} dL \psi(x). \quad (7.85)$$

Consequently we have

$$|a \cdot \nabla_y \psi(y)| \leq \frac{1}{S_n} \oint_{BS} \frac{n+1}{|x+y|^n} |dL| |\psi(x)|. \quad (7.86)$$

Now we consider the case when ψ is bounded, implying that $|\psi(x)|$ never exceeds a given value. Taking the BS to large radius $r = |x|$, we will find that the right-hand side falls off with $1/r$. This is all that is needed to prove that the derivative of ψ must vanish in all directions and that the only monogenic function that is bounded over all space is a constant.

If we want to propagate a function in Euclidean space subject to some boundary conditions, this can also be done with the help of Eq. (7.79). Suppose we want to propagate ψ off the surface of the unit ball, subject to the condition that it falls off to zero at infinity. We assume ψ to be of the form

$$\psi = \sum_{l=-\infty}^{\infty} \alpha_l r^l \psi_l. \quad (7.87)$$

where ψ_l are independent of r . The derivative of ψ can be written as

$$\nabla \psi = \sum_{l=-\infty}^{\infty} (\alpha_l \sigma_r l r^{l-1} \psi_l + r^l \nabla \psi_l). \quad (7.88)$$

It follows that each of the ψ_l is an angular monogenic verifying the condition

$$x \wedge \nabla \psi = -l \psi. \quad (7.89)$$

Only the negative powers have their integral closed over the exterior region and the result has the form

$$\psi = \sum_{l=1}^{\infty} \alpha_{-l} r^{-l} \psi_{-l}, \quad r > 1. \quad (7.90)$$

Similarly, only the positive powers contribute to the interior solution.

7.8 Spacetime propagators

Monogenic functions in mixed signature spaces are somewhat different from their Euclidean counterparts. Liouville's theorem does not apply and one can easily construct

bounded solutions which are non-singular over all space. Three-dimensional plane wave solutions or their analogue in higher dimensional spaces are an example of such functions, which will be of great interest for us. There are also *characteristic surfaces* or surfaces from which one cannot propagate a monogenic function; these are surfaces whose normal lies on the surface itself. Suppose the space dimension is n and we have a set of tangent vectors on a surface of dimension $n - 1$, $\{g_1, \dots, g_{n-1}\}$. For the normal n to lie on the surface we must have

$$n \wedge (g_1 \wedge g_2 \wedge \dots \wedge g_{n-1}) = 0. \quad (7.91)$$

But this implies

$$(I^{-1}n) \wedge n = I^{-1}n \cdot n = 0. \quad (7.92)$$

A characteristic surface therefore has $n^2 = 0$, so characteristic surfaces are always null surfaces.

Consider a two-dimensional Minkowskian space with basis vectors $\{\sigma_0, \sigma_1\}$, $(\sigma_1)^2 = -(\sigma_0)^2 = 1$, and unit pseudoscalar $I = \sigma_{01}$. The monogenic equation is $\nabla\psi = 0$, where ψ is an even-grade multivector with scalar and pseudoscalar terms. We define the null vectors

$$n_{\pm} = \sigma_0 \pm \sigma_1. \quad (7.93)$$

We note that $In_+ = n_+$ and $In_- = -n_-$. Pre-multiplying the monogenic equation by n_+ we get

$$n_+ \cdot \nabla\psi = -n_+ \wedge \nabla\psi = I(n_+I) \cdot \nabla\psi = -In_+ \cdot \nabla\psi. \quad (7.94)$$

It follows that

$$(1 + I)n_+ \cdot \nabla\psi = 0, \quad (7.95)$$

and similarly

$$(1 - I)n_- \cdot \nabla\psi = 0. \quad (7.96)$$

We can decompose ψ into $\psi = \psi_+ + \psi_-$, with

$$\psi_{\pm} = \frac{1}{2}(1 \pm I)\psi. \quad (7.97)$$

The two components separate, having vanishing derivatives along the respective null vectors. Propagation of ψ from an initial surface is straightforward. The function is split into ψ_{\pm} and the values of these are transported along the respective null vectors, dispensing with complicated contour integrals.

Since the values of ψ are carried along characteristic surfaces, any surface where the initial values are specified must cut the characteristic surface only once. Otherwise the initial values will not normally be consistent with the differential equation. For the monogenic condition $\nabla\psi = 0$, one can, for instance, specify the initial conditions along the σ_1 axis. The fundamental theorem involves integrals around closed loops; the theorem can still be applied in Minkowski spacetime and we can find out what happens to the

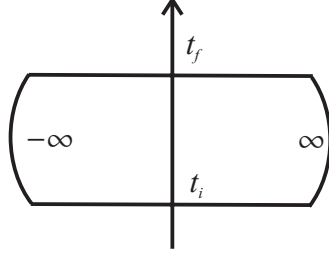


Figure 7.2: A spacetime contour, closed at spatial infinity.

boundary data if we try to construct an interior solution with arbitrary boundary data. We must start by setting up the Green's function. This can be done via its Fourier transform. With $x = \sigma_0 x^0 + \sigma_1 x^1$ we find

$$\begin{aligned} G(x) &= i \int \frac{d\omega}{2\pi} \frac{d\kappa}{2\pi} \frac{\omega\sigma_0 + \kappa\sigma_1}{\kappa^2 - \omega^2} e^{i(\omega x^0 - \kappa x^1)} \\ &= \frac{i}{2} \int \frac{d\omega}{2\pi} \frac{d\kappa}{2\pi} \left(\frac{\sigma_0 + \sigma_1}{\kappa - \omega} + \frac{\sigma_1 - \sigma_0}{\omega + \kappa} \right) e^{i(\omega x^0 - \kappa x^1)} \\ &= \frac{x^0}{4|x^0|} [\delta(x^0 - x^1)(\sigma_0 + \sigma_1) + \delta(x^0 + x^1)(\sigma_1 - \sigma_0)]. \end{aligned} \quad (7.98)$$

To apply the fundamental theorem we consider the contour of Fig. 7.2, which runs parallel to the σ_1 axis for two different values of x^0 , t_i and t_f , with $t_i < t_f$. The function ψ that is being propagated is assumed to fall off to zero at infinity. The fundamental theorem then gives¹

$$\begin{aligned} \psi(y) &= I \int_{-\infty}^{\infty} d\lambda G(t_i \sigma_0 + \lambda \sigma_1 - y) \sigma_1 \psi(t_i; \lambda) \\ &\quad - I \int_{-\infty}^{\infty} d\lambda G(t_f \sigma_0 + \lambda \sigma_1 - y) \sigma_1 \psi(t_f; \lambda) \\ &= \frac{1}{4} (1 + I) [\psi(t_i, y^1 - y^0 + t_i) + \psi(t_f, y^1 - y^0 + t_f)] \\ &\quad - \frac{1}{4} (1 - I) [\psi(t_i, y^0 - y^1 + t_i) + \psi(t_f, y^0 - y^1 + t_f)]. \end{aligned} \quad (7.99)$$

We see that for the function $\psi_+(y)$, for instance, we form the null vector n_+ through y . The value at y is then the average value at the two intersections with the boundary, and similarly for $\psi_-(y)$. Much like the Euclidean case, only the part of the function on the boundary that is consistent with the monogenic equation is propagated to the interior.

¹Check the signs in this equation.

8 Curved manifolds

In this chapter we will extend the previous results to spaces that cannot be spanned by a constant frame with the same dimension as that of the space itself. The simplest example is that of a 2-sphere, which is a 2-dimensional surface but can only be spanned by a constant frame if we resort to 3 dimensions; this process is called embedding. We will assume that embedding is always possible, so that we can always resort to the higher dimension embedding space for our arguments, even if later on we find means of working on the curved space itself. Curved spaces embedded in higher dimensional flat space will frequently be designated by embedded surfaces, by extrapolation of the embedding of a 2-dimensional surface in 3-space.

We will use a restrictive definition for *vector manifold*, as being a set of points labelled by vectors laying in a geometric algebra of arbitrary dimension and signature. In practice this definition will prove sufficient for all our needs and will allow us to simplify derivations when compared to a more general one. If we consider a path on the surface $x(\lambda)$, the tangent vector is defined naturally by

$$\dot{x} = \left. \frac{\partial x(\lambda)}{\partial \lambda} \right|_{\lambda_0} = \lim_{\varepsilon \rightarrow 0} \frac{x(\lambda_0 + \varepsilon) - x(\lambda_0)}{\varepsilon}. \quad (8.1)$$

The notation \dot{x} is used for derivatives with respect to a parameter in general, so that \dot{x} can later be used for derivatives with respect to an affine parameter. Note that the numerator on the right-hand side fraction will not normally belong to the surface but it will belong to the embedding space. We next define the path length by the integral formula

$$s = \int_{\lambda_1}^{\lambda_2} (\dot{x} \cdot \dot{x})^{1/2} d\lambda. \quad (8.2)$$

The embedded surface inherits a metric from the surrounding embedding flat space. In many applications the embedding space is used only conceptually for the derivations and one works entirely in the embedded surface; in other applications, however, the embedding space can be used with great advantage. It is useful to define a displacement vector dx along the path by

$$dx = \frac{\dot{x}}{|\dot{x}|} d\lambda, \quad (8.3)$$

so that the path length may also be written as

$$s = \int_{\lambda_1}^{\lambda_2} (dx \cdot dx)^{1/2}. \quad (8.4)$$

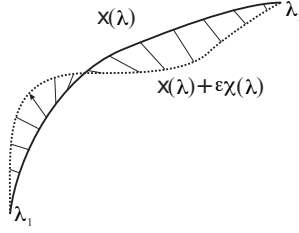


Figure 8.1: The geodesic is the path that provides a stationary, usually minimum, length between the endpoints.

8.1 Geodesics

Suppose now that n paths pass through a point x , n being the dimension of the embedded surface; the paths define a set of tangent vectors $\{g_1, \dots, g_n\}$; we assume the paths to be chosen so that those vectors are independent, so they form a basis for the n -dimensional tangent space at point x . The reciprocal frame is denoted, as usual by $\{g^t\}$. This can be done for all points on the surface so, for a general path $x(\lambda)$, the tangent vector can be expanded in terms of the local frame as

$$\dot{x} = g_t g^t \cdot \dot{x} = g_t \dot{x}^t. \quad (8.5)$$

The path length can now also be written in terms of coordinates as

$$s = \int_{\lambda_1}^{\lambda_2} (g_{t\kappa} \dot{x}^t \dot{x}^\kappa)^{1/2} d\lambda = \int_{\lambda_1}^{\lambda_2} (2L)^{1/2} d\lambda, \quad (8.6)$$

where L is called the *Lagrangian* and is defined by

$$L(x^t, \dot{x}^t) = \frac{1}{2} g_{t\kappa} \dot{x}^t \dot{x}^\kappa. \quad (8.7)$$

The Lagrangian is an explicit scalar function of the coordinates through the metric tensor $g_{t\kappa}$ and of the coordinate derivatives with respect to the parameter λ .

A *geodesic* is defined as the curve passing through two points on the surface, whose length is stationary with respect to path variations. Usually a geodesic provides the shortest length between two points but in special cases this may not be so. Therefore we assume that $x(\lambda)$ is a geodesic and envisage a slightly distorted path

$$x(\lambda) + \varepsilon \chi(\lambda),$$

where ε is small and the function χ specifies the distortion. The ends of the path remain fixed, so it must be $\chi(\lambda_1) = \chi(\lambda_2) = 0$. The path length is now a function of ε and we

require, for an arbitrary distortion, as in Fig. 8.1;

$$\left. \frac{ds(\varepsilon)}{d\varepsilon} \right|_{\varepsilon=0} = 0. \quad (8.8)$$

But the path length is now given by

$$s(\varepsilon) = \int_{\lambda_1}^{\lambda_2} [2L(x + \varepsilon\chi, \dot{x} + \varepsilon\dot{\chi})]^{1/2} d\lambda. \quad (8.9)$$

Deriving with respect to ε we get

$$\begin{aligned} \left. \frac{ds(\varepsilon)}{d\varepsilon} \right|_{\varepsilon=0} &= \left[\int_{\lambda_1}^{\lambda_2} \frac{1}{\sqrt{2L}} \left(\frac{\partial L}{\partial \dot{x}^i} \dot{\chi}^i + \frac{\partial L}{\partial x^i} \chi^i \right) d\lambda \right]_{\varepsilon=0} \\ &= \left[\int_{\lambda_1}^{\lambda_2} \left[\frac{d}{d\lambda} \left(-\frac{1}{\sqrt{2L}} \frac{\partial L}{\partial \dot{x}^i} \right) + \frac{1}{\sqrt{2L}} \frac{\partial L}{\partial x^i} \right] \chi^i d\lambda \right]_{\varepsilon=0}. \end{aligned} \quad (8.10)$$

Note that the first term on the right-hand side can be written

$$\int_{\lambda_1}^{\lambda_2} \frac{1}{\sqrt{2L}} \frac{\partial L}{\partial \dot{x}^i} \dot{\chi}^i d\lambda = \int_{\lambda_1}^{\lambda_2} \frac{\partial(\sqrt{2L})}{\partial \dot{x}^i} \dot{\chi}^i d\lambda. \quad (8.11)$$

This can be integrated by parts

$$\int_{\lambda_1}^{\lambda_2} \frac{\partial(\sqrt{2L})}{\partial \dot{x}^i} \dot{\chi}^i d\lambda = \left[\frac{\partial(\sqrt{2L})}{\partial \dot{x}^i} \chi^i \right]_{\lambda_1}^{\lambda_2} - \int_{\lambda_1}^{\lambda_2} \frac{d}{d\lambda} \left(\frac{\partial(\sqrt{2L})}{\partial \dot{x}^i} \right) \chi^i d\lambda. \quad (8.12)$$

The first term on the right-hand side is zero because χ^i vanishes for the end points; replacing in Eq. (8.10)

$$\left. \frac{ds(\varepsilon)}{d\varepsilon} \right|_{\varepsilon=0} = \frac{1}{\sqrt{2}} \int_{\lambda_1}^{\lambda_2} \left[\frac{d}{d\lambda} \left(-\frac{1}{\sqrt{L}} \frac{\partial L}{\partial \dot{x}^i} \right) + \frac{1}{\sqrt{L}} \frac{\partial L}{\partial x^i} \right] \chi^i d\lambda. \quad (8.13)$$

The right-hand side must be zero for arbitrary distortion functions χ^i , so we conclude that the following set of n simultaneous equations must be verified

$$\frac{d}{d\lambda} \left(\frac{1}{\sqrt{L}} \frac{\partial L}{\partial \dot{x}^i} \right) = \frac{1}{\sqrt{L}} \frac{\partial L}{\partial x^i}; \quad (8.14)$$

these are the Euler-Lagrange equations defining a geodesic which does not intersect or belong to a characteristic surface.

In practice we usually choose λ to be an *affine parameter*, that is, one for which the Lagrangian L is constant along the curve. Actually we may even use the measured length

along the curve s as parameter. When this choice is made $L = 1/2$ and the equations simplify to

$$\frac{d}{ds} \left(\frac{\partial L}{\partial \dot{x}^\iota} \right) = \frac{\partial L}{\partial x^\iota}. \quad (8.15)$$

One further advantage of this option is that $L = 1/2$ is always available as *first integral* of these equations, simplifying their solution.

Using the Lagrangian definition (8.7) we see that

$$\frac{\partial L}{\partial \dot{x}^\iota} = g_{\iota\kappa} \dot{x}^\kappa, \quad \frac{\partial L}{\partial x^\iota} = \frac{1}{2} \frac{\partial g_{\kappa\lambda}}{\partial x^\iota} \dot{x}^\kappa \dot{x}^\lambda. \quad (8.16)$$

The Euler-Lagrange equations can then be written

$$2 \frac{d}{ds} (g_{\iota\kappa} \dot{x}^\kappa) - \frac{\partial g_{\kappa\lambda}}{\partial x^\iota} \dot{x}^\kappa \dot{x}^\lambda = 0. \quad (8.17)$$

Performing the s -derivation we get

$$g_{\iota\kappa} \ddot{x}^\kappa + \frac{1}{2} \left(\frac{\partial g_{\iota\alpha}}{\partial x^\beta} + \frac{\partial g_{\iota\beta}}{\partial x^\alpha} - \frac{\partial g_{\alpha\beta}}{\partial x^\iota} \right) \dot{x}^\alpha \dot{x}^\beta = 0. \quad (8.18)$$

This equation is usually written in a slightly modified form as

$$\ddot{x}^\iota + \Gamma_{\kappa\lambda}^\iota \dot{x}^\kappa \dot{x}^\lambda = 0 \quad (8.19)$$

with $\Gamma_{\kappa\lambda}^\iota$ the three index Christoffel connection symbols defined by

$$\Gamma_{\kappa\lambda}^\iota = \frac{1}{2} g^{\iota\alpha} \left(\frac{\partial g_{\kappa\alpha}}{\partial x^\lambda} + \frac{\partial g_{\lambda\alpha}}{\partial x^\kappa} - \frac{\partial g_{\kappa\lambda}}{\partial x^\alpha} \right). \quad (8.20)$$

The geodesics are naturally independent of the choice of coordinates but we need these if we want to express geodesics as a set of scalar simultaneous differential equations.

Example 8.1

Suppose we want to find the geodesics of a 2-sphere, which we know beforehand to be great circles. We adopt spherical polar coordinates θ and ϕ to describe the sphere. When each of these is varied in turn, a basis for tangent space $\{g_\theta, g_\phi\}$ is defined, according to Eq. (4.49). The metric tensor has just two non-zero elements:

$$g_{\theta\theta} = r^2, \quad g_{\phi\phi} = r^2 \sin^2(\theta) \quad (8.21)$$

and the geodesic Lagrangian is given by

$$L = \frac{1}{2} r^2 \left[\dot{\theta}^2 + \sin^2(\theta) \dot{\phi}^2 \right] = 1/2. \quad (8.22)$$

Because the Lagrangian is independent of ϕ we conclude that

$$r^2 \sin^2(\theta) \dot{\phi} = J, \quad (8.23)$$

with J constant. Replacing in the Lagrangian we get another equation

$$r^2 \dot{\theta}^2 = 1 - \frac{J^2}{r^2 \sin^2(\theta)}. \quad (8.24)$$

Since the left-hand side is a square we see that we must have $\sin(\theta) > |J|/r$. These equations can be integrated but we don't go into that, because this is not a very practical way of dealing with great circles on the sphere.

Example 8.2

We now consider the case in 3-dimensional space, where the frame vectors are given by

$$g_m(x) = n(x) \sigma_m, \quad (8.25)$$

with $n(x)$ a scalar function of position and $\{\sigma_m\}$ an orthonormed basis in tangent space and not the usual basis of Euclidean flat space. We work here on the embedded surface, without explicit consideration of the higher-dimension embedding space. For an arbitrary path $x(\lambda)$ the distance between two positions is given by

$$s = \int_{\lambda_1}^{\lambda_2} n \left[\sum_{m=1}^3 (\dot{x}^m)^2 \right]^{1/2} d\lambda. \quad (8.26)$$

The geodesic linking points $x(\lambda_1)$ and $x(\lambda_2)$ is the path that makes this length stationary. But now assume that the parameter λ represents time and that $n(x)$ is an optical refractive index and the geodesic equation becomes a statement of Fermat's principle of geometric optics.

We could even extend the principle to non-isotropic materials by redefining the frame vectors, as in

$$g_m = n^o_m \sigma_o, \quad (8.27)$$

producing the metric tensor

$$g_{mn} = \sum_{o=1}^3 n^o_m n^o_n. \quad (8.28)$$

The geodesics of this space would still reproduce the paths followed by light rays. It is useful to compare the equation above with Eq. (3.117), which was written for a constant frame.

In mixed signature spaces the integrand in Eq. (8.2) can have zero or even imaginary values. Extending the geodesic definition to the cases of imaginary integrand presents no problem, but when the integrand is zero we need a new approach. The condition

$$(\dot{x} \cdot \dot{x})^{1/2} = 0 \quad (8.29)$$

defines what is known by a *null path* and the set of all null paths defines a *null surface*, also known as *characteristic surface*. All paths lying on the null surface have zero length, so we cannot use length to define *null geodesics*. If we are working with coordinates, the null path condition can be written as

$$g_{\iota\kappa} \dot{x}^\iota \dot{x}^\kappa = 0. \quad (8.30)$$

Suppose that one frame vector is normal to all the remaining frame vectors; if this particular frame vector is g_σ , we have

$$g_{\sigma\iota} = g_{\sigma\sigma} \delta_{\sigma\iota}. \quad (8.31)$$

We can then rewrite the null path condition as

$$\dot{x}^\sigma = \sqrt{\frac{-1}{g_{\sigma\sigma}}} g_{\iota\kappa} \dot{x}^\iota \dot{x}^\kappa, \quad \iota, \kappa \neq \sigma. \quad (8.32)$$

This allows the definition of a null path length by

$$s = \int_{\lambda_1}^{\lambda_2} \dot{x}^\sigma d\lambda = \int_{\lambda_1}^{\lambda_2} \sqrt{\frac{-1}{g_{\sigma\sigma}}} g_{\iota\kappa} \dot{x}^\iota \dot{x}^\kappa d\lambda, \quad \iota, \kappa \neq \sigma. \quad (8.33)$$

Provided that the right-hand side of the equation is independent from x^σ , the null path length can now be used for the definition of null surface geodesics in the usual way. We will encounter situations where more than one coordinate can be chosen to play the role of x^σ . When this happens the geodesic paths are independent of the choice that is made but the physical interpretation of the resulting equations is influenced by that choice. This fact can prove very useful in providing different perspectives over a given problem of physics.

8.2 The pseudoscalar and projection

The exterior product of the embedded surface's frame vectors $\{g_1, \dots, g_n\}$ defines a pseudoscalar for the surface; the unit pseudoscalar is then given by

$$I = \frac{g_1 \wedge g_2 \wedge \dots \wedge g_n}{|g_1 \wedge g_2 \wedge \dots \wedge g_n|}. \quad (8.34)$$

The pseudoscalar I has the orientation specified by the order of the tangent vectors in the numerator. The pseudoscalar will satisfy $I^2 = \pm 1$, with the sign dependent on the dimension and signature. This definition for the unit pseudoscalar implies that the denominator in the equation is non-zero, which may not happen in certain mixed signature spaces.

The pseudoscalar $I(x)$ is position dependent; this can be clearly understood by applying the definition to the 2-sphere. It contains all the geometric information about the surface and unites its intrinsic and extrinsic properties. We will assume that $I(x)$ is continuous and differentiable over the whole surface, that is, has the same grade everywhere and is single-valued. We will also have to assume that the manifold is orientable; surfaces like the Möbius strip cannot be handled in this way.

Any multivector defined on the embedding space can be projected onto the space tangent to the surface by a projector defined with the help of the pseudoscalar $I(x)$. This is the projector's definition:

$$P(A, x) = A(x) \cdot I(x) I^{-1}(x) = A \cdot I I^{-1}, \quad (8.35)$$

for any multivector A of grade lower than the surface dimension n . Every projector must reproduce itself when operated twice over the argument, which happens with the operator just defined:

$$P^2(A) = P[P(A)] = P(A). \quad (8.36)$$

If applied to a vector a , the projector P selects the component that lies entirely on tangent space at point x . This component is said to be *intrinsic* to the manifold. The complement

$$P_{\perp}(a) = a - P(a), \quad (8.37)$$

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

If $A(x)$ is a multivector field, not necessarily intrinsic to the manifold but defined in some region of the latter, given a vector a in the tangent space, the directional derivative along a is defined as usual

$$a \cdot \nabla A(x) = \lim_{\varepsilon \rightarrow 0} \frac{A(x + \varepsilon a) - A(x)}{\varepsilon}. \quad (8.38)$$

The derivative operator $a \cdot \nabla$ is quite simply the derivative in embedding space, contracted with a vector in tangent space. Replacing a with the tangent frame vectors $\{g_i\}$ we can define a vector derivative ∂ , designated as the *intrinsic derivative* because it is intrinsic to the manifold, by

$$\partial = g^i g_i \cdot \nabla = P(\nabla). \quad (8.39)$$

The definition of ∂ requires the existence of a reciprocal frame $\{g^i\}$ and this is why we need the existence of the pseudoscalar. The projection of ∂ naturally reproduces itself, and the contraction with a tangent vector $a \cdot \partial = a \cdot \nabla$ is the directional derivative along

the direction of a . The concept of monogenic functions is extended to curved manifolds by means of the intrinsic derivative. In a curved manifold, a monogenic function verifies the equation

$$\partial\psi = g^t\partial_t\psi = 0. \quad (8.40)$$

As we shall see, the intrinsic derivative of a function defined on the manifold may lie outside the manifold, but this poses no problem, since here we are dealing with null derivatives, which don't lie anywhere.

8.3 Directed integration for embedded surfaces

We now seek to extend the theory developed in Sec. 7.4 to embedded surfaces and this can be done without many changes. The surface to be integrated over is again triangulated into a series of simplices, the only difference being that now the pseudoscalar varies with position. The relation

$$\oint dL = 0 \quad (8.41)$$

maintains its validity over the closed boundary of a simply connected vector manifold.

The interpolation results used for arriving at Eq. (7.44) are still valid because we have an embedding space; the reciprocal frame required in the equation exists because we required that the pseudoscalar $I(x)$ is well defined. The only change needed is the replacement of the embedding space derivative ∇ by its projection on the manifold. The most general statement of the fundamental theorem is then

$$\oint_{BS} \mathbf{L}(dL) = \int_S \vec{L}(\overleftarrow{\partial} dX) = \int_S \vec{L}(\overleftarrow{\nabla} \cdot dX). \quad (8.42)$$

Of the two forms for the surface integral, the one with ∂ is more useful as it forms a geometric product with the surface element. The linear function \mathbf{L} can be any linear function and it need not be restricted to lie in tangent space. One feature of this general form for the fundamental theorem is not immediately apparent. In effect, if we write $dX = I|dX|$ we recognize that the element dX is position dependent. However this position-dependence is not differentiated in the integral; it is only the integrand that gets differentiated.

The divergence theorem can now be generalized to curved spaces. We write again

$$\mathbf{L}(A) = \langle JAI^{-1} \rangle, \quad (8.43)$$

where J is now a vector field in tangent space and I is the unit pseudoscalar for the n -dimensional curved space. The fundamental theorem now gives

$$\oint_{BS} n \cdot J |dL| = \int_S \left(\partial \cdot J + \langle J \overrightarrow{\partial} \overleftarrow{I}^{-1} I \rangle \right) |dX|, \quad (8.44)$$

where $|dX| = I^{-1}dL$ and $n|dL| = I^{-1}dL$. The final term in the integral must be zero because

$$\langle J \vec{\partial} \overleftarrow{I} I \rangle = \frac{1}{2} \langle J \vec{\partial} (\overleftarrow{I} I + I \overleftarrow{I}) \rangle = \frac{1}{2} \langle J \partial(I^2) \rangle = 0 \quad (8.45)$$

and $I^{-1} = \pm I$. It follows that the divergence theorem in curved space has the same form as in flat space, so

$$\int_S \partial \cdot J |dX| = \oint_{BS} n \cdot J |dL|. \quad (8.46)$$

As an application of the fundamental theorem we can derive Stokes' theorem in three dimensions. Suppose that S is an open, connected surface in three dimensions, with boundary BS . The linear function L takes a vector as its argument and is defined by

$$L(a) = J \cdot a. \quad (8.47)$$

Applying the fundamental theorem we get

$$\oint_{BS} J \cdot dL = \int_S \langle \vec{J} \overleftarrow{\nabla} \cdot dX \rangle = - \int_S (\nabla \wedge J) \cdot dX, \quad (8.48)$$

where we have used the form of the integral theorem involving ∇ because the embedding is specified. The vector n normal to the surface is defined by

$$dX = i n |dX|, \quad (8.49)$$

with i the three-dimensional right-handed unit pseudoscalar. The direction of n depends on the orientation of dX . Around the border we can denote the tangent vector at the boundary by l and the vector pointing into the surface by m . Then dX has the orientation specified by $l \wedge m$ and from Eq. (8.49) we see that $\{l, m, n\}$ must form a right-handed set. This process extends inwards to define n over the whole surface. We now have

$$\oint_{BS} J \cdot dL = \int_S - (i \nabla \wedge J) \cdot n |dX| = \int_S \text{curl}(J) \cdot n |dX|, \quad (8.50)$$

which is the familiar Stokes' theorem in three dimensions.

8.4 Intrinsic and extrinsic geometry

If a is a tangent vector, the directional derivative $a \cdot \partial$ of a tangent vector field results in a vector which may well not lie entirely on tangent space. This effect can be illustrated with the case of a circle in the plane. In this case the derivative of a tangent vector with constant magnitude around the circle is a radial vector, lying entirely outside the circle. We need a new derivative operator which produces results intrinsic to the manifold; this is the *covariant derivative* defined by

$$a \cdot DA(x) = P[a \cdot \partial A(x)]. \quad (8.51)$$

So, the operator aD acts on multivectors in the tangent space, returning a new multivector field which is itself intrinsic to the manifold. The covariant derivative aD verifies Leibniz's rule:

$$a \cdot D(AB) = P[a \cdot \partial(AB)] = (a \cdot DA) + A(a \cdot DB). \quad (8.52)$$

The vector operator D is then defined in the obvious way by

$$D = g^t g_t \cdot D. \quad (8.53)$$

For a grade- r multivector A_r , $r < n$, we can write

$$DA_r = g^t (g_t \cdot DA_r) = P(\partial A_r). \quad (8.54)$$

This is the product of a vector and multivector and the result can be decomposed as

$$\begin{aligned} D \cdot A_r &= \langle DA_r \rangle_{r-1}, \\ D \wedge A_r &= \langle DA_r \rangle_{r+1}. \end{aligned} \quad (8.55)$$

So, like ∇ , the operators ∂ and D have the algebraic properties of vectors. The intrinsic and covariant derivatives produce the same result when acting on a scalar field $\alpha(x)$ defined over the manifold, so

$$\partial \alpha(x) = D \alpha(x). \quad (8.56)$$

Suppose now that we want to know how the unit pseudoscalar I changes along the direction of tangent vector a ; we expect this variation to lie partially outside the manifold. So, let $\{\sigma_t\}$ denote an orthonormal basis intrinsic to the manifold and I be defined by

$$I = \sigma_1 \sigma_2 \cdots \sigma_n = \sigma_{12 \dots n}. \quad (8.57)$$

It follows that

$$\begin{aligned} a \cdot \partial I I^{-1} &= \sum_{t=1}^n \sigma_1 \cdots [a \cdot D \sigma_t + P_{\perp}(a \cdot \partial \sigma_t)] \cdots \sigma_n I^{-1} \\ &= a \cdot D I I^{-1} + P_{\perp}(a \cdot \partial \sigma_t) \wedge \sigma^t. \end{aligned} \quad (8.58)$$

In the first term, $a \cdot D I$ must remain in the tangent space, so it must be a multiple of I . It follows that

$$(a \cdot D I) I = \langle (a \cdot D I) I \rangle = \frac{1}{2} \langle a \cdot D(I^2) \rangle = 0, \quad (8.59)$$

so

$$a \cdot D I = 0. \quad (8.60)$$

We conclude that the unit pseudoscalar is a covariant constant over the manifold. We can then write

$$a \cdot \partial I = P_{\perp}(a \cdot \sigma_t) \wedge \sigma^t I = -S(a) I. \quad (8.61)$$

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

$$a \cdot \partial I = I \times S(a) \quad (8.62)$$

and

$$S(a) \cdot I = S(a) \wedge I = 0. \quad (8.63)$$

The fact that $S(a) \cdot I = 0$ implies that $P[S(a)] = 0$, confirming that $S(a)$ lies partly outside the manifold

The shape tensor $S(a)$ can be seen as an angular momentum of $I(x)$ when this is moved along the manifold; it encodes the change on the orientation of the unit pseudoscalar. The shape tensor also provides a compact relation between intrinsic and covariant derivatives, as we will show next. Start with

$$b \cdot S(a) = b^i P_{\perp}(a \cdot \sigma_i) = P_{\perp}(a \cdot \partial b), \quad (8.64)$$

where both a and b are tangent vectors. Then we have

$$a \cdot \partial b = P(a \cdot \partial b) + P_{\perp}(a \cdot \partial b) = a \cdot Db + b \cdot S(a), \quad (8.65)$$

which can be rearranged to

$$a \cdot Db = a \cdot b + S(a) \cdot b. \quad (8.66)$$

Applying this result to the geometric product bc we get

$$\begin{aligned} a \cdot D(bc) &= (a \cdot \partial b)c + S(a) \cdot bc + b(a \cdot \partial c) + bS(a) \cdot c \\ &= a \cdot \partial(bc) + S(a) \times (bc). \end{aligned} \quad (8.67)$$

Consequently, we have for any multivector taking its values in the tangent space,

$$a \cdot DA = a \cdot \partial A + S(a) \times A. \quad (8.68)$$

Since $S(a)$ is a bivector, the grade of $S(a) \times A$ is the same as that of A . Equation (8.62) is just a special case of this. If we now write

$$a \cdot \partial b = a \cdot \partial P(b) = a \cdot \overrightarrow{\partial} \overleftarrow{P}(b) + P(a \cdot \partial b) = a \cdot \overrightarrow{\partial} \overleftarrow{P}(b) + a \cdot Db, \quad (8.69)$$

we establish further that

$$a \cdot \overrightarrow{\partial} \overleftarrow{P}(b) = b \cdot S(a), \quad (8.70)$$

for any pair of tangent vectors a and b .

8.5 Coordinates and derivatives

The intrinsic and covariant derivatives were introduced in the previous section without recourse to a coordinate system, however in most practical situations we will need to express differential equations as a set of scalar relations, for which we will need to use coordinates. This happened already when we discussed geodesics and will now be extended to the formulation of derivatives. In a region of the manifold we introduce coordinates x^t and define the frame vectors

$$g_t = \frac{\partial x}{\partial x^t}. \quad (8.71)$$

Using the definition of the intrinsic derivative we find that $g^t = \partial x^t$. The frame vectors $\{g_t\}$ are tangent vectors and the reciprocal frame vectors are usually known as *cotangent vectors* or *1-forms*. A global coordinate system may not be possible, as in the case of a 2-sphere; when this happens we patch together a series of local coordinate systems. The covariant derivative along a frame vector, g_t , satisfies

$$g_t \cdot DA = D_t A = g_t \cdot \partial A + S(g_t) \times A = \partial_t A + S_t \times A. \quad (8.72)$$

Besides defining the symbols D_t and S_t , the previous equation shows that ∂_t is more than a conventional representation for a partial derivative; it is also a component of intrinsic derivative vector.

For the tangent frame vectors we can write

$$\partial_t g_\kappa - \partial_\kappa g_t = (\partial_t \partial_\kappa - \partial_\kappa \partial_t)x = 0. \quad (8.73)$$

This relation can be projected onto the manifold, resulting in

$$D_t g_\kappa - D_\kappa g_t = 0. \quad (8.74)$$

It can also be projected out of the manifold, producing

$$g_t \cdot S_\kappa = g_\kappa \cdot S_t. \quad (8.75)$$

Extending to arbitrary tangent vectors a and b we have

$$a \cdot S(b) = b \cdot S(a). \quad (8.76)$$

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

$$S(a) = g^\kappa \wedge P_\perp(a \cdot \partial g_\kappa) \quad (8.77)$$

It follows that

$$S_t = g^\kappa \wedge P_\perp(\partial_t g_\kappa) = g^\kappa \wedge P_\perp(\partial_\kappa g_t). \quad (8.78)$$

The tangent frame vectors therefore satisfy

$$\partial \wedge g_\iota = g^\kappa \wedge [P(\partial_\kappa g_\iota) + P_\perp(\partial_\kappa g_\iota)] = D \wedge g_\iota + S_\iota. \quad (8.79)$$

Now for a general vector, we can make the decomposition $a = a^\iota g_\iota$ and write generally

$$\partial \wedge a = D \wedge a + S(a). \quad (8.80)$$

A possible interpretation of the shape tensor is that it picks up the component of the curl of a tangent vector lying outside the manifold. Because we can write

$$\partial \wedge a = \partial \wedge [P(a)] \vec{\partial} \wedge \overleftarrow{P}(a) + P(\partial \wedge a) = D \wedge a + \vec{\partial} \wedge \overleftarrow{P}(a) + P(\partial \wedge a), \quad (8.81)$$

it also true that

$$\vec{\partial} \wedge \overleftarrow{P}(a) + P(\partial \wedge a) = S(a), \quad (8.82)$$

yet another perspective over the shape tensor.

If the previous result is applied to the gradient of a scalar we get

$$\partial \wedge \partial \phi = P(\nabla) \wedge P(\nabla \phi) = P(\nabla \wedge \nabla \phi) + \vec{\partial} \wedge \overleftarrow{P}(\nabla \phi). \quad (8.83)$$

The embedding space derivative, however, satisfies $\nabla \wedge \nabla = 0$, so we have

$$D \wedge (D\phi) = 0. \quad (8.84)$$

This result can be applied to the coordinate scalars and we see that

$$D \wedge (Dx^\iota) = D \wedge g^\iota = 0, \quad (8.85)$$

which is consistent with Eq. (8.74). Applying the result to a general vector $a = a_\iota g^\iota$ we find that

$$D \wedge a = D \wedge (a_\kappa g^\kappa) = g^\iota \wedge g^\kappa (\partial_\iota a_\kappa) = \frac{1}{2} g^\iota \wedge g^\kappa (\partial_\iota a_\kappa - \partial_\kappa a_\iota). \quad (8.86)$$

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

8.6 Riemannian geometry

In the previous sections we used the embedding space to define various concepts intrinsic to the manifold. Although this procedure simplifies the exposure, it is seldom useable in practice, as we have already seen in the example of Fermat's principle in page 147. In the majority of practical cases one has to work entirely in the manifold and this is the subject of Riemannian geometry. Most textbooks characterize the manifold by means of the metric tensor, which we express in terms of the $\{g_\iota\}$ coordinate frame as

$$g_{\iota\kappa} = g_\iota \cdot g_\kappa. \quad (8.87)$$

The derivations that follow are valid irrespective of the signature, although strictly the designation Riemannian geometry should only apply to spaces with Euclidean signature. Applications in physics usually require spaces with Minkowskian signature and these will thus be of greatest interest for us.

In addition to the metric tensor, we need to know how the frame vectors change with position and we need to do this without resorting to their expression in terms of an embedding space frame. The Christoffel connection symbols that were introduced in Sec. 8.1 give us that information. We will now give an alternative definition for those symbols and then we will show that it is equivalent to the previous one. The new definition is

$$\Gamma_{\kappa\lambda}^{\iota} = (D_{\kappa}g_{\lambda}) \cdot g^{\iota}. \quad (8.88)$$

The connection symbols give us the components of the covariant derivative applied to the frame vectors, that is, they tell us precisely how the frame vectors change with position. A connection on its own does not imply that a space is curved; an example is given by spherical polar coordinates, whose coordinate vectors change with position in spite of the space being flat.

Our earlier definition of the connection symbols was made directly from the metric tensor; to show that the two definitions are equivalent let us first show that the $\Gamma_{\kappa\lambda}^{\iota}$ symbols are symmetric in the lower indices. This comes from

$$\Gamma_{\kappa\lambda}^{\iota} - \Gamma_{\lambda\kappa}^{\iota} = (D_{\kappa}g_{\lambda} - D_{\lambda}g_{\kappa}) \cdot g^{\iota} = 0, \quad (8.89)$$

by virtue of Eq. (8.74). We also need the curl of a coordinate vector,

$$D \wedge g_{\iota} = D \wedge (g_{\iota\kappa}g^{\kappa}) = (Dg_{\iota\kappa}) \wedge g^{\kappa}. \quad (8.90)$$

This allows us to write

$$\begin{aligned} \Gamma_{\kappa\lambda}^{\iota} &= \frac{1}{2} g^{\iota} \cdot (D_{\kappa}g_{\lambda} + D_{\lambda}g_{\kappa}) \\ &= \frac{1}{2} g^{\iota} \cdot [g_{\kappa} \cdot (Dg_{\lambda\alpha} \wedge g^{\alpha}) + g_{\lambda} \cdot (Dg_{\kappa\alpha} \wedge g^{\alpha}) + Dg_{\kappa\lambda}] \\ &= \frac{1}{2} g^{\iota} \cdot (\partial_{\kappa}g_{\lambda\alpha}g^{\alpha} + \partial_{\lambda}g_{\kappa\alpha}g^{\alpha} - Dg_{\kappa\lambda}) \\ &= \frac{1}{2} g^{\iota\alpha} (\partial_{\kappa}g_{\lambda\alpha} + \partial_{\lambda}g_{\kappa\alpha} - \partial_{\alpha}g_{\kappa\lambda}), \end{aligned} \quad (8.91)$$

which recovers Eq. (8.20).

Example 8.3

We may want to calculate the Christoffel connection symbols for the the 2-sphere, using

polar coordinates. In practice it is frequently easier to resort to Eq. (8.19) than to the definition.

Using the Lagrangian from example 6.1 we derive the two Euler-Lagrange equations

$$\begin{aligned}\ddot{\theta} - \sin(\theta) \cos(\theta) \dot{\phi}^2 &= 0, \\ \ddot{\phi} + 2\cot(\theta) \dot{\theta} \dot{\phi} &= 0.\end{aligned}\tag{8.92}$$

Comparing with Eq. (8.19) we realize immediately that the only two non-zero components are

$$\begin{aligned}\Gamma_{\phi\phi}^{\theta} &= -\sin(\theta) \cos(\theta), \\ \Gamma_{\theta\phi}^{\phi} &= \Gamma_{\phi\theta}^{\phi} = \cot(\theta).\end{aligned}\tag{8.93}$$

Example 8.4

We now look at the connection symbols behind the geometry of Fermat's principle used in geometrical optics. For this we recover the coordinate frame from example 6.2 and write the metric tensor components

$$g_{mn} = n^2 \text{diag}(1, 1, 1).\tag{8.94}$$

Going directly to the definition we see that all the indices ι, κ, λ and α must be equal for the second member to be non-zero. We then write the non-zero symbols as

$$\Gamma_{mm}^m = \frac{1}{2n} \partial_m n.\tag{8.95}$$

Euler-Lagrange equations can now be written in a form more appropriate to geometric algebra formalism. We will allow the Lagrangian to change along a geodesic, which essentially means we can use a non-affine parameter to describe the geodesic; this is useful if we want the Lagrangian to represent the total energy of a physical system. Returning to Eq. (8.7) we multiply the Lagrangian by a factor of 2, which just amounts to including another factor of 2 in the parameter λ , leaving the Euler-Lagrange equations unaltered. The Lagrangian then becomes

$$L = g_{\iota\kappa} \dot{x}^{\iota} \dot{x}^{\kappa} = (g_{\iota} \dot{x}^{\iota}) \cdot (g_{\kappa} \dot{x}^{\kappa}).\tag{8.96}$$

We can then define two vectors, p and \dot{x} , by the equations

$$p = p_{\iota} g^{\iota} = (p_{\iota} g^{\iota\kappa}) (g_{\kappa\iota} g^{\iota}) = g_{\kappa} \dot{x}^{\kappa},\tag{8.97}$$

$$\dot{x} = g_{\iota} \dot{x}^{\iota}.\tag{8.98}$$

The two vectors are one and the same, but they are expressed in the reciprocal and direct frames, respectively. The coordinates of vector p are usually known as *generalized momenta*. The Lagrangian now becomes simply

$$L = p \cdot \dot{x} = p_\iota \dot{x}^\iota. \quad (8.99)$$

Euler-Lagrange equations (8.14) can now be written in a coordinate-free form as

$$\frac{d}{d\lambda} \left(\frac{1}{\sqrt{L}} p \right) = \frac{1}{\sqrt{L}} \partial p \cdot \dot{x}. \quad (8.100)$$

This form of the equations is truly coordinate-free; notice, in particular, that it uses the intrinsic derivative ∂ and not the covariant derivative D , which avoids the recourse to connection symbols.

We have seen that the existence of a connection does not imply a curvature of the manifold; therefore we need a method for encoding the intrinsic curvature of the manifold. Suppose we form the commutator of the two derivatives

$$\begin{aligned} [D_\iota, D_\kappa]A &= \partial_\iota(\partial_\kappa A + S_\kappa \times A) + S_\iota \times (\partial_\kappa A + S_\kappa \times A) \\ &\quad - \partial_\kappa(\partial_\iota A + S_\iota \times A) + S_\kappa \times (\partial_\iota A + S_\iota \times A) \\ &= (\partial_\iota S_\kappa - \partial_\kappa S_\iota) \times A + (S_\iota \times S_\kappa) \times A, \end{aligned} \quad (8.101)$$

where we made use of the Jacobi identity of Eq. (3.58). All the derivatives of multivector A have disappeared from the equation. The result can be further simplified by noting that

$$\begin{aligned} (\partial_\iota S_\kappa - \partial_\kappa S_\iota) &= -\partial_\iota(\partial_\kappa I I^{-1}) + \partial_\kappa(\partial_\iota I I^{-1}) \\ &= -S_\kappa I S_\iota I^{-1} + S_\iota I S_\kappa I^{-1} \\ &= -2S_\iota \times S_\kappa. \end{aligned} \quad (8.102)$$

The anti-commutativity between $S(a)$ and I was used in the derivation. Substituting this expression into the previous equation we obtain the simple result

$$[D_\iota, D_\kappa]A = -(S_\iota \times S_\kappa) \times A. \quad (8.103)$$

The commutator of covariant derivatives defines the *Riemann tensor*, denoted by $R(g_\iota \wedge g_\kappa)$, such that

$$R(g_\iota \wedge g_\kappa) \times A = [D_\iota, D_\kappa]A. \quad (8.104)$$

This is a bivector-valued function of its bivector argument.

The Riemann tensor is usually expressed in terms of 4-index components. To achieve this we first write

$$R(g_\iota \wedge g_\kappa) \cdot g_\lambda = [D_\iota, D_\kappa] \cdot g_\lambda = D_\iota(\Gamma_{\kappa\lambda}^\alpha g_\alpha) - D_\kappa(\Gamma_{\iota\lambda}^\alpha g_\alpha). \quad (8.105)$$

So, we get the components

$$\begin{aligned} R_{\iota\kappa\lambda}{}^\alpha &= R(g_\iota \wedge g_\kappa) \cdot (g_\lambda \wedge g^\alpha) \\ &= \partial_\iota \Gamma_{\kappa\lambda}^\alpha - \partial_\kappa \Gamma_{\iota\lambda}^\alpha + \Gamma_{\kappa\lambda}^\beta \Gamma_{\iota\beta}^\alpha - \Gamma_{\iota\lambda}^\beta \Gamma_{\kappa\beta}^\alpha, \end{aligned} \quad (8.106)$$

recovering the standard definition of Riemannian geometry. The Riemann tensor has n^4 components but the many symmetries reduce substantially the number of independent components. Because $R(a \wedge b)$ is a bivector-valued linear function of a bivector, the number of degrees of freedom is immediately reduced to $n^2(n-1)^2/4$.

A further symmetry of the Riemann tensor can be found by

$$\begin{aligned} R(g_\iota \wedge g_\kappa) \cdot g_\lambda &= D_\iota D_\kappa g_\lambda - D_\kappa D_\iota g_\lambda \\ &= D_\iota D_\lambda g_\kappa - D_\kappa D_\lambda g_\iota \\ &= [D_\iota, D_\lambda] g_\kappa - [D_\kappa, D_\lambda] g_\iota + D_\lambda (D_\iota g_\kappa - D_\kappa g_\iota) \\ &= R(g_\iota \wedge g_\lambda) \cdot g_\kappa - R(g_\kappa \wedge g_\lambda) \cdot g_\iota. \end{aligned} \quad (8.107)$$

Consequently

$$a \cdot R(b \wedge c) + c \cdot R(a \wedge b) + b \cdot R(c \wedge a) = 0, \quad (8.108)$$

for any three vectors a , b and c , belonging to tangent space. This equation tells us that there is a vector quantity that vanishes for all trivectors $a \wedge b \wedge c$, which provides a set of $n^2(n-1)(n-2)/6$ scalar equations. The number of independent degrees of freedom is then given by the formula

$$\frac{1}{4} n^2(n-1)^2 - \frac{1}{6} n^2(n-1)(n-2) = \frac{1}{12} n^2(n^2-1). \quad (8.109)$$

This gives 1, 6, 20 and 50 independent components for 2, 3, 4 and 5 dimensions, respectively.

Most directed integration theorems involve elements extrinsic to the manifold; the most significant theorem involving only intrinsic quantities is a generalization of Stokes' theorem, applicable to a grade- r multivector and an open surface S of dimension $r+1$. In this case we have

$$\oint_B S A_r \cdot dL = \int_S (\vec{A}_r \wedge \overleftarrow{\partial} \cdot X = (-1)^r \int_S (D \wedge A_r) \cdot dX. \quad (8.110)$$

A particular case is when $r = n-1$, which recovers the divergence theorem.

9 Electromagnetism

The choice of electromagnetism as the field of physics where we first will apply the consequences of geometric relations in 5-dimensional spacetime is mainly justified by two considerations. First of all, this is a field with extremely well established concepts and familiar to most people; our task will be to show that electromagnetic relations are indeed rooted in geometry. The second consideration is the comparative mathematical simplicity of electromagnetism compared to gravitational dynamics. This is contrary to classical thinking but the reader will probably find it acceptable if he realizes that, classically, dynamics is rooted in geometry by means of the general relativity theory, considerably more complex than electromagnetism. Electromagnetism and geometry have already been placed in contact by the works of Kaluza, later complemented by Klein, in what is usually known as Kaluza-Klein theory.

What we will do in this chapter is similar in many ways to the Kaluza approach, the main difference being the point of departure. We will accept that the 5-dimensional spacetime manifold can exhibit curvature, this curvature being caused by sources. In the case of electromagnetism this statement will lead us to Maxwell's equations, because we impose restrictions to the allowed curvature. Later in the book we will relax the restrictions on curvature to find that gravitation is indeed just a different manifestation of the same foundational principle as electromagnetism.

9.1 Vacuum equations

Virtually every textbook on electrodynamics starts with an introductory chapter presenting Maxwell's equations in vectorial form; the notation used may vary but in essence there will always be two divergence equations

$$\operatorname{div} \mathbf{D} = \rho, \quad \operatorname{div} \mathbf{B} = 0, \quad (9.1)$$

where \mathbf{D} is the electric displacement vector, ρ is electric charge density and \mathbf{B} the magnetic induction vector. There will also be two curl equations

$$\operatorname{curl} \mathbf{E} = -\frac{\partial}{\partial t} \mathbf{B}, \quad \operatorname{curl} \mathbf{H} = \frac{\partial}{\partial t} \mathbf{D} + \mathbf{j}, \quad (9.2)$$

with \mathbf{E} the electric field vector, \mathbf{H} the magnetic field vector and \mathbf{j} the electric current vector. Finally there will be two equations characterizing the medium

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P}, \quad \mathbf{H} = \frac{1}{\mu_0} \mathbf{B} - \mathbf{M}, \quad (9.3)$$

with ε_0 the electrical permittivity of vacuum, \mathbf{P} the polarization vector, μ_0 the permeability of vacuum and \mathbf{M} the magnetization vector. Note that we introduced boldface characters to represent conventional 3-dimensional vectors, mainly because we are not using geometric algebra equations; we will revert to our usual conventions shortly.

Our aim in this chapter is more ambitious than to rewrite Maxwell's equations in the formalism of geometric algebra; we want to derive all the equations of electromagnetism and electrodynamics from a principle that can later be generalized to gravitation and quantum mechanics. We will accept that space can have curvature induced by sources and we will study monogenic functions in such curved spaces. In the case of electromagnetism we associate coordinate x^0 to time and coordinates x^m to space; we also define the potential vector $A = A_\mu \sigma^\mu$ and allow curvature to be defined by the frame vectors

$$g_\mu = \sigma_\mu + A_\mu \sigma_4, \quad g_4 = \sigma_4, \quad (9.4)$$

where A is dependent on coordinates x^μ but not on coordinate x^4 . This is in direct relation with the reciprocal frame

$$g^\mu = \sigma^\mu, \quad g^4 = \sigma^4 - A = \sigma^4 - A_\mu \sigma^\mu. \quad (9.5)$$

What we have just done is equivalent to Kaluza's approach with null scalar field [4, 7] and no gravity field. In fact evaluating the metric tensor $g_{\alpha\beta} = g_\alpha \cdot g_\beta$ we get

$$g_{\alpha\beta} = \begin{pmatrix} \sigma_\mu \cdot \sigma_\nu + A_\mu A_\nu & A_\mu \\ A_\mu & 1 \end{pmatrix}. \quad (9.6)$$

The frame is dependent on x^μ through A but does not depend on x^4 . The frame independence on x^4 in Kaluza's theory was one of the reasons for the modifications introduced by Klein [4], assuming that x^4 was a compact coordinate; we don't need to go into those subtleties in this chapter, for we will only be concerned with macroscopic electromagnetism.

In Eq. (8.85) it was established that the exterior derivative of reciprocal frame vectors must be zero, so we must have

$$D \wedge g^4 = 0. \quad (9.7)$$

Inserting the definitions (9.5) we get

$$\begin{aligned} D \wedge \sigma^\mu &= 0, \\ D \wedge \sigma^4 + D \wedge A &= 0. \end{aligned} \quad (9.8)$$

The first equation tells us that the four σ^μ vectors are invariant under covariant transformations. Now we assume that the potential vector is independent from coordinate x^4 . Using Eq. (8.86) we write

$$\frac{1}{2} \sigma^{\mu\nu} (\partial_\mu A_\nu - \partial_\nu A_\mu) = -D \wedge \sigma^4. \quad (9.9)$$

This can be rewritten in a more conventional and compact form if we use the fact that A is independent from x^4 ; it is

$$\mu\nabla \wedge A = \nabla \wedge A = F, \quad (9.10)$$

where F is called the *Faraday bivector*, defined as

$$F = -2\nabla \wedge \sigma^4. \quad (9.11)$$

We are free to establish anything we want about the divergence of the potential vector but the simplest assumption is that of zero divergence, that is,

$$\nabla A = \nabla \cdot A + \nabla \wedge A = \nabla \wedge A. \quad (9.12)$$

Under this assumption Eq. (9.10) can be written in the simpler form

$$\nabla A = F. \quad (9.13)$$

This equation relates space curvature induced by the vector potential to the Faraday bivector which incorporates both electric and magnetic fields, as we shall soon see.

We know that the Faraday bivector is a general bivector lying on the 4×4 -space spanned by $\sigma^{\mu\nu}$; it is possible to encode the usual three-dimensional electromagnetic field vectors, \mathbf{E} and \mathbf{B} , in F by

$$F = \sigma^0 \mathbf{E} - i\sigma^{04} \mathbf{B}. \quad (9.14)$$

In this way we make sure that the electric and magnetic field components are always kept separate. If now we take the derivative of the Faraday bivector, $\nabla F = \nabla^2 A$, it must return a vector, because ∇^2 is a Laplacian operator that is grade preserving. We then write

$$\nabla F = \nabla^2 A = J, \quad (9.15)$$

where J is called the *current vector* and encodes charge and current densities, as in

$$J = -\rho\sigma^0 + \mathbf{j}. \quad (9.16)$$

Equation (9.15) is the most compact form of Maxwell's equations in vacuum that can be written but it incorporates all the features of the traditional 3-dimensional vector version.

Equation (9.15) can now be split into its vector and trivector parts

$$\nabla \cdot F = J, \quad (9.17)$$

$$\nabla \wedge F = 0, \quad (9.18)$$

which we will examine separately. If we consider the Faraday bivector split in Eq. (9.14), the divergence equation produces

$$\begin{aligned}\nabla \cdot F &= -\partial_0 \mathbf{E} - {}^m\nabla \cdot \mathbf{E} \sigma^0 \\ &+ (\partial_2 B^3 - \partial_3 B^2) \sigma^1 + (\partial_3 B^1 - \partial_1 B^3) \sigma^2 + (\partial_1 B^2 - \partial_2 B^1) \sigma^3 = J.\end{aligned}\quad (9.19)$$

This is equivalent to the two 3-dimensional equations

$$\operatorname{div} \mathbf{E} = \rho, \quad \operatorname{curl} \mathbf{B} = \partial_0 \mathbf{E} + \mathbf{j}, \quad (9.20)$$

which we will immediately recognize as two of Maxwell's equations for vacuum, as long as coordinate x^0 is assigned to time.

Taking now Eq. (9.18) and inserting the Faraday bivector split we get

$$\nabla \wedge F = -\sigma^0 {}^m\nabla \wedge \mathbf{E} + \mathrm{i}\sigma^4 \partial_0 \mathbf{B} - \mathrm{i}\sigma^{04} {}^m\nabla \cdot \mathbf{B} = 0. \quad (9.21)$$

From this we derive the two conditions

$$\begin{aligned}{}^m\nabla \cdot \mathbf{B} &= 0, \\ \sigma^0 {}^m\nabla \wedge \mathbf{E} &= \mathrm{i}\sigma^4 \partial_0 \mathbf{B}.\end{aligned}\quad (9.22)$$

The second equation can be multiplied on the left by $\mathrm{i}\sigma^4$ to result in

$$\mathrm{i}\sigma^{04} {}^m\nabla \wedge \mathbf{E} = -\partial_0 \mathbf{B}. \quad (9.23)$$

This equation and the first equation in (9.22) produce the remaining two 3-dimensional vector equations

$$\operatorname{div} \mathbf{B} = 0, \quad \operatorname{curl} \mathbf{E} = -\partial_t \mathbf{B}. \quad (9.24)$$

The set formed by Eqs. (9.20) and (9.24) is indeed the set of Maxwell's equations applicable to vacuum, since in non-dimensional units $\varepsilon_0 = \mu_0 = 1$.

Charge conservation is obtained from Eq. (9.15) by left multiplication by ∇ ; this amounts to applying a Laplacian to the Faraday bivector $\nabla^2 F$ and the result must contain only bivector terms; so

$$\nabla \cdot J = 0. \quad (9.25)$$

In the more conventional 3D notation we would write

$$\partial_t \rho + \operatorname{div} \mathbf{j} = 0; \quad (9.26)$$

that is, total charge must be conserved.

In the absence of electric charges and currents we must have $J = 0$ and Eq. (9.15) exhibits solutions in the form of plane waves; this becomes apparent if the vector derivative is split into its time and space parts

$$(\sigma^0 \partial_0 + \sigma^m \partial_m) F = 0. \quad (9.27)$$

We try solutions of the form

$$F = F_0 e^{\pm i h (\omega x^0 \pm k_m x^m + \theta)}, \quad (9.28)$$

where h is a unitary hermitian multivector. Upon closer examination we see that F_0 gets multiplied by a scalar term and a term containing ih ; since the product $F_0 ih$ must be expandable into electric and magnetic field bivectors, the only hermitian element that meets the requirement is $h = \sigma^4$. When inserted into Eq. (9.27) this solution produces

$$(\omega \sigma^0 \pm k_m \sigma^m) F_0 i \sigma^4 e^{\pm i h (\omega x^0 \pm k_m x^m + \theta)} = 0. \quad (9.29)$$

For this to be verified F_0 must contain a nilpotent vector factor equal to the one in parenthesis, that is

$$\sum_m (k_m)^2 = \omega^2. \quad (9.30)$$

For simplicity we can assume propagation along σ_3 , thus making $k_1 = k_2 = 0$; we can also zero the phase angle θ by proper choice of the time origin. With these simplifications the Faraday bivector becomes

$$F = \omega (\sigma^0 \pm \sigma^3) n e^{\pm i \omega \sigma^4 (x^0 \pm x^3)}, \quad (9.31)$$

with n a vector on the σ^{12} plane, so that F acquires a bivector character; without loss of generality we can set $n = \sigma^1$. The exponential produces a rotation on the σ^{12} plane, which affects both electric and magnetic field; this can be verified as follows

$$\begin{aligned} (\sigma^{01} + \sigma^{31}) e^{\pm i \sigma^4 \alpha} &= (\sigma^{01} + \sigma^{31}) (\cos \alpha \pm i \sigma^4 \sin \alpha) \\ &= (\cos \alpha \sigma^{01} \mp \sin \alpha \sigma^{02}) + (\cos \alpha \sigma^{31} \pm \sin \alpha \sigma^{23}); \end{aligned} \quad (9.32)$$

both parenthesis are bivectors that rotate with α . Referring to the overall sign in the exponent, we call left/right rotation to the $+/-$ signs, respectively. The $+/-$ signs within the parenthesis in the exponent obviously refer to backward/forward propagating waves.

Considering that Faraday bivector is the derivative of the vector potential, Eq. (9.13), a circularly polarized electromagnetic wave is associated with a circularly polarized vector potential wave, whose 0th derivative produces the electric field while the 3rd derivative produces the magnetic field. This affects the frame vectors defined in Eqs. (9.4) and (9.5). In the reciprocal frame it is g^4 that spins around σ^4 on the σ^{12} plane; the direct frame has both g_1 and g_2 affected, sharing a σ_4 component which is continuously passed between them from one to the other and back.

9.2 Electrodynamics

In this section we study the dynamics of a charged point particle. The concept of a point charge is the same as usually accepted in physics literature, although it proves somewhat inconsistent with our geometric approach. This is because we would like to incorporate the charge as a deviation from the geometry of vacuum; for the time being, however, we will have to use this less than perfect approach.

For a particle with charge q and mass m we will modify the frame vectors multiplying the field vector A by q/m . We will also make use of the fact that a particle's wavefunction verifies the monogenic condition, as already established in Refs. [8, 9]; this is particularly applicable to the dynamics of an electron. We recall that the electron's electric charge in non-dimensional units is $q \approx -0.302822$ and its mass is $m \approx 1.049038 \times 10^{-22}$. The definition of a monogenic function must now use the intrinsic derivative and a particle's wavefunction must verify.

$$\partial\psi = \left[\sigma^\mu \partial_\mu + \left(\sigma^4 - \frac{q}{m} A \right) \partial_4 \right] \psi = 0. \quad (9.33)$$

In Eq. (6.19) we attributed physical meaning to the partial derivatives as follows:

- energy is associated with 0th derivative by $\partial_0\psi = iE\psi$,
- momentum is associated with the 3 spatial derivatives by $\sigma^m \partial_m \psi = i\mathbf{p}\psi$,
- finally rest mass is associated with the fourth derivative by $\partial_4\psi = im\psi$.

If we call p to the 5D vector defined by $p = E\sigma^0 + \mathbf{p} + m\sigma^4 - qA$, the monogenic condition implies that it is a nilpotent vector. The particle's dynamics is determined by the evolution of vector p ; since all frame vectors are independent from x^4 , this is the best choice for an affine parameter. We then define a Lagrangian by the equation

$$L = (E\sigma^0 + \mathbf{p} - qA)^2 = m^2. \quad (9.34)$$

Using Euler-Lagrange equation (8.100) to find the evolution of p we get

$$\dot{E}\sigma^0 + \dot{\mathbf{p}} = -q\partial A \cdot \dot{x}, \quad (9.35)$$

where we used an acute accent to denote total derivative with respect to x^4 and made $\dot{x} = \sigma_\mu \dot{x}^\mu$. Considering Eq. (9.12)

$$\dot{E}\sigma^0 + \dot{\mathbf{p}} = -q\dot{x} \cdot F. \quad (9.36)$$

The expansion of the right hand side inner product produces

$$\dot{x} \cdot F = -\dot{x}^0 \mathbf{E} + \sigma^{123} \dot{\mathbf{x}} \wedge \mathbf{B} + \dot{\mathbf{x}} \cdot \mathbf{E} \sigma^0; \quad (9.37)$$

and we can equate

$$\dot{\mathbf{p}} = q(\dot{x}^0 \mathbf{E} - \sigma^{123} \dot{\mathbf{x}} \wedge \mathbf{B}), \quad (9.38)$$

$$\dot{E} = q \dot{\mathbf{x}} \cdot \mathbf{E}. \quad (9.39)$$

The first equation in this pair is the relativistic Lorentz force and the second one is a relativistic effect which accounts for the increase in energy when the electric field is partially aligned with the particle's velocity.

9.3 Electromagnetism in media

The usual approach to the formulation of electromagnetic equations in media considers polarization and magnetization vectors which are functions of the applied electric and magnetic fields respectively. In first order approximation one takes this dependence to be linear, resulting in the definition of proportionality constants generically called susceptibility, electric or magnetic respectively. In our geometric approach we must carry the medium influence to the frame vectors, if we want to be consistent. We introduce a linear function \mathbf{X} , called susceptibility, incorporating the perturbation of the reciprocal frame vectors, by

$$g^\mu = \sigma^\mu + \mathbf{X}(\sigma^\mu). \quad (9.40)$$

The exterior derivative of A must now be written with the covariant derivative and can be split into two terms, as in

$$D \wedge (A) = \nabla \wedge A + \mathbf{X}(\sigma^\mu) \partial_\mu \wedge A = F + P; \quad (9.41)$$

with F the Faraday bivector and P the polarization bivector.

The susceptibility is a linear function that will transform any vector into another vector, but it is convenient to consider two special cases: electrical susceptibility, $\mathbf{X}(\sigma^0)$ and magnetic susceptibility, $\{\mathbf{X}(\sigma^m)\}$. The transformation of a vector into another vector is always the result of a rotation and a dilation, so, we can write electrical and magnetic susceptibilities as

$$\begin{aligned} \mathbf{X}(\sigma^0) &= \varepsilon \Omega \sigma^0 \tilde{\Omega}, \\ \mathbf{X}(\sigma^m) &= \frac{1}{\mu} \Omega \sigma^m \tilde{\Omega}, \end{aligned} \quad (9.42)$$

where ε is the electrical permittivity, μ is the magnetic permeability and Ω is a rotor. In the most general cases the magnetic permeability and the associated rotor can be different for the 3 frame vectors σ^m ; the simplest cases, however, are those of isotropic materials, for which there is only direction independent dilation. For isotropic materials we need consider only position dependent ε and μ . We note that the vector potential A ,

affecting g^4 can also be seen as a manifestation of the susceptibility function acting on σ^4 .

Electromagnetic equations are obtained from the following generalization of Eq. (9.10):

$$\nabla(F + P) = J_A + J_D. \quad (9.43)$$

In this equation J_A is the previous current vector, incorporating charge density and applied currents, while J_D is a new displacement current which incorporates charge density and currents generated in the media by the applied field. Explicitly the displacement current is given by

$$J_D = -X(\sigma^\mu)\partial_\mu(F + P). \quad (9.44)$$

The extension of Eq. (9.17) is trivial but for Eq. (9.18) we note that because $\nabla \wedge F$ is nilpotent so must also be $\nabla \wedge P$. In a similar way to what we did for the Faraday bivector, we can separate the polarization bivector into its electric and magnetic parts

$$P = \mathbf{P}\sigma^0 - \sigma^{123}\mathbf{M}; \quad (9.45)$$

the minus sign on the right hand side is needed for consistency with the traditional form of Maxwell's equations. Separating Eq. (9.43) into its four equivalent 3D equations we obtain

$$\nabla \cdot (\mathbf{E} + \mathbf{P}) = (J_A + J_D)_0 = \rho; \quad (9.46)$$

$$\nabla \cdot \mathbf{B} = 0; \quad \nabla \cdot \mathbf{M} = 0; \quad (9.47)$$

$$\nabla \wedge \mathbf{E} - \sigma^{123}\partial_0\mathbf{B} = 0; \quad \nabla \wedge \mathbf{P} - \sigma^{123}\partial_0\mathbf{M} = 0; \quad (9.48)$$

$$\partial_0(\mathbf{E} + \mathbf{P}) + \sigma^{123}\nabla \wedge (\mathbf{B} - \mathbf{M}) = (J_A + J_D)_m \sigma^m = \mathbf{j}. \quad (9.49)$$

This set of equations is equivalent to Maxwell's equations in their usual 3D form, but we get two extra equations resulting from $\nabla \wedge P = 0$. These indicate that the media has the ability to support lattice waves, which we designate by phonons.

9.4 The field of a moving charge

A moving charge is fixed in its own frame, so it produces no magnetic field and the potential vector must be directed along coordinate x'^0 . The field vector of a single point charge in its own frame can then be expressed by

$$A = \frac{q}{4\pi d} \sigma'^0. \quad (9.50)$$

We use primed coordinates and frame vectors to refer to the charge's frame, in order to distinguish them from the lab's frame; d is the distance from the point charge's location to the point where the field is evaluated. This formulation for the field vector is

unsuitable for a geometric transformation into the lab frame's field due to the presence of distance d ; we must therefore replace this with a vector formulation, which will then be transformable between frames.

We are here dealing with electromagnetic interaction, which is mediated by photons, travelling in 5D spacetime along constant x^4 lines. Because 5D displacements are nilpotent vectors, we have for photons $(dx'^0)^2 + \sum_m (dx'^m)^2 = 0$. It follows that a photon travelling between the charge and a point at distance d has a total displacement given by the vector

$$X = d(\sigma'_0 + \sigma'_r), \quad (9.51)$$

where σ'_r represents the unit vector in the direction linking the two positions, normal to both σ'_0 and σ'_4 . We can then obtain the distance by the inner product $\sigma'^0 \cdot X$; this is a vector formula perfectly suitable for frame transformations. The field vector is then

$$A = \frac{q}{\sigma'^0 \cdot X} \sigma'^0. \quad (9.52)$$

The particle's frame vector σ'^0 can be expressed in terms of its lab frame components because we know it is the result of a Lorentz transformation applied to σ^0 . The Lorentz transformation is described by the rotor

$$L = e^{B\alpha/2}, \quad (9.53)$$

with $B = \sigma^0 n$ a unitary bivector on a plane containing σ^0 . The transformation of a generic vector a is achieved by the formula

$$a \rightarrow La\tilde{L}. \quad (9.54)$$

Applying this to σ^0 we see that σ'^0 is

$$\sigma'^0 = \cosh(\alpha)\sigma^0 + \sinh(\alpha)n = \nu; \quad (9.55)$$

ν is the well known relativistic 4-velocity; this is to be expected, since the particle is stationary in its own frame, that is, the 4-velocity coincides with σ'^0 . The transformed field vector is known as Liénard-Wiechert potential [10, 11]

$$A = \frac{q}{4\pi} \frac{\nu}{\nu \cdot X}. \quad (9.56)$$

In order to find the Faraday bivector associated with the potential vector above we will need to evaluate the derivative of X ; for this purpose we write X as the difference between a fixed lab position, x and the particle's position parameterized by x^4 , $p(x^4)$:

$$X = x - p(x^4). \quad (9.57)$$

Now it is easy to verify that 4-velocity is given by

$$\partial_4 p = v, \quad (9.58)$$

so, we can write for X

$$\nabla X = \nabla x - \nabla x^4 v. \quad (9.59)$$

Now we need the gradient of coordinate x^4 , but we can find it by evaluating the gradient of $X^2 = 0$; it is then

$$0 = \nabla X \cdot X = \nabla x \cdot X - \nabla x^4 (v \cdot X). \quad (9.60)$$

It follows that

$$\nabla x^4 = \frac{X}{X \cdot v}. \quad (9.61)$$

The gradient of x^4 is directed along X , but X is itself directed along x^4 . When we introduced the gradient we called to the reader's attention that in spaces of mixed signature it was not necessarily directed along the direction of steepest increase and this is just one such situation.

Now we need to concentrate on the derivative of Xv , which appears in the denominator of the potential vector. We find that

$$\begin{aligned} \nabla(X \cdot v) &= \overrightarrow{\nabla} \overleftarrow{X} \cdot v + \nabla x^4 X \cdot (\partial_4 v) \\ &= v - \nabla x^4 + \nabla x^4 X \cdot \check{v}, \end{aligned} \quad (9.62)$$

where $\check{v} = \partial_4 v$. We can finally evaluate ∇A ; it is

$$\begin{aligned} \nabla A &= \frac{q}{4\pi} \left[\frac{\nabla v}{X \cdot v} - \frac{1}{(X \cdot v)^2} \nabla(X \cdot v) v \right] \\ &= \frac{q}{4\pi} \left[\frac{X \check{v}}{(X \cdot v)^2} - \frac{1}{(X \cdot v)^2} - \frac{(X X \cdot \check{v} - X)v}{(X \cdot v)^3} \right] \\ &= \frac{q}{4\pi} \left[\frac{X \wedge \check{v}}{(X \cdot v)^2} + \frac{X \wedge v - X \cdot \check{v} X \wedge v}{(X \cdot v)^3} \right]. \end{aligned} \quad (9.63)$$

It is clear that all the terms in the result are bivectors, so, $\nabla \cdot A = 0$, which was expected.

To better interpret the Faraday bivector we note that

$$X \cdot v X \cdot \check{v} X \wedge v = -X[X \cdot (\check{v} \wedge v)] = \frac{1}{2} X \check{v} \wedge v X, \quad (9.64)$$

using the fact that $X^2 = 0$. Now, the bivector $\Omega_v = \check{v} \wedge v$ induces a rotation on vector v and is called the acceleration bivector of the particle; with this we arrive at the final compact formula

$$F = \frac{q}{4\pi} \frac{X \wedge v + \frac{1}{2} X \Omega_v X}{(X \cdot v)^3}. \quad (9.65)$$

We could proceed with the detailed analysis of the Faraday bivector but that is not our intent here; there are excellent books on electromagnetism where that analysis is carefully made and there is nothing new we could propose.

9.5 Fermat's principle

We call optics to the field of physics which deals with the propagation of electromagnetic waves, usually in non-homogeneous materials. In Sec. 9.1 we established that electromagnetic waves in vacuum were the solutions of Eq. (9.27) and we found that these propagated in a straight line that laid entirely in 3-dimensional space; this was found consistent with the unitary 3-dimensional velocity of electromagnetic waves. Optical media affect the propagation in several ways; particularly they can affect the velocity and, consequently, deviate waves from straight line propagation, and they can originate propagation modes in special geometries called waveguides. Interference and diffraction resulting from obstacles placed on the path of electromagnetic waves are also important subjects in optics.

The inclusion of some sections on optics at this stage is justified because we have all the necessary background, but it serves the purpose of establishing the foundations for the chapters on gravitation and quantum mechanics. There are strong links between light rays and trajectories of massive particles, as well as between the modes of optical waveguides and energy states of an Hydrogen atom. The mathematical methods we must use for addressing these challenging problems of physics can be seen as an upgrade from the methods used in optics, so, the optics' sections will give us an invaluable experience before we venture into more demanding areas. We will not consider the microscopic explanation for the behaviour of optical media and we will also not deal with non-linear behaviour of some media.

Isotropic materials were characterized above as those having $g^0 = \varepsilon\sigma^0$ and $g^m = \sigma^m/\mu$, with ε and μ electrical permittivity and magnetic permeability, respectively. The last reciprocal frame vector, $g^4 = \sigma^4 + A$, incorporates the potential vector. Looking for electromagnetic waves we search for bivector solutions of Eq. (9.43) with $J_A = 0$. Note that both the polarization derivative, ∇P , and displacement current, J_D , are different from zero, showing that the medium reacts to the wave. We then have

$$\nabla F = 0. \quad (9.66)$$

The equation becomes easier to solve if we multiply on the left by ∂ and expand the terms to obtain

$$-\varepsilon \partial_{00}F + \frac{1}{\mu} \sum_{m=1}^3 \partial_{mm}F = 0. \quad (9.67)$$

It is useful to define the refractive index,

$$n = \sqrt{\varepsilon\mu}, \quad (9.68)$$

so that the wave equation becomes

$$\sum_{m=1}^3 \partial_{mm}F = n^2 \partial_{00}F. \quad (9.69)$$

This equation admits quasi-plane waves given by

$$F = F_0 e^{(i\sigma^4 p_\mu x^\mu + \theta)}, \quad (9.70)$$

as long as the derivatives of coefficients p_μ can be neglected with respect to the derivatives of F . When this happens, inserting the solution into Eq. (9.70) we get

$$-n^2(p_0)^2 + \sum_{m=1}^3 (p_m)^2 = 0. \quad (9.71)$$

This suggests that p_μ are the components of a nilpotent vector in the reciprocal frame

$$g^0 = \sigma^0, \quad g^m = \frac{1}{n} \sigma^m, \quad (9.72)$$

allowing the solution (9.70) to be written as

$$F = F_0 e^{i(\sigma^4 p \cdot x + \theta)}. \quad (9.73)$$

The situation is now similar to the one we found in Sec. 6.1, with one less dimension and quasi-plane waves rather than truly plane waves.

Constancy of F implies that the exponent remains constant. This can happen if we consider displacements dx normal to p , in which case we are describing the plane of a wavefront. On the other hand, F will also be constant for nilpotent displacements, that is, for

$$-(dx^0)^2 + n^2 \sum_{m=1}^3 (dx^m)^2 = 0. \quad (9.74)$$

we now assign coordinate x^0 to time, as we have done before, and isolate $(dt)^2$ on the left hand side to get

$$(dt)^2 = n^2 \sum_{m=1}^3 (dx^m)^2. \quad (9.75)$$

The travel time between two points can then be evaluated by the integral

$$t = \int_{t_1}^{t_2} n \left[\sum_{m=1}^3 (\dot{x}^m)^2 \right]^{1/2} dt. \quad (9.76)$$

Fermat's principle states that the path followed by light between two points is one that makes the travel time an extremum, usually a minimum. Now compare this principle with the problem of finding a geodesic in example 8.2 and you will see that the two problems are formally identical. We can then take the refractive index as being formally identical to 3-dimensional space curvature, at least if only light paths are of interest.

Example 9.1

Consider the case of an axially-symmetric refractive index given by

$$n = e^{a/r}, \quad (9.77)$$

with a constant and r the distance to the symmetry axis. The refractive index is singular on the axis but that should not concern us, because we will be working with large r , in a region where n is not very large. For simplicity we align the symmetry axis with σ_3 and adopt cylindrical polar coordinates. The frame vectors are then

$$g_0 = \sigma_0, \quad g_r = n \sigma_r, \quad g_\phi = nr \sigma_\phi, \quad g^3 = \sigma^3. \quad (9.78)$$

Using the definition (8.7) we define the Lagrangian

$$L = \frac{n^2}{2} \left[\dot{r}^2 + r^2 \dot{\phi}^2 + (\dot{x}^3)^2 \right] = \frac{1}{2}. \quad (9.79)$$

Since the Lagrangian is independent from x^3 and ϕ , we derive two propagation constants from the Euler-Lagrange equations:

$$n^2 \dot{x}^3 = v, \quad (9.80)$$

$$n^2 r^2 \dot{\phi} = J. \quad (9.81)$$

The former of these equations tells us simply that the propagation velocity along the symmetry axis is dependent on the distance to the axis. The latter equation is a sort of angular momentum conservation. The two constants can be inserted into the Lagrangian and we obtain the one-dimensional equation

$$n^2 \dot{r}^2 + \frac{J^2}{n^2 r^2} + \frac{v^2}{n^2} = 1. \quad (9.82)$$

We can search for points of inflexion by solving for $\dot{r} = 0$. Inserting this condition in the previous equation and considering Eq. (9.77), we get

$$r = \frac{J}{\sqrt{e^{2a/r} - v^2}}. \quad (9.83)$$

Depending on the constants' values, the equation can have two solutions, in which case the wave becomes trapped in an helix at a distance from the axis oscillating between the two solutions found. If the two solutions coincide the helix has a circular profile and if there are no solutions the wave is untrapped.

This analysis is valid only as far as the quasi-plane wave approximation holds. This means that the derivative of the refractive index must be small compared to the derivatives of the wave function or, briefly, that $a/r^2 \ll \omega$, with ω the angular frequency. In optics this is known as the *geometric optics criterion*.

9.6 Optical waveguides

Fermat's principle can be applied only when the the geometric optics criterion holds; if this is not the case we must solve Eq. (9.69) without the quasi-plane wave assumption. In practice this means that we can no longer think of the solution as being a wave with wavefronts whose normal defines an optical ray. Consequently we are no longer taken into a problem of finding a geodesic. There are many situations where we are faced with this type of considerations and in this section we will examine the case of optical waveguides.

In principle any index of refraction which exhibits an elongated region of high index values surrounded by regions with lower index can act as an optical waveguide, but we will concentrate our attention in axially-symmetric distributions of refractive index, because these include optical fibres and also because they establish a good background for the forthcoming discussion of the Hydrogen atom. We will then consider a refractive index given by the function $n(r)$, with r the radial coordinate in cylindrical polar coordinates and we look for bivector solutions of Eq. (9.69). We are interested in solutions that are waves propagating along σ^3 and we assume that the solutions are separable into radial and angular parts. Our solution will then have the general form

$$F = R(r)Y(\phi)e^{i\sigma^4(\omega x^0 - \beta x^3)}. \quad (9.84)$$

Introducing into Eq. (9.69) we get

$$\nabla^2(RY) + [\beta^2 - (n\omega)^2]RY = 0. \quad (9.85)$$

Since the refractive index has a radial dependence we can use 2-dimensional null-Laplacian functions to find the angular dependence factor, $Y(\phi)$. In cylindrical coordinates this condition is written as

$$\left(\partial_{rr} + \frac{1}{r} \partial_r + \frac{1}{r^2} \partial_{\phi\phi} \right) \Psi = 0. \quad (9.86)$$

The solutions can have the form

$$\Psi_s = (r)^s \psi_0 e^{s\sigma^{r\phi}(\phi+\alpha)} = (r)^s \psi_0 Y_s, \quad (9.87)$$

where s is an integer different from -1 . Inserting Y_s into Eq. (9.85) we see that R must verify the equation

$$r^2 R'' + rR' - [s^2 - (r\beta)^2 + (r n\omega)^2]R = 0. \quad (9.88)$$

The simplest waveguides have $n > 1$ inside a cylinder of radius r_0 and $n = 1$ outside the cylinder. For such waveguides n is the step function

$$n = \begin{cases} n_0 & \text{for } r \leq r_0, \\ 1 & \text{for } r > r_0. \end{cases} \quad (9.89)$$

For a guided wave we define the two real quantities

$$\begin{aligned}\kappa &= \sqrt{(n_0\omega)^2 - \beta^2}, \\ \eta &= \sqrt{\beta^2 - \omega^2}.\end{aligned}\tag{9.90}$$

Equation (9.88) is then split into two equations as follows:

$$\begin{aligned}r^2 R'' + rR' - (s^2 - r^2\kappa^2)R &= 0, \\ r^2 R'' + rR' - (s^2 + r^2\eta^2)R &= 0.\end{aligned}\tag{9.91}$$

The solutions to these equations have the form of Bessel or modified Bessel functions; they are

$$R_s = \begin{cases} C_i J_s(\kappa r), & r \leq r_0, \\ C_o K_s(\eta r), & r > r_0; \end{cases}\tag{9.92}$$

with J_s the s -order Bessel function of the first kind and K_s the modified Bessel function of the second kind and C_i and C_o are integration constants. The radial function, R_s , must be smooth at r_0 , implying the continuity of both R_s and R'_s . As a result one determines the ratio C_o/C_i and the value of the propagation constant β . As a result only certain propagation modes are possible but we will not give a detailed discussion of such modes here. For a complete treatment of optical waveguides, using conventional vector calculus, see Okamoto [12].

We must still ensure that the product RY is a bivector so that F can retain its Faraday bivector character. The only vector content allowed for the radial function includes components along σ^0 and σ^3 because these are independent from the azimuthal angle. For the angular function, on the other hand, we can expect components along σ^r and σ^ϕ . In Eq. (9.87) we left an undetermined constant Ψ_0 , which we now set equal to any unitary vector on the $\sigma^{r\phi}$ plane, for instance $\Psi_0 = \sigma^1$. For the radial function, we can include a vector in constant C_i , and this can be either σ^0 or σ^3 . This choice gives rise to two separate families of modes, known as *transverse electric (TE)* and *transverse magnetic (TM)*, respectively.

9.7 Solitons and photons

Physicists have had great difficulty in proposing the mathematical description for a photon, due mainly to the necessity of describing a wave packet that remains stable both in space and time duration. It is known from diffraction theory that a beam of light, that is, spatially localized electromagnetic radiation, can be described by a superposition of plane waves. However, every light beam expands as it propagates, so, the process fails to provide spatial localization. In a similar fashion, a light pulse can be constructed by superposition of plane waves with varying frequency and phase but it will also expand, thus

frustrating the desired pulse length stabilization. Mathematically, one can produce wave packets that remain stable while propagating by means of solitons. The problem here is that solitons solutions exist only for non-linear equations and it is difficult to justify some sort of non-linear behaviour for vacuum. Solitons can be produced in non-linear optical fibres but any sort of vacuum non-linearity seems to come out of nowhere.

At this point it is useful to return to electrodynamics equation (9.33) and examine it under the perspective of a gauge theory.

10 Gravitation

Einstein's general theory of relativity formulates the dynamics of massive test particles in terms of geometry alone; this means that one is able to predict the motion of a tiny massive particle by simply letting it fall freely along one geodesic of space. Space being modelled by its mass and energy densities, the shape of geodesics determines the paths followed by particles in free fall. This fits very nicely into the objectives of this book, which are, as we recall, to derive the equations of physics from geometric criteria. Of course this is no coincidence; it was the success of general relativity as a geometric theory that led the authors to try and extend the method to the whole of physics. We chose to make the presentation in a logical way, starting with the basics of geometric algebra and progressing in steps to the equations of physics but mental processes are not always logical and the book took shape in authors' minds very frequently from end to beginning. In this chapter we will formulate gravitational dynamics following a procedure similar to what was used for electrodynamics.

10.1 Gravitational dynamics

The effect of mass and energy distributions in space is assumed to induce space curvature. We can always choose a coordinate system that leaves one of the frame vectors with unit norm, so, we make $g_4 = \sigma_4$ to signify this fact. We will also assume that the unit pseudoscalar of 5-dimensional spacetime stays unaltered by curvature and can be used as a complex imaginary. This can then be decomposed as follows:

$$\mathbf{i} = \frac{g_0 \wedge \cdots \wedge g_3}{|g_0 \wedge \cdots \wedge g_3|} \wedge \sigma_4 = I \wedge \sigma_4. \quad (10.1)$$

We are thus using I to represent the unit pseudoscalar of 4-dimensional spacetime algebra, projected from the embedding 5-dimensional spacetime algebra. Using Eq. (8.62) we see that the shape tensor verifies the equation

$$g_\mu \cdot \partial I = I \times S(g_\mu). \quad (10.2)$$

The shape tensor is a bivector that encodes matter and energy contents of space.

Example 10.1

In the case of electromagnetism in vacuum the frame is given by Eq. (9.4) the unit pseudoscalar of 4D spacetime is defined by

$$I = \frac{ig^4}{|g^4|}. \quad (10.3)$$

Bibliography

- [1] G. Ross, *Grand Unified Theories* (Westview Press, Oxford, U.K., 2003).
- [2] D. Hestenes and G. Sobczyk, *Clifford Algebras to Geometric Calculus. A Unified Language for Mathematics and Physics*, Fundamental Theories of Physics (Reidel, Dordrecht, 1989).
- [3] P. Lounesto, *Clifford Algebras and Spinors*, vol. 286 of *London Mathematical Society Lecture Note Series* (Cambridge University Press, Cambridge, U.K., 2001), 2nd ed.
- [4] M. J. Duff, *Kaluza-Klein theory in perspective*, 1994, arXiv: hep-th/9410046.
- [5] J. W. Goodman, *Introduction to Fourier Optics* (Roberts & Company, Greenwood Village, USA, 2005), 3rd ed.
- [6] G. O. Reynolds, J. B. DeVelis, G. B. Parrent Jr., and B. J. Thompson, *The New Physical Optics Notebook: Tutorialas in Fourier Optics* (SPIE, Bellingham, USA, 1989).
- [7] T. Kaluza, *On the problem of unity in physics*, Sitzungsber. Preuss. Akad. Wiss. Berlin. (Math. Klasse) pp. 966–972, 1921.
- [8] J. B. Almeida, *Hidden geometric character of relativistic quantum mechanics*, J. Math. Phys. **49**, 012301, 2007, arXiv: quant-ph/0606123.
- [9] J. B. Almeida, *A geometric algebra approach to the hydrogen atom*, 2006, arXiv: physics/0602116.
- [10] C. Doran and A. Lasenby, *Geometric Algebra for Physicists* (Cambridge University Press, Cambridge, U.K., 2003).
- [11] J. D. Jackson, *Classical Electrodynamics* (J. Wiley & Sons, N. York, 1998), 3rd ed.
- [12] K. Okamoto, *Fundamentals of Optical Waveguides*, Optics and Photonics (Academic Press, San Diego, USA, 2001).