IEEE Press Series on Electromagnetic Wave Theory
Andreas Cangellaris, *Series Editor*

Understanding Geometric Algebra for Electromagnetic Theory

John W. Arthur

WILEY
IEEE
IEEE PRESS
WWW. LINK AVAILABLE

Understanding Geometric Algebra for Electromagnetic Theory

IEEE Press
445 Hoes Lane
Piscataway, NJ 08854

Understanding Geometric Algebra for Electromagnetic Theory

John W. Arthur
Honorary Fellow, School of Engineering, University of Edinburgh

IEEE Antennas and Propagation Society, *Sponsor*

The IEEE Press Series on Electromagnetic Wave Theory
Andreas C. Cangellaris, *Series Editor*

IEEE PRESS

A John Wiley & Sons, Inc., Publication

Library of Congress Cataloging-in-Publication Data:

Arthur, John W., 1949–
Understanding geometric algebra for electromagnetic theory / John W. Arthur.
p. cm.—(IEEE Press series on electromagnetic wave theory ; 38)
Includes bibliographical references and index.
ISBN 978-0-470-94163-8
1. Electromagnetic theory—Mathematics. 2. Geometry, Algebraic. I. Title.
QC670.A76 2011
530.14′10151635—dc22

2011005744

Printed in the United States of America

oBook ISBN: 978-1-118-07854-3
ePDF ISBN: 978-1-118-07852-5
ePub ISBN: 978-1-118-07853-2

10 9 8 7 6 5 4 3 2 1

. . . it is a good thing to have two ways of looking at a subject, and to admit that there are two ways of looking at it.

James Clerk Maxwell, on addressing the question of two versions of electromagnetic theory, one due to Michael Faraday and the other to Wilhelm Weber, in a paper on Faraday's lines of force presented at Cambridge University, February 1856 [1, p. 67].

Contents

Preface

Geometric algebra provides an excellent mathematical framework for physics and engineering, particularly in the case of electromagnetic theory, but it can be difficult for someone new to the subject, in particular a nonspecialist, to penetrate much of the available literature. This book has therefore been addressed to a fairly broad readership among scientists and engineers in any of the following categories:

- those who are interested in electromagnetic theory but mainly wish to see a new approach to it, either with or without special relativity;
- those who already have some knowledge of geometric algebra but wish to see how it is applied to electromagnetic theory;
- those who wish for further explanation on the application of geomagnetic algebra in order to access more advanced material; and
- those who may simply wish to gain some understanding of the role of special relativity in electromagnetic theory.

It is the aim of this work to provide an introduction to the subject together with a tutorial guide to its application to electromagnetic theory. Its readers are likely to be familiar with electromagnetic theory by way of traditional methods, that is to say, vector analysis including linear vector spaces, matrix algebra, gradient, divergence, curl, and the like. Knowledge of tensors, however, is not required. Because the emphasis is on understanding how geometric algebra benefits electromagnetic theory, we need to explore what it is, how it works, and how it is applied.

The new ideas are introduced gradually starting with background and concepts followed by basic rules and some examples. This foundation is then built upon by extending and generalizing the basics, and so on. Equations are worked out in considerable detail and ample time is spent on discussing rationale and points of interest. In addition, before moving on to the next level, care is taken over the explanation of topics that tend to be difficult to grasp. The general intent has been to try to keep the presentation self-contained so as to minimize the need for recourse to external material; nevertheless, several key works are regularly cited to allow the interested reader to connect with the relevant literature.

The mathematical content is addressed to those who prefer to use mathematics as a means to an end rather than to those who wish to study it for its own sake. While formality in dealing with mathematical issues is kept to a minimum, the aim has nevertheless been to try to use the most appropriate methods, to try to take a line that is obvious rather than clever, and to try to demonstrate things to a

reasonable standard rather than to prove them absolutely. To achieve simplicity, there have been a few departures from convention and some changes of emphasis:

- The use of indices is kept to a minimum, for example, basis elements are written as $\boldsymbol{x}, \boldsymbol{xy}\ldots$ rather than $\boldsymbol{e}_1, \boldsymbol{e}_{12}\ldots$
- The basic intuitive ideas of parallel and perpendicular are exploited wherever this may be advantageous.
- The term "translation" is introduced to describe a mapping process between spacetime and 3D as distinct to the spacetime split.
- A notation is introduced whereby a vector underscored with a tilde, for example, $\underset{\sim}{\boldsymbol{u}}, \underset{\sim}{\boldsymbol{R}}$, is to be identified as a purely spatial vector. Since such vectors are orthogonal to a given time vector, this contributes to the aim of exploiting parallel and perpendicular.
- To maximize the readability of equations, a system is introduced whereby SI units are retained but equations are simplified in a way similar to the mathematical physicist's convention taking the speed of light to be 1.

A geometric algebra is a vector space in which multiplication and addition applies to all members of the algebra. In particular, multiplication between vectors generates new elements called multivectors. And why not? Indeed, it will be seen that this creates valuable possibilities that are absent in the theory of ordinary linear vector spaces. For example, multivectors can be split up into different classes called grades. Grade 0 is a scalar, grade 1 is a vector (directed line), grade 2 is a directed area, grade 3 is a directed volume, and so on. Eventually, at the maximum grade, an object that replaces the need for complex arithmetic is reached.

We begin with a gentle introduction that aims to give a feel for the subject by conveying its basic ideas. In Chapters 2–3, the general idea of a geometric algebra is then worked up from basic principles without assuming any specialist mathematical knowledge. The things that the reader should be familiar with, however, are vectors in 3D, including the basic ideas of vector spaces, dot and cross products, the metric and linear transformations. We then look at some of the interesting possibilities that follow and show how we can apply geometric algebra to basic concepts, for example, time t and position $\mathbf{r}$ may be treated as a single multivector entity $t + \mathbf{r}$ that gives rise to the idea of a (3+1)D space, and by combining the electric and magnetic fields $\mathbf{E}$ and $\mathbf{B}$ into a multivector $\boldsymbol{F}$, they can be dealt with as a single entity rather than two separate things. By this time, the interest of the reader should be fully engaged by these stimulating ideas.

In Chapter 4, we formalize the basic ideas and develop the essential toolset that will allow us to apply geometric algebra more generally, for example, how the product of two objects can be written as the sum of inner and outer products. These two products turn out to be keystone operations that represent a step-down and step-up of grades, respectively. For example, the inner product of two vectors yields a scalar result akin to the dot product. On the other hand, the outer product will create a new object of grade 2. Called a bivector, it is a 2D object that can represent an

area or surface density lying in the plane of the two vectors. Following from this is the key result that divergence and curl may be combined into a single vector operator that appears to be the same as the gradient but which now operates on any grade of object, not just a scalar.

Armed with this new toolset, in Chapter 5 we set about applying it to fundamental electromagnetics in the situation that we have called (3+1)D:

- In free space, Maxwell's four equations reduce to just one, $(\boldsymbol{\nabla} + \partial_t)\boldsymbol{F} = \boldsymbol{J}$.
- Circularly polarized plane electromagnetic waves are described without either introducing complex numbers or linearly polarized solutions.
- The electromagnetic energy density and momentum density vector fall out neatly from $\frac{1}{2}\boldsymbol{F}\boldsymbol{F}^{\dagger}$.
- The vector and scalar potentials unite.
- The steady-state solution for the combined electric and magnetic fields of a moving charge distribution has a very elegant form that curiously appears to be taken directly from the electrostatic solution.

Once the basic possibilities of the (3+1)D treatment of electromagnetic theory have been explored, we then prepare the ground for a full 4D treatment in which space and time are treated on an equal footing, that is to say, as spacetime vectors. Geometric algebra accommodates the mathematics of spacetime extremely well, and with its assistance, we discover how to tackle the electromagnetic theory of moving charges in a systematic, relativistically correct, and yet uncomplicated way.

A key point here is that it is not necessary to engage in special relativity in order to benefit from the spacetime approach. While it does open the door to special relativity on one level, on a separate level, it may simply be treated as a highly convenient and effective mathematical framework. Most illuminatingly, we see how the whole of electromagnetic theory, from the magnetic field to radiation from accelerating charges, falls out of an appropriate but very straightforward spacetime treatment of Coulomb's law. The main features of the (3+1)D treatment are reproduced free of several of its inherent limitations:

- A single *vector* derivative ∇ replaces the multivector form $\boldsymbol{\nabla} + \partial_t$.
- Maxwell's equation in free space is now simply $\boldsymbol{\nabla}\boldsymbol{F} = \boldsymbol{J}$.
- The Lorentz force reduces to the form $\boldsymbol{f} = q\boldsymbol{\upsilon}\cdot\boldsymbol{F}$.
- Maxwell's equations for polarizable media can be encoded neatly so as to eliminate the bound sources through an auxiliary field $\boldsymbol{G}$ that replaces both $\mathbf{D}$ and $\mathbf{H}$.
- The proper velocity υ plays a remarkable role as the time vector associated with a moving reference frame.
- The phase factor for propagating electromagnetic waves is given by the simple expression $\boldsymbol{r}\cdot\boldsymbol{k}$ where the vectors $\boldsymbol{r}$ and $\boldsymbol{k}$ represent time + position and frequency + wave vector, respectively.

- The general solution for the electromagnetic field of charges in motion follows directly from Coulomb's law.
- The significance of Maxwell's equation taking the form $\nabla F = J$ is that the range of analytic solutions of Maxwell's equation is extended from 2D electrostatics and magnetostatics into fully time-dependent 3D electromagnetics.

The relationship between (3+1)D and spacetime involves some intriguing subtleties, which we take time to explain; indeed, the emphasis remains on the understanding of the subject throughout. For this reason, in Chapters 7–12, we try to give a reasonably self-contained primer on the spacetime approach and how it fits in with special relativity. This does not mean that readers need to understand, or even wish to understand, special relativity in any detail, but it is fairly certain that at some point, they will become curious enough about it to try and get some idea of how it underpins the operational side of the spacetime geometric algebra, that is to say, where we simply use it as a means of getting results. Nevertheless, even on that level, readers will be intrigued to discover how well this toolset fits with the structure of electromagnetic theory and how it unifies previously separate ideas under a single theme. In short,

$$\text{Coulomb's Law} + \text{Spacetime} = \sum \text{Classical Electromagnetic Theory}.$$

The essentials of this theme are covered in Chapter 11, and in Chapter 12 we work through in detail the electromagnetic field of an accelerating charge. This provides an opportunity to see how the toolset is applied in some depth. Finally, we review the overall benefits of geometric algebra compared with the traditional approach and briefly mention some of its other features that we did not have time to explore. There are exercises at the end of most chapters. These are mostly straightforward, and their main purpose is to allow the readers to check their understanding of the topics covered. Some, however, provide results that may come in very useful from time to time. Worked solutions are available from the publisher (e-mail ieeeproposals@wiley.com for further information). The book also includes seven appendices providing explanatory information and background material that would be out of place in the main text. In particular, they contain a glossary of key terms and symbols. It is illustrated with 21 figures and 10 tables and cites some 51 references.

Finally, I wish to express my sincere thanks to my wife, Norma, not only for her painstaking efforts in checking many pages of manuscript but also for her patience and understanding throughout the writing of this book. I am also most grateful to Dr. W. Ross Stone who has been an unstinting source of help and advice.

JOHN W. ARTHUR

Reading Guide

The benefit of tackling a subject in a logical and progressive manner is that it should tend to promote better understanding in the long run, and so this was the approach adopted for the layout of the book. However, it is appreciated that not all the material in the book will be of interest to everyone, and so this reading guide is intended for those who wish to take a more selective approach.

- Chapters 1–2 provide some background on geometric algebra before exploring the basic ideas of what it is and how it works. The basic ideas of parallel and perpendicular are exploited. Readers who prefer a more axiomatic approach can refer to Appendix 14.4.
- Chapter 3 addresses how geometric algebra fits in with electromagnetic theory and how we begin to apply it.
- Chapter 4 develops the idea of a geometric algebra more fully. Although the ideas of parallel and perpendicular are still referred to, more formality is introduced. The aim is to provide a grounding on the essential mathematical tools, structures, and characteristics of geometric algebra. This chapter may be skimmed at first reading and referred to for further information as and when necessary. Appendix 14.4 may also be used for ready reference or as supporting material.
- Chapter 5 is sufficient to show how some of the key topics in electromagnetic theory can be dealt with using (3+1)D geometric algebra. Readers should be able to form their own opinion as to its superiority compared with traditional methods.
- Chapter 6 recaps what has been achieved thus far. This may be a convenient stopping or resting point for some readers, but hopefully, their curiosity will be sufficiently aroused to encourage them to carry on to the spacetime approach.
- Chapters 7–8 provide an introduction to the spacetime geometric algebra with minimal involvement in special relativity. Readers are encouraged to attempt at least these sections before starting Chapter 11. It is also recommended that Chapter 4 should have been studied beforehand.
- Chapters 9–10 deal with different frames and transforming between them. These sections are primarily intended for those readers who are interested in the underlying physics and wish to get a better appreciation of the spacetime approach to special relativity. It is not essential reading for other readers who

may simply prefer to refer to these chapters only when they require more information.

- Chapter 11 covers the treatment of key topics in electromagnetic theory through the spacetime geometric algebra. Readers who have not covered Chapters 9–10 in any detail may need to refer to these sections as and when necessary. The results are to be compared with the (3+1)D results of Chapter 5 where, despite the many successes, we were encouraged to hope for more. It is hoped that by this point, readers will feel obliged to conclude that spacetime approach is not only superior to the (3+1)D but in some sense it is also an ideal fit to the subject in hand.
- For those who wish to go the whole way, Chapter 12 covers the process of differentiating the spacetime vector potential as the means of obtaining the radiated electromagnetic field of an accelerating point charge. Being familiar with Chapters 9–10 beforehand is fully recommended.
- Appendices 14.1–14.7 include explanatory information and background material that would be out of place in the main text. In particular, it opens with a glossary that provides a ready reference to key terms and notation.
- Several chapters have exercises. These are mostly straightforward and their main purpose is to allow readers to check their understanding of the topics covered. Some, however, provide results that may come in very useful from time to time.

Chapter 1

Introduction

Sooner or later, any discussion of basic electromagnetic theory is certain to come to the issue of how best to categorize the vectors **B**, the magnetic induction, and **H**, the magnetic field strength. Polar or axial is the central issue [2]. From an elementary physical perspective, taking **B** as an axial vector seems appropriate since, as far as we know, all magnetism originates from currents (see Appendix 14.2). From a mathematical standpoint, however, taking **H** as a polar (or true) vector seems a better fit with an integral equation such as Ampere's law, $\int \mathbf{H} \cdot d\mathbf{l} = \mu_0 I$, particularly in relation to the subject of differential forms [3–5]. But taking the view that **B** can be one sort of vector while **H** is another seems to be at odds with an equation such as $\mathbf{B} = \mu\mathbf{H}$ in which the equality implies that they should be of the same character. A separate formal operator is required in order to get around this problem, for example, by writing $\mathbf{B} = \mu * \mathbf{H}$ where $*$ converts a true vector to an axial one and vice versa, but for most people, any need for this is generally ignored.

Geometric algebra provides a means of avoiding such ambiguities by allowing the existence of entities that go beyond vectors and scalars. In 3D, the additional entities include the bivector and the pseudoscalar. Here the magnetic field is represented by a bivector, which cannot be confused with a vector because it is quite a different kind of entity. Multiplication with a pseudoscalar, however, conveniently turns the one into the other. But the new entities are far from arbitrary constructs that have simply been chosen for this purpose, they are in fact generated inherently by allowing a proper form of multiplication between vectors, not just dot and cross products.

Using geometric algebra, Maxwell's equations and the Lorentz force are expressed in remarkably succinct forms. Since different types of entities, for example vectors and bivectors, can be combined by addition, the field quantities, the sources, and the differential operators can all be represented in a way that goes quite beyond simply piecing together matrices. While multiplication between entities of different sorts is also allowed, the rules all stem from the one simple concept of vector multiplication, the geometric product. Multiplication of a vector by itself results in a scalar, which provides the basis for a metric. Inner and outer products are very simply extracted from the geometric product of two vectors, the inner

Understanding Geometric Algebra for Electromagnetic Theory, First Edition. John W. Arthur.

product being the scalar part of the result whereas the outer product is the bivector part. Given that the product of a vector with itself is a scalar, inner and outer products are directly related to the ideas of parallel and perpendicular. The bivector therefore represents the product of two perpendicular vectors and has the specific geometric interpretation of a directed area. The pseudoscalar in 3D corresponds to a trivector, the product of three vectors, and can be taken to represent a volume. This hierarchy gives rise to the notion that a geometric algebra is a graded algebra, scalars being of grade 0, vectors grade 1, bivectors grade 2, and so on. Crucially, objects of different grades may be added together to form a general form of object known as a multivector. Just how this is possible will be explained in due course, but an example is $t+\mathbf{r}$, which has pretty much the same meaning as writing $(t,\mathbf{r})$ in normal vector algebra where, for example, this is the way we would normally write the time and position parameters of some given variable, for example, $\mathbf{E}(t,\mathbf{r})$. Why not $\mathbf{E}(t+\mathbf{r})$?

In the geometric algebras we shall be dealing with, pseudoscalars always have a negative square, a property that leads to complex numbers being superfluous. It is also found that the inner and outer products may generally be considered to be step-down and step-up operations, respectively. Provided the results are nonzero, the inner product of an object of grade n with another of grade m creates an object of grade $|m-n|$, whereas their outer product creates an object of grade $m+n$.

In addition to the novel algebraic features of geometric algebra, we also find that it is easy to turn it to calculus. In fact, the vector derivative $\boldsymbol{\nabla}$ provides all the functions of gradient, divergence, and curl in a unified manner, for, as the name suggests, it behaves like other vectors and so we can operate not only on scalars but also on vectors and objects of any other grade. While the inner and outer products with $\boldsymbol{\nabla}$ relate respectively to divergence and curl, the salient point is that we can use $\boldsymbol{\nabla}$ as a *complete entity* rather than in separate parts. Although the time derivative still requires to be dealt with separately by means of the usual scalar operator ∂_t, this no longer needs to stand entirely on its own, for just as we can combine time and position in the multivector form $t+\mathbf{r}$, we can do the same with the scalar and vector derivatives, in particular by combining the time and space derivatives in the form $\partial_t+\boldsymbol{\nabla}$. The everyday tools of electromagnetic theory are based on standard vector analysis in which time and space are treated on separate 1D and 3D footings, but here we have a more unified approach, which though not quite 4D may be appropriately enough referred to as (3+1)D where

$$(3+1)\text{D} = 3\text{D (space)} + 1\text{D (time)}$$

For example, we can write novel-looking equations such as $(\partial_t+\boldsymbol{\nabla})(t+\mathbf{r})=4+\mathbf{v}$ and $(\partial_t+\boldsymbol{\nabla})\mathbf{r}^2=2(\mathbf{r}+\mathbf{r}\cdot\mathbf{v})$, and many more besides, but we do have to be careful about what such equations might mean. Note however that (3+1)D properly refers to the physical model, where vectors represent space and scalars represent time, whereas the geometric algebra itself is 3D and should be strictly referred to as such. For the same reason, (3+1)D cannot be equated to 4D—time is treated as a scalar here, whereas it would properly require to be a vector

in order to contribute a fourth dimension. When we do opt for a full 4D treatment, however, this is found to provide a very elegant and fully relativistic representation of spacetime. This has even more significance for the representation of electromagnetic theory because it unravels the basic mystery as to the existence of the magnetic field. It simply arises from a proper treatment of Coulomb's law so that there is no separate mechanism by which a moving charge produces a magnetic field. In fact, this was one of the revolutionary claims put forward in 1905 by Albert Einstein (see, e.g., Reference 2).

The aim of this work is to give some insight into the application of geometric algebra to electromagnetic theory for a readership that is much more familiar with the traditional methods pursued by the great majority of textbooks on the subject to date. It is our primary intention to focus on understanding the basic concepts and results of geometric algebra without attempting to cover the subject in any more mathematical detail than is strictly necessary. For example, although quaternions and the relationship between a 2D geometric algebra and complex numbers are important subjects, we discuss them only by way of background information as they are not actually essential to our main purpose.

We have also tried to avoid indulging in mathematics for its own sake. For example, we do not take the axiomatic approach; rather, we try to make use of existing ideas, extending them as and when necessary. Wherever it helps to do so, we draw on the intuitive notions of parallel and perpendicular, often using the symbols $\perp$ and // as subscripts to highlight objects to which these attributes apply. On the whole, the approach is also practical with the emphasis being on physical insight and understanding, particularly when there is an opportunity to shed light on the powerful way in which geometric algebra deals with the fundamental problems in electromagnetic theory.

The reader will find that there are some excellent articles that give a fairly simple and clear introduction to basic geometric algebra, for example, in Hestenes [6, 7] and in the introductory pages of Gull et al. [8], but in general, the literature on its application to electromagnetic theory tends to be either limited to a brief sketch or to be too advanced for all but the serious student who has some experience of the subject. The aim of this work is therefore to make things easier for the novice by filling out the bare bones of the subject with amply detailed explanations and derivations. Later on, however, we consider the electromagnetic field of an accelerating point charge. While this may be seen as an advanced problem, it is worked out in detail for the benefit of those readers who feel it would be worth the effort to follow it through. Indeed, geometric algebra allows the problem to be set up in a very straightforward and elegant way, leaving only the mechanics of working through the process of differentiation and setting up the result in the observer's rest frame.

Even if the reader is unlikely to adopt geometric algebra for routine use, some grasp of its rather unfamiliar and thought-provoking ideas will undoubtedly provide a better appreciation of the fundamentals of electromagnetics as a whole. Hopefully, any reader whose interest in the subject is awakened will be sufficiently encouraged to tackle it in greater depth by further reading within the cited references. It is only necessary to have a mind that is open to some initially strange ideas.

We start with a brief examination of geometric algebra itself and then go on to take a particular look at it in (3+1)D, which we may also refer to as the Newtonian world inasmuch as it describes the everyday intuitive world where time and space are totally distinct and special relativity does not feature. In his *Principia* of 1687, Newton summarized precisely this view of space and time which was to hold fast for over two centuries: "I will not define time, space, place and motion, as being well known to all" [9]. We then embark on finding out how to apply it to the foundations of basic electromagnetics, after which we briefly review what has been achieved by the process of restating the traditional description of the subject in terms of geometric algebra—what has been gained, what if anything has been lost, what it does not achieve, and what more it would be useful to achieve. This then leads to exploring the way in which the basic principles may be extended by moving to a 4D non-Euclidean space referred to as spacetime, in which time is treated as a vector in an equivalent but apparently somewhat devious manner to spatial vectors in that its square has the opposite sign. The concept of spacetime was originated by Hermann Minkowski in a lecture given in 1909, the year of his death. "Raum und Zeit," the title of the lecture, literally means "space and time" whereas the modern form, spacetime, or space-time, came later. After covering the basics of this new geometric algebra, we learn how it relates to our ordinary (3+1)D world and in particular what must be the appropriate form for the spacetime vector derivative.

Once we have established the requisite toolset of the spacetime geometric algebra, we turn once again to the basic electromagnetic problems and show that not only are the results more elegant but also the physical insight gained is much greater. This is a further illustration of the power of geometric algebra and of the profound effect that mathematical tools in general can have on our perception of the workings of nature. It would be a mistake to have the preconception that the spacetime approach is difficult and not worth the effort; in fact, the reverse is true. Admittedly, many relativity textbooks and courses may give rise to such apprehensions. Even if the reader is resolutely against engaging in a little special relativity, they need not worry since the spacetime approach may simply be taken at face value without appealing to relativity. Only a few simple notions need to be accepted:

- Time can be treated as a vector.
- The time vector of any reference frame depends in a very simple way on its velocity.
- The square of a vector may be positive, zero, or negative.
- An observer sees spacetime objects projected into (3+1)D by a simple operation known as a spacetime split that depends only on the time vector of the chosen reference frame.

Again we draw on the notions of parallel and perpendicular, and, as a further aid, we also introduce a notation whereby underscoring with a tilde, ~, indicates that any vector marked in this way is orthogonal to some established time vector. That is to say, given $\boldsymbol{t}$ as the time vector, we can express any vector $\boldsymbol{u}$ in the form $u_t\boldsymbol{t} + \underset{\sim}{\boldsymbol{u}}$

where $u_t \boldsymbol{t} /\!/ \boldsymbol{t}$ and $\underset{\sim}{\boldsymbol{u}} \perp \boldsymbol{t}$. As a result, $\underset{\sim}{\boldsymbol{u}}$ may be interpreted as being a purely spatial vector. This has many advantages, an obvious one being that $\boldsymbol{u} = u_t \boldsymbol{t} + u_x \boldsymbol{x} + u_y \boldsymbol{y} + u_z \boldsymbol{z}$ can be written more simply as $\boldsymbol{u} = u_t \boldsymbol{t} + \underset{\sim}{\boldsymbol{u}}$.

It is quite probable that many readers may wish to skip the two chapters that mainly cover themes from special relativity, but it is also just as probable that they will refer to them later on if and when they feel the need to look "under the lid" and investigate how spacetime works on a physical level. This may be a useful approach for those readers who do initially skip these chapters but later decide to tackle the radiated field of an accelerating charge. On the other hand, those intrepid readers who wish from the outset to embark on the full exposé will probably find it best to read through the chapters and sections in sequence, even if this means skimming from time to time. With or without the benefit of special relativity, it is to be hoped that all readers should be about to put geometric algebra into practice for themselves and to appreciate the major themes of this work:

- The electric and magnetic fields are not separate things, they have a common origin.
- The equations governing them are unified by geometric algebra.
- In general, they are also simplified and rendered in a very compact form.
- This compactness is due to the ability of geometric algebra to encode objects of different grades within a single multivector expression.
- The grade structure is the instrument by which we may "unpack" these multivector expressions and equations into a more traditional form.
- Coulomb's law + spacetime = Σ classical electromagnetic theory.

While SI units are used throughout this book, in the later stages we introduce a convention used by several authors in which constants such as c, ε_0, and μ_0 are suppressed. This is often seen in the literature where natural units have been used so that $c = 1$. As a result, the equations look tidy and their essential structures are clearer. However, this need not be taken as a departure from SI into a different set of units; in our case, it is just a simple device that promotes the key points by abstracting superfluous detail. Restoring their conventional forms complete with the usual constants is fairly straightforward.

We use the familiar bold erect typeface for 3D vectors; for example, the normal choice of orthonormal basis vectors is $\mathbf{x}, \mathbf{y}, \mathbf{z}$, whereas for spacetime, we switch to bold italic simply to make it easier to distinguish the two when they are side by side. The usual spacetime basis is therefore taken as $\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$. Vectors may be expressed in component form as $u_x \boldsymbol{x} + u_y \boldsymbol{y} \ldots$ and so on. While this is a departure from the notation typically seen in the literature, in our view, many readers will be more comfortable with this rather than having to deal with the use of indexed basis elements and the same typeface for scalars and vectors alike. When indexed basis elements are required, we use $\mathbf{e}_x, \mathbf{e}_y \ldots$ to mean the same thing as $\mathbf{x}, \mathbf{y} \ldots$ and so on. This is no different from using numerical indices since, after all, indices are only labels. This makes it possible to use the summation sign, for example,

$\mathbf{u} = \sum_{k=x,y,z} u_k \mathbf{e}_k$. Other notational forms and symbols are kept to a minimum, being introduced as and when necessary either because they are standard and need to be learned or simply because they fulfill a specific purpose such as readability; for example, we use the less familiar but neater form ∂_u rather than $\partial/\partial u$ for derivatives. Finally, the glossary of Appendix 14.1 provides an *aide memoire* to key terms and notation, and the other appendices provide a little more detail on some issues that, though of interest, would have simply been a digression in the main text.

Chapter 2

A Quick Tour of Geometric Algebra

As has been made clear, this book is not intended as the basis of a mathematical treatment of geometric algebra; rather, it is directed at understanding its application to a physical problem like classical electromagnetic theory. The reader is therefore assumed to be familiar with the conventional rules of linear vector spaces [10, 11; Appendix 14.4.1] and vector analysis [12]. Most of the basic rules and principles still apply, and so we will give our attention only to the main extensions and differences under geometric algebra. We will, however, take care to go into sufficient detail for the unfamiliar reader to be able to get to grips with what is going on. If there is any sort of catch to geometric algebra, it may be that by dispensing with so much of the comparative complexity of the traditional equations, there occasionally seems to be a lack of detail to get to grips with! For example, we shall eventually see Maxwell's equations in free space in the strikingly simple form $\boldsymbol{\nabla F} = \boldsymbol{J}$, where $\boldsymbol{J}$ is a vector comprising both charge and current source densities, $\boldsymbol{F}$ is a multivector comprising both the electric and magnetic fields, and $\boldsymbol{\nabla}$ is the vector derivative involving both space and time. Notwithstanding the question of what this new equation actually means, it is not hard to appreciate the amazing degree of rationalization of the usual variables and derivatives, namely ρ, $\mathbf{J}$, $\mathbf{E}$, $\mathbf{B}$, $\boldsymbol{\nabla}\cdot$, $\boldsymbol{\nabla}\times$, and ∂_t, into just three. By comparison, tools such as components, coordinates, matrices, tensors, and traditional vector analysis seem like low-level languages, whereas geometric algebra also works as a higher level one. As can be seen from this example, geometric algebra can sweep away the minutiae to reveal what is a common underlying structure, a feature which has often been referred to as "encoding." Four separate equations in four variables are therefore found to be encoded in the simple form $\boldsymbol{\nabla F} = \boldsymbol{J}$. Not only that, this tells us that Maxwell's four separate equations are simply different manifestations of a common process encapsulated in $\boldsymbol{\nabla F} = \boldsymbol{J}$.

Maxwell's original equations were actually written in terms of simultaneous differential equations for each component of the field quantities in turn without even using subscripts; for example, the field vector $\mathbf{E}$ is represented by the three separate

Understanding Geometric Algebra for Electromagnetic Theory, First Edition. John W. Arthur.

variables P, Q, and R [13]. In his later treatise [14], he revised his approach by employing the quaternions of William Rowan Hamilton [15], which bear some similarity to a geometric algebra (see Appendix 14.5). Maxwell's equations were cast in their familiar form by Oliver Heaviside [16] based on the now conventional 3D vector analysis developed by J. Willard Gibbs [17]. Although the theory of dyadics and tensors [18–21] was well developed by the end of the nineteenth century and the theory of differential forms [3, 22] was introduced by Elié Cartan[1] in the early twentieth century, these have generally been regarded as the province of theoretical physics rather than as a tool for general use. Geometric algebra, however, is much older but does have several points of similarity, notably provision for the product of vectors and the new entities so created. Its ideas were originated by Hermann Grassmann [23] and developed by William Clifford [24] around the mid-nineteenth century, but it languished as "just another algebra" until relatively recently when it was revived by David Hestenes [25] and promoted by him [6] and others [26–28]. In common with many other specialized mathematical techniques, such as group theory, it combines some points of difficulty in the mathematics with great usefulness in its application. Nevertheless, both disciplines provide important physical interpretations, which are surprisingly intuitive and relatively easy to master. Once the mind has become used to concepts that at first seem unfamiliar or even illogical, things begin to get easier, and a clearer picture begins to emerge as the strangeness begins to wane and by and by key points fit into place.

We have already assumed the reader will be adequately familiar with linear vector spaces and the traditional rules of manipulating the vectors within them (linear algebra, matrices, inner products, cross products, curl, divergence, etc.). These mathematical rules, stated in Appendix 14.4.1 for reference, allow the inner product as the only general notion of multiplication between vectors. The vectors $\mathbf{u}$ and $\mathbf{v}$ may be "multiplied" insofar as the inner (or dot) product $\mathbf{u}\cdot\mathbf{v}$ determines the length of the projection of $\mathbf{u}$ onto $\hat{\mathbf{v}}$. The result is a scalar, the main function of which is the provision of a metric over the vector space to which both $\mathbf{u}$ and $\mathbf{v}$ belong. The metric allows us to attribute a specific length to a vector and to measure the angle between any pair of vectors. For any such vector space, a basis of mutually orthogonal vectors of unit length can be found, for example, the familiar $\hat{\mathbf{x}},\hat{\mathbf{y}},\hat{\mathbf{z}}$ or $\mathbf{i},\mathbf{j},\mathbf{k}$, with the result that any vector in the space is equivalent to some linear combination of these vectors. The number of such basis vectors required to span any given vector space is unique and equal to the dimension of the space. But we must step outside this linear algebra over vector spaces in order to multiply vectors in any other way. In the specific case of 3D, often also denoted by $\mathbb{R}^3$ or $\mathcal{E}^3$, we have an extension in the form of the cross product denoted by $\mathbf{u}\times\mathbf{v}$. The result is represented by a third "vector" $\mathbf{w}$ that is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$. Its length corresponds to the area formed by sweeping the one vector along the length of the other, while its direction is given by the right-hand screw rule. When $\mathbf{u}$ and $\mathbf{v}$ are expressed in terms of the orthonormal basis vectors $\mathbf{x},\mathbf{y},\mathbf{z}$, $\mathbf{w}$ may be written as though it were a determinant in the form

[1] Henri Cartan, the author of Reference 22, is in fact the son of Elié Cartan.

$$\mathbf{w} = \mathbf{u} \times \mathbf{v} = \begin{vmatrix} \mathbf{x} & \mathbf{y} & \mathbf{z} \\ u_x & u_y & u_z \\ v_x & v_y & v_z \end{vmatrix} \tag{2.1}$$

There are, however, conceptual problems with the cross product. Not only does it stand outside the usual framework of vector spaces, but as the product of just two vectors, it is also difficult to see how Equation (2.1) could be generalized to spaces of higher dimensions. Also, $\mathbf{u} \times \mathbf{v}$ is not a physical vector in the same sense that $\mathbf{u}$ and $\mathbf{v}$ are. If the basis vectors are inverted, then the components of all the vectors in the space, $\mathbf{u}$ and $\mathbf{v}$ included, are likewise inverted, whereas it is readily observed that the components of $\mathbf{u} \times \mathbf{v}$ are not. Although the magnitude of $\mathbf{u} \times \mathbf{v}$ represents an area rather than a length, given that it has both magnitude and direction, it still manages to qualify as a vector. This analogy between directed areas and vectors arises only in 3D and is the cause of much confusion. We have already discussed the conflicting notions about the characters of $\mathbf{B}$ and $\mathbf{H}$, but current density seems to be a case that would equally well merit being treated as a true vector or an axial one. On the one hand, taken as being $\rho\mathbf{v}$ where ρ is charge density (a scalar) and $\mathbf{v}$ is velocity (a true vector), it may be considered to be a true vector. On the other hand, taken as the scalar current $\mathcal{I}$ flowing through an orientated unit area $\mathbf{A}$, it is manifestly an axial vector, just like $\mathbf{A}$ itself. While these issues may be more mathematical rather than physical, the last point is nevertheless a concern to us in that it implies that in a given space there could be two vectors that appear to be equivalent yet actually behave differently under some simple linear transformations. While this need not be a physical issue so long as we are careful to differentiate between true and axial vectors, it tends to suggest that we are missing some underlying principle on which the classification and multiplication of vectors should more generally depend.

Geometric algebra provides this missing principle in that it allows for a proper multiplication process between vectors, not so as to produce more vectors but rather new entities called *bivectors*. Admittedly, geometric algebra is not unique in this respect, and as previously mentioned, there are other systems that allow similar ideas, but the arithmetic of geometric algebra is different in that it also allows for the addition of any combination of different objects under the umbrella of what is called a general multivector. Starting from a real scalar and multiplying each time by a vector, the resulting scheme of objects is scalar, vector, bivector, trivector, and so on. We will clarify exactly what is involved in this particular multiplication process in due course. Each step here is called a grade, with scalars being of grade 0, vectors grade 1, and so on. Systematically, therefore, the objects in a geometric algebra may be referred to as n-vectors, where n refers to the grade of the object. Some care is needed here because all the objects in a geometric algebra are also vectors in the broadest meaning of the word, that is to say, they are all members of a vector space. On the other hand, when we refer to a vector without such qualification we usually mean a 1-vector.

In 3D, a scalar may represent any quantity that can be reduced to just a number, for example, temperature and voltage; a vector is a directed line; a bivector is an

orientated area, and a trivector is an orientated volume, as depicted in Figure 2.1(a)–(c). As a result, ambiguities such as between true vectors and axial vectors simply do not arise since axial vectors are replaced by *bivectors*, which are a different class of object altogether. The 3D trivector, however, represents a volume and so would seem to be similar to a scalar, but nevertheless if the three vectors involved in the product are known, it may also be ascribed an orientation, as shown. Figure 2.2 is intended as an aid to visualizing the basic structures in a 3D geometric algebra and illustrates some of the ways in which they may be combined, while Table 2.1(a) summarizes the general hierarchy of structures within a geometric algebra and Table 2.1(b) gives a specific example for 4D.

Beyond that, for any given dimension of space, the number of distinct entities is limited, that is to say, there is a maximum grade. Closure arises because the highest-grade entity in any geometric algebra is the so-called pseudoscalar, which in 3D is the same as the trivector, and multiplying any entity by a pseudoscalar simply results in a different but previously generated entity. The unit pseudoscalar, which we will represent by I, has the property that I^2 is scalar, and, in the geometric algebras of interest to us at least, it also obeys $I^2 = -1$ and behaves much like, but not exactly like, the imaginary unit j. More about the relationship between geometric algebra and complex numbers is to be found in Appendix 14.3, but for present purposes, the main point to stress is that although I and j appear to be similar, they cannot actually be treated as being the same thing.

Having created new entities that are not vectors, how do we add them? The new entities can themselves be treated as vectors in a vector space, and so for any specific type, we may add them, multiply by scalars, and so on, as usual. But how about adding different types such as a bivector to a vector or a scalar? The *only* difference from adding two vectors is that the result is a multivector, an entity that is generally composed of different grades, that is to say, different types. Objects of the same grade add together arithmetically, while objects of different grades remain separate. This is easier to see if we introduce a basis. If the basis vectors are, say, $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \ldots$, we know that the sum of two vectors $\boldsymbol{u} = u_x \boldsymbol{x} + u_y \boldsymbol{y} + \ldots$ and $\boldsymbol{v} = v_x \boldsymbol{x} + v_y \boldsymbol{y} + \ldots$ is given by

$$\boldsymbol{u} + \boldsymbol{v} = \left(u_x + v_x\right)\boldsymbol{x} + \left(u_y + v_y\right)\boldsymbol{y} + \ldots \tag{2.2}$$

We simply extend this principle by letting the basis elements include scalars, bivectors, trivectors, and so on. When we say basis here, we mean all such elements, whereas when we say basis vectors, we mean only the vectors themselves, from which, of course, all of the other basis elements can be generated, even scalars. Given a full complement of basis elements $\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3 \ldots$, any multivector $\boldsymbol{U}$, that is to say any object in the geometric algebra, may be written in the form

$$\boldsymbol{U} = U_0 + U_1 \boldsymbol{X}_1 + U_2 \boldsymbol{X}_2 + \ldots \tag{2.3}$$

Note that for completeness, we include the scalar part along with all the others; that is, we take $\boldsymbol{X}_0 \equiv 1$, and if $\boldsymbol{X}_0, \boldsymbol{X}_1, \boldsymbol{X}_2 \ldots$ are to qualify as a proper basis, there must be no smaller set of basis elements that will do the job. The rules here are

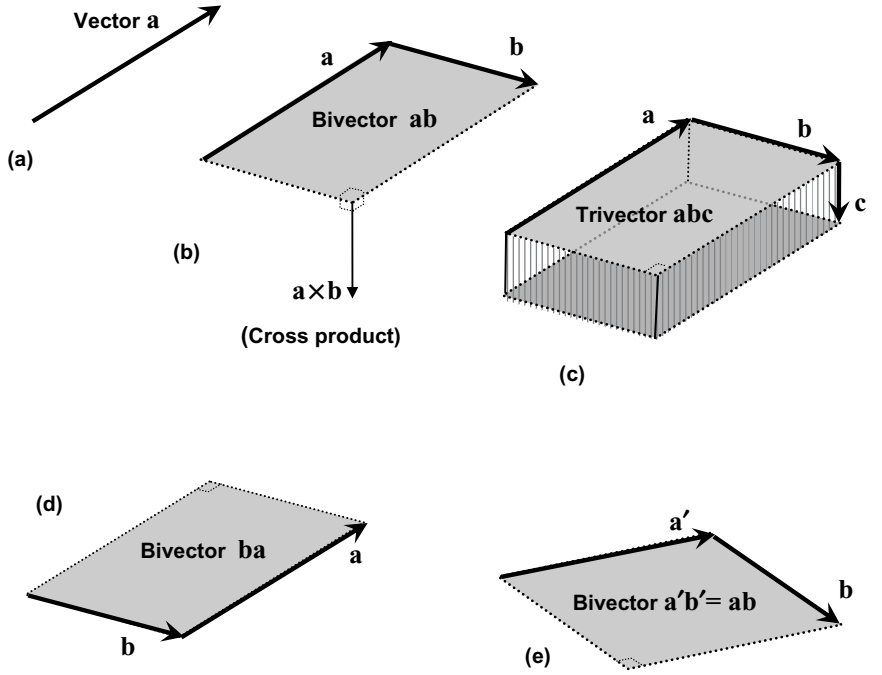

Figure 2.1 Visualizing the vector, bivector, and trivector. The vectors **a**, **b**, and **c** are all mutually orthogonal, and a, b, and c represent their magnitudes. A bold erect font is customary for 3D vectors, but we could equally well use $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ if we have no specific dimension of space in mind.

(a) We start with the vector **a**.

(b) The bivector **ab** is then produced by multiplying **a** on the right with **b**. It may be interpreted as a parallelogram with edges **a** and **b** and area given by ab. The sense of **ab** can be taken from the path of the arrows, that is, **a** then **b**. The axial vector $\mathbf{a}\times\mathbf{b}$ is shown for comparison. A bivector may also be more generally expressed as an outer product as in item (f).

(c) Finally, on multiplying by a third mutually orthogonal vector **c**, the result is now the trivector **abc**. We can take the ordered triple **a**, **b**, **c** as representing an orientated volume. More generally, a trivector may be expressed as $\mathbf{a}\wedge\mathbf{b}\wedge\mathbf{c}$ where **a**, **b**, and **c** need not be orthogonal. Note also that the triple product $\mathbf{a}\times\mathbf{b}\cdot\mathbf{c}$ is the nearest comparison to the trivector. Although both are interpreted as the volume enclosed by the three given vectors, the triple product is scalar rather than pseudoscalar.

(d) The bivector **ba** is shown for comparison with **ab** in item (b). Since it has the opposite sense, that is to say along **b** then **a**, this demonstrates $\mathbf{ba}=-\mathbf{ab}$.

(e) Care is required with the interpretation of vector products as orientated figures. Although the edges of the figure can be taken from the vectors making the product, these are in no way unique. For example, in the case of the bivector **ab**, it is easy to find any number of pairs of orthogonal vectors, say $\mathbf{a}'$ and $\mathbf{b}'$, that would produce an algebraically equivalent result. It is only necessary to rotate both **a** and **b** by the same angle in the **ab** plane to show this. Moreover, we may give the bivector any shape we choose as long as it has the same area, lies in the same plane, and has the same sense. The trivector can also be made up in any number of ways, all of which may have different shapes and orientations. The only unique orientational information is the handedness. For example, in a right-handed system, $abcI$ corresponds to a right-handed triple, whereas $-abcI$ corresponds to a left-handed one.

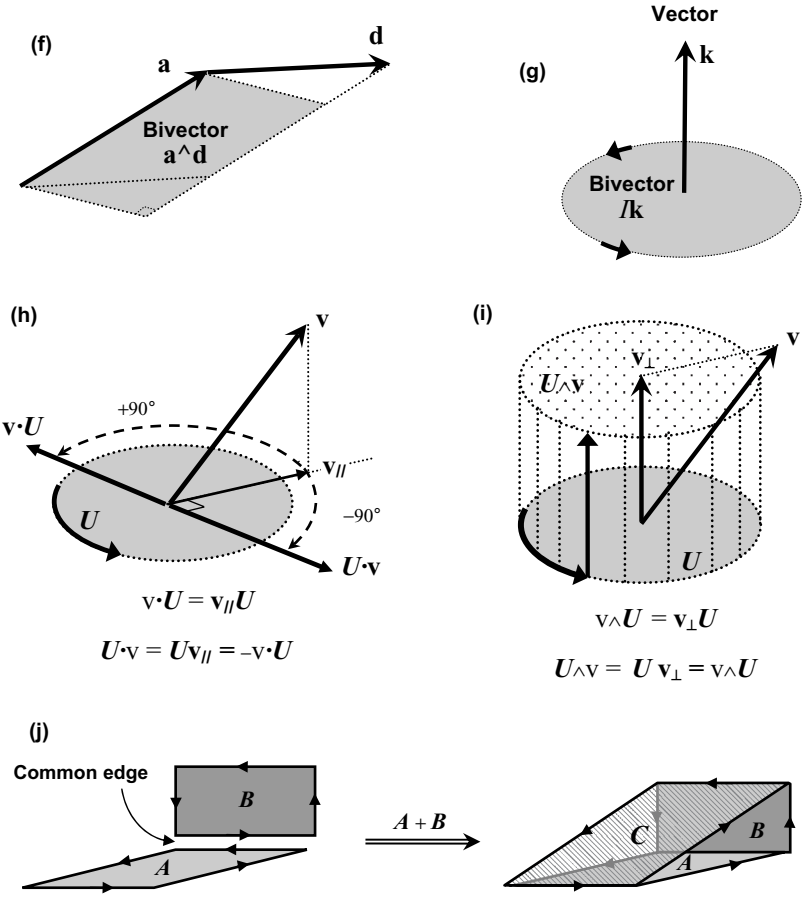

Figure 2.1 Continued

(f) In contrast to item (b), a bivector is formed here by the outer product of two vectors **a** and **d** that are not orthogonal. The area of $\mathbf{a} \wedge \mathbf{d}$ is equal to that of the parallelogram formed by the closed path formed by stepping through **a**, **d**, **−a**, then **−d**. The sense of the circuit thus formed gives the sense of the bivector.

(g) Here we have a bivector generated by the dual of the vector **k**. Note that there is no point trying to ascribe some shape to $I\mathbf{k}$, but we still have an orientation given by the right-handed screw rule.

(h) The figure shows a geometric interpretation of $\boldsymbol{U} \cdot \mathbf{v}$. We start with the bivector $\boldsymbol{U}$ shown in item (g) and any given vector **v**. The projection of **v** onto the bivector plane gives us $\mathbf{v}_{/\!/}$, which is then rotated by 90° in the *opposite* sense to the bivector. Once we have the orientation, the magnitude of the resulting vector is given by $Uv_{/\!/}$, which evaluates as $\sqrt{-\boldsymbol{U}^2\mathbf{v}_{/\!/}^2}$. Note that $\mathbf{v} \cdot \boldsymbol{U}$ would be found by rotating $\mathbf{v}_{/\!/}$ by 90° in the *same* sense as $\boldsymbol{U}$.

(i) In contrast to item (h), here we have the geometric interpretation of $\boldsymbol{U} \wedge \mathbf{v}$ in which we are effectively multiplying the bivector $\boldsymbol{U}$ with $\mathbf{v}_\perp$, the part of **v** that is perpendicular to the bivector plane.

(j) The figure shows how two 3D bivectors add geometrically when they are represented by rectangular plane elements with one common edge. In 3D, it is always possible to find a common edge.

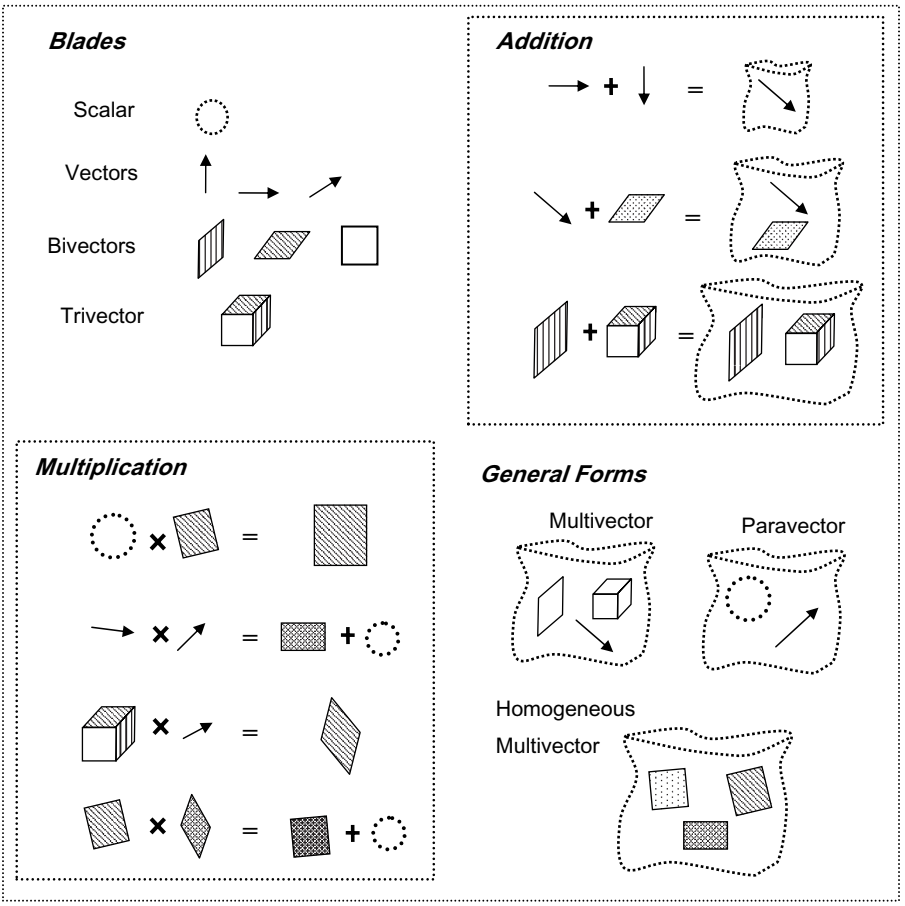

Figure 2.2 Schematic overview of the generation and classification of objects in a 3D geometric algebra. This representation attempts to illustrate some basic ideas about the classification of objects and the terms used to describe them. It is certainly not meant to be taken literally:

- Vectors and scalars are represented by individual lines and blobs, respectively.
- The size of each object determines its magnitude on some arbitrary scale.
- Vectors, bivectors, and pseudoscalars can convey different orientations.
- Although geometric multiplication of objects is symbolically indicated here by $\times$, in algebraic expressions, geometric multiplication is shown as a simple juxtaposition of symbols, for example, $\mathbf{ab}$ rather than $\mathbf{a} \times \mathbf{b}$.
- Addition of like objects is the same as vector addition, but addition of dissimilar objects may be thought of as simply placing them together—like putting them in a bag. The objects within in the bag can then be sorted out so as to form new linear combinations of the basis elements.

Table 2.1 General Hierarchy of the Entities in a Geometric Algebra

***N*-dimensional Geometric algebra**

Grade	Entities	Quantity
N	Pseudoscalar	1
...	...	...
4	4-Vector	$\binom{N}{4}$
3	Trivector	$\binom{N}{3}$
2	Bivector	$\binom{N}{2}$
1	Vector	N
0	Real scalar	1

(a)

4D (spacetime)

Grade	Entities	Quantity
4	$I = xyzt$	1
3	zyx, tzy, txz, tyx	4
2	xt, yt, zt yz, zx, xy	6
1	t, x, y, z	4
0	1	1

(b)

(a) The general hierarchy of the elements. The term n-vector provides a systematic alternative way of labeling each of the entities by grade, for example, 2-vector instead of bivector.

(b) The example of a 4D algebra with four orthogonal basis vectors taken as being $\boldsymbol{t},\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}$ so that we can see how the higher-grade entities are formed. Note that the pseudoscalar and 4-vector are the same thing, but beware that within the tensor formulation of special relativity, the term 4-vector has a different meaning.

Table 2.2 Some Common Forms of Notation for the 3D Basis Elements

3D basis element	This work	Hestenes, Doran and Lasenby, Gull et al.		Lounesto
Vector	$\mathbf{x},\mathbf{y},\mathbf{z}$	$\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3$	$\sigma_1,\sigma_2,\sigma_3$	$\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3$
Bivector	$\mathbf{yz}$, $\mathbf{zx}$, $\mathbf{xy}$	$\mathbf{e}_2\mathbf{e}_3,\mathbf{e}_3\mathbf{e}_1,\mathbf{e}_1\mathbf{e}_2$	$\sigma_2\sigma_3,\sigma_3\sigma_1,\sigma_1\sigma_2$	$\mathbf{e}_{23},\mathbf{e}_{31},\mathbf{e}_{12}$
Trivector/ Pseudoscalar	$I = \mathbf{xyz}$	$I = \mathbf{e}_1\mathbf{e}_2\mathbf{e}_3$	$I = \sigma_1\sigma_2\sigma_3$	$\mathbf{e}_{123}$

Hestenes [e.g., Reference 6] and Gull et al. [8] employed a notation similar to Doran and Lasenby [27], but they used the symbol $\boldsymbol{\sigma}$ rather than $\mathbf{e}$ to make it consistent with their notation for the even subalgebra of spacetime (see Table 7.1). Lounesto [28] used the single symbol $\mathbf{e}$ for all grades together with subscripts to indicate the implied product; for example, $\mathbf{e}_{123} \equiv \mathbf{e}_1\mathbf{e}_2\mathbf{e}_3$. Whenever we need to use indices in this text, we use $\mathbf{e}_k$ or $\boldsymbol{e}_k$. For example, $\mathbf{e}_x = \mathbf{x}$ or $\boldsymbol{e}_t = \boldsymbol{t}$. These examples of notation are only representative and are not intended to be either exclusive or exhaustive.

therefore just the same as for vectors and so addition of multivectors proceeds in just the same way, that is to say,

$$\boldsymbol{U}+\boldsymbol{V}=(U_0+V_0)+(U_1+V_1)\boldsymbol{X}_1+(U_2+V_2)\boldsymbol{X}_2+\ldots \tag{2.4}$$

The only difference is that we have an expanded set of basis elements of different types, or more precisely, different grades.

We shall always take our basis elements to be an orthonormal set of vectors together with the set of unique elements that is generated by their multiplication. The spatial basis vectors will always form a right-handed set. The basis elements for 3D are shown in Table 2.2, both in our notation and two other forms that are commonly seen in the literature for comparison. We have already shown the structure of the 4D basis elements in Table 2.1(b).

Finally, some further points about terminology and notation. We have adopted the term (3+1)D as a way of describing a physical view of the world that is consistent with the theme of space plus time. The 3D geometric algebra allows us to express this succinctly in multivector expressions such as $t+\mathbf{r}$. Strictly speaking, (3+1)D only refers to the physical model, whereas mathematically speaking, the geometric algebra itself is simply 3D. It would seem pedantic to adhere to the observation of this distinction in every instance. The reader has only to be aware that if the context is mathematical rather than physical, then 3D is the intended meaning.

Since it is already standard practice in traditional 3D vector algebra to use boldface characters for the representation of vectors, in geometric algebra it is conventional to reserve this notation specifically for 3D vectors. Normal weight italic characters are then used for the symbols of all other entities. This rule helps to distinguish the (3+1)D form of any equation from its spacetime counterpart, as we shall see. There is also a preference toward writing only multivectors in uppercase, for example, u may be either a scalar or vector while $\boldsymbol{U}$ is generally a multivector of some sort, but this is not hard and fast and certainly those vectors that are conventionally labeled in uppercase such as $\mathbf{E}$, $\mathbf{B}$, $\mathbf{J}$ are left as they stand. However, since this work is intended for a readership that will be initially unfamiliar with geometric algebra, we have made a slight concession to these rules. To avoid confusion between scalars and vectors, we will stay with the familiar rule whereby the former will be shown in normal weight while the latter will be in bold. However, it will make sense to extend the general principle here by representing pseudoscalars and scalars in the same way, while all other classes of object will be represented in boldface. Examples are therefore

- scalars and pseudoscalars — a, I, Ia, $\mathfrak{E}$, $\mathfrak{I}$, $\mathfrak{K}$, ρ
- vectors in general — $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{t}$, $\boldsymbol{r}$, $\boldsymbol{R}$
- vectors specific to 3D — $\mathbf{d}$, $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$, $\mathbf{r}$, $\mathbf{E}$
- bivectors and multivectors — $\boldsymbol{B}$, $\boldsymbol{U}$, $\boldsymbol{V}$, $\boldsymbol{\Gamma}$
- multivector expressions — $t+\mathbf{r}$, $\mathbf{E}+\boldsymbol{B}$, $\rho-\mathbf{J}$

Note here the scalar t versus the vector $\boldsymbol{t}$, the vector $\mathbf{B}$ versus its bivector form $\boldsymbol{B}$, and the (3+1)D vector $\mathbf{r}$ versus the general form $\boldsymbol{r}$. The reason for such distinctions is that it will actually prove convenient to use the same symbol in different contexts. When it comes to writing equations by hand, one may choose to write $\boldsymbol{z}$ as either $\mathbb{z}$ or $\bar{z}$ and, in the case of a bivector, $\boldsymbol{N}$ as either $\mathbb{N}$ or $\bar{\bar{N}}$. An alternative strategy is simply to associate each symbol with only one constant or variable, making it no longer necessary to distinguish between different classes of object. The disadvantages with this approach, however, are that there never seems to be a sufficient supply of convenient symbols and also that it is necessary to remember what type of object each one represents.

We also refrain from using the symbol i for either the unit pseudoscalar or $\sqrt{-1}$, and instead use I for the former and j, the engineering preference, for the latter, should it be required. Because of the prevalent use of the symbol I for the unit pseudoscalar, we will need to use the script form $\mathcal{I}$ as the symbol for current. Finally, the use of the caret or "hat" to identify unit basis vectors will be avoided and we shall simply use, for example, $\mathbf{x},\mathbf{y},\mathbf{z}$. The caret itself will be reserved as a normalization operator, as in $\hat{\mathbf{n}} \equiv |\mathbf{n}|^{-1}\mathbf{n}$. There are some exceptions to these rules, but in the rare circumstances in which these occur, both the reason and the intent should be clear from the context.

2.1 THE BASIC RULES OF A GEOMETRIC ALGEBRA

All the basic rules of a geometric algebra are summarized for reference in Appendix 14.4.2. Our aim here, however, is to develop these rules gradually starting from the basic concepts. First of all, let us introduce the basic properties of multiplication for any geometric algebra. Taking any pair of vectors, that is to say 1-vectors, $\boldsymbol{u}$ and $\boldsymbol{v}$, we proceed by postulating multiplication between them as follows:

- Unless stated to the contrary, scalars are always real.
- The geometric product of $\boldsymbol{u}$ and $\boldsymbol{v}$ is written as $\boldsymbol{uv}$. This may also be referred to as the direct product.
- For $\boldsymbol{u} /\!/ \boldsymbol{v}$, then $\boldsymbol{uv}$ is a scalar and $\boldsymbol{uv} = \boldsymbol{vu}$. We would recognize this result as being the same as $\boldsymbol{u}\cdot\boldsymbol{v}$ for parallel vectors in a conventional linear algebra.
- Conversely, for $\boldsymbol{u} \perp \boldsymbol{v}$, $\boldsymbol{uv}$ is a new object called a 2-vector, or bivector, and $\boldsymbol{uv} = -\boldsymbol{vu}$.
- Multiplication is therefore not necessarily commutative, but it is always associative, that is to say, no brackets are needed.
- In general, the product of n mutually orthogonal vectors forms an n-vector.
- The length, or measure, of $\boldsymbol{u}$ will be denoted by $|\boldsymbol{u}|$ where $|\boldsymbol{u}| = |\boldsymbol{u}^2|^{1/2}$.

These rules therefore directly relate the commutation properties of objects in a geometric algebra to the notions of parallel and perpendicular. Under multiplication, parallel vectors commute whereas orthogonal vectors anticommute. This will turn out be a recurrent theme that we try to exploit wherever possible because most of us have an intuitive idea of what perpendicular and parallel should mean. Taking a

geometric algebra as being a completely abstract concept, we could define parallel and perpendicular from the commutation properties alone with no geometric interpretation attached. On the other hand, we can take our geometric interpretation of parallel and perpendicular and show that it fits in with the idea that multiplication of parallel vectors gives a scalar result, from which we can establish a measure of length, while the product of perpendicular vectors defines a new object, a bivector, that can be identified with an orientated area as shown in Figure 2.1(b), (d), and (e). This in turn fits back in with the commutation properties. When the order of multiplication is reversed, in one case we must have the same result, for example, $\boldsymbol{u}$ multiplied by itself, while for the other case where $\boldsymbol{u} \perp \boldsymbol{v}$, the change in sign between $\boldsymbol{uv}$ and $\boldsymbol{vu}$ is to be associated with the opposite orientation of the object.

The basic rules imply the following:

- The product of any two vectors results in a scalar plus a bivector.
- $\boldsymbol{u}^2$, the product of any vector $\boldsymbol{u}$ with itself, is a scalar.
- There is therefore no need to define a separate inner product as the basis of the metric.
- There is the possibility of $\boldsymbol{u}^2$ being positive, negative, or even zero.
- For any two vectors that are orthogonal, $\boldsymbol{uv} + \boldsymbol{vu} = 0$. This is also a relationship that does not require an inner product to be defined and yet it is comparable with $\boldsymbol{u} \cdot \boldsymbol{v} = 0$.
- Since $\boldsymbol{u}^2$ is a scalar, any vector $\boldsymbol{u}$ possesses an inverse equal to $\left(1/\boldsymbol{u}^2\right)\boldsymbol{u}$ provided that $\boldsymbol{u}^2 \neq 0$.
- Under geometric multiplication, an N-dimensional vector space becomes an N-dimensional geometric algebra.
- The original vectors together with all the n-vectors that can be generated form a larger vector space of dimension 2^N.
- The highest possible grade object in an N-dimensional geometric algebra is an N-vector.

If we treat $\boldsymbol{u}^2$ as being the same thing as $\boldsymbol{u} \cdot \boldsymbol{u}$, the possibility of $\boldsymbol{u}^2$ being positive, negative, or zero also occurs in linear algebras over the complex numbers. However, in the geometric algebras that we will be dealing with the scalars will always be real. In the 3D geometric algebra, all vectors will have positive squares. In the 4D spacetime algebra, this will not be the case. For any given geometric algebra, this is determined by choosing a set of basis vectors with the required properties according to whether each has a positive or negative square. This choice is referred to as the metric signature.

2.2 3D GEOMETRIC ALGEBRA

We now use 3D as an example with which to demonstrate the application of the basic principles laid down so far. Given an orthonormal basis $\mathbf{x}, \mathbf{y}, \mathbf{z}$, by virtue of the simple multiplication rules and the given orthonormality of the basis vectors,

we have $\mathbf{x}^2 = \mathbf{y}^2 = \mathbf{z}^2 = 1$ from the parallel rule, while from the perpendicular rule, $\mathbf{yx} = -\mathbf{xy}$, $\mathbf{zy} = -\mathbf{yz}$, $\mathbf{zx} = -\mathbf{xz}$. From these products, we may then select three independent bivectors $\mathbf{yz}, \mathbf{zx}, \mathbf{xy}$. Note that the order of multiplication affects only the sign (we have taken the pairings in cyclic order but this is purely a matter of convenience) and that self-multiplication produces only a scalar. Going to the next stage so as to create trivectors, we can start with $\mathbf{x}(\mathbf{yz})$ to get $\mathbf{xyz}$, which is what we are looking for, whereas in contrast, $\mathbf{y}(\mathbf{zy}) = -\mathbf{yyz} = -\mathbf{y}^2\mathbf{z} = -\mathbf{z}$ is only a vector. By reducing such products to the simplest possible form, it is clear that the result is a vector when any two of the vectors involved are the same, for we can reorder the product so that this pair of vectors must be multiplied together, with the consequence that their product reduces to a scalar. To create a trivector, therefore, each of the three basis vectors must be different. Since again the order of multiplication affects only the sign, the only distinct trivector that can be created is $\mathbf{xyz}$. There is therefore just one trivector. Now, for the reasons just stated, we have $\mathbf{xyz} = -\mathbf{zyx}$ so that $(\mathbf{xyz})^2 = -\mathbf{xyzzyx}$, which simply reduces to -1. Probing further, it is clear that in 3D the product of any four basis vectors must have at least two vectors the same so that the result can at most be a bivector. In 3D, therefore, we cannot go past the trivector. As can be readily verified, it turns out that the 3D trivector commutes with everything. It is an intriguing fact that every geometric algebra must have at least one such "top" entity I that is similar to a unit scalar. Care must be taken, however, because

- I is not a true scalar;
- in some geometric algebras, I does not commute with everything;
- nor need we necessarily have $I^2 = -1$; and
- other entities may also have a negative square, for example, a bivector such as $\mathbf{xy}$.

It should now be obvious why we chose to avoid the same symbol as for the imaginary unit. This point very clearly demonstrates that the rules for 3D cannot be simply extrapolated into general principles. In particular, it should also be noted that not all spaces are Euclidean so that instead of having $\boldsymbol{u}^2$ being positive for every vector $\boldsymbol{u}$, we can have it being negative or zero, as is the case in spacetime. Nevertheless, this causes no great difficulty, and the basic rules still apply with very little modification as we shall see. Finally, one last point about notation for basis elements. As shown in Table 2.2, different forms of notation may be encountered in 3D. They are all variations on a theme. We generally avoid indices in this book except for those very few occasions when it is necessary to write a summation. For example, as in $\sum_k u_k \mathbf{e}_k$ where $\mathbf{e}_x, \mathbf{e}_y \ldots$ means the same as $\mathbf{x}, \mathbf{y} \ldots$ and so on. Lounesto [28] did not write out the higher-grade basis elements as products, as for example in $\mathbf{e}_3\mathbf{e}_1$ and $\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3$, but for compactness, he represented them as $\mathbf{e}_{31}, \mathbf{e}_{123}$, and so on. The letters designating the basis vectors vary. Although the $\sigma_1, \sigma_2, \sigma_3$ variety has its roots in quantum mechanics, it is commonly seen in the general literature of geometric algebra, along with its spacetime counterpart $\gamma_0, \gamma_1, \gamma_2, \gamma_3$, as, for example, in the case of the authors indicated.

2.3 DEVELOPING THE RULES

The results of multiplication for all entities in 3D can be worked out algebraically from the rules given so far, and it is something that interested readers should attempt for themselves, but it is also instructive to look at how it all turns out as a whole as shown, for example, in Table 2.3. Scalars are omitted since their multiplication with all entities is trivial, affecting only magnitude and sign. While the **x**-**y**-**z** notation helps to make the structure of the multiplication table quite clear, note that it is not necessary to have such a straightforward relationship between the bivectors and basis vectors, nor is it even necessary that the basis vectors be orthonormal; it just happens to be very convenient because it makes the relationships as simple as possible. As in any vector space, we can choose different sets of basis vectors based on linear combinations of any given set, and, not only that, we can apply this independently to the other basis elements as well. We could therefore regard the set of basis elements we have chosen, in which each element is simply the product of n basis vectors (where n ranges from 0 to the dimension of the space), as a sort of canonical form.

Table 2.3 The 3D Multiplication Table

U \ V	**x**	**y**	**z**	**yz**	**zx**	**xy**	I = **xyz**
x	1	**xy**	**−zx**	I	**−z**	**y**	**yz**
y	**−xy**	1	**yz**	**z**	I	**−x**	**zx**
z	**zx**	**−yz**	1	**−y**	**x**	I	**xy**
yz	I	**−z**	**y**	−1	**−xy**	**zx**	**−x**
zx	**z**	I	**−x**	**xy**	−1	**−yz**	**−y**
xy	**−y**	**x**	I	**−zx**	**yz**	−1	**−z**
I = **xyz**	**yz**	**zx**	**xy**	**−x**	**−y**	**−z**	−1

Legend

	$\boldsymbol{UV} = \boldsymbol{U} \cdot \boldsymbol{V}$ and $\boldsymbol{U} \wedge \boldsymbol{V} = 0$
	$\boldsymbol{UV} = \boldsymbol{U} \wedge \boldsymbol{V}$ and $\boldsymbol{U} \cdot \boldsymbol{V} = 0$
	$\boldsymbol{UV} = \boldsymbol{U} \times \boldsymbol{V}$ and $\boldsymbol{U} \cdot \boldsymbol{V} = \boldsymbol{U} \wedge \boldsymbol{V} = 0$

Multiplication by scalars is trivial and is therefore not shown. Note how multiplication among bivectors is the same as multiplication among vectors with just a change of sign. Multiplication by the pseudoscalar exchanges vectors with bivectors and also scalars with pseudoscalars. Reading across from any object $\boldsymbol{U}$ in the leftmost column, and down from any object $\boldsymbol{V}$ in the top row, then the geometric product $\boldsymbol{UV}$ may be read of from the intersecting cell. The shading of the cells indicates special cases such as $\boldsymbol{UV} = \boldsymbol{U} \cdot \boldsymbol{V}$ and $\boldsymbol{UV} = \boldsymbol{U} \wedge \boldsymbol{V}$, as explained in the legend.

2.3.1 General Rules

To expand on the basic ideas, we begin by making use of the intuitive notion that given some vector $\boldsymbol{u}$, any other vector $\boldsymbol{v}$ may be written as the sum of two parts, say $\boldsymbol{v}_{/\!/}$ and $\boldsymbol{v}_{\perp}$, which are respectively parallel and perpendicular to $\boldsymbol{u}$. If any proof is needed, see Theorem (6) in Appendix 14.4.2. The product $\boldsymbol{uv}$ may therefore be written as $\boldsymbol{uv}_{/\!/} + \boldsymbol{uv}_{\perp}$. From the discussion of Section 2.1, however, $\boldsymbol{uv}_{/\!/}$ is a scalar while $\boldsymbol{uv}_{\perp}$ is a bivector. This leads to some of the following more general rules:

- The product of any two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ results in a scalar plus a bivector.
  - This may be written formally as

$$\boldsymbol{uv} = \boldsymbol{u} \cdot \boldsymbol{v} + \boldsymbol{u} \wedge \boldsymbol{v} \tag{2.5}$$

  - The scalar $\boldsymbol{u} \cdot \boldsymbol{v}$ is equal to $\boldsymbol{uv}_{/\!/}$, the product of the parallel parts of $\boldsymbol{u}$ and $\boldsymbol{v}$. This is referred to as the **inner product**.
  - The bivector $\boldsymbol{u} \wedge \boldsymbol{v}$ is equal to $\boldsymbol{uv}_{\perp}$, the product of the orthogonal parts. This is referred to as the **outer product**. Figure 2.1(f) illustrates this way of forming bivectors when the two vectors involved are not orthogonal.
  - Since $\boldsymbol{uv}_{/\!/} = \boldsymbol{v}_{/\!/}\boldsymbol{u}$ for any two parallel vectors such as $\boldsymbol{u}$ and $\boldsymbol{v}_{/\!/}$, it then follows that in general, $\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{u}$.
  - Since $\boldsymbol{uv}_{\perp} = -\boldsymbol{v}_{\perp}\boldsymbol{u}$ for any two orthogonal vectors such as $\boldsymbol{u}$ and $\boldsymbol{v}_{\perp}$, it follows that in general, $\boldsymbol{u} \wedge \boldsymbol{v} = -\boldsymbol{v} \wedge \boldsymbol{u}$.
- The properties given above result in the following standard formulas for the inner and outer products between any two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$:

$$\begin{aligned} \boldsymbol{u} \cdot \boldsymbol{v} &= \tfrac{1}{2}(\boldsymbol{uv} + \boldsymbol{vu}) \\ \boldsymbol{u} \wedge \boldsymbol{v} &= \tfrac{1}{2}(\boldsymbol{uv} - \boldsymbol{vu}) \end{aligned} \tag{2.6}$$

- The product of n mutually orthogonal vectors always generates an object of grade n.
  - But unless they are actually orthogonal, objects of other grades will be generated too.
- The outer product of n linearly independent vectors, irrespective of whether or not they are actually orthogonal, will also generate an object of grade n.
- If any two multivector quantities are equal, then each of their separate parts of each grade (scalar, vector, bivector, etc.) must be equal.
  - The grade filtering function $\langle \boldsymbol{U} \rangle_k$ returns the part of $\boldsymbol{U}$ that is of grade k.
  - By convention, if the subscript has been omitted, then the scalar part is implied, that is to say $\langle \boldsymbol{U} \rangle \equiv \langle \boldsymbol{U} \rangle_0$.
  - We can therefore say $\boldsymbol{U} = \boldsymbol{V} \Leftrightarrow \langle \boldsymbol{U} \rangle_k = \langle \boldsymbol{V} \rangle_k$ for $0 \le k \le N$.
  - In addition, $\boldsymbol{u} \cdot \boldsymbol{v} = \langle \boldsymbol{uv} \rangle_0$ and $\boldsymbol{u} \wedge \boldsymbol{v} = \langle \boldsymbol{uv} \rangle_2$ provides a different form of Equation (2.6).

- Any two multivector quantities may be multiplied by first decomposing each of them into separate parts of each grade, then multiplying them part by part according to the rules (distributive law).
  - One simple way of applying this is to express the multivectors involved as linear combinations of the chosen basis elements.
- Like addition, multiplication is associative so that

$$\boldsymbol{U}(\boldsymbol{VW}) = (\boldsymbol{UV})\boldsymbol{W} = \boldsymbol{UVW} .$$

- But it is not necessarily either commutative or anticommutative.
- An object $\boldsymbol{U}$ is said to be null if $\boldsymbol{U}^2 = 0$.
- Every object that is not null may possess an inverse, for example, $(t+\boldsymbol{u})^{-1} = (t-\boldsymbol{u})/(t^2 - \boldsymbol{u}^2)$ provided that $\boldsymbol{u}^2 \neq t^2$.

Figure 2.2 attempts to give a pictorial interpretation of how a geometric algebra works under such rules. It is not intended to be taken too literally, but it does give some idea of how multiplication and addition between objects of different grades can be handled. The inner and outer products, however, are particularly important and deserve specific comment on their interpretation.

2.3.2 3D

For the specific case of the 3D geometric algebra, the general rules above imply the following:

- Pseudoscalars commute with everything.
- The unit pseudoscalar I obeys $I^2 = -1$ and fills a role comparable to the imaginary unit j.
- The product of a vector $\mathbf{u}$ with a bivector $\boldsymbol{V}$ results in a vector plus a pseudoscalar.
  - The vector part, $\langle \mathbf{u}\boldsymbol{V} \rangle_0$, is the inner product and is written as $\mathbf{u} \cdot \boldsymbol{V}$, and

$$\mathbf{u} \cdot \boldsymbol{V} = -\boldsymbol{V} \cdot \mathbf{u}$$

  - The pseudoscalar part, $\langle \mathbf{u}\boldsymbol{V} \rangle_3$, is the outer product and is written as $\mathbf{u} \wedge \boldsymbol{V}$, and

$$\mathbf{u} \wedge \boldsymbol{V} = \boldsymbol{V} \wedge \mathbf{u}$$

- Any bivector can be replaced by the product of a pseudoscalar with a vector, for example, $\boldsymbol{P} = I\boldsymbol{p}$.
  - Bivectors therefore have negative squares.

- Bivectors can therefore be thought of as the counterparts of axial vectors.
  - Like an axial vector, a bivector is unaffected by inversion.
  - However, this is no longer information that we need to carry in our head.
  - Also, the geometric interpretation is quite different, being the plane area itself rather than the vector normal to it (Figure 2.1b).
- Given the relationship $\boldsymbol{P} = I\boldsymbol{p}$, a bivector may be referred to as a pseudovector.
  - A pseudovector is always the dual of a vector, so that in spaces of any other dimension it will not be a bivector; for example, in 4D a pseudovector is a trivector.

As a result, in 3D we need to use only vector multiplication. It is only necessary to write all bivectors as pseudovectors and treat I as though it were a scalar, employing the rule $I^2 = -1$ whenever necessary. Take as an example the product of two different multivectors expressed in terms of the usual basis elements as $(a\mathbf{x} + b\mathbf{xy})$ and $(c + \mathbf{yz})$, say. Then their product may be evaluated as

$$\begin{aligned}(a\mathbf{x} + b\mathbf{xy})(c + \mathbf{yz}) &= (a\mathbf{x} + bI\mathbf{z})(c + I\mathbf{x}) \\ &= ac\mathbf{x} + bcI\mathbf{z} + bI^2\mathbf{zx} + aI\mathbf{x}^2 \\ &= \underset{\text{vector}}{ac\mathbf{x}} + \underbrace{I(bc\mathbf{z} - b\mathbf{y})}_{\text{bivector}} + \underset{\text{pseudoscalar}}{Ia}\end{aligned} \tag{2.7}$$

This approach often helps to simplify multivector products in any dimension of space. Other geometric algebras may obey similar rules, but each needs to be examined as a special case to see how the general rules should apply.

2.3.3 The Geometric Interpretation of Inner and Outer Products

We have seen that the multiplication of two mutually orthogonal vectors, **a** and **b**, creates an object **ab** of grade 2, and that multiplication by a third mutually orthogonal vector **c** creates an object **abc** that is of grade 3. The formation of these objects has a geometric interpretation, as shown, for example, in Figure 2.1(a)–(c). Starting from **a** and then multiplying on the right by **b**, followed in turn by **c**, at each stage we are increasing the dimension of the object by 1. If the vectors are not all orthogonal but are at least all linearly independent, then the outer products $\mathbf{a} \wedge \mathbf{b}$ and $\mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c}$ achieve the same thing, for this is equivalent to multiplying only the mutually orthogonal parts of the vectors. Figure 2.1(f) illustrates this for the simple case of constructing a bivector from two nonorthogonal 3D vectors, **a** and **d**. In general, the outer product of two vectors such as **a** and **b** can result in either of two things: if **b** is linearly dependent on **a**, the result is zero; otherwise, it gives rise to a new object of grade 2. The only difference in taking $\mathbf{b} \wedge \mathbf{a}$ rather than $\mathbf{a} \wedge \mathbf{b}$ is a change of sign. If **b** is linearly dependent on **a**, this effectively means that **b** is parallel to **a**. Otherwise, **b** must include some part that is orthogonal to **a**. It follows that we

may split $\mathbf{b}$ into $\mathbf{b}_{/\!/} + \mathbf{b}_{\perp}$ such that $\mathbf{b}_{/\!/}$ lies within the same 1D space as $\mathbf{a}$, whereas $\mathbf{b}_{\perp}$ lies outside it, that is to say $\mathbf{b}_{\perp}$ is orthogonal to $\mathbf{a}$. It then must be the case that on the one hand, $\mathbf{a} \wedge \mathbf{b}_{/\!/} = 0$, while on the other, $\mathbf{a} \wedge \mathbf{b}_{\perp} = \mathbf{a}\mathbf{b}_{\perp} = \mathbf{a} \wedge \mathbf{b}$. Therefore, if the *outer* product of vector $\mathbf{a}$ with vector $\mathbf{b}$ effectively adds a new dimension to $\mathbf{a}$, we see that the new dimension comes from the part of $\mathbf{b}$ that lies *outside* the space occupied by $\mathbf{a}$. This also extends to higher grades when we take the outer product of $\mathbf{a} \wedge \mathbf{b}$ with $\mathbf{c}$, effectively multiplying $\mathbf{a} \wedge \mathbf{b}$ by $\mathbf{c}_{\perp}$, where $\mathbf{c}_{\perp}$ is mutually perpendicular to both $\mathbf{a}$ and $\mathbf{b}$. Provided $\mathbf{c}_{\perp} \neq 0$, that is to say $\mathbf{c}$ is not linearly dependent on $\mathbf{a}$ and $\mathbf{b}$, this adds a new dimension to the bivector $\mathbf{a} \wedge \mathbf{b}$ to form the trivector $\mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c}$. If, on the other hand, $\mathbf{c}_{\perp} = 0$, that is to say $\mathbf{c}$ *is* linearly dependent on $\mathbf{a}$ and $\mathbf{b}$, then the result vanishes, there being no new dimension to add.

The outer product of an object with a vector may therefore be seen as a "step-up" operation, but what about the inner product? Let us begin with a bivector $\boldsymbol{U}$ that happens to take the convenient form $\mathbf{a} \wedge \mathbf{b}$. Following the same process described above, we may equally well represent $\boldsymbol{U}$ as $\mathbf{a}\mathbf{b}_{\perp}$. If we multiply $\mathbf{a}\mathbf{b}_{\perp}$ on its right by $\mathbf{b}_{\perp}$, the result will be $\mathbf{a}\mathbf{b}_{\perp}^2 = b_{\perp}^2\mathbf{a}$, bringing us back to an object of grade 1, in fact a vector that is parallel to $\mathbf{a}$. If instead we multiply on the left, as a result of anticommuting $\mathbf{b}_{\perp}$ past $\mathbf{a}$ (to which it is of course orthogonal), we find $\mathbf{b}_{\perp}\mathbf{a}\mathbf{b}_{\perp} = -b_{\perp}^2\mathbf{a}$ so that, apart from a change of sign, the effect is just the same in that the object $\boldsymbol{U}$ still loses a dimension and results in another vector parallel to $\mathbf{a}$. More specifically, it loses the dimension associated with the vector $\mathbf{b}_{\perp}$ that we multiplied it by, and, instead of saying that the resulting vector is parallel to $\mathbf{a}$, we could just as well say that it is orthogonal to $\mathbf{b}_{\perp}$.

It is clear that similar results would be obtained by multiplying $\boldsymbol{U}$ with $\mathbf{a}$ or, in fact, by multiplying it with any other vector $\mathbf{v}_{/\!/}$ that lies in the plane spanned by $\mathbf{a}$ and $\mathbf{b}$. The outcome would always be a vector that is, first of all, perpendicular to $\boldsymbol{v}_{/\!/}$, the stripped-out dimension and, second, lies in the $\mathbf{a} \wedge \mathbf{b}$ plane. While this process works for a vector lying in the plane spanned by $\mathbf{a}$ and $\mathbf{b}$, what about the general case for a vector $\mathbf{v} = \mathbf{v}_{/\!/} + \mathbf{v}_{\perp}$, which has a part, $\boldsymbol{v}_{\perp}$, that is perpendicular to the bivector plane? By simply multiplying $\boldsymbol{U}$ with $\mathbf{v}$, in addition to generating $\boldsymbol{U}\mathbf{v}_{/\!/}$, we would also generate $\boldsymbol{U}\boldsymbol{v}_{\perp}$, which is, in fact, the same as trivector $\boldsymbol{U} \wedge \mathbf{v}$. To avoid this problem, the specific process for reducing a grade is more systematically represented by the *inner* product rather than the basic geometric product. For example, $\mathbf{v} \cdot \boldsymbol{U}$ and $\boldsymbol{U} \cdot \mathbf{v}$ respectively result in $\mathbf{v}_{/\!/}\boldsymbol{U}$ and $\boldsymbol{U}\mathbf{v}_{/\!/}$ so that both of these inner products reduce the grade of $\boldsymbol{U}$ by removing the dimension contributed by $\mathbf{v}_{/\!/}$, that is to say, the part of $\mathbf{v}$ that lies within the plane defined by $\boldsymbol{U}$. Consequently, as illustrated in Figure 2.1(h), the resulting object, a vector, must be orthogonal to $\mathbf{v}_{/\!/}$. Figure 2.1(i), on the other hand, shows the contrasting process of forming the trivector $\boldsymbol{U} \wedge \mathbf{v}$ from $\boldsymbol{U}\boldsymbol{v}_{\perp}$.

While this discussion provides some introduction to the geometric meaning of these two important types of operation and goes some way to illustrate their underlying principles, the general rules for inner and outer products between vectors and any sort of object will be set out in Section 4.5. Following on from that, this is extended in Section 4.6 to inner and outer products between two objects of any grade. Despite the fact that the concept becomes harder to visualize with objects of

higher dimension, the outer product is always some sort of step-up operation, adding one or more dimensions, whereas the inner product is always a "step-down," or grade reduction, operation that strips out one or more dimensions from an object.

2.4 COMPARISON WITH TRADITIONAL 3D TOOLS

Table 2.3 shows that multiplication of the basis vectors follows two distinct patterns. For the off-diagonal cases, it is much like the cross product, but for those on the diagonal, it is like the dot product. This is consistent with the following useful relationship:

$$\mathbf{uv} = \mathbf{u} \cdot \mathbf{v} + I\mathbf{u} \times \mathbf{v} \tag{2.8}$$

That is to say, we could actually define the cross product through $\mathbf{u} \times \mathbf{v} = -I\,\mathbf{u} \wedge \mathbf{v}$. It is permissible to use the cross product here since the 3D context is evident from the use of bold erect symbols for $\mathbf{u}$ and $\mathbf{v}$. Taking the example $\mathbf{u} = \mathbf{v} = \mathbf{x}$, Equation (2.8) gives $\mathbf{xx} = x^2 = 1$ as required, while for $\mathbf{u} = \mathbf{x}, \mathbf{v} = \mathbf{y}$ we get $\mathbf{xy} = I\mathbf{z}$, which also proves correct. Expressing any pair of vectors $\mathbf{u}$ and $\mathbf{v}$ as linear combinations of the basis vectors proves this result in a general way. It is to be noted that some authors are against the use of the cross product, preferring instead to use the outer product alone, but the cross product is fairly well embedded throughout the disciplines of physics and engineering and, at the least, it is useful to be able to translate to the one from the other. It is true to say, however, that we should treat both the cross product and the axial vector with due caution.

While thus far the 3D geometric algebra seems to be simply a restatement of traditional ideas in different terms, we have gained the ability to represent directed areas and volumes. Also, the notion that a vector cross product transfers into a 3D geometric algebra as a bivector rather than a vector seems to resolve the ambiguity that exists between true and axial vectors at once. This is more than just semantics because in an ordinary vector space, true and axial vectors have identical basis elements, $\mathbf{x},\mathbf{y},\mathbf{z}$, while in a geometric algebra, vectors and bivectors have the distinct sets of basis elements $\mathbf{x},\mathbf{y},\mathbf{z}$ and $I\mathbf{x}, I\mathbf{y}, I\mathbf{z}$. Note that here we have made use of the dual form to represent the bivectors, as shown in Figure 2.1(g). Furthermore, I is a trivector and not simply the imaginary unit j. As we shall see in due course, there will be no need for complex scalars. Geometric algebra has its own way of dealing with equations such as $\boldsymbol{U}^2 = -1$. Indeed, equations such as this have new meaning, for we do not need to say in advance what sort of entity $\boldsymbol{U}$ might be. Finally, we also have a formalism that will work in any dimension of space, not just 3D, a theme we will develop in Chapter 4.

2.5 NEW POSSIBILITIES

It is worthwhile taking some time to develop this point a little further. Sticking at present with 3D, we begin by writing $\boldsymbol{U}$ in the most general form that a multivector may

take with components of every possible grade, namely $\boldsymbol{U} = U_0 + U_1\mathbf{u} + U_2 I\mathbf{v} + U_3 I$ where $U_0, \ldots U_3$ are real scalars, while $\mathbf{u}$ and $\mathbf{v}$ are any two unit vectors (so clearly $I\mathbf{v}$ is a bivector). It is interesting to see what happens when we form expressions under the new rules. For example, when we expand $\boldsymbol{U}^2$ and collect together terms of the same sort we find

$$\begin{aligned}\boldsymbol{U}^2 &= (U_0 + U_1\mathbf{u} + U_2 I\mathbf{v} + U_3 I)^2 \\ &= \underbrace{(U_0^2 + U_1)^2 - (U_2^2 + U_3^2)}_{\text{scalar}} + \underbrace{2U_0U_1\mathbf{u} - 2U_2U_3\mathbf{v}}_{\text{vector}} \\ &\quad + \underbrace{2U_0U_2 I\mathbf{v} + 2U_1U_3 I\mathbf{u}}_{\text{bivector}} + \underbrace{2(U_0U_3 + U_1U_2\mathbf{u}\cdot\mathbf{v})I}_{\text{pseudoscalar}}\end{aligned} \tag{2.9}$$

Note that, as already mentioned, whereas I commutes with everything, the vectors $\mathbf{u}$ and $\mathbf{v}$ do not in general commute. Also, $\mathbf{u}^2 = \mathbf{v}^2 = 1$ by assumption, and we have made use of Equation (2.6) to replace $\mathbf{uv} + \mathbf{uv}$ with $2\mathbf{u}\cdot\mathbf{v}$, a scalar quantity. From inspection, we can see that $\boldsymbol{U}^2$ can only result in a scalar if its vector, bivector and pseudoscalar parts all vanish separately. This results in three sorts of solutions to the equation $\boldsymbol{U}^2 = \lambda$:

$$\begin{aligned}&\text{(i)} && \boldsymbol{U} = \pm\sqrt{\lambda} \text{ for } 0 \le \lambda \\ &\text{(ii)} && \boldsymbol{U} = \pm I\sqrt{-\lambda} \text{ for } \lambda \le 0 \\ &\text{(iii)} && \boldsymbol{U} = U_1\mathbf{u} + U_2 I\mathbf{v} \text{ where } U_1^2 - U_2^2 = \lambda \text{ and } \mathbf{u} \perp \mathbf{v}\end{aligned} \tag{2.10}$$

While the first two solutions are easily grasped, the third solution is apparently in new territory. In particular, for $\lambda = 0$, we have $\boldsymbol{U} = U_1(\mathbf{u} \pm I\mathbf{v})$, which, as we shall see, is the form taken by the electromagnetic field of a plane wave. The concept of null multivectors extends to vectors themselves in other geometric algebras such as 4D spacetime. Another interesting case is $\lambda = -1$, since we can have the obvious solution $\boldsymbol{U} = \pm I$, or given $\mathbf{u} \perp \mathbf{v}$, a more unusual solution is $\boldsymbol{U} = \sinh(x)\mathbf{u} + \cosh(x)I\mathbf{v}$ for any value of x.

Beyond these simple yet intriguing results, it is clear that more complicated multivector equations, such as a quadratic, may provide even more unusual mathematics to investigate—but that is not what we are here to do.

Having come this far, it will now be both possible and useful to gain some appreciation of how geometric algebra benefits the study of electromagnetic theory. The rules that we already have for 3D will be nearly sufficient for this purpose; we will only need to take into consideration one or two subtleties, for example, the meaning of the inner product between two bivectors. Once the basis of the practical relationship between electromagnetic theory and geometric algebra has been established, we will go on in Chapter 4 to generalize the basic rules so as to make the new mathematical toolset sufficient to all our likely needs. Not only will this allow us to work with geometric algebras of any dimension, it will also help us to better understand the rules by gaining insight into the regularity of their overall structure, in particular as far as inner and outer products are concerned. In the meantime, we

will make the best of the basics with the benefit of a little further explanation as and when necessary.

2.6 EXERCISES

All questions employ the notational rules given above.

1. Evaluate $(\mathbf{yz})^2,(\mathbf{zx})^2,(\mathbf{xy})^2$ and $\mathbf{x}(\mathbf{yx})(\mathbf{zx})\mathbf{x}$.
2. Express $I\mathbf{x}, I\mathbf{y}, I\mathbf{z}, I\mathbf{yz}, I\mathbf{zx}, I\mathbf{xy}$ in their simplest forms without using I.
3. All of the following apply to the 3D geometric algebra.
   (a) Show that if the bivector $\boldsymbol{U} = I\mathbf{u}$ can also be expressed as $\boldsymbol{U} = \mathbf{vw}$, then $\mathbf{u}\cdot\mathbf{v} = \mathbf{u}\cdot\mathbf{w} = \mathbf{v}\cdot\mathbf{w} = 0$.
   (b) Show that any bivector may be written as the product of two vectors.
   (c) Show that this then means that it may also be written as the product of two bivectors.
   (d) Show that in 3D any pair of bivectors $\boldsymbol{A}$ and $\boldsymbol{B}$ have a common factor $\mathbf{u}$ (a vector) such that $\boldsymbol{A} = \mathbf{au}$ and $\boldsymbol{B} = \mathbf{bu}$.
   (e) Finally, show that this leads to a simple geometric interpretation of bivector addition.
4. Given orthonormal basis vectors $\mathbf{x},\mathbf{y},\mathbf{z}$, show that any bivector may be written as a linear combination of the three bivectors $\mathbf{yz},\mathbf{zx},\mathbf{xy}$.
5. Prove Equation (2.6).
6. Show that if the vectors $\mathbf{a}'$ and $\mathbf{b}'$ are related to the vectors $\mathbf{a}$ and $\mathbf{b}$ through the same rotation in the $\mathbf{a}\wedge\mathbf{b}$ plane, then $\mathbf{a}\cdot\mathbf{b} = \mathbf{a}'\cdot\mathbf{b}'$ and $\mathbf{a}\wedge\mathbf{b} = \mathbf{a}'\wedge\mathbf{b}'$.
7. (a) Show that $\boldsymbol{A}$, the bivector representing the directed area of a triangle with vertices $\mathbf{a},\mathbf{b},\mathbf{c}$, taken in that order (i.e., the edges are taken in turn as the vectors from $\mathbf{a}$ to $\mathbf{b}$, $\mathbf{b}$ to $\mathbf{c}$, and $\mathbf{c}$ to $\mathbf{a}$) is given by $2\boldsymbol{A} = \mathbf{a}\wedge\mathbf{b} + \mathbf{b}\wedge\mathbf{c} + \mathbf{c}\wedge\mathbf{a}$.
   (b) Extend this result to the case of simple polygons with ordered vertices $\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3\ldots\mathbf{a}_N$. $2\boldsymbol{A} = \mathbf{a}_1\wedge\mathbf{a}_2 + \mathbf{a}_2\wedge\mathbf{a}_3\ldots + \mathbf{a}_N\wedge\mathbf{a}_1$.
   (c) What happens when $\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3\ldots\mathbf{a}_N$ are not actually coplanar?
8. Based on the representation of a polygon as bivector, justify the representation of an orientated circular disc as a bivector, for example, as shown, in Figure 2.1(g). Given any two non-collinear radii of the disc, say $\mathbf{a}$ and $\mathbf{b}$, find the associated bivector.
9. Find a scalar expression for $(\mathbf{a}\wedge\mathbf{b})^2$.
10. By defining $e^{\boldsymbol{U}}$ as $1+\boldsymbol{U}+\dfrac{\boldsymbol{U}^2}{2!}\cdots+\dfrac{\boldsymbol{U}^n}{n!}+\cdots$, show that $e^{\mathbf{a}} = \cosh a + \sinh a\,\hat{\mathbf{a}}$ and $e^{I\mathbf{a}} = \cos a + \sin a\, I\hat{\mathbf{a}}$ where $a = |\mathbf{a}|$. Is it true that $e^{\mathbf{a}}e^{\mathbf{b}} = e^{\mathbf{a}+\mathbf{b}}$?
11. Find an expression for the vector that is the common factor of two given bivectors.

Chapter 3

Applying the Abstraction

The mathematical construct of a geometric algebra is all very well, but how do we apply it? And even if we can apply it, will it describe physical phenomena in any better way? Let us therefore look at some basic examples.

3.1 SPACE AND TIME

First, we may express the time and location of any event as a multivector $\boldsymbol{R}$ that is the combination of scalar and a vector in the form $\boldsymbol{R} = t + \mathbf{r}$. From this, it follows that if we are referring to some particle moving with velocity $\mathbf{v}$, then by differentiating $\boldsymbol{R}$ with respect to time we get $\boldsymbol{V} = \partial_t \boldsymbol{R} = 1 + \mathbf{v}$ as some sort of "velocity multivector" (recall $\partial_t \boldsymbol{R} \equiv \partial \boldsymbol{R} / \partial t$). There is even some freedom here since we could equally well have $\boldsymbol{R} = ct + \mathbf{r}$ or even $\boldsymbol{R} = -t + \mathbf{r}$, and so on. This approach is interesting because it means that there are new ways of dealing with old concepts and combining things we would normally keep separate. For example, we have already seen that $t + \mathbf{r}$ means the same sort of thing as $(t, \mathbf{r})$, which is regularly used as the argument of a function that depends on both time and position, that is to say $\boldsymbol{U}(t + \mathbf{r}) \equiv \boldsymbol{U}(t, \mathbf{r})$. It is clear that, even if it only amounts to being a simple scalar or vector, $\boldsymbol{U}$ itself must be some sort of multivector. We need only make sure that any such multivector function has a valid meaning within the laws of physics.

Any equation between multivector expressions must result in a valid equation between the terms of each grade on either side of the equation. For example, if we have $t + \mathbf{r}$ on one side of an equation, it does not matter that t and $\mathbf{r}$ are different types of entity and have different physical dimensions of time and length respectively, but the grades and dimensions of the entities on the other side of the equation must match exactly. In this example, we would need a scalar of dimension time and a vector of dimension length, that is to say $t + \mathbf{r} = a + \mathbf{b}$ is equivalent to two separate equations $t = a$ and $\mathbf{r} = \mathbf{b}$. In general, therefore a single multivector equation can represent as many separate equations as there are grades involved. The word "encoding" is often used in the literature to describe how geometric algebra represents a physical or mathematical process, and this case is an example of encoding a number

Understanding Geometric Algebra for Electromagnetic Theory, First Edition. John W. Arthur.

of separate but logically related equations in a single unified form. The ability of geometric algebra to combine time and space as a single multivector entity makes the description (3+1)D that we have been using seem quite appropriate. It reflects the way that geometric algebra represents or, better still, encodes the Newtonian world. Although he did not use the acronym (3+1)D, credit for this idea must actually go to Hamilton who saw this very same potential in his related quaternion algebra ([15] and Appendix 14.5):

> *my real* [the scalar part] *is the representative of a sort of fourth dimension, inclined equally to all lines in space* [20, Vol. 3, p. 254],

and, when he later realized that this fourth dimension could be time,

> *in technical language* [the quaternion] *may be said to be "time plus space", or "space plus time", and in this sense it has, or at least it involves, a reference to four dimensions.* [29, Vol. 3, p. 635]

The use of (3+1) dimensions as a form of mathematical shorthand for this idea came very much later in the context of special relativity when Hermann Weyl claimed "The world is a (3+1)-dimensional metrical manifold" [30, p. 283]. Eddington also adopted this [31, section 1.9] but, in the end, it was the notion of calling it a 4D world that actually took hold. It therefore seems quite appropriate to reuse the idea of (3+1)D here where we mean 3D space plus 1D time in the sense that Hamilton saw it,[1] where they are separate vector and scalar entities bound within a common mathematical framework without any particular regard to special relativity. That is to say, for us (3+1)D will be an essentially Newtonian model, in contrast with 4D, which we will associate with spacetime and gives a view consistent with special relativity.

3.2 ELECTROMAGNETICS

3.2.1 The Electromagnetic Field

Let us now examine some examples of how the electromagnetic field is encoded. The electric field vector $\mathbf{E}$, being a true vector, carries straight over into a geometric algebra since we can express it directly in terms of the unit basis vectors as $E_x\mathbf{x} + E_y\mathbf{y} + E_z\mathbf{z}$. But we will recall that an axial vector such as $\mathbf{B}$ behaves differently, and so it must be replaced by a bivector, which, according to our convention in which the form $\mathbf{B}$ is reserved for a vector, we write italicized as $\boldsymbol{B}$. As we have seen, the vector and bivector forms can be related through $\boldsymbol{B} = I\mathbf{B}$, but some care is required here because in geometric algebra, $\mathbf{B}$ will always represent a true

[1] Indeed, this notion fired his imagination so much that in 1846 he wrote a sonnet to the astronomer Sir John Herschel, the key lines of which are "One of Time, of Space the three, Might in the chain of symbol girdled be." [29, Vol. 2, p. 525]

vector, that is to say $\mathbf{B} = B_x\mathbf{x} + B_y\mathbf{y} + B_z\mathbf{z}$, because there is simply no separate form for an axial vector. The problem lies within the axial vector concept itself because it uses the same basis vectors as for true vectors. For consistency with the inversion test, as an axial vector $\mathbf{B}$ should really be written in the form $\mathbf{B} = B_x(\mathbf{y}\times\mathbf{z}) + B_y(\mathbf{z}\times\mathbf{x}) + B_z(\mathbf{x}\times\mathbf{y})$. Geometric algebra gets round this problem because the bivectors and vectors have separate basis elements in terms of which $\boldsymbol{B}$ is represented by $B_x\mathbf{yz} + B_y\mathbf{zx} + B_z\mathbf{xy}$. It is clear, however, that the vector $\mathbf{B}$ and the bivector $\boldsymbol{B}$ share the same components from

$$\begin{aligned}\boldsymbol{B} &= I\mathbf{B}\\ &= I\left(B_x\mathbf{x} + B_y\mathbf{y} + B_z\mathbf{z}\right)\\ &= B_xI\mathbf{x} + B_yI\mathbf{y} + B_zI\mathbf{z}\\ &= B_x\mathbf{yz} + B_y\mathbf{zx} + B_z\mathbf{xy}\end{aligned} \tag{3.1}$$

In going from one to the other, therefore, although the components stay the same, the basis elements change. To convert an axial vector into a bivector we simply keep the same components and substitute the bivector basis elements $\mathbf{yz}, \mathbf{zx}, \mathbf{xy}$ for $\mathbf{x}, \mathbf{y}, \mathbf{z}$.

Now, it may at first seem it is an imposition to have $\mathbf{E}$ being one sort of thing and $\mathbf{B}$ another, but in fact it is a convenience, for while it makes little sense to add $\mathbf{E}$ and $\mathbf{B}$, adding $\mathbf{E}$ and $\boldsymbol{B}$ poses no problem at all since, as a result of the difference in grades, the multivectors so formed will retain their separate characters. In fact, putting aside for the moment the question of units, we may define a new multivector field $\boldsymbol{F}$ that combines the two:

$$\boldsymbol{F} \equiv \mathbf{E} + c\boldsymbol{B} = \mathbf{E} + Ic\mathbf{B} \tag{3.2}$$

The constant c simply gives $\boldsymbol{F}$ the single dimension of Vm^{-1}, the same as for $\mathbf{E}$. We can go further and introduce a multivector electromagnetic source density $\boldsymbol{J}$ that combines both charge and current density. Let us propose the form

$$\boldsymbol{J} \equiv \tfrac{1}{\varepsilon_0}\rho - Z_0\mathbf{J} \tag{3.3}$$

where $Z_0 = (\mu_0/\varepsilon_0)^{\frac{1}{2}}$ is the characteristic impedance of free space. In this instance, we therefore have a combination of scalar and vector, each with dimensions of Vm^{-2}. The justification for the particular form it takes will come later in Section 5.2 when we find the equation that relates $\boldsymbol{F}$ to $\boldsymbol{J}$ but, in the meantime, it is of interest to note that if we link the current density $\boldsymbol{J}$ to the motion of the charge density through $\mathbf{J} = \rho\mathbf{v}$, which is the case when only a single type of charge is involved, we find $\boldsymbol{J} = \varepsilon_0^{-1}\rho(1 - \mathbf{v}/c)$. This suggests that if we start with a static charge density ρ, multiplying it by the factor $\varepsilon_0^{-1}(1 - \mathbf{v}/c)$ somehow transforms it to the moving source density $\boldsymbol{J}$. This, of course begs the question, if ρ gives rise to an electric field $\mathbf{E}$, will $\boldsymbol{J}$ then give rise to an electromagnetic field $\boldsymbol{F}$ that is directly related to $\mathbf{E}(1 - \mathbf{v}/c)$ or perhaps $(1 - \mathbf{v}/c)\mathbf{E}$? We will return to this idea in due course.

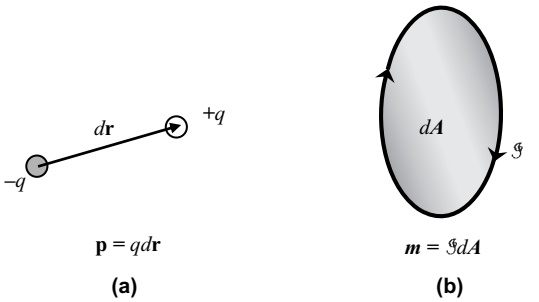

Figure 3.1 Comparison of polar and solenoidal dipoles. (a) The polar dipole, $\mathbf{p} = q\,d\mathbf{r}$, is unequivocally represented by a vector. (b) The solenoidal or current dipole is represented by a bivector. Given an orientated element of area $d\boldsymbol{A}$ around which a positive current $\mathfrak{I}$ runs in the same sense as the bivector, then $\boldsymbol{m} = \mathfrak{I}\,d\boldsymbol{A}$. However, we are probably more accustomed to the magnetic dipole in axial vector form, $\mathbf{m} = \mathfrak{I}\,d\mathbf{a}$, a representation that only helps to confuse it with the polar type.

3.2.2 Electric and Magnetic Dipoles

Since the electric dipole moment $\mathbf{p}$ resulting from two equal and opposite charges $+q$ and $-q$ separated by an infinitesimal displacement $d\mathbf{r}$ is given by $\mathbf{p} = qd\mathbf{r}$, it follows that $\mathbf{p}$ is of the same character as $d\mathbf{r}$, that is to say a vector, as illustrated in Figure 3.1(a).

In order to treat the magnetic dipole in a fundamental way, we cannot simply say (as is sometimes done either through analogy or even by default) that like its polar cousin, it too must be a vector. If the current dipole is the correct model, then the dipole moment produced by a current $\mathfrak{I}$ circulating around an infinitesimal loop of directed area $d\boldsymbol{A}$ is given by $\boldsymbol{m} = \mathfrak{I}d\boldsymbol{A}$, as in Figure 3.1(b), which has the character of a bivector just like $d\boldsymbol{A}$ itself. When $d\boldsymbol{A}$ is represented as $d\mathbf{u} \wedge d\mathbf{v}$ and $\mathfrak{I}$ is positive, then the flow of the current is along $d\mathbf{u}$ into $d\mathbf{v}$. When we traditionally represent a magnetic dipole by an axial vector we are simply referring to it through $\mathbf{m}$, its dual, that is to say $\mathbf{m} = -I\boldsymbol{m}$. Figure 2.1(g) illustrates the relationship between a vector and its bivector dual, which of course also works the other way around as is the case here. However, we must be clear that within the context of a geometric algebra, the magnetic dipole itself originates from the physical model *as a bivector*. Swapping it for its vector dual may be a convenient way of dealing with it, for example, as a means of manipulating it in equations, but we should refrain from doing so if we wish to retain its physical character; otherwise, we simply make its origin ambiguous.

The separate natures of electric and magnetic dipoles are therefore quite distinct in a geometric algebra. The same distinction must therefore apply to the volume electric and magnetic polarizations $\mathbf{P}$ and $\boldsymbol{M}$. Note that in this context, these quantities represent the total dipole moment within an infinitesimal volume dV, a pseudoscalar quantity, divided by the *magnitude* of that volume, $|dV|$, which is a scalar quantity. This means that $\mathbf{P}$ and $\boldsymbol{M}$ have exactly the same characters as do $\mathbf{p}$ and $\boldsymbol{m}$, that is to say, they are vector and bivector, respectively.

We can check these characters in physical equations such as for the torque $\boldsymbol{\Gamma}$ acting on a dipole placed in a static electromagnetic field. Torque is conventionally expressed as the axial vector $\boldsymbol{\Gamma} = \mathbf{d} \times \mathbf{f}$, where $\mathbf{d}$ and $\mathbf{f}$ are both vectors, $\mathbf{d}$ being the directed distance and $\mathbf{f}$ being the force applied at that distance. It is appropriate, however, that we should convert this equation to the bivector form $\boldsymbol{\Gamma} = I\boldsymbol{\Gamma} = I\mathbf{d} \times \mathbf{f} = \mathbf{d} \wedge \mathbf{f}$. For the electric dipole, we therefore have $\boldsymbol{\Gamma}_e = \mathbf{p} \wedge \mathbf{E} = \frac{1}{2}(\mathbf{pE} - \mathbf{Ep})$, where $\mathbf{p}$ takes the place of $\mathbf{d}$ and $\mathbf{E}$ takes the place of $\mathbf{f}$. Another way of putting this is that $\boldsymbol{\Gamma}_e = \langle \mathbf{pE} \rangle_2$, that is to say, the bivector part of $\mathbf{pE}$. Recall that the $\langle\ \rangle_k$ function introduced in Section 2.3.1 is the grade filter such that when $k = 2$, $\langle \boldsymbol{U} \rangle_k$ returns the bivector part of any multivector $\boldsymbol{U}$.

The case of the magnetic dipole, however, is not just as straightforward. We will find out in due course that the outer product of two 3D bivectors such as $\boldsymbol{B}$ and $\boldsymbol{m}$ must always vanish so that we cannot take $\boldsymbol{B} \wedge \boldsymbol{m}$ as meaning the same thing as $\frac{1}{2}(\boldsymbol{Bm} - \boldsymbol{mB})$, which is actually the form we require. However, we can still say that $\boldsymbol{\Gamma}_m$ is given by the bivector part of $\boldsymbol{Bm}$ in the same way that $\boldsymbol{\Gamma}_e$ is given by the bivector part of $\mathbf{pE}$. The difference in the order of the terms that arises here occurs because $\boldsymbol{Bm} - \boldsymbol{mB}$ is equivalent to $(I\mathbf{B})(I\mathbf{m}) - (I\mathbf{m})(I\mathbf{B})$, which in turn reduces to $I^2(\mathbf{Bm} - \mathbf{mB}) = \mathbf{mB} - \mathbf{Bm}$. While this appears to be something of a nuance, geometric algebra is in fact very systematic and consistent in the way it deals with objects, it is just a bit different from ordinary linear algebra.

It is significant that this result implies that if $\boldsymbol{\Gamma}_m$ and $\boldsymbol{m}$ are both inherently bivector in character, then $\boldsymbol{B}$ must also be a bivector. This was a fact we simply adopted earlier on because we were familiar with the notion of $\mathbf{B}$ being an axial vector. Recalling that two sides of a multivector equation should match grade by grade, $\boldsymbol{B}$ must be a bivector. We know that it cannot be a scalar or pseudoscalar on physical grounds, and if it were a vector, then a product such as $\mathbf{B}\boldsymbol{m}$ would, according to the rules, furnish only a vector plus a trivector (pseudoscalar). Only if $\boldsymbol{B}$ is a bivector can $\boldsymbol{Bm}$ include a bivector so as to match the grade of $\boldsymbol{\Gamma}_m$.

In a similar vein, the usual constitutive relation $\mu_0 \boldsymbol{B} = \boldsymbol{H} + \boldsymbol{M}$ implies that the auxiliary magnetic field $\boldsymbol{H}$ must also be a bivector. Not only would it be a bit odd if we made it somehow different from $\boldsymbol{B}$ and $\boldsymbol{M}$ by replacing it with its dual form $I\mathbf{H}$, there would be little point in doing so. As we have already pointed out, switching to a dual form is purely a matter of convenience; the form we should be interested in here is the one that derives from the first principles, in this case, the bivector form. While we could have used $\mathbf{m} \wedge \mathbf{B}$ as a means of representing the bivector part of $\boldsymbol{Bm}$, this is just an expedient that can be avoided by using a different sort of product, the commutator product, to work directly between the bivectors $\boldsymbol{B}$ and $\boldsymbol{m}$ (see Exercise 4.8.11d).

Finally, let us observe that despite this awkwardness, given a dipole that has both an electric moment $\mathbf{p}$ and magnetic dipole moment $\boldsymbol{m}$, the total torque resulting from an electromagnetic field $\boldsymbol{F} = \mathbf{E} + \boldsymbol{B}$ may be expressed quite simply as

$$\begin{aligned} \boldsymbol{\Gamma} &= \langle \mathbf{pE} + \boldsymbol{Bm} \rangle_2 \\ &= \langle (\mathbf{p} - \boldsymbol{m})(\mathbf{E} + \boldsymbol{B}) \rangle_2 \\ &= \langle (\mathbf{p} - \boldsymbol{m})\boldsymbol{F} \rangle_2 \end{aligned} \tag{3.4}$$

Equation (3.4) therefore works because the grades of the cross terms $\mathbf{p}\boldsymbol{B}$ and $-\boldsymbol{m}\mathbf{E}$ from $(\mathbf{p}-\boldsymbol{m})(\mathbf{E}+\boldsymbol{B})$ must be odd and so no part of them may be of grade 2. On the other hand, the grades of $\mathbf{pE}$ and $-\boldsymbol{mB}$ must be even and so they may contribute to a bivector. This interesting result shows that we may express this interaction between a single multivector dipole moment $\mathbf{p}-\boldsymbol{m}$ and the electromagnetic field $\boldsymbol{F}$ as a whole, whereas previously it was necessary to consider the individual moments being acted on by $\boldsymbol{E}$ and $\mathbf{B}$ separately. Even if we refer to them as $I\mathbf{m}$ and $I\mathbf{B}$, the fundamentally bivector characters of $\boldsymbol{m}$ and $\boldsymbol{B}$ should always be borne in mind. Although they play similar roles to $\mathbf{p}$ and $\mathbf{E}$, they are not vectors. It would be wrong to forget this and to refer instead to $\mathbf{m}$ and $\mathbf{B}$ just as though they originated from the same sort of concepts as do $\mathbf{p}$ and $\mathbf{E}$, which is clearly not the case. Furthermore, it is only because of these different characters that it is possible to create multivectors such as $\mathbf{p}-\boldsymbol{m}$ and $\mathbf{E}+\boldsymbol{B}$, and to give any real meaning to an equation such as Equation (3.4).

It would seem, therefore, that Equation (3.4) does much to establish the value of using geometric algebra as a mathematical framework for *encoding* electromagnetic theory, but there is, as it were, an encore to this. If the grade 2 part of the multivector $\boldsymbol{U}=(\mathbf{p}-\boldsymbol{m})\boldsymbol{F}$ represents the torque on an arbitrary electromagnetic dipole, what meaning, if any, does this multivector have as a whole? It is clear that the cross terms represented by $\mathbf{p}\boldsymbol{B}-\boldsymbol{m}\mathbf{E}$ have no physical meaning and that they give rise only to terms of odd grade, that is to say grades 1 and 3. This therefore leaves the scalar $\langle\boldsymbol{U}\rangle_0$ to account for: $\langle\boldsymbol{U}\rangle_0=\langle(\mathbf{p}-\boldsymbol{m})(\mathbf{E}+\boldsymbol{B})\rangle_0=\langle\mathbf{pE}-\boldsymbol{mB}\rangle_0$. If we temporarily revert to writing $I\mathbf{m}$ and $I\mathbf{B}$ for $\boldsymbol{m}$ and $\boldsymbol{B}$ (which we do only in order to avoid introducing a scalar product for bivectors just at present), we find $\langle\boldsymbol{U}\rangle_0=\mathbf{p}\cdot\mathbf{E}+\mathbf{m}\cdot\mathbf{B}$, which is recognizable as $-U$ where U is the potential energy of the generalized dipole $\mathbf{p}-\boldsymbol{m}$ in the field $\boldsymbol{F}$. We may therefore express the interaction between dipole and field more completely as

$$\begin{aligned}\boldsymbol{U}&=-U+\boldsymbol{\Gamma}\\&=\mathbf{pE}+\boldsymbol{Bm}\\&=\langle(\mathbf{p}-\boldsymbol{m})\boldsymbol{F}\rangle_{0,2}\end{aligned}\tag{3.5}$$

where we understand that $\langle\boldsymbol{U}\rangle_{0,2}$ means that grades 0 and 2 should both be selected.

The fact that two or more associated physical relationships may be encoded in a single equation is one of the more remarkable features of geometric algebra; in fact, it is a recurrent theme not only in the treatment of electromagnetic theory but also in other branches of physics. Geometric algebra is therefore much more than just a different scheme of mathematical notation that replaces cross products, axial vectors and imaginary numbers with outer products, bivectors, and pseudoscalars.

3.3 THE VECTOR DERIVATIVE

Let us begin by considering the continuity equation for electric charge:

$$\partial_t \rho + \boldsymbol{\nabla} \cdot \mathbf{J} = 0 \tag{3.6}$$

Here ρ and $\mathbf{J}$ appear separately, but is it possible that we will now be able to write the continuity equation in terms of $\boldsymbol{J}$, the combined electromagnetic source multivector that we introduced in Equation (3.3)? After all, there would seem to be little point in having such a multivector source density if we can only use it by splitting it back up into separate scalar and vector parts.

Now, although we are not quite ready to tackle this particular question at present, it does raise the more basic question as to what $\boldsymbol{\nabla}$ should mean in the context of a geometric algebra. Equation (2.8) suggests that the well known forms $\boldsymbol{\nabla} \cdot \mathbf{f}$ and $\boldsymbol{\nabla} \times \mathbf{f}$ of traditional vector analysis may be combined into the single form $\boldsymbol{\nabla}\mathbf{f} = \boldsymbol{\nabla} \cdot \mathbf{f} + I\boldsymbol{\nabla} \times \mathbf{f}$ simply by treating $\boldsymbol{\nabla}$ as a vector. For 3D, therefore, if we write $\boldsymbol{\nabla}$ as $\mathbf{x}\partial_x + \mathbf{y}\partial_y + \mathbf{z}\partial_z$ and $\mathbf{f}$ as $f_x\mathbf{x} + f_y\mathbf{y} + f_z\mathbf{z}$, $\boldsymbol{\nabla}\mathbf{f}$ is found simply by multiplying out $(\mathbf{x}\partial_x + \mathbf{y}\partial_y + \mathbf{z}\partial_z)(f_x\mathbf{x} + f_y\mathbf{y} + f_z\mathbf{z})$. Provided the basis itself does not vary with position, differentiation and vector multiplication can be carried out in either order. By collecting all the scalar and bivector terms, the suggestion that $\boldsymbol{\nabla}\mathbf{f}$ is equivalent to $\boldsymbol{\nabla} \cdot \mathbf{f} + I\boldsymbol{\nabla} \times \mathbf{f}$ is readily proved. However, we would not actually wish to use this as a means of defining $\boldsymbol{\nabla}\mathbf{f}$, for it is better to write $\boldsymbol{\nabla} \equiv \mathbf{x}\partial_x + \mathbf{y}\partial_y + \mathbf{z}\partial_z$ and let geometric algebra take care of the rest. There is therefore nothing to stop the form $\mathbf{f}\boldsymbol{\nabla}$ from arising, but it is implicit that each derivative always acts on an object *to its right* rather than to its left. This form is therefore best interpreted as modifying $\boldsymbol{\nabla}$ to form some new differential operator. In conventional vector algebra, we have this possibility, for example, $\mathbf{f} \cdot \boldsymbol{\nabla} = f_x\partial_x + f_y\partial_y + f_z\partial_z$. With geometric algebra, this principle now simply extends to

$$\mathbf{f}\boldsymbol{\nabla} = (f_x\partial_x + f_y\partial_y + f_z\partial_z) + \mathbf{yz}(f_y\partial_z - f_z\partial_y) + \mathbf{zx}(f_z\partial_x - f_x\partial_z) + \mathbf{yz}(f_x\partial_y - f_y\partial_x) \tag{3.7}$$

For this reason, therefore, without some modification of the rules we cannot use Equation (2.6) to give us $\boldsymbol{\nabla} \cdot \mathbf{f}$ and $\boldsymbol{\nabla} \wedge \mathbf{f}$; rather, for the time being, we must use $\boldsymbol{\nabla} \cdot \mathbf{f} = \langle \boldsymbol{\nabla}\mathbf{f} \rangle_0$ and $\boldsymbol{\nabla} \wedge \mathbf{f} = \langle \boldsymbol{\nabla}\mathbf{f} \rangle_2$. Applying $\boldsymbol{\nabla}$ to a multivector function has results that depend on the grade of the multivector, or of its separate parts, as follows:

- For any scalar function f, $\boldsymbol{\nabla}f$ is a vector identical to the traditional gradient of f.
- For any vector function $\mathbf{f}$, $\boldsymbol{\nabla}\mathbf{f}$ comprises a scalar plus a bivector.
- In particular, the standard divergence and curl are given respectively by $\boldsymbol{\nabla} \cdot \mathbf{f}$ and $-I\boldsymbol{\nabla} \wedge \mathbf{f}$.
- In general, application of a vector operator to a multivector function follows the same rules as for left multiplication (premultiplication) by an ordinary vector.

By the simple process of direct multiplication, $\boldsymbol{\nabla}^2$ is readily shown to be equivalent to the scalar operator $\partial_x^2 + \partial_y^2 + \partial_z^2$. In traditional vector analysis, however,

- $\boldsymbol{\nabla}^2$ needs to be denoted by ∇^2 so as to make it overtly scalar and
- the separate vector form $\boldsymbol{\nabla}^2$ that is defined through $\boldsymbol{\nabla}^2\mathbf{f} \equiv \boldsymbol{\nabla}\boldsymbol{\nabla}\cdot\mathbf{f} - \boldsymbol{\nabla}\times\boldsymbol{\nabla}\times\mathbf{f}$ is actually redundant in geometric algebra.

Within the context of a geometric algebra, $\boldsymbol{\nabla}$ is more generally referred to as the vector derivative. Note that by convention, $\boldsymbol{\nabla}$ strictly refers to 3D, just as with other vectors shown in erect boldface, whereas in spaces of other dimensions it is more generally written simply as ∇. While the properties of the vector derivative so far may seem fairly pedestrian, it is a really striking fact that it actually has an inverse, in complete contrast to traditional 3D vector analysis where only the gradient has a definite inverse and the curl and divergence do not. As an example of how $\boldsymbol{\nabla}$ works in electromagnetic theory, let us evaluate $\boldsymbol{\nabla}\mathbf{E}$ in free space. This readily resolves into its scalar and bivector parts as follows

$$\begin{aligned}\boldsymbol{\nabla}\mathbf{E} &= \boldsymbol{\nabla}\cdot\mathbf{E} + I\boldsymbol{\nabla}\times\mathbf{E} \\ &= \rho/\varepsilon_0 - I\partial_t\mathbf{B} \\ &= \rho/\varepsilon_0 - \partial_t\boldsymbol{B}\end{aligned} \tag{3.8}$$

This gives us two of Maxwell's equations in one! We will, however, wait until Section 5.2 to see just how far we can go with this idea.

3.4 THE INTEGRAL EQUATIONS

The usual integral equations of electromagnetics to be found in textbooks are

$$\begin{aligned}&\int_{\partial V}\mathbf{D}\cdot d\mathbf{A} = \int_V \rho dV && \text{(i)} \\ &\int_{\partial V}\mathbf{B}\cdot d\mathbf{A} = 0 && \text{(ii)} \\ &\int_{\partial A}\mathbf{E}\cdot d\mathbf{l} = -\int_A \partial_t\mathbf{B}\cdot d\mathbf{A} && \text{(iii)} \\ &\int_{\partial A}\mathbf{H}\cdot d\mathbf{l} = \int_A (\mathbf{J} + \partial_t\mathbf{D})\cdot d\mathbf{A} && \text{(iv)}\end{aligned} \tag{3.9}$$

Here V refers to the volume enclosed by a surface ∂V, while A refers to the area enclosed by a boundary ∂A. For the moment, it is only the form of the equations that we want to consider, in particular (iii) and (iv), which, apart from the presence of $\mathbf{J}$ in the latter, are similar in appearance. As mentioned in the introduction, it has often been asserted that $\mathbf{H}$ must be like $\mathbf{E}$, a true vector in character. Likewise, $\mathbf{J}$ and $\mathbf{D}$ must be like $\mathbf{B}$, axial vectors in character, and this set of integral equations are a source of the usual arguments given for this. Now, in a polar theory of magnetism, there would be justification for drawing a parallel between $\mathbf{H}$ and $\mathbf{E}$, since $-\mathbf{H}\cdot d\mathbf{l}$ would correspond to the work done in displacing a unit pole through $d\mathbf{l}$ in

a magnetic field $\mathbf{H}$, just as $-\mathbf{E}\cdot d\mathbf{l}$ gives the work done on displacing a unit charge in an electric field. This sounds reasonable, but in the absence of magnetic poles, it is quite unfounded. These equations may look similar, and indeed in conventional vector analysis there is indeed no way to make the difference clear, but represented in a geometric algebra they turn out quite differently. Since $\mathbf{E}\cdot d\mathbf{l}$ is the inner product of two vectors, it is therefore a scalar. Here, as discussed in Section 3.2.2, we assert that $\mathbf{H}$ should actually be a bivector like $\boldsymbol{B}$, and consequently, $\mathbf{H}\cdot d\mathbf{l}$ should actually be replaced by $\boldsymbol{H}\wedge d\mathbf{l}$. Now the outer product of a bivector with a vector results in a *trivector*, as shown in Figure 2.1(c), which is also referred to as a pseudoscalar. On the right-hand side of the equation, therefore, the result should also be a pseudoscalar. But the axial vector $d\mathbf{A}$ is the outward normal of an element of area of magnitude dA, so as per our previous discussion it should be properly represented by $d\boldsymbol{A}$, the *bivector* for the directed area itself, as again depicted in Figure 2.1(b). In order for the result to be a pseudoscalar, the operation must again be the outer product of a vector with a bivector. Consequently, with $d\boldsymbol{A}$ being the *bivector*, $\mathbf{J}+\partial_t\mathbf{D}$ must be the required vector, meaning that $\mathbf{J}$ and $\mathbf{D}$, unlike $\boldsymbol{H}$, are both vectors. The proper forms of equations (i)–(iv) are therefore

$$\begin{aligned}
&\int_{\partial V}\mathbf{D}\wedge d\boldsymbol{A}=\int_{V}\rho dV && \text{(i) pseudoscalar}\\
&\int_{\partial V}\boldsymbol{B}\cdot d\boldsymbol{A}=0 && \text{(ii) scalar}\\
&\int_{\partial A}\mathbf{E}\cdot d\mathbf{l}=\int_{A}\partial_t\boldsymbol{B}\cdot d\boldsymbol{A} && \text{(iii) scalar}\\
&\int_{\partial A}\boldsymbol{H}\wedge d\mathbf{l}=\int_{A}(\mathbf{J}+\partial_t\mathbf{D})\wedge d\boldsymbol{A} && \text{(iv) pseudoscalar}
\end{aligned} \tag{3.10}$$

With a little effort, the reader may be convinced that (i)–(iii) are also in the correct form given the proper term by term assignments of vector and bivector. Note that the element of area $d\boldsymbol{A}$ can be written in the symbolic form $d\mathbf{u}\wedge d\mathbf{v}$, which is automatically a bivector, while an element of volume dV can be written as $d\mathbf{x}d\mathbf{y}d\mathbf{z}$ (or $d\mathbf{x}\wedge d\mathbf{y}\wedge d\mathbf{z}$) and is therefore a *trivector* or pseudoscalar. No other combination of vector and bivector can result in a scalar and so $\mathbf{D}\wedge d\boldsymbol{A}$, also a pseudoscalar, is the only option for the left-hand side of (i).

In Equation (3.10) (ii), multiplying the two bivectors on the left can only give rise to some combination of scalar and bivector parts; for example, $\mathbf{xyyx}=1$ is a scalar while $\mathbf{xyxz}=-\mathbf{yz}$ is a bivector. In the first example, the planes of the two bivectors are parallel, while in the second they are orthogonal. Considering the meaning of Gauss' law, it is therefore the scalar result that is appropriate. We will find out later that this is exactly what the inner product of two bivectors should produce.

In Equation (3.10) (iii) the situation is quite clear since $\mathbf{E}\cdot d\mathbf{l}$ is the inner product of two vectors and should be a scalar with the same meaning as in Equation

(3.9). Since we have already concluded that $\boldsymbol{B}\cdot d\boldsymbol{A}$ is a scalar, the term $\partial_t \boldsymbol{B}\cdot d\boldsymbol{A}$ on the right-hand side of (iii) must also be scalar, but note the change of sign as compared with $-\partial_t \mathbf{B}\cdot d\mathbf{A}$ in the traditional form. This simply follows from $\boldsymbol{B}d\boldsymbol{A} = (I\mathbf{B})(Id\mathbf{A}) = I^2\mathbf{B}d\mathbf{A} = -\mathbf{B}d\mathbf{A}$.

To conclude, we can use the grade structure of geometric algebra to make it clear that Equation (3.10) (iv) is characteristically different from Equation (3.10) (iii), and so there is no basis for comparing **H** with **E**. The equations are made consistent by **E**, **D**, and **J** being vectors while $\boldsymbol{H}$ and $\boldsymbol{B}$ are different, being bivectors as discussed in Section 3.2. Geometric algebra therefore encodes these equations in a much clearer way than is possible using the notion of true and axial vectors.

3.5 THE ROLE OF THE DUAL

We have frequently made the point that the classification of a given quantity as being either a vector or bivector should be made on physical grounds. This is usually obvious, as the preceding examples tend to show, but that does not mean to say the answer is always clear-cut. Take for example the current density **J**, which we have established as being a vector in keeping with the idea that it represents a moving charge density as in the equation $\mathbf{J} = \rho\mathbf{v}$. But on the other hand we can view it as a flux that is associated with the rate at which charge Q, a scalar quantity, passes through a unit area, as in the equation $\partial_t Q = \boldsymbol{J}\cdot d\boldsymbol{A}$, where it is clearly the bivector representation that seems more appropriate. It is also apparent from Equation (3.10) that expressions like $\boldsymbol{H}\wedge d\mathbf{l}$ can be modified using the dual form; in fact, we will find out in due course that $(I\mathbf{H})\wedge d\mathbf{l} = I(\mathbf{H}\cdot d\mathbf{l})$. The key issue is therefore to distinguish between valid physical forms and those that are mere expedients.

The examples we have been discussing draw attention to the role of the dual. In fact, if every element in a geometric algebra were to be replaced with its dual, we would still have a geometric algebra. For example, replacing the basis elements of 3D by their duals gives us

$$1;\mathbf{x},\mathbf{y},\mathbf{z};I\mathbf{x},I\mathbf{y},I\mathbf{z};I \quad\mapsto\quad I;I\mathbf{x},I\mathbf{y},I\mathbf{z};-\mathbf{x},-\mathbf{y},-\mathbf{z};-1 \tag{3.11}$$

Clearly, the rules of multiplication are still obeyed, but equations and expressions will generally appear different. For example, $\boldsymbol{U} = \mathbf{xy}$ becomes $\boldsymbol{U}' = -I\boldsymbol{x}'\boldsymbol{y}'$ where $\boldsymbol{x}' = I\mathbf{x}$, $\boldsymbol{y}' = I\mathbf{y}$ and $\boldsymbol{U}' = I\boldsymbol{U}$. We can therefore always find an alternative representation of a system of equations where everything is replaced by its dual. This is not all that remarkable since we get somewhat similar mappings if we replace each element by its negative, inverse, or even reverse. The equations may look different to the ones we are used to, but all the same it serves to illustrate that the way in which physical quantities are represented is not entirely sacrosanct. The main criteria for choosing the one representation or the other may be summed up as

- achieving a proper correspondence with the physical model,
- maintaining consistency throughout the complete system of equations, and

- remembering the difference between an alternative representation and the native form.

3.6 EXERCISES

Assume a 3D geometric algebra in all of the following:

1. Evaluate $\nabla\mathbf{r}$, $\nabla(\mathbf{ar})$, $\nabla(\mathbf{a}\cdot\mathbf{r})$, $\nabla(\mathbf{a}\wedge\mathbf{r})$, $\nabla\mathbf{r}^2$, $\nabla^2\mathbf{r}$, $\nabla^2\mathbf{r}^2$, ∇r, and ∇r^{-1} where $\mathbf{r}=x\mathbf{x}+y\mathbf{y}+z\mathbf{z}$, $\mathbf{a}$ is a constant vector and $r=|\mathbf{r}|$.
2. Evaluate $\nabla\mathbf{u}$, $\nabla\cdot\mathbf{u}$, $\nabla\wedge\mathbf{u}$, and ∇u where $\mathbf{u}=u_x\mathbf{x}+u_y\mathbf{y}+u_z\mathbf{z}$.
3. Justify why $\nabla(I\mathbf{u})$ may be written as $I\nabla\mathbf{u}$.
4. Confirm that the grades of the expressions on each side of Equation (3.10) (i)–(iv) are as shown.
5. Under what circumstances, if any, might it be possible to represent the following by bivectors: (a) the electric field, (b) angle, and (c) motion?
6. Find a meaning for $(\mathbf{u}\cdot\nabla)f$ and $(\mathbf{u}\cdot\nabla)\mathbf{f}$. In particular, what does $\mathbf{x}\cdot\nabla$ mean?
7. The multivector $\boldsymbol{R}=t+\mathbf{r}$ specifies position $\mathbf{r}$ at time t. The equation of motion of a particle moving with constant velocity $\mathbf{v}$ may then be written as $\boldsymbol{R}_p=\mathbf{r}_0+\boldsymbol{V}t$. What is the meaning of $\boldsymbol{V}$ here? Following similar lines, construct the equation of motion for a uniformly accelerating particle.

Chapter 4

Generalization

So far, we have covered the basics of geometric algebra largely by means of using 3D as an illustration. We now develop these ideas in a more general way that will be applicable to spaces of any dimension and, in particular, to the case of 4D, which is essential to a complete understanding of classical electromagnetic theory. The new information to be encountered in this chapter will allow the reader to appreciate the following:

- Geometric algebras have a simple, regular structure.
- The geometric product is the common underlying principle.
- The classification of objects into grades is essential.
- The geometric product of any object with a vector decomposes into the sum of an inner and outer product.
- Provided the result is nonzero:
  - The inner product reduces the grade of an object by 1.
  - The outer product increases the grade of an object by 1.
- The generalized formulas for the inner and outer products are nevertheless fairly simple.
  - How they work always depends on the classification of the objects involved.
- The motivation for using inner and outer products is that the classification of the resulting product is readily maintained.

Provided that the reader can get a firm grip of these principles, they may skim over the fine detail at a first reading and return to it, when necessary, for reference purposes such as clarifying the meaning of an expression. Take note that since we will be discussing geometric algebras in general rather than just the specific 3D variety, we will mostly be using the form $\boldsymbol{u}$ rather than $\mathbf{u}$ to denote a vector. The need for this distinction will become clear in due course when we have to deal with both the 4D and 3D forms of the same vector side by side. Appendix 14.4.2 provides a more formal summary of the rules of geometric algebra and includes a number of

Understanding Geometric Algebra for Electromagnetic Theory, First Edition. John W. Arthur.

theorems that help to extend the basic properties. Here, however, we continue with our less formal discussion-based approach.

4.1 HOMOGENEOUS AND INHOMOGENEOUS MULTIVECTORS

To accommodate vector spaces of dimension greater than 3 within the framework of a geometric algebra, it is necessary to classify all objects in a more general way according to their grade. We therefore write $\boldsymbol{V}_k$ to imply that an object $\boldsymbol{V}$ is a k-vector, that is to say, it has the specific grade k. The maximum value of k is determined by N, the dimension of the space. A k-vector is often referred to as a homogeneous multivector, particularly when we do not know what grade to assign to it. On the other hand, if the parts of an object are of different grades, the object is a general, or inhomogeneous, multivector that we would normally simply refer to as $\boldsymbol{V}$, that is to say, with no subscript.

4.2 BLADES

We have already seen that the multiplication of orthogonal basis vectors generates new elements that are not vectors. This process gives rise to the natural extension of the space of 1-vectors into an entire geometric algebra, the basis elements of which comprise the unit scalar, the original basis vectors, and the set of all those unique elements that can be generated by their multiplication. These new basis elements are a special type of multivector called a blade, a name that unfortunately seems to convey very little about their rather important function. As an example, if $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$ are mutually orthogonal vectors, then $\boldsymbol{x}$, $\boldsymbol{xy}$, and $\boldsymbol{xyz}$, are blades of grades 1, 2, and 3, respectively. We may therefore consider blades to be the simplest elements belonging to any given grade, including the scalars, which are allocated to grade 0. Since we can manufacture all the basis elements in the form of blades, every multivector must be expressible as a sum of blades.

If a set of vectors, say $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$ . . . , are not mutually orthogonal, then we can still form blades from their outer products rather than their direct geometric products. The outer product of two vectors is equivalent to multiplying only their orthogonal parts, and so it turns out in general that the outer product of n linearly independent vectors is the same as the geometric product of their n mutually orthogonal parts. Recalling the grade filtering function $\langle \boldsymbol{U} \rangle_n$ that returns the part of $\boldsymbol{U}$ that is of grade n, another way to look at this is $\boldsymbol{u} = \langle \boldsymbol{u} \rangle_1$, $\boldsymbol{u} \wedge \boldsymbol{v} = \langle \boldsymbol{uv} \rangle_2$, $\boldsymbol{u} \wedge \boldsymbol{v} \wedge \boldsymbol{w} = \langle \boldsymbol{uvw} \rangle_3$, and so on. Provided that none of the results is 0, we may therefore state more generally that $\boldsymbol{u}$, $\boldsymbol{u} \wedge \boldsymbol{v}$, and $\boldsymbol{u} \wedge \boldsymbol{v} \wedge \boldsymbol{w}$ are examples of blades of grade 1, 2, and 3, respectively.

Note, however, that similar appearing expressions such as $\boldsymbol{u} \cdot \boldsymbol{v} \cdot \boldsymbol{w}$ or $(\boldsymbol{u} \cdot \boldsymbol{v}) \cdot \boldsymbol{w}$ involving the inner product of three or more vectors all vanish since any inner product with a scalar vanishes by definition. For $1 < n$, a blade of grade n may therefore be generally defined as a nonzero object formed by the outer product of n vectors. For completeness sake, we may allow the blades to include the special cases

of scalars and vectors, which are blades of grades 0 and 1, respectively. Since the direct product of orthogonal vectors yields the same result as their outer product, our earlier examples of blades are still perfectly valid.

Blades need not have the appearance of a simple product such as $\boldsymbol{xy}$ or $\boldsymbol{xyz}$; it is readily confirmed that the bivector $\boldsymbol{xz}+\boldsymbol{yx}+\boldsymbol{yz}$ is a blade since it can be formed by $(\boldsymbol{x}+\boldsymbol{y})\wedge(\boldsymbol{x}+\boldsymbol{z})$. On the other hand, if we were seeking a set of basis elements, this is unlikely to be the sort of blade we would choose. While a blade is a homogeneous multivector (n-vector), not all homogeneous multivectors are blades [27, p. 90], at least in the case of spaces of dimension higher than 3. For example, in 4D with the orthonormal basis $\boldsymbol{w},\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}$, there is no way to represent the bivector $\boldsymbol{xy}+\boldsymbol{zw}$ (which is an illustration of a 2-vector, or homogeneous multivector of grade 2) as the product of two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. By assuming $\boldsymbol{a}$ to be a vector and then expressing it in terms of the basis vectors $\boldsymbol{w},\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}$, it is straightforward to show that $\boldsymbol{b}$ must include a trivector part in order to satisfy the proposition $\boldsymbol{xy}+\boldsymbol{zw}=\boldsymbol{ab}$.

While n-vectors and homogeneous vectors amount to the same thing, it is important to remember the distinction between

- blades versus n-vectors, some of which may only be expressible as a sum of blades;
- homogeneous multivectors versus general multivectors comprising any mixture of grades; and
- forming blades by direct multiplication of orthogonal vectors versus forming blades with general vectors where it is necessary to use the outer product.

Consider the construction of blades from a set of N orthogonal basis vectors. The entire set of blades is readily obtained by taking all possible combinations of two or more basis vectors and forming their products. Since orthogonal vectors anticommute, it is clear that the order in which we multiply them can only affect the sign of the result and so only one blade arises per combination. Given that there is also one blade of grade 0 and N distinct blades of grade 1 that account for the scalars and the vectors respectively, it is clear that there must therefore be $\binom{N}{k}$ distinct blades for each grade k so that the total number of blades, or unique basis elements, must be $\sum_{k=0}^{N}\binom{N}{k}=2^N$. The resulting general hierarchy is shown in Table 2.1(a), while Table 2.1(b) gives an example for $N=4$, which readily confirms that we have $\binom{4}{1}=4$ unique vectors, $\binom{4}{2}=6$ unique bivectors, $\binom{4}{3}=4$ unique trivectors, and a single scalar and pseudoscalar, totaling $1+4+6+4+1=16=2^4$ unique basis elements in all.

There is yet another method of generating blades with the geometric product when the set of vectors is not orthogonal but merely linearly independent. Following the same scheme as above, it is only necessary to multiply together any k-vectors chosen from the set as before but then we must *select the part of the result that is of grade k*. For example, we multiply two vectors and discard the scalar part of the result. The remaining part of grade 2, if any, will be a blade, in fact a bivector. If $\boldsymbol{u}$ is a vector and $\boldsymbol{U}_k$ is some blade of grade k that we have already generated, then $\langle \boldsymbol{u}\boldsymbol{U}_k \rangle_{k+1}$ will provide a blade of grade $k+1$, once again as long as the result is

nonzero. If the result turns out to be zero, then either $\boldsymbol{u}$ was previously chosen in the generation of $\boldsymbol{U}_k$ or the original set of vectors was not linearly independent. This idea can be applied inductively until, at some point, we cannot find any new blades.

As we have already seen, another method of generating blades from a set of linearly independent vectors is to employ the outer product rather than the direct product, the only difference being that the grade selection filter is no longer necessary because a blade of grade k will always result from the outer product of k linearly independent vectors. For practical purposes, however, it will often be convenient to continue to generate blades from a set of orthogonal vectors so that the process for each grade is just a simple matter of finding the $\binom{N}{k}$ distinct products of the N basis vectors taken k at a time.

4.3 REVERSAL

It is convenient to have a way of expressing any given product of vectors in reverse order. This operation is indicated by the superscript symbol † so that if, say, $\boldsymbol{U} = \boldsymbol{def}$, then the reverse of $\boldsymbol{U}$ is simply $\boldsymbol{U}^\dagger = \boldsymbol{fed}$. If $\boldsymbol{U}$ happens to be a general multivector, then it will amount to a sum of terms, each of which is either a product of vectors or may be multiplied out and further broken down until this is the case. This is simply how we would go about expressing any multivector in its simplest terms. We now have a sum of products to which the reverse can be individually applied. For example, suppose $\boldsymbol{U}$ may be reduced to the form $a + \boldsymbol{b} + \boldsymbol{cd} + \boldsymbol{ef} + \boldsymbol{ghk}$, we then have

$$\begin{aligned}\boldsymbol{U}^\dagger &= (a + \boldsymbol{b} + \boldsymbol{cd} + \boldsymbol{ef} + \boldsymbol{ghk})^\dagger \\ &= a + \boldsymbol{b}^\dagger + (\boldsymbol{cd})^\dagger + (\boldsymbol{ef})^\dagger + (\boldsymbol{ghk})^\dagger \\ &= a + \boldsymbol{b} + \boldsymbol{dc} + \boldsymbol{fe} + \boldsymbol{khg}\end{aligned}$$

Reversal has the following useful properties:

- For any two multivectors $\boldsymbol{U}$ and $\boldsymbol{V}$, $(\boldsymbol{U}+\boldsymbol{V})^\dagger = \boldsymbol{U}^\dagger + \boldsymbol{V}^\dagger$ and $(\boldsymbol{UV})^\dagger = \boldsymbol{V}^\dagger\boldsymbol{U}^\dagger$.
- Scalars and vectors are unaffected by reversal so that $(a+\boldsymbol{u})^\dagger = (a+\boldsymbol{u})$.
- For any product of vectors $\boldsymbol{U} = \boldsymbol{u}_1\boldsymbol{u}_2\cdots\boldsymbol{u}_n$ we have $\boldsymbol{U}\boldsymbol{U}^\dagger = \boldsymbol{U}^\dagger\boldsymbol{U} = \boldsymbol{u}_1^2\boldsymbol{u}_2^2\cdots\boldsymbol{u}_n^2$, a scalar.
  - This clearly includes blades.
  - When $\boldsymbol{U}$ is a not a product of vectors, however, $\boldsymbol{U}\boldsymbol{U}^\dagger$ may not be scalar and it may not even be equal to $\boldsymbol{U}^\dagger\boldsymbol{U}$. Assuming $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$, and $\boldsymbol{w}$ are all orthonormal, the example $\boldsymbol{w}+\boldsymbol{xyz}$ demonstrates this since $(\boldsymbol{w}+\boldsymbol{xyz})(\boldsymbol{w}+\boldsymbol{xyz})^\dagger = 2+2\boldsymbol{xyzw}$, whereas $(\boldsymbol{w}+\boldsymbol{xyz})^\dagger(\boldsymbol{w}+\boldsymbol{xyz}) = 2-2\boldsymbol{xyzw}$.
- If $\boldsymbol{U}_k$ is a k-vector, then $\boldsymbol{U}_k^\dagger = \pm\boldsymbol{U}_k$ where the sign is positive when $\lfloor k/2 \rfloor$ is even, that is, for $k = 1,4,5,8,9\ldots$ and negative when $\lfloor k/2 \rfloor$ is odd, that is, for $k = 2,3,6,7\ldots$
  - The measure of $\boldsymbol{U}_k$ may be defined through $|\boldsymbol{U}_k|^2 = \left|\left\langle \boldsymbol{U}_k\boldsymbol{U}_k^\dagger \right\rangle\right| = \left|\left\langle \boldsymbol{U}_k^2 \right\rangle\right|$

- For the unit pseudoscalar, $I^\dagger = \left(I^2\right) I$.
  - When $I^\dagger = I$, the criterion for k-vectors given above therefore implies that I^2 must be positive in any Euclidean space.

While it is clear that the result $\boldsymbol{U}_k^\dagger = \pm \boldsymbol{U}_k$ implies that all bivectors change sign under reversal, be careful to observe that $(\boldsymbol{cd})^\dagger = -\boldsymbol{cd}$ only if $\boldsymbol{c} \perp \boldsymbol{d}$, that is to say $\boldsymbol{cd}$ is a pure bivector. Another point of caution is that *vectors with a negative square cause exceptions to some of the above properties*. This matters when we come to spacetime where we find that I^2 turns out to be negative despite the fact that $I^\dagger = I$. All we really need to remember here is $I^2 = -1$, just as in 3D.

From these properties it follows that:

- any product of vectors must have an inverse,
  - an exception to this, however, is when null vectors are allowed, for example, in spacetime;
- in particular, all blades have inverses;
- if we replace every element in a geometric algebra by its reverse, then the result is also a geometric algebra.

A point of caution, however, is that different authors employ a variety of superscripts for reversal, involution (inversion) and conjugate based on some permutation of the symbols †, ~, and *. The † notation for reversal is that of Hestenes, Doran, and coworkers except in spacetime where they switch to ~. Since we will rarely, if ever, have need for involution or conjugate, we will keep to the symbol † for reversal as a matter of consistency. Lounesto, on the other hand, adheres to the symbol ~ throughout. There seems to be no way to avoid these differences, and so when comparing expressions from different sources, it is always advisable to check the notation being used.

4.4 MAXIMUM GRADE

While the dimension of the space N is equal to the number of *basis vectors* alone, as we have seen, the number of all the *basis elements* that are required to span the entire geometric algebra is 2^N. In the case of 3D, therefore, $2^3 = 8$ entities are needed as the basis of a geometric algebra, comprising, in order of increasing grade, 1 scalar, 3 vectors, 3 bivectors, and 1 pseudoscalar. We saw earlier on that in multiplying a number of orthogonal basis vectors together, the result depends on the number of times each basis vector occurs in the product. If it occurs an even number of times, then it simply disappears from the result (reduction) while for an odd number of times it remains, but with just a single occurrence, for example, $\boldsymbol{yxzyzxzy} = \boldsymbol{zy}$. Any product of orthogonal basis vectors therefore reduces to a blade, and in addition, the highest grade of object in a geometric algebra must be the blade that includes each basis vector exactly once in the product. This means that the highest possible

grade is N, that is to say, $0 \le k \le N$ where k is the grade. A similar argument leads to the result that the number of distinct blades in each grade is $\binom{N}{k}$, as we have already seen.

4.5 INNER AND OUTER PRODUCTS INVOLVING A MULTIVECTOR

We have already introduced the notion of inner and outer products based on a geometric interpretation in which the inner product is formed by multiplying the parallel parts of a pair of vectors whereas the outer product is formed by multiplying their perpendicular parts. This led to the standard formulas for the inner and outer products of any two vectors (Equation 2.6). In addition, in Section 2.3.3, we began to explore by example the notion of inner and outer products between other grades of object with a vector, showing that

- provided the result is nonzero, the outer product $\boldsymbol{U}_2 \wedge \mathbf{v}$ adds a dimension to the bivector $\boldsymbol{U}_2$, that is to say it creates an object of grade 3, a trivector;
- the inner product $\boldsymbol{U}_2 \cdot \mathbf{v}$ removes a dimension, reducing the bivector to grade 1, a vector;
- the dimension added or removed has a specific relationship to $\mathbf{v}$;
- for the outer product, the dimension added is along the part of $\mathbf{v}$ that is orthogonal to $\boldsymbol{U}$; and
- for the inner product, the dimension removed is along the part of $\mathbf{v}$ that is parallel to $\boldsymbol{U}$.

We will now review these two products on a more formal basis and extend the idea to objects other than vectors.

We have seen that a blade of any given grade k may be expressed as a product of k orthogonal basis vectors. The basis vectors chosen to form the blade must all be different since the product would reduce to a lower grade if any two of the vectors were the same. Given any such blade $\boldsymbol{V}_k$, any vector $\boldsymbol{u}$ may be written as $\boldsymbol{u}_+ + \boldsymbol{u}_-$ where $\boldsymbol{u}_+$ is *outside* the space spanned by the basis vectors that form $\boldsymbol{V}_k$, while $\boldsymbol{u}_-$ is *within* this space. This is clearly just a generalization of the notation $\mathbf{u} = \mathbf{u}_\perp + \mathbf{u}_{/\!/}$ that we used earlier when the objects were restricted to 3D. The product $\boldsymbol{u}_k \boldsymbol{V}_k$ may therefore be written as $\boldsymbol{u}_+ \boldsymbol{V}_k + \boldsymbol{u}_- \boldsymbol{V}_k$. The first product, $\boldsymbol{u}_+ \boldsymbol{V}_k$, must then be of grade $k+1$ since $\boldsymbol{u}_+$ is, by choice, orthogonal to every basis vector used in forming $\boldsymbol{V}_k$. On the other hand, $\boldsymbol{u}_- \boldsymbol{V}_k$ must be of grade $k-1$ since $\boldsymbol{u}_-$ can be split into a sum of parts, each one of which is of the form $a\boldsymbol{b}$, where a is a scalar and $\boldsymbol{b}$ is any of the basis vectors used in forming $\boldsymbol{V}_k$. If this were not the case, then $\boldsymbol{u}_-$ would not be within the space spanned by these basis vectors. Any term of the form $a\boldsymbol{b}\boldsymbol{V}_k$ must then be of grade $k-1$ since a reduction will inevitably occur given that vector $\boldsymbol{b}$ already occurs somewhere in the product making up $\boldsymbol{V}_k$. Since this applies to any

choice of $\boldsymbol{b}$, it must apply to all, and if all the parts of $\boldsymbol{V}_k$ are of grade $k-1$, then so must be the whole.

The entire argument may be repeated for a product of the form $\boldsymbol{V}_k\boldsymbol{u}$. In addition, any n-vector may be written as a sum of blades of grade n, and so it can be concluded in general that the product of a vector with any object of grade k must result in an object of grade $k-1$ plus an object of grade $k+1$. This property is fundamental to the introduction of inner and outer products as a generalization of Equations (2.5) and (2.6) to objects of higher grade.

We consider the lower-grade part of $\boldsymbol{u}\boldsymbol{V}_k$ to be the inner product of $\boldsymbol{u}$ and $\boldsymbol{V}_k$, written as $\boldsymbol{u}\cdot\boldsymbol{V}_k$, whereas the higher-grade part is the outer product, written as $\boldsymbol{u}\wedge\boldsymbol{V}_k$. If $k=N$ where N is the dimension of the space, then $\boldsymbol{V}_k$ is a pseudoscalar. Since there can be no higher-grade object, we must have $\boldsymbol{u}\wedge\boldsymbol{V}_N\equiv 0$ for any vector $\boldsymbol{u}$. Equation (2.5) may therefore be replaced with the more general formulas:

$$\begin{aligned}\boldsymbol{u}\boldsymbol{V}_k &= \boldsymbol{u}\cdot\boldsymbol{V}_k+\boldsymbol{u}\wedge\boldsymbol{V}_k\\ \boldsymbol{V}_k\boldsymbol{u} &= \boldsymbol{V}_k\cdot\boldsymbol{u}+\boldsymbol{V}_k\wedge\boldsymbol{u}\end{aligned}\tag{4.1}$$

While for the purposes of Equation (4.1) $\boldsymbol{u}$ has to be a vector or, in the trivial case, a scalar (see Exercise 4.8.11b), we may drop the subscript k from $\boldsymbol{V}_k$ because the result is true for any grade k, and so Equation (4.1) actually works with any multivector $\boldsymbol{V}$:

$$\begin{aligned}\boldsymbol{u}\boldsymbol{V} &= \boldsymbol{u}\cdot\boldsymbol{V}+\boldsymbol{u}\wedge\boldsymbol{V}\\ \boldsymbol{V}\boldsymbol{u} &= \boldsymbol{V}\cdot\boldsymbol{u}+\boldsymbol{V}\wedge\boldsymbol{u}\end{aligned}\tag{4.2}$$

From the previous discussion, it can also be seen how the terms inner and outer product arise. Restricting ourselves to the case where $\boldsymbol{V}_k$ is a blade, the inner product is formed between $\boldsymbol{u}_{/\!/}$ and $\boldsymbol{V}_k$ where $\boldsymbol{u}_{/\!/}$ is *within* the space spanned by the basis vectors involved in the construction of $\boldsymbol{V}_k$, that is to say the *inner space* of $\boldsymbol{V}_k$, whereas the outer product involves $\boldsymbol{u}_\perp$, which is *outside* the space spanned by these basis vectors; that is, it is in the *orthogonal* or *outer* space of $\boldsymbol{V}_k$. We may therefore equally well write Equation (4.1) in the form

$$\begin{aligned}\boldsymbol{u}\boldsymbol{V}_k &= \boldsymbol{u}_{/\!/}\boldsymbol{V}_k+\boldsymbol{u}_\perp\boldsymbol{V}_k\\ \boldsymbol{V}_k\boldsymbol{u} &= \boldsymbol{V}_k\boldsymbol{u}_{/\!/}+\boldsymbol{V}_k\boldsymbol{u}_\perp\end{aligned}\tag{4.3}$$

where $\boldsymbol{u}_{/\!/}\wedge\boldsymbol{V}_k=0$ and $\boldsymbol{u}_\perp\cdot\boldsymbol{V}_k=0$. From this we get a more general notion of "parallel to" and "orthogonal to." These terms are used in the same sense that a line can be either parallel to a plane or orthogonal to it. We could equally use "inside" and "outside" in a similar context, particularly for objects of higher dimensions. For example, it may appear more meaningful to say that $\boldsymbol{x}$ is in the space of $\boldsymbol{xyz}$ rather than parallel to it. Things, however, are less clear when $\boldsymbol{u}$ is not a vector. For example, $\boldsymbol{xy}$ could be said to be orthogonal to $\boldsymbol{yz}$, but they nevertheless share a

common subspace spanned by $\boldsymbol{y}$. We would need to evaluate $(\boldsymbol{xy})\cdot(\boldsymbol{yz})$ to know for sure (we will find out how to do this in Section 4.6). For the same reason, it is not generally possible to use commutation properties as a way of defining parallel and perpendicular except in the simple case where one of the objects is a vector. Even here, the situation is awkward because commutation with an object means parallel when the higher-grade object is of odd grade, whereas anticommutation means parallel when it is of even grade and vice versa for orthogonal. For example, in the case of $\boldsymbol{x} /\!/ \boldsymbol{xy}$, we have $\boldsymbol{x}(\boldsymbol{xy}) = -(\boldsymbol{xy})\boldsymbol{x}$, while for $\boldsymbol{z} \perp \boldsymbol{xy}$, we have $\boldsymbol{z}(\boldsymbol{xy}) = +(\boldsymbol{xy})\boldsymbol{z}$. It is therefore much simpler to base the definition of parallel and perpendicular on

$$\begin{aligned} \boldsymbol{u} /\!/ \boldsymbol{V} &\Leftrightarrow \boldsymbol{u} \wedge \boldsymbol{V} = 0 \\ \boldsymbol{u} \perp \boldsymbol{V} &\Leftrightarrow \boldsymbol{u} \cdot \boldsymbol{V} = 0 \end{aligned} \tag{4.4}$$

Given their respective symbols, the operations $\cdot$ and $\wedge$ are commonly spoken of as "dot" and "wedge" respectively. When it is necessary to distinguish these special products from the basic geometric product, we shall specifically refer to the latter as the geometric or direct product.

The 3D dot and cross products are consequently superseded by the more general notion of inner and outer products. While the inner product between vectors is the same as the dot product, we must think differently when one of the objects is not a pure vector. A side-by-side comparison of some simple 3D vector products is shown in Table 4.1, first using the rules of geometric algebra and then using dot and cross products. Note that similar-looking products do not necessarily correspond in meaning.

The rules for defining inner and outer products may be stated more formally using the notion of stepping-up and stepping-down of grades. The $\langle \ldots \rangle_m$ function introduced in Section 2.3.1 is once again very helpful. We must particularly note that the result will always be 0 if m is outside the range $0 \le m < N$:

$$\begin{aligned} \boldsymbol{u} \cdot \boldsymbol{V}_k &= \langle \boldsymbol{u}\boldsymbol{V}_k \rangle_{k-1} \\ \boldsymbol{u} \wedge \boldsymbol{V}_k &= \langle \boldsymbol{u}\boldsymbol{V}_k \rangle_{k+1} \end{aligned} \tag{4.5}$$

Table 4.1 Comparison of Various Simple Vector Products in 3D

Traditional 3D vector toolset		3D geometric algebra	
$\mathbf{u} \cdot \mathbf{v}$	Scalar	$\mathbf{u} \cdot \mathbf{v}$	Scalar
$\mathbf{u} \times \mathbf{v}$	Axial vector	$\mathbf{u} \wedge \mathbf{v}$	Bivector
$\mathbf{u} \times (\mathbf{v} \times \mathbf{w})$	Vector	$\mathbf{u} \cdot (\mathbf{v} \wedge \mathbf{w})$	Vector
$\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$	Scalar	$\mathbf{u} \wedge (\mathbf{v} \wedge \mathbf{w})$	Pseudoscalar

Although some expressions on a given row may look different, they nevertheless have closely associated meanings.

The same rules apply if the order of $\boldsymbol{u}$ and $\boldsymbol{V}_k$ is reversed. This turns out to be a useful approach, but it does have the disadvantage that, being rather more abstract, it does not readily afford the geometric interpretation described in Section 2.3.3.

Equation (4.1) leads to another simple set of rules for evaluating inner and outer products. These generalizations of Equation (2.6) may be expressed as

$$\boldsymbol{u}\cdot\boldsymbol{V}_k=\begin{Bmatrix}\frac{1}{2}(\boldsymbol{u}\boldsymbol{V}_k+\boldsymbol{V}_k\boldsymbol{u}) & \left|\begin{matrix}k\text{ odd}\\k\text{ even}\end{matrix}\right.\\ \frac{1}{2}(\boldsymbol{u}\boldsymbol{V}_k-\boldsymbol{V}_k\boldsymbol{u})\end{Bmatrix}\qquad \boldsymbol{u}\wedge\boldsymbol{V}_k=\begin{Bmatrix}\frac{1}{2}(\boldsymbol{u}\boldsymbol{V}_k-\boldsymbol{V}_k\boldsymbol{u}) & \left|\begin{matrix}k\text{ odd}\\k\text{ even}\end{matrix}\right.\\ \frac{1}{2}(\boldsymbol{u}\boldsymbol{V}_k+\boldsymbol{V}_k\boldsymbol{u})\end{Bmatrix} \tag{4.6}$$

It seems peculiar that the rules for inner and outer product are almost identical, with only a sign that alternates depending on the parity of k (evenness or oddness) to distinguish them. Only by defining them in this way, however, can one consistently be a step-down operator while the other is consistently the step-up kind for any grade of operand. Similarly, note that the commutation properties of dot and wedge also change with the parity of k as follows:

$$\boldsymbol{u}\cdot\boldsymbol{V}_k=\begin{Bmatrix}\boldsymbol{V}_k\cdot\boldsymbol{u} & \left|\begin{matrix}k\text{ odd}\\k\text{ even}\end{matrix}\right.\\ -\boldsymbol{V}_k\cdot\boldsymbol{u}\end{Bmatrix}\qquad \boldsymbol{u}\wedge\boldsymbol{V}_k=\begin{Bmatrix}-\boldsymbol{u}\wedge\boldsymbol{V}_k & \left|\begin{matrix}k\text{ odd}\\k\text{ even}\end{matrix}\right.\\ \boldsymbol{u}\wedge\boldsymbol{V}_k\end{Bmatrix} \tag{4.7}$$

Equation (4.7) may be used to show that

- in a geometric algebra of odd dimension, the unit pseudoscalar commutes with everything, whereas
- in a geometric algebra of even dimension, the unit pseudoscalar commutes only with those entities that are of even grade. It anticommutes with the rest, in particular the vectors.

An important property of the outer product is that it is associative, that is to say an expression such as $\boldsymbol{v}_1\wedge(\boldsymbol{v}_2\cdots\wedge\boldsymbol{v}_n)$ or $(\boldsymbol{v}_1\wedge\boldsymbol{v}_2\wedge\cdots)\wedge\boldsymbol{v}_n$ needs no brackets and may simply be written as $\boldsymbol{v}_1\wedge\boldsymbol{v}_2\wedge\cdots\wedge\boldsymbol{v}_n$, that is to say we may carry out the outer products in any manner we choose as long as we do not change the order of any of the terms. This means that expressions such as $(\boldsymbol{v}_1\wedge\cdots\wedge\boldsymbol{v}_k)\wedge(\boldsymbol{v}_{k+1}\wedge\cdots\wedge\boldsymbol{v}_n)$ must be valid, and since both $(\boldsymbol{v}_1\wedge\cdots\wedge\boldsymbol{v}_k)$ and $(\boldsymbol{v}_{k+1}\wedge\cdots\wedge\boldsymbol{v}_n)$ form blades, this leads to the conclusion that it is possible to have an outer product between blades of any grade.

Another important property specific to the outer product is that $\boldsymbol{v}_1\wedge\boldsymbol{v}_2\wedge\cdots\wedge\boldsymbol{v}_n$ must vanish if $\boldsymbol{v}_1,\boldsymbol{v}_2,\ldots\boldsymbol{v}_n$ are linearly dependent (see Exercise 4.8.12). This clearly includes the important case where any two of the vectors are the same, or the same to within some scalar factor. On the other hand, the direct product $\boldsymbol{v}_1\boldsymbol{v}_2\cdots\boldsymbol{v}_n$ does not vanish in this way and although the inner product $\boldsymbol{v}_1\cdot\boldsymbol{v}_2\cdot\ldots\cdot\boldsymbol{v}_n$ vanishes for $2<n$, this is just a trivial case given that the inner product of any vectors with a scalar is by definition 0. We will have frequent need of these properties of the inner and outer products.

4.6 INNER AND OUTER PRODUCTS BETWEEN HIGHER GRADES

So far we have not mentioned specific rules for what happens in the case of $\boldsymbol{U}_m \cdot \boldsymbol{V}_k$ and $\boldsymbol{U}_m \wedge \boldsymbol{V}_k$ where $\boldsymbol{U}_m$ and $\boldsymbol{U}_k$ are *both* homogeneous multivectors of grades higher than 1, as in the case of $\boldsymbol{B} \cdot d\boldsymbol{A}$ in Equation (3.10) (ii). It turns out that while Equation (4.6) cannot be extrapolated to such cases, they may be addressed by extending Equation (4.5) in the following manner:

$$\begin{aligned} \boldsymbol{U}_m \cdot \boldsymbol{V}_k &= \langle \boldsymbol{U}_m \boldsymbol{V}_k \rangle_{|k-m|} \\ \boldsymbol{U}_m \wedge \boldsymbol{V}_k &= \langle \boldsymbol{U}_m \boldsymbol{V}_k \rangle_{k+m} \end{aligned} \tag{4.8}$$

If either $k = 0$ or $m = 0$, then $\boldsymbol{U}_m \cdot \boldsymbol{V}_k$ vanishes, while $\boldsymbol{U}_m \wedge \boldsymbol{V}_k$ vanishes for $N < k + m$. In general, the product $\boldsymbol{U}_m \boldsymbol{V}_k$ comprises a spectrum of n-vectors ranging from grades $|k-m|$ to $(k+m)_{\mathrm{mod}\,N+1}$ in steps of 2. This may be deduced inductively by first of all assuming that $\boldsymbol{U}_m$ and $\boldsymbol{V}_k$ are both blades of grades m and k respectively; however, it is more simply demonstrated when the blades are written as the geometric product of orthogonal basis vectors. The grade of the result will be determined by the number of basis vectors that the products representing $\boldsymbol{U}_m$ and $\boldsymbol{V}_k$ have in common. If they share l basis vectors in common, then there are $k + m - 2l$ remaining basis vectors between them that are distinct, and so the grade of their product $\boldsymbol{U}_m \boldsymbol{V}_k$ must be $k + m - 2l$. The maximum number of common basis vectors is $\min(k,m)$, whereas the minimum number is determined by the fact that $k + m - 2l$, the maximum number of distinct basis elements, cannot exceed N. These limits therefore imply that the grade of the result may take any value from $|k-m|$ to $(k+m)_{\mathrm{mod}\,N+1}$ in steps of 2. This result is not affected if $\boldsymbol{U}_m$ and $\boldsymbol{V}_k$ have to be expressed as a sum of blades, for it applies to every resulting product of blades that is formed when $\boldsymbol{U}_m$ and $\boldsymbol{V}_k$ are multiplied out. When k and m are chosen to maximize the number of possible grades in $\boldsymbol{U}_m \boldsymbol{V}_k$, they are either both equal to $N/2$ when N is even, or one is equal to $(N-1)/2$ while the other is equal to $(N+1)/2$ when N is odd. Depending on whether N is even or odd, the spectrum of possible grades in the result is therefore either $0,2,4,\cdots N$ or $1,3,5,\cdots N$. The inner and outer products can pick out only the lowest and highest of these grades, and it is for this reason that Equation (4.2) holds as an identity only when one of the objects involved is a vector.

In the important case of two bivectors such as $\boldsymbol{B}$ and $d\boldsymbol{A}$

$$\begin{aligned} \boldsymbol{B}\, d\boldsymbol{A} &= \langle \boldsymbol{B}\, d\boldsymbol{A} \rangle_0 + \langle \boldsymbol{B}\, d\boldsymbol{A} \rangle_2 + \langle \boldsymbol{B}\, d\boldsymbol{A} \rangle_4 \\ &= \boldsymbol{B} \cdot d\boldsymbol{A} + \langle \boldsymbol{B}\, d\boldsymbol{A} \rangle_2 \end{aligned} \tag{4.9}$$

$\boldsymbol{B} \wedge d\boldsymbol{A} = \langle \boldsymbol{B}\, d\boldsymbol{A} \rangle_4$ vanishes in 3D because 4 exceeds the dimension of the space, but be aware that it does not necessarily vanish in a 4D space, where the result would be a pseudoscalar.

On the other hand, $\boldsymbol{B}\cdot d\boldsymbol{A}$ is a scalar, confirming the assignment that was offered in Section 3.4. It was also pointed out in Section 3.2.2 that $\boldsymbol{B}\wedge\boldsymbol{m}$ must vanish, and here we have the reason why—$\boldsymbol{B}$ and $\boldsymbol{m}$ are both bivectors. The bivector term $\langle \boldsymbol{B}\, d\boldsymbol{A}\rangle_2$ is associated with neither the inner nor the outer product. However, it is often referred to as the *commutator product* of $\boldsymbol{B}$ and $d\boldsymbol{A}$. The commutator product and some other types of product, each of which offers a different sort of spectrum of blades in the result, do occasionally feature in the literature. In this introductory text, however, it is more useful to focus purely on the inner and outer product, and so these other forms are mentioned only briefly so that the reader will at least be able to recognize them. For any two multivectors $\boldsymbol{U}$ and $\boldsymbol{V}$

- the commutator product, symbol $\times$, is defined as $\boldsymbol{U}\times\boldsymbol{V}\equiv\frac{1}{2}(\boldsymbol{UV}-\boldsymbol{VU})$;
- the symmetric product is defined as $\frac{1}{2}(\boldsymbol{UV}+\boldsymbol{VU})$; and
- the scalar product, symbol $*$, is defined as $\boldsymbol{U}*\boldsymbol{V}=\langle \boldsymbol{UV}\rangle_0$.

However, some key points worth remembering are the following:

- We have used a bold $\times$ here for the commutator product so that it will not be confused with the cross product for vectors. Other authors simply use $\times$ for both.
- Despite appearances, in general, $\boldsymbol{U}\times\boldsymbol{V}$ is not the same thing as $\boldsymbol{U}\wedge\boldsymbol{V}$.
- Similarly, $\frac{1}{2}(\boldsymbol{UV}+\boldsymbol{VU})$ and $\boldsymbol{U}*\boldsymbol{V}$ both generally differ from $\boldsymbol{U}\cdot\boldsymbol{V}$.
- And, in general, $\boldsymbol{UV}\neq\boldsymbol{U}\cdot\boldsymbol{V}+\boldsymbol{U}\wedge\boldsymbol{V}$.
  - A particular example of this occurs when both $\boldsymbol{U}$ and $\boldsymbol{V}$ are bivectors so that $\boldsymbol{UV}$ results in a scalar $\boldsymbol{U}\cdot\boldsymbol{V}$, plus a 4-vector $\boldsymbol{U}\wedge\boldsymbol{V}$, *plus a bivector* that neither of these terms can account for. The fact that $\boldsymbol{U}\wedge\boldsymbol{V}$ vanishes in 3D makes no difference.
- In certain special cases however, equalities between some of the above products do occur (see Exercise 4.8.11).
- Formulas similar to Equation (4.6), such as $\boldsymbol{U}\cdot\boldsymbol{V}=\frac{1}{2}(\boldsymbol{UV}\pm\boldsymbol{VU})$ or $\boldsymbol{U}\wedge\boldsymbol{V}=\frac{1}{2}(\boldsymbol{UV}\mp\boldsymbol{VU})$, **cannot be used**. These only apply if one of $\boldsymbol{U}$ and $\boldsymbol{V}$ is a vector.
- Since Equations (4.8) apply only on a grade-by-grade basis, they cannot be used when either $\boldsymbol{U}_m$ or $\boldsymbol{V}_k$ is a general multivector.
- The inner and outer products of general multivectors such as $\boldsymbol{U}=\boldsymbol{U}_0+\boldsymbol{U}_1\cdots+\boldsymbol{U}_l$ and $\boldsymbol{V}=\boldsymbol{V}_0+\boldsymbol{V}_1\cdots+\boldsymbol{V}_n$ therefore require to be evaluated by applying Equation (4.8) to every individual term of the form $\boldsymbol{U}_m\boldsymbol{V}_k$ that occurs in their product.
- There is a general rule for expressions without brackets. Inner products should be formed first, then outer products, and finally direct products, for example, $\boldsymbol{u}\wedge\boldsymbol{v}\cdot\boldsymbol{wz}=\boldsymbol{u}\wedge(\boldsymbol{v}\cdot\boldsymbol{w})\boldsymbol{z}$. However, it is often less confusing to use brackets, as will generally be the case here.
- Where appropriate, orthogonality between multivectors $\boldsymbol{U}$ and $\boldsymbol{V}$ may be defined by

$$\boldsymbol{V} \perp \boldsymbol{U} \Leftrightarrow \boldsymbol{V} \cdot \boldsymbol{U} = 0 \tag{4.10}$$

- Parallelism between objects of different grades has limited scope. While it can be seen to apply to a vector and bivector when the vector lies in the bivector plane, in general, it is more meaningful to say $\boldsymbol{U}$ is within the space of $\boldsymbol{V}$ rather than try to say $\boldsymbol{U} \,/\!/\, \boldsymbol{V}$.
- Two objects $\boldsymbol{U}$ and $\boldsymbol{V}$ of the same nonzero grade, however, may be defined as being parallel through

$$\boldsymbol{V} \,/\!/\, \boldsymbol{U} \Leftrightarrow \boldsymbol{V}\boldsymbol{U} = \langle \boldsymbol{V}\boldsymbol{U} \rangle \tag{4.11}$$

  that is to say, $\boldsymbol{V}\boldsymbol{U}$ is a scalar.

4.7 SUMMARY SO FAR

The vector, dot and cross products are replaced in geometric algebra by systematic inner and outer products that depend on the grades of object involved. The inner product results in a step-down in grade, whereas the outer product results in a step-up. Measure originates directly from the geometric product itself rather than from the inner product, which can be viewed as a secondary construct. Given a set of basis vectors, a closed set of elements called blades can be generated by multiplication so as to span the entire geometric algebra. There are two special blades, the scalar and the pseudoscalar, that have grades 0 and N, respectively, where N is the maximum possible grade. Since the outer product always increases grade, it is just the operation needed to form blades out of other objects, in particular from nonorthogonal vectors.

In working with expressions in a geometric algebra, results can always be determined by first expressing the objects in terms of the basis elements (scalars, vectors, and other blades) then carrying out the required operations according to the rules. In particular, inner and outer products can always be reduced to a sum of direct products, for example, by using Equation (4.6) where applicable. Because of the antisymmetry of the outer product for vectors, we also find that the usual rules for reordering products of orthogonal vectors given in Section 2.1 apply equally to the outer products of vectors in general. But remember the following:

- The direct product of two vectors can only be commuted or anticommuted if they are respectively either parallel or orthogonal.
- Direct products generally produce multivectors rather than blades.
- Provided the result is nonzero, the outer product of any two blades of grades m and n produces a third blade of grade $m+n$.
- Homogeneous multivectors are the same as n-vectors.
- Blades and n-vectors, however, are not necessarily the same thing. Blades can always be written as a product of vectors, whereas n-vectors may require a sum of such products.

- A pseudoscalar cannot be simply factored out of an inner or outer product since it changes the grade of one of the objects involved. For example, $\boldsymbol{u}\cdot(I\boldsymbol{v}) = I(\boldsymbol{u}\wedge\boldsymbol{v})$ if the dimension of the space is odd, in which case pseudoscalars commute with vectors; otherwise, we get $\boldsymbol{u}\cdot(I\boldsymbol{v}) = -I(\boldsymbol{u}\wedge\boldsymbol{v})$.
- Equations (4.1) through (4.7) apply only when one of the objects is a vector.
- Nevertheless, the rule given in Equation (4.8) provides a general method of evaluating inner and outer products for objects of any grade.
- But the rule *can only be applied* where the objects concerned are each of a specific grade.
- General multivectors must be expressed as a sum of separate grades upon which the rule may be applied to each term in the resulting product.

There is much more to the study of geometric algebra than this, but it is beyond the scope of this work to give more than a brief outline of the essentials, some feeling for its potential, and a reasonable idea of how it works. Enough at least, we hope, to be able to appreciate the fresh perspective it gives to the mathematical representation of electromagnetic theory. Topics of further interest to the reader, such as the relationship between geometric algebra and the complex numbers in 2D and quaternions in 3D, are to be found in Appendices 14.3 and 14.5. These topics, together with such things as applications to mechanics and computer graphics, are also discussed in References 6–8, 26–28, 32, and 33. In addition, starting from the properties of vector spaces, Appendix 14.4.2 summarizes all the main rules for geometric algebras. At the end of the Appendix there are also some simple theorems that may prove useful.

4.8 EXERCISES

1. For any three vectors **u**, **v**, and **w** in a 3D geometric algebra:
   (a) What is the grade of $\mathbf{u}\cdot(\mathbf{v}\wedge\mathbf{w})$?
   (b) With reference to Figure 2.1, what is the geometric interpretation of $\mathbf{u}\cdot(\mathbf{v}\wedge\mathbf{w})$ when **u** is first parallel to the $\mathbf{v}\wedge\mathbf{w}$ plane and then perpendicular to it?
   (c) Show that $\mathbf{u}\cdot(\mathbf{v}\wedge\mathbf{w}) = -\mathbf{u}\times(\mathbf{v}\times\mathbf{w})$.

2. For any three vectors $\boldsymbol{u}$, $\boldsymbol{v}$, and $\boldsymbol{w}$, show the following:
   (a) $\boldsymbol{u}\cdot(\boldsymbol{u}\cdot(\boldsymbol{u}\wedge\boldsymbol{v}\wedge\boldsymbol{w})) = 0$,
   (b) $\boldsymbol{u}\cdot I = \boldsymbol{u}I$, and
   (c) $\boldsymbol{u}\wedge I = 0$.
   Give a geometric interpretation for each of the results in each case.

3. Prove $\nabla\cdot(I\mathbf{u}) = I\nabla\wedge\mathbf{u}$ and $\nabla\wedge(I\mathbf{u}) = I\,\nabla\cdot\mathbf{u}$.

4. **(a)** If the vector **u** is parallel to the $\mathbf{v}\wedge\mathbf{w}$ plane, show that $\mathbf{u}\cdot(\mathbf{v}\wedge\mathbf{w}) = (\mathbf{w}\wedge\mathbf{v})\mathbf{u}$ and hence show that if **u** lies in the plane of some unit bivector $\boldsymbol{N}$, then $(\alpha-\beta\boldsymbol{N})\mathbf{u}$ rotates **u** through an angle ϕ in the $\boldsymbol{N}$-plane in the same sense as $\boldsymbol{N}$ provided that $\alpha = \cos\phi$ and $\beta = \sin\phi$. While this combination of scalar plus bivector is referred to as a rotor, note that this property is limited to vectors lying in the plane of the bivector.

(b) Show that $(\alpha+\beta\boldsymbol{N})^2=\alpha^2-\beta^2+2\alpha\beta\boldsymbol{N}$ and $(\alpha-\beta\boldsymbol{N})(\alpha+\beta\boldsymbol{N})=1$ *without* resorting to multiplying out these expressions.

(c) Show that $(\alpha-\beta\boldsymbol{N})\mathbf{u}(\alpha+\beta\boldsymbol{N})$ rotates *any* vector $\mathbf{u}$ by an angle 2ϕ in the $\boldsymbol{N}$-plane (hint: consider the commutation properties of suitably defined $\mathbf{u}_\perp$ and $\mathbf{u}_{/\!/}$ with $\boldsymbol{N}$).

5. (a) For any two homogeneous multivectors $\boldsymbol{U}$ and $\boldsymbol{V}$ in a space of dimension $\boldsymbol{N}$, under what conditions do $\boldsymbol{U}\cdot(I\boldsymbol{V})=I\boldsymbol{U}\wedge\boldsymbol{V}$ and $\boldsymbol{U}\wedge(I\boldsymbol{V})=I\boldsymbol{U}\cdot\boldsymbol{V}$ hold good?

(b) If $\boldsymbol{UV}=0$, what can be said about $\boldsymbol{U}\cdot\boldsymbol{V}$ and $\boldsymbol{U}\wedge\boldsymbol{V}$?

(c) If $\boldsymbol{U}\cdot\boldsymbol{V}=0$ and $\boldsymbol{U}\wedge\boldsymbol{V}=0$, what can be said about $\boldsymbol{UV}$?

(d) In the case of 3D, compare $(I\boldsymbol{U})\cdot\boldsymbol{V}$, $\boldsymbol{U}\cdot(I\boldsymbol{V})$, and $I(\boldsymbol{U}\cdot\boldsymbol{V})$.

(e) Show that $(I\boldsymbol{U})\cdot\boldsymbol{V}$ represents the intersection of the bivectors $\boldsymbol{U}$ and $\boldsymbol{V}$.

6. Confirm that the outer product is indeed associative; that is, $(\boldsymbol{a}\wedge\boldsymbol{b})\wedge\boldsymbol{c}=\boldsymbol{a}\wedge(\boldsymbol{b}\wedge\boldsymbol{c})$.

7. For any three vectors $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{w}$:

(a) Reduce the direct product $\boldsymbol{uvw}$ to a sum of separate grades (hint: write the direct products using only inner and outer products and assess the grades of each term using the step-up and -down concept).

(b) If $\boldsymbol{U}=\boldsymbol{u}\wedge\boldsymbol{w}$ and $\boldsymbol{V}=\boldsymbol{v}\wedge\boldsymbol{w}$, evaluate $\boldsymbol{UV}$, $\boldsymbol{U}\cdot\boldsymbol{V}$, and $\boldsymbol{U}\wedge\boldsymbol{V}$. Contrast these results with the case when both $\boldsymbol{U}$ and $\boldsymbol{V}$ are vectors.

(c) Show that $\boldsymbol{a}\cdot(\boldsymbol{b}\wedge\boldsymbol{c})+\boldsymbol{b}\cdot(\boldsymbol{c}\wedge\boldsymbol{a})+\boldsymbol{c}\cdot(\boldsymbol{a}\wedge\boldsymbol{b})=0$.

(d) Simplify $(\boldsymbol{u}+\boldsymbol{v})(\boldsymbol{v}+\boldsymbol{w})(\boldsymbol{w}+\boldsymbol{u})$.

(e) Simplify the result for (b) in the case where the three vectors are mutually orthogonal.

8. (a) Given $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, and $\boldsymbol{d}$ are vectors and $\boldsymbol{U}=(\boldsymbol{a}\wedge\boldsymbol{b})\wedge(\boldsymbol{c}\wedge\boldsymbol{d})$, find $\boldsymbol{U}^\dagger$.

(b) Find inverses for $\boldsymbol{a}+\boldsymbol{b}$, $\boldsymbol{a}\wedge\boldsymbol{b}$ (hint: $\boldsymbol{a}\wedge\boldsymbol{b}=\boldsymbol{ab}-\boldsymbol{a}\cdot\boldsymbol{b}$), and $\boldsymbol{a}\wedge\boldsymbol{d}+\boldsymbol{c}\wedge\boldsymbol{d}$.

9. Prove that pseudoscalars commute with vectors in a geometric algebra of odd dimension, whereas they anticommute if the dimension is even (hint: use the basic properties of inner and outer products).

10. (a) Prove that for any vector $\boldsymbol{u}$ in an N-dimensional geometric algebra, the grade of $I\boldsymbol{u}$ is $N-1$.

(b) If $\boldsymbol{U}=\boldsymbol{a}_1\wedge\boldsymbol{a}_2\wedge\cdots\boldsymbol{a}_n$, show that $\boldsymbol{U}^\dagger=-1^{n(n-1)/2}\boldsymbol{U}$.

(c) Hence show that in a space of dimension N, $I^\dagger=-1^m I$ where $m=N/2$ rounded down to the nearest integer. Compare the results for spaces of dimension 3 and 4.

11. (a) In the case of homogeneous multivectors $\boldsymbol{U}_m$ and $\boldsymbol{V}_k$, discuss a way of finding how many different grades are involved in $\boldsymbol{U}_m\boldsymbol{V}_k$.

(b) Show that in general $\boldsymbol{U}_m\boldsymbol{V}_k\not\equiv\boldsymbol{U}_m\cdot\boldsymbol{V}_k+\boldsymbol{U}_m\wedge\boldsymbol{V}_k$ unless at least one of m and k is less than 2. (This is the reason why in Equation 4.1, $\boldsymbol{u}$ is restricted to being either a scalar or a vector.)

(c) For vectors, $\boldsymbol{U}\times\boldsymbol{V}$ and $\boldsymbol{U}*\boldsymbol{V}$ are the same as the outer and inner products, what is the case when $\boldsymbol{U}$ and $\boldsymbol{V}$ are bivectors?

(d) Discuss whether it is possible to use the commutator product to express the total torque acting on a body in an electromagnetic field $\boldsymbol{F}$ when it has both electric and magnetic dipole moments.

12. Given any set of k linearly independent vectors $\boldsymbol{p},\boldsymbol{q},\boldsymbol{r},\boldsymbol{s}\ldots$, show that $\boldsymbol{U}=\boldsymbol{p}\wedge\boldsymbol{q}\wedge\boldsymbol{r}\wedge\boldsymbol{s}\wedge\ldots$ will form a blade.

13. If $\boldsymbol{N}$ is any bivector and $\boldsymbol{u}$ is any vector, show the following:

(a) $\boldsymbol{Nu} = \boldsymbol{uN} \Leftrightarrow \boldsymbol{u} \perp \boldsymbol{N}$, whereas $\boldsymbol{Nu} = -\boldsymbol{uN} \Leftrightarrow \boldsymbol{u} /\!/ \boldsymbol{N}$.

(b) If $\boldsymbol{a}$ and $\boldsymbol{b}$ are any two vectors that are both orthogonal to $\boldsymbol{N}$, then $(\boldsymbol{a} \wedge \boldsymbol{b} \wedge \boldsymbol{N})^{\dagger} = -\boldsymbol{a} \wedge \boldsymbol{b} \wedge \boldsymbol{N}$.

(c) If $\boldsymbol{a}$ and $\boldsymbol{b}$ are any two vectors parallel to $\boldsymbol{N}$, then $\boldsymbol{N} /\!/ \boldsymbol{a} \wedge \boldsymbol{b}$ provided $\boldsymbol{a} \wedge \boldsymbol{b} \neq 0$.

14. If $\boldsymbol{u}$ is any vector and $\boldsymbol{B}$ is a bivector, show that $\boldsymbol{w} = \boldsymbol{u} \cdot \boldsymbol{B}$ is a vector such that $\boldsymbol{w} \perp \boldsymbol{u}$.

15. Any vector $\boldsymbol{u}$ that is given in terms of an orthonormal basis as $u_x\boldsymbol{x} + u_y\boldsymbol{y} \ldots + u_w\boldsymbol{w}$ may be expressed in matrix form by

$$\begin{bmatrix} \boldsymbol{x} & \boldsymbol{y} & \cdots & \boldsymbol{w} \end{bmatrix} \begin{bmatrix} u_x \\ u_y \\ \vdots \\ u_w \end{bmatrix}$$

(a) Show that

$$\boldsymbol{uv} = \begin{bmatrix} u_x & u_y & \cdots & u_w \end{bmatrix} \begin{bmatrix} 1 & \boldsymbol{xy} & \cdots & \boldsymbol{xw} \\ \boldsymbol{yx} & 1 & & \boldsymbol{yw} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{wx} & \boldsymbol{wy} & \cdots & 1 \end{bmatrix} \begin{bmatrix} v_x \\ v_y \\ \vdots \\ v_w \end{bmatrix}$$

(b) How would $\boldsymbol{u} \cdot \boldsymbol{v}$ and $\boldsymbol{u} \wedge \boldsymbol{v}$ be represented?

(c) How should a bivector such as $U_{xy}\boldsymbol{xy} + U_{yz}\boldsymbol{yz} \ldots$ be expressed?

(d) Can a trivector be represented? Consider the case of the unit pseudoscalar in 3D.

16. Show that for any two bivectors $\boldsymbol{A}$ and $\boldsymbol{B}$ in 3D,

(a) $\boldsymbol{A}$ and $\boldsymbol{B}$ share $I\boldsymbol{A} \times \boldsymbol{B}$ as a common edge;

(b) $I\boldsymbol{A} \times \boldsymbol{B} = (I\boldsymbol{A}) \times (I\boldsymbol{B})$; and

(c) if $\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{C}$, then $\boldsymbol{A}^2 + \boldsymbol{B}^2 = \boldsymbol{C}^2 \Leftrightarrow \boldsymbol{A} \perp \boldsymbol{B}$.

17. If an even multivector is the sum of objects of even grade, while an odd multivector is the sum of objects of purely odd grades, show that the product of two even or two odd multivectors must be an even multivector. Similarly, the product of an even and an odd multivector can only result in an odd multivector.

18. For any two non-null multivectors $\boldsymbol{U}$ and $\boldsymbol{V}$:

(a) Discuss the possible meanings of orthogonality.

(b) Show that the condition $\boldsymbol{U} \cdot \boldsymbol{V} = 0$ is not in general equivalent to $\boldsymbol{U} \wedge \boldsymbol{V} \neq 0$.

(c) Explain what is meant by saying that $\boldsymbol{U}$ is within the space of $\boldsymbol{V}$.

Chapter 5

(3+1)D Electromagnetics

We should now be in a position to put geometric algebra to some practical use in (3+1)D electromagnetic theory. The chosen topics have been selected as a representative cross section of the fundamental electromagnetic building blocks—the Lorentz force; Maxwell's equations; the potential; the field of a moving point charge; charge conservation; plane waves; the energy and momentum densities; polarizable media; and finally, the boundary conditions at an interface. The reader is expected to be reasonably familiar with all of these subjects; all we are attempting to do here is to show them in a fresh light by bringing a new toolset to bear. Once we are well under way, a simpler way of writing equations will be introduced so as to make the task a little easier by reducing key results to their essential form.

5.1 THE LORENTZ FORCE

The Lorentz force is the force $\mathbf{f}$ acting on a particle of charge q while it moves with velocity $\mathbf{v}$ through an electric field $\mathbf{E}$ and magnetic field $\mathbf{B}$. Conventionally, the scalar, true vector and axial vector quantities involved are brought together in the well-known relationship written as

$$\mathbf{f} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \tag{5.1}$$

Is there any better way to express this equation using geometric algebra?

As before, the true vectors, here $\mathbf{f}$, $\mathbf{E}$, and $\mathbf{v}$, carry straight over into a geometric algebra, but we already know that an axial vector such as $\mathbf{B}$ behaves differently and must be replaced by the only other possibility, a bivector, which we write, according to our convention, as $\boldsymbol{B}$. As discussed in Chapter 3, the bivector and vector forms are duals that are simply related by $\boldsymbol{B} = I\mathbf{B}$. However, if a bivector is involved, it is clear that the magnetic contribution to the Lorentz force must take some form other than $q\mathbf{v} \times \mathbf{B}$. The result we require must be a vector, and so, counter intuitive though it may initially seem, the cross product $\mathbf{v} \times \mathbf{B}$ is replaced by the inner product $\boldsymbol{B} \cdot \mathbf{v}$, which follows from using Equation (2.8) to write $\mathbf{v} \times \mathbf{B}$ as $-I\mathbf{v} \wedge \mathbf{B}$. Note that even though I commutes with everything, it cannot be

Understanding Geometric Algebra for Electromagnetic Theory, First Edition. John W. Arthur.

simply introduced into, or factored out of, an inner or outer product since this alters the grades of the terms involved in the product. Any change of grade then affects the form of the product as shown in Equation (4.6), and so we find $-I\mathbf{v}\wedge\mathbf{B}=-\frac{1}{2}I\left(\mathbf{vB}-\mathbf{Bv}\right)=\frac{1}{2}\left(I\mathbf{Bv}-\mathbf{v}I\mathbf{B}\right)=\boldsymbol{B}\cdot\mathbf{v}$. It is clear from Equation (4.7) that the inner product between vector and bivector preserves the antisymmetric form of the cross product, while from Equation (4.5), it also picks out the lowest-grade part of the direct product, a vector, both of which are required characteristics of the result. The order of the terms in $\boldsymbol{B}\cdot\mathbf{v}$ is different from $\mathbf{v}\times\mathbf{B}$ only so as to avoid using a negative sign, but note that it is consistent with $\boldsymbol{B}\wedge\boldsymbol{m}$, which we encountered in the case of the torque on a magnetic dipole (Section 3.2.2). There is nothing surprising in the electric term, however, and so for the time being, we may write

$$\mathbf{f}=q\left(\mathbf{E}+\boldsymbol{B}\cdot\mathbf{v}\right) \tag{5.2}$$

It seems that little of significance has changed, only the notation. At least the systematic formalism of geometric algebra does make quite clear the respective vector and bivector characters of the key variables involved. If $\boldsymbol{B}$ were mistakenly taken as a vector, neither $\mathbf{v}\cdot\mathbf{B}$ nor $\mathbf{v}\wedge\mathbf{B}$, and not even $\mathbf{vB}$, would work since none of these expressions results in a pure vector. But is there a rule that we can use to evaluate $\boldsymbol{B}\cdot\mathbf{v}$ without reverting to the cross product? It will be helpful to refer to Figure 2.1 and in particular diagrams (b) and (h). First, we need to identify the orientation of $\boldsymbol{B}$. If, for example, $\boldsymbol{B}=\mathbf{xy}$, we travel along $\mathbf{x}$ and then along $\mathbf{y}$, which takes us anticlockwise in the $\mathbf{xy}$ plane. Rotate $\mathbf{v}_{/\!/}$, the part of $\mathbf{v}$ that lies in the plane of $\boldsymbol{B}$, by 90° *in the opposite sense*. Multiply the result by $|\boldsymbol{B}|$ to get $\boldsymbol{B}\cdot\mathbf{v}$. Perhaps evaluating it geometrically in this way is less easy than using $\mathbf{v}\times\mathbf{B}$, but that is because we are already so familiar with the cross product. Evaluating it algebraically, however, there is no real difference.

While this may be mildly encouraging, it does not seem possible at this stage to write the Lorentz force simply in terms of a multivector electromagnetic field such as $\boldsymbol{F}$ in Equation (3.2) since the inner product with $\mathbf{v}$ applies to $\boldsymbol{B}$ alone. While we could write $\mathbf{f}=q\langle\boldsymbol{F}\left(1+\mathbf{v}/c\right)\rangle_1$ (see Exercise 5.10.5), the awkwardness of this form is overcome in 4D spacetime, as will be seen in due course!

5.2 MAXWELL'S EQUATIONS IN FREE SPACE

The differential form of Maxwell's equations in free space hardly needs introduction:

$$\begin{aligned}\boldsymbol{\nabla}\cdot\mathbf{E} &= \frac{\rho}{\varepsilon_0}\\ \boldsymbol{\nabla}\times\mathbf{E} &= -\partial_t\mathbf{B}\\ \boldsymbol{\nabla}\cdot\mathbf{B} &= 0\\ \boldsymbol{\nabla}\times\mathbf{B} &= \mu_0\mathbf{J}+\frac{1}{c^2}\partial_t\mathbf{E}\end{aligned} \tag{5.3}$$

These are the *microscopic* equations in which *all* sources of charge and current are taken into account. By the addition of a constitutive relation, they yield the more usual *macroscopic* equations that include the auxiliary fields **D** and **H**, which take into account the *bound* electric and magnetic sources within matter, leaving only the *free* charge and current sources to be dealt with directly [2]. The impact of using geometric algebra on Maxwell's equations, however, is best appreciated through the microscopic form shown here. We will address the macroscopic equations a little later on in Section 5.9.

In Chapter 3, we discussed the integral form of Maxwell's equations and the nature of electric and magnetic dipoles, from which we concluded that the electric field was best represented by a vector, **E**, whereas the magnetic field was best represented by a bivector, $\boldsymbol{B}$. Recall also from Chapter 3 that in 3D, the vector derivative acting on a vector such as **E** may be written in terms of divergence and curl as $\nabla\mathbf{E} = \nabla\cdot\mathbf{E} + I\nabla\times\mathbf{E}$. In the case of a bivector like $\boldsymbol{B}$, however, we have $\nabla\boldsymbol{B} = \nabla(I\mathbf{B}) = I\nabla\mathbf{B}$ where **B** is the vector that is dual to $\boldsymbol{B}$. Here we have treated ∇ just like any other vector and used the fact that I commutes with all 3D vectors.[1] We may now apply this to Maxwell's equations as follows:

$$\begin{aligned} \nabla\mathbf{E} &= \nabla\cdot\mathbf{E} + I\nabla\times\mathbf{E} = \frac{\rho}{\varepsilon_0} - I\partial_t\mathbf{B} \\ \nabla(cI\mathbf{B}) &= Ic\nabla\cdot\mathbf{B} + I^2c\nabla\times\mathbf{B} = -c\mu_0\mathbf{J} - \frac{1}{c}\partial_t\mathbf{E} \end{aligned} \tag{5.4}$$

where $c\mu_0 = Z_0$, the characteristic impedance of free space. Taking first the sum and then the difference of these two equations provides us with

$$\begin{aligned} \nabla(\mathbf{E} + c\boldsymbol{B}) &= \left(\frac{\rho}{\varepsilon_0} - Z_0\mathbf{J}\right) - \frac{1}{c}\partial_t(\mathbf{E} + c\boldsymbol{B}) \quad \text{(i)} \\ \nabla(\mathbf{E} - c\boldsymbol{B}) &= \left(\frac{\rho}{\varepsilon_0} + Z_0\mathbf{J}\right) + \frac{1}{c}\partial_t(\mathbf{E} - c\boldsymbol{B}) \quad \text{(ii)} \end{aligned} \tag{5.5}$$

As before, the bivector $\boldsymbol{B}$ is equal to $I\mathbf{B}$. From version (i), we may write

$$\left(\nabla + \frac{1}{c}\partial_t\right)(\mathbf{E} + c\boldsymbol{B}) = \frac{\rho}{\varepsilon_0} - Z_0\mathbf{J} \tag{5.6}$$

or, in terms of the multivector fields $\boldsymbol{F}$ and $\boldsymbol{J}$ that were introduced in Section 3.2.1,

$$\left(\nabla + \frac{1}{c}\partial_t\right)\boldsymbol{F} = \boldsymbol{J} \tag{5.7}$$

where it is now clear that the constants involved were chosen to be consistent with Equation (5.6), that is to say

[1] This is in contrast to cases where an inner or outer product is involved. As pointed out in Section 5.1, I cannot be simply factored out of an inner or outer product without due regard to the change of grade that results for one or other of the items involved in the product.

$$\boldsymbol{F} = \mathbf{E} + c\boldsymbol{B} \quad \text{and} \quad \boldsymbol{J} = \frac{\rho}{\varepsilon_0} - Z_0\mathbf{J} \tag{5.8}$$

On the other hand, from version (ii), we could equally well write

$$\left(\boldsymbol{\nabla} - \frac{1}{c}\partial_t\right)\boldsymbol{F}' = \boldsymbol{J}' \tag{5.9}$$

with $\boldsymbol{F}' = \mathbf{E} - c\boldsymbol{B}$ and $\boldsymbol{J}' = \dfrac{\rho}{\varepsilon_0} + Z_0\mathbf{J}$. Clearly, the two versions are equivalent and in fact they are related simply by spatial inversion (involution). It is therefore necessary to select one of them to be the standard form, and it is conventional to use (i), that is to say Equations (5.7) and (5.8).

The multivector variables $\boldsymbol{F}$ and $\boldsymbol{J}$ chosen above are the same as those put forward in Equations (3.2) and (3.3). They lead to a rendition of Maxwell's equations that is amazingly compact while at the same time being completely free of any "trick" such as replacing matrices with algebraic symbols. In fact, we should refer to Equation (5.7) simply as Maxwell's equation! Any attempt to find solutions of the four traditional equations usually results in a pair of second-order differential equations, but here we have just a single first order one. The impression is given that somehow the mathematics of geometric algebra fits very well the underlying mechanisms that govern the electromagnetic field. By comparison, traditional vector analysis provides only a partial and somewhat inelegant fit.

In spite of this success, however, there are a number of points that seem to detract from what would have been an otherwise seemingly perfect picture.

First, we do not meet with quite the same level of compactness for Maxwell's equations in a polarizable medium, in which either the dielectric constant or relative magnetic permeability is different from unity. We may have expected that bringing in an auxiliary electromagnetic field multivector, such as $\boldsymbol{G} = \mathbf{D} + I\mathbf{H}$, would make it possible to express Maxwell's equation in terms of $\boldsymbol{F}$, $\boldsymbol{G}$ and the *free* electromagnetic source density $\boldsymbol{J}_{free}$ alone, but, as will be discussed in Section 5.9, this cannot be achieved without using a grade selection filter, just as was the case when we tried to express the Lorentz force in terms of $\boldsymbol{F}$ and $q\mathbf{v}$ alone. This by no means invalidates the outcome; it just does not have the same compelling simplicity as in the case of free space. We could therefore describe this as "just a cosmetic problem," but all the same, it seems to suggest that there is some fundamental limitation with our present attempt to encode the physical model using geometric algebra.

Second, a more subtle point is that the two derivatives in Equation (5.7) stand apart, one being a vector and the other a scalar. Why then do we not combine the time and space derivatives under one symbol? This is indeed possible, resulting in a multivector derivative $\boldsymbol{\nabla} + \frac{1}{c}\partial_t$ being referred to by a single symbol such as $\boldsymbol{\nabla}$. But this would be a departure from the concept of ∇ in general being a *vector* differential operator. We have paid attention to $\mathbf{E}$, $\boldsymbol{B}$, $\mathbf{J}$, and ρ as though they were entities that depend on a 3D position vector and a separate scalar time, with the consequence that the loose time derivative is left in Equation (5.7). On this basis, the description (3+1)D does indeed seem quite appropriate to this form of treatment.

On reflection, however, there seems to be no logical reason why time has not been treated as a vector just like **x**, **y**, and **z**.

While it is clear that some progress has been made, it is also clear that if a total fit between the mathematics of geometric algebra and the mechanisms of electrodynamics is going to be possible, we must go further. The problem with the separate time derivative is perhaps a clue to the fact that we will require to empoly a full 4D treatment where time is the fourth basis vector. It remains to be seen whether or not this will fix the other two minor "cosmetic problems".

5.3 SIMPLIFIED EQUATIONS

To simplify matters from now on, we could begin to measure time and distance with the same meterstick. If we adopted 1 m of time as being the same thing as ~3.33564095 ns, we would find that the speed of light would be dimensionless and equal to unity. If we then take $Z_0 = 1$, this would have the effect of normalizing ε_0 and μ_0 as well. With few exceptions, all the usual fundamental constants would disappear from the equations we require, and this is therefore a common convention to be seen in the literature [27, 34], especially in the physics of spacetime. However, it is still possible to gain the advantages of this idea without departing from SI units by means of a form of shorthand that simply amounts to suppressing those physical constants that monotonously recur in equation after equation. In fact, this can be regularized by understanding the variables in any such equation to be modified in a very simple way, for it is only necessary to incorporate an appropriate constant within each variable as shown in Table 5.1, resulting in what we may call "modified variables". In most cases, the equations are so familiar that this is something we hardly need to think about.

As an example, it is easy to see how this works on Equation (5.6):

$$\left(\boldsymbol{\nabla} + \frac{1}{c}\partial_t\right)(\mathbf{E} + c\boldsymbol{B}) = \frac{\rho}{\varepsilon_0} - Z_0\mathbf{J} \quad \rightleftarrows \quad (\boldsymbol{\nabla} + \partial_t)(\mathbf{E} + \boldsymbol{B}) = \rho - \mathbf{J} \tag{5.10}$$

We could have identified all the modified variables by using different letters or by priming them but as long as we are consistent in their use there seems little point of doing so. The obvious absence of the usual physical constants tells us when this convention is being used and, while it is obviously a little more tricky to translate back to the normal form, it is only a matter of spotting the time derivative, magnetic field, charge and current in any given expression. A second-order time derivative will clearly involve $1/c^2$ and so on. Table 5.1 may be used if necessary by simply replacing the shorthand form with the corresponding complete form. Importantly, all shorthand equations must remain dimensionally correct in terms of the modified variables.

From time to time, for example when we introduce a new equation or expression, it may be appropriate to highlight any constants that would normally be suppressed in the shorthand form. When we do so, any such constants will be shown

Table 5.1 Representation of Frequently Used Expressions in Simplified Equations

Complete SI form	Units	Modified form	Complete Gaussian form	Units
x, y, z	m	x, y, z	x, y, z	cm
ct	m	t	ct	cm
$\mathbf{v}/c$	1	$\mathbf{v}$	$\mathbf{v}/c$	1
$\frac{1}{c}\partial_t$	m^{-1}	∂_t	$\frac{1}{c}\partial_t$	cm^{-1}
$\mathbf{E}$	Vm^{-1}	$\mathbf{E}$	$\mathbf{E}$	$\mathrm{statVcm}^{-1}$
$c\boldsymbol{B}$	Vm^{-1}	$\boldsymbol{B}$	$\boldsymbol{B}$	$\mathrm{statVcm}^{-1}$
ρ/ε_0	Vm^{-2}	ρ	ρ	$\mathrm{statCcm}^{-3}$
$Z_0\mathbf{J}$	Vm^{-2}	$\mathbf{J}$	$\frac{1}{c}\mathbf{J}$	$\mathrm{statCcm}^{-3}$
Φ	V	Φ	Φ	statV
$c\mathbf{A}$	V	$\mathbf{A}$	$\mathbf{A}$	statV
$\mathscr{E}/\varepsilon_0$	$\mathrm{V}^2\mathrm{m}^{-2}$	$\mathscr{E}$	$\mathscr{E}$	$\mathrm{statV^2cm}^{-2}$
$\mathbf{D}/\varepsilon_0, \mathbf{P}/\varepsilon_0$	Vm^{-1}	$\mathbf{D}, \mathbf{P}$	$\mathbf{D}, \mathbf{P}$	$\mathrm{statVcm}^{-1}$
$Z_0\mathbf{H}, Z_0\mathbf{M}$	Vm^{-1}	$\mathbf{H}, \mathbf{M}$	$\mathbf{H}, \mathbf{M}$	$\mathrm{statVcm}^{-1}$

In simplified equations, the usual variables are modified by the inclusion of an appropriate physical constant so that the symbols in the middle column now replace the expressions in the far left column in all the standard equations. For readers who are accustomed to Gaussian units, they also replace the expressions on the far right. As a consequence, SI and Gaussian equations look the same, but the main advantage is that quantities having different units may be combined into multivector form without having to repeatedly write out all the constants involved. This notation therefore allows us to concentrate on the form of equations rather than the detail, as for example in $(\boldsymbol{\nabla} + \partial_t)(\mathbf{E} + \boldsymbol{B}) = \rho - \mathbf{J}$ compared with $\left(\boldsymbol{\nabla} + \frac{1}{c}\partial_t\right)(\mathbf{E} + c\boldsymbol{B}) = (\rho/\varepsilon_0) - Z_0\mathbf{J}$.

in square brackets, for example, $\mathbf{D} = [\varepsilon_0]\mathbf{E} + \mathbf{P}$ spells out what is meant by the modified form $\mathbf{D} = \mathbf{E} + \mathbf{P}$.

Finally, note that since it would be impracticable to do so, this system does not apply to numerical factors and transcendental constants such as $\frac{1}{2}$ and 4π.

5.4 THE CONNECTION BETWEEN THE ELECTRIC AND MAGNETIC FIELDS

We now return to the notion introduced in the closing remarks of Section 3.2.1, which implied that there should in principle be a relationship between the magnetic field of a moving charge distribution and its electric field, and that this should be simply expressed in terms of geometric algebra. Whereas we know that an electric current is simply the result of charges in motion, classical electromagnetic theory treats charge and current as separate things from the outset. The main relationship linking the two comes in the form of Equation (3.6), the continuity equation. Although it is implicit in Maxwell's equations, this equation is frequently considered to be a separate requirement with the consequence that electrostatics and magnetostatics are almost universally treated on separate footings. In the present formulation, we have a multivector electromagnetic source density $\boldsymbol{J} = \rho - \mathbf{J}$ (using the modified

form just introduced) that combines charge and current. The dimensions of the scalar and vector parts of $\boldsymbol{J}$ are compatible (Vm^{-2}), as shown in Table 5.1. Fundamentally, however, we should recognize $\boldsymbol{J}$ as being given by

$$\boldsymbol{J}(\mathbf{r}) = \sum_i \frac{\rho_i(\mathbf{r})}{[\varepsilon_0]}\left(1 - \frac{\mathbf{v}_i(\mathbf{r})}{[c]}\right) \tag{5.11}$$

We have made the constants reappear so as to draw attention to the factor $1/c$ that accompanies $\mathbf{v}_e(\mathbf{r})$, and while in principle the sum requires to be over all the different types of charge carrier involved, we can simplify the discussion by focusing on the net positive and negative charge distributions, which we designate by $i = +$ or $i = -$ as appropriate. Equation (5.11) makes it quite clear that under normal circumstances, that is to say $v_i \ll c$, the motion of the charge is a secondary factor. The magnetic field of an unbalanced charge is minuscule in comparison with its electric field so that the motion represents a tiny perturbation to the field arising from a static charge distribution $\rho_+ + \rho_-$. There is a commonplace exception to this, however, when $\rho_+ + \rho_- = 0$. Now the primary effect, the electric field, is eliminated due to the exact balance of positive and negative charges resulting in a completely neutral charge distribution. But this situation does not preclude motion within the charge distributions so that $\boldsymbol{J}$ may still be nonzero, in which case it will give rise to *the secondary effect* on its own, a magnetic field that is now readily observable—the electrically neutral current-carrying conductor being the ubiquitous example. The magnetic field may therefore arguably be said to arise out of some modification of the electric field due to the motion of the charge and so we should therefore expect a closer link between magnetostatics and electrostatics than the classical theory tends to show.

This preamble therefore leads us to the question, can we derive the magnetic field due to a moving charge distribution simply starting from the electric field of the static distribution? Let us assume the quasistatic condition $\partial_t \to 0$ as is the case for any steady flow of current around a circuit (or where the end points of the flow are at infinity). The starting point is therefore the equation for the electrostatic field produced by charge density ρ:

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi}\int_V \frac{(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3}\rho(\mathbf{r}')d^3r' \tag{5.12}$$

where the volume V includes the entire charge distribution, and, by our convention, ε_0 is suppressed from the denominator since ρ now embodies the relevant factor. For the benefit of those readers who are unfamiliar with Green's functions [19, chapter 12], Equation (5.12) simply represents the electric field of a point charge extended to the case of a charge distribution. Consider the field

$$\frac{dq}{4\pi}\frac{(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3}$$

due to an infinitesimal amount of charge dq located at $\mathbf{r}'$. In the case of a charge distribution, dq may be replaced by $\rho(\mathbf{r}')dV$, being the amount of charge in the

volume element dV located at $\mathbf{r}'$. However, every element of the charge distribution contributes to the field, and so we require only to integrate all such contributions throughout V in order to get the total field.

We are now simply going to appeal to the mathematics of geometric algebra and say that this equation is incomplete. Where we have ρ, we should now have $\boldsymbol{J}$, the *total* source distribution, and where we have $\mathbf{E}$, we should have the complete field $\boldsymbol{F} = \mathbf{E} + \boldsymbol{B}$, so that our conjecture is simply stated as

$$\boldsymbol{F}(\mathbf{r}) = \frac{1}{4\pi}\int_V \frac{(\mathbf{r}-\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3}\boldsymbol{J}(\mathbf{r}')d^3r' \tag{5.13}$$

Note the placing of $\boldsymbol{J}$ on the *right* of $(\mathbf{r}-\mathbf{r}')$ in the integrand. Because of the geometric product, the choice of right or left obviously affects the result. The product between a vector on the left and a scalar + vector on the right gives a result that takes the form of a scalar + vector + bivector. It is easy enough to work out that only the sign of the bivector part is affected by the order of multiplication, but in any case, the rationale for the choice we have made will soon be apparent. On splitting $\boldsymbol{J}$ back into ρ and $-\mathbf{J}$, it is clear that if Equation (5.13) is to be valid, then the terms involving $-\mathbf{J}$ must contribute to $\boldsymbol{B}$ alone. Following this through, we must then find

$$\begin{aligned}\boldsymbol{B} &= \frac{-1}{4\pi}\int_V \frac{(\mathbf{r}-\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3}\mathbf{J}(\mathbf{r}')d^3r' \\ &= \frac{-1}{4\pi}\int_V \frac{(\mathbf{r}-\mathbf{r}')\cdot\mathbf{J}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3}d^3r' + \frac{-1}{4\pi}\int_V \frac{(\mathbf{r}-\mathbf{r}')\wedge\mathbf{J}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3}d^3r'\end{aligned} \tag{5.14}$$

Now, as shown in Appendix 14.6, it turns out that the first integral on the second line of Equation (5.14) vanishes when $\partial_t \to 0$, as we have already assumed, and the source distribution is limited to some finite region of space, no matter how large. The second integral, however, will be better recognized in axial vector and cross product form, and so by dividing both sides of the equation by the unit pseudoscalar I, we find

$$\mathbf{B}(\mathbf{r}) = \frac{1}{4\pi}\int_V \frac{\mathbf{J}(\mathbf{r}')\times(\mathbf{r}-\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3}d^3r' \tag{5.15}$$

where as usual $\boldsymbol{B} = I\mathbf{B}$. This result is identifiable with that given by Stratton [35, p. 232, equation 12]. Equation (5.13) is therefore valid for the purposes of both electrostatics and magnetostatics. The consequences of this are powerful, for taken with Equation (5.11) and put in the form

$$\boldsymbol{F}_i(\mathbf{r}) = \frac{1}{4\pi}\int_V \frac{(\mathbf{r}-\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3}\rho_i(\mathbf{r}')(1-\mathbf{v}_i(\mathbf{r}'))d^3r' \tag{5.16}$$

with the total field $\boldsymbol{F}(\mathbf{r})$ being given by $\boldsymbol{F}_+(\mathbf{r}) + \boldsymbol{F}_-(\mathbf{r})$, it states that, as postulated in Section 3.2.1, the electromagnetic field as a whole is determined solely by the

properties of the charge distribution (this statement not only applies to the quasistatic limit we are dealing with here but, as will be seen later, it is also true in general). The charge distribution itself determines the electric part of the field while its motion determines the magnetic part. In fact, if $\mathbf{v}_i(\mathbf{r})$ is constant over the whole charge distribution, we may infer from Equation (5.16) that for each type of charge, we must have $\boldsymbol{B}_i(\mathbf{r}) = \mathbf{v}_i \wedge \mathbf{E}_i(\mathbf{r})$. We are left with the outer product here since any terms arising from $\mathbf{v}_i \cdot \mathbf{E}_i(\mathbf{r})$, a scalar, cannot survive in the final result. That is to say, the magnetic field of an individual charge is directly related to both its electric field and its velocity. Any magnetic field of this sort clearly has no existence separate from the electric field, and must therefore be a manifestation of the electric field that is invoked by the velocity of the sources with respect to the observer.

In the case of intrinsic magnetic sources where there is no obvious current, for example, in a bar magnet, the concept of a classical charge distribution at the atomic level has to be replaced by its quantum mechanical equivalent. We can still envisage an orbiting electron as a current source, and even conceive of electron spin as being related in some way to a spinning distribution of charge. The fact that the full picture has to be obtained through quantum mechanics is of no concern here.

It can now be explained why ρ and $\boldsymbol{J}$ were placed on the right of the integrand in Equations (5.12–5.14). Going back to Maxwell's equation taken in its static limit $\boldsymbol{\nabla}(\mathbf{E} + \boldsymbol{B}) = \rho - \mathbf{J}$, we may speculate that, as a vector, $\boldsymbol{\nabla}$ will have an inverse within geometric algebra, which would then allow us to write $\mathbf{E} + \boldsymbol{B} = \boldsymbol{\nabla}^{-1}(\rho - \mathbf{J})$. But this is exactly what we do have in Equation (5.13), where $\boldsymbol{J}$ is placed on the right simply to underline the fact that the operator $\boldsymbol{\nabla}^{-1}$ is to be given in integral form through

$$\boldsymbol{\nabla}^{-1}\boldsymbol{U} \equiv \frac{1}{4\pi}\int_V d^3r' \frac{(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3}\boldsymbol{U}(\mathbf{r}') \tag{5.17}$$

The symbol d^3r' has been moved next to the integral sign (as is sometimes the fashion) in order to make it slightly more apparent that the integrand must be completed by attaching to $\boldsymbol{U}$ before integration takes place. Beware, however, that the result is a function of $\mathbf{r}$ rather than of $\mathbf{r}'$.

One of the powerful features of geometric algebra is that it allows us to bring together different but related quantities, here $\mathbf{E}$ and $\mathbf{B}$ for one such pair and ρ and $\mathbf{J}$ for another, which we can then treat as individual entities called multivectors between which we can form novel and revealing relationships. Previously, as reference to any of the classic textbooks shows, they were treated as separate entities with $\mathbf{J}$ and $\mathbf{B}$ obeying different rules from ρ and $\mathbf{E}$. In addition, we see that the Green's function for the magnetic field arises purely from the electric field of a point charge (from Equation 5.12) with the charge density being localized at the single point $\mathbf{r}'$. The physical implication is that even in the steady state, the electric and magnetic fields cannot be regarded as separate phenomena; indeed, the magnetic field arises out of the electric field.

From a mathematical standpoint, when the Green's function approach is put in operator form, the operator concerned is simply the inverse of the vector

derivative ∇. It is therefore to be regretted that as yet our version of ∇ does not include time, since that would have allowed us to produce a fully time-dependent solution of Maxwell's equation for any given source distribution. In 4D spacetime, however, this will actually become possible.

5.5 PLANE ELECTROMAGNETIC WAVES

We now address plane electromagnetic wave solutions to Maxwell's equation in free space (Equation 5.7), but we will require some discussion beforehand as to how exponential functions are to be handled.

First, we define a bivector $I\mathbf{k}$ that is independent of $\mathbf{r}$ and from which we can generate a pseudoscalar factor $-I\mathbf{k}\cdot\mathbf{r}$ that will determine the phase of some wavefront traveling along $\mathbf{k}$. As a bivector, $I\mathbf{k}$ is actually associated with the plane of the wavefront, whereas $\mathbf{k}$ points along the axis of propagation and is therefore perpendicular to the wavefront. Now, evaluating the vector derivative of $-I\mathbf{k}\cdot\mathbf{r}$ (see Exercises 3.6.2 and 3.6.3), we find $\nabla(-I\mathbf{k}\cdot\mathbf{r}) = -I\mathbf{k}$, so that more generally

$$\begin{aligned}\nabla(-I\mathbf{k}\cdot\mathbf{r})^n &= n(-I\mathbf{k}\cdot\mathbf{r})^{n-1}\nabla(-I\mathbf{k}\cdot\mathbf{r}) \\ &= -nI\mathbf{k}(-I\mathbf{k}\cdot\mathbf{r})^{n-1}\end{aligned} \tag{5.18}$$

Applied to $e^{-I\mathbf{k}\cdot\mathbf{r}} = 1 - I\mathbf{k}\cdot\mathbf{r} + \frac{1}{2}(-I\mathbf{k}\cdot\mathbf{r})^2\cdots$, this reveals

$$\begin{aligned}\nabla e^{-I\mathbf{k}\cdot\mathbf{r}} &= \nabla\left(1 - I\mathbf{k}\cdot\mathbf{r} + \tfrac{1}{2}(-I\mathbf{k}\cdot\mathbf{r})^2\cdots\right) \\ &= -I\mathbf{k}\left(0 + 1 - I\mathbf{k}\cdot\mathbf{r} + \tfrac{1}{2}(-I\mathbf{k}\cdot\mathbf{r})^2\cdots\right) \\ &= -I\mathbf{k}e^{-I\mathbf{k}\cdot\mathbf{r}}\end{aligned} \tag{5.19}$$

This therefore works in just the same way as if we were using imaginary numbers rather than pseudoscalars. It will be immediately obvious that the same is true for $+I\mathbf{k}\cdot\mathbf{r}$ so that $e^{\pm I\mathbf{k}\cdot\mathbf{r}}$ defines a traveling plane wave frozen at some instant of time. Unfortunately, there is no simple way to introduce time and frequency as an integral part of this process. By employing multivectors such as $\boldsymbol{R} = t - \mathbf{r}$ and $\boldsymbol{K} = I(\omega + \mathbf{k})$, we could use $\langle \boldsymbol{R}\boldsymbol{K}\rangle_3$ to refer to the required pseudoscalar part but, while this would work formally, it is of little practical advantage. We therefore simply introduce the time-dependent phase term $\pm I\omega t$ separately so as to reproduce the complete time and space dependency of a traveling wave. The phase therefore takes the form $L(\omega t - \mathbf{k}\cdot\mathbf{r})$ where $L = \pm I$. This is just the usual form but with the important difference that the pseudoscalar L replaces $\pm\mathrm{j}$. It must be emphasized, however, that we are not simply making do with this as some sort of substitute for the usual complex form, $e^{L(\omega t - \mathbf{k}\cdot\mathbf{r})}$ actually gives it new meaning.

It will be clear from the analysis leading to Equation (5.19) that there is no problem in using the function $e^{\boldsymbol{\Psi}}$ where $\boldsymbol{\Psi}$ is any pseudoscalar. Any plane electromagnetic wave can therefore be written in the form $\boldsymbol{F} = \boldsymbol{F}_0 e^{L(\omega t - \mathbf{k}\cdot\mathbf{r})}$ where $\mathbf{k}$ is the wave vector, $I\mathbf{k}$ represents the directed wavefront, and $\boldsymbol{F}_0$ is some constant

multivector denoting the strength of the electromagnetic field. We may now apply this to Maxwell's equation in free space in the absence of local sources by letting $\boldsymbol{F}_0 = \mathbf{E}_0 + I\mathbf{B}_0$ and setting $\boldsymbol{J}$ to 0. Using Equation (5.19) and noting that $\partial_t e^{L\omega t} = L\omega e^{L\omega t}$ we have

$$\begin{aligned} &(\boldsymbol{\nabla} + \partial_t)\boldsymbol{F} = L(\omega - \mathbf{k})\boldsymbol{F} = 0 \\ \Rightarrow\ &(\omega - \mathbf{k})(\mathbf{E}_0 + I\mathbf{B}_0) = 0 \\ \Rightarrow\ &\underbrace{-\mathbf{k}\cdot\mathbf{E}_0}_{\text{scalar}} + \underbrace{\omega\mathbf{E}_0 - I\mathbf{k}\wedge\mathbf{B}_0}_{\text{vector}} - \underbrace{\mathbf{k}\wedge\mathbf{E}_0 - \omega I\mathbf{B}_0}_{\text{bivector}} - \underbrace{I\mathbf{k}\cdot\mathbf{B}_0}_{\text{pseudo-scalar}} = 0 \end{aligned} \tag{5.20}$$

Now, as before, when an entire expression vanishes, then the collected terms of each grade within the expression must separately vanish so that

$$\begin{aligned} &\mathbf{k}\cdot\mathbf{E}_0 = 0 \\ &\omega\mathbf{E}_0 = I\mathbf{k}\wedge\mathbf{B}_0 = \mathbf{k}\times\mathbf{B}_0 \\ &\omega\mathbf{B}_0 = -I\mathbf{k}\wedge\mathbf{E}_0 = -\mathbf{k}\times\mathbf{E}_0 \\ &\mathbf{k}\cdot\mathbf{B}_0 = 0 \end{aligned} \quad \text{(a)} \qquad \begin{aligned} &\mathbf{k}\cdot\mathbf{E}_0 = 0 \\ &\omega\mathbf{E}_0 = \mathbf{k}\cdot\boldsymbol{B}_0 \\ &\omega\boldsymbol{B}_0 = \mathbf{k}\wedge\mathbf{E}_0 \\ &\mathbf{k}\wedge\boldsymbol{B}_0 = 0 \end{aligned} \quad \text{(b)} \tag{5.21}$$

Here (a) and (b) are alternative versions with the magnetic field in axial vector and bivector form respectively. All the usual conditions for plane waves in free space therefore come tumbling out from either equation—but we do need to remember to include the factor c if we wish to restore the modified variables $\mathbf{B}_0$ and $\boldsymbol{B}_0$ to their standard form.

While for the purposes of the preceding discussion we simply assumed a plane wave solution, the necessary wave equations for $\mathbf{E}$ and $\mathbf{B}$ may be obtained from Maxwell's equations in the usual way. We may get even more insight, however, by turning to $(\boldsymbol{\nabla} - \partial_t)(\boldsymbol{\nabla} + \partial_t)\boldsymbol{F}$ since we know that $(\boldsymbol{\nabla} - \partial_t)(\boldsymbol{\nabla} + \partial_t) = \boldsymbol{\nabla}^2 - \partial_t^2$ must be a scalar operator and therefore should produce a simple result. Indeed,

$$\begin{aligned} &(\boldsymbol{\nabla} + \partial_t)\boldsymbol{F} = \boldsymbol{J} \\ \Rightarrow\ &\left(\boldsymbol{\nabla}^2 - \partial_t^2\right)\boldsymbol{F} = (\boldsymbol{\nabla} - \partial_t)\boldsymbol{J} \end{aligned} \tag{5.22}$$

Here we have produced the scalar wave equation for $\boldsymbol{F}$, that is to say, for $\mathbf{E}$ and $\boldsymbol{B}$ jointly. In any region free of sources, or at least where $(\boldsymbol{\nabla} - \partial_t)\boldsymbol{J}$ vanishes, this gives

$$\begin{aligned} \left(-k^2 + \omega^2\right)\boldsymbol{F} = 0 \quad &\Leftrightarrow \quad \omega^2 = [c]^2\mathbf{k}^2 \\ &\Leftrightarrow \quad \omega\ = [c]k \end{aligned} \tag{5.23}$$

where $k = |\mathbf{k}|$. Apart from the choice of sign, this result is otherwise equivalent to the standard free space dispersion relation $\omega = [c]k$. Note that it may also be deduced from Equation (5.21), but here we obtain it directly. Again the power of the approach is demonstrated in its compactness. In spite of this achievement, there has been no need to introduce the separate concept of complex numbers as the

pseudoscalars automatically provide the required facility. While one might be tempted to say that they are the same thing by another name, the pseudoscalars have a stronger physical interpretation. First, from our original definition, it must be recognized that the electric and magnetic fields are always the vector and bivector parts of $\boldsymbol{F}$, respectively. In a manner analogous to complex exponentials, we may write $e^{\Psi} = \alpha + I\beta$ where $\alpha = \cos(\omega t - \mathbf{k}\cdot\mathbf{r})$ and $\beta = \pm\sin(\omega t - \mathbf{k}\cdot\mathbf{r})$ with β being taken to have the same sign as L. $\boldsymbol{F}$ may be therefore decomposed into the appropriate electric and magnetic parts as follows:

$$\begin{aligned}\boldsymbol{F} &= (\mathbf{E}_0 + I\mathbf{B}_0)(\alpha + I\beta) \\ &= \underbrace{(\alpha\mathbf{E}_0 - \beta\mathbf{B}_0)}_{\mathbf{E}} + I\underbrace{(\alpha\mathbf{B}_0 + \beta\mathbf{E}_0)}_{\mathbf{B}}\end{aligned} \tag{5.24}$$

Any seeming analogy with the traditional use of complex arithmetic here must be treated with care. Remember that I obeys the rules of geometric multiplication and causes a change of grade in the process. It can be seen that it would have been quite wrong to have assumed that a familiar form like $\mathbf{E} = \mathbf{E}_0(\alpha + I\beta)$ would be applicable since to do so would give $\mathbf{E}$ both vector and bivector parts, contrary to the definition of $\mathbf{E}$ as a pure vector. Instead, Equation (5.24) deals with $\mathbf{E}$ and $\mathbf{B}$ together so that the analogy applies to $\boldsymbol{F}$ on its own. Going back to the relations given in Equation (5.21b), however, we can write $\mathbf{B}_0 = -(I/\omega)\mathbf{k}\wedge\mathbf{E}_0$ and given that $\mathbf{k}\cdot\mathbf{E}_0 = 0$, this is the same thing as $\mathbf{B}_0 = -(I/\omega)\mathbf{k}\mathbf{E}_0$. Putting this back into Equation (5.24) yields

$$\begin{aligned}\boldsymbol{F} &= (\mathbf{E}_0 + I\mathbf{B}_0)(\alpha + I\beta) \\ &= (\mathbf{E}_0 + (\mathbf{k}/\omega)\mathbf{E}_0)(\alpha + I\beta) \\ &= \left(1 + \hat{\mathbf{k}}\right)\mathbf{E}_0(\alpha + I\beta)\end{aligned} \tag{5.25}$$

Here the factor $1 + \hat{\mathbf{k}}$ allows us to eliminate $\mathbf{B}$, which, after all, depends directly on $\mathbf{E}$ and $\hat{\mathbf{k}}$. Now, just for the time being, if we imagine vector and bivector to be the equivalent of real and imaginary, the factor $\alpha + I\beta$ rotates $\left(1 + \hat{\mathbf{k}}\right)\mathbf{E}_0$ in space in the same way that $\alpha + \mathrm{j}\beta$ would rotate a number in the complex plane. On this basis, therefore, we do have an analogy with complex arithmetic, but one which works in a plane in space determined by $I\mathbf{k}$.

In order that we may pursue this further, let us return to Equation (5.24). Using once more the relationship $\mathbf{B}_0 = -(I/\omega)\mathbf{k}\mathbf{E}_0$ and the reciprocal form $\mathbf{E}_0 = (I/\omega)\mathbf{k}\mathbf{B}_0$, we find

$$\begin{aligned}\boldsymbol{F} &= (\mathbf{E}_0 + I\mathbf{B}_0)(\alpha + I\beta) \\ &= (\alpha + I\beta)(\mathbf{E}_0 + I\mathbf{B}_0) \\ &= (\alpha\mathbf{E}_0 - \beta\mathbf{B}_0) + I(\alpha\mathbf{B}_0 + \beta\mathbf{E}_0) \\ &= (\alpha\mathbf{E}_0 + \beta(I/\omega)\mathbf{k}\mathbf{E}_0) + I(\alpha\mathbf{B}_0 + \beta(I/\omega)\mathbf{k}\mathbf{B}_0) \\ &= \left(\alpha + I\beta\hat{\mathbf{k}}\right)\mathbf{E}_0 + I\left(\alpha + I\beta\hat{\mathbf{k}}\right)\mathbf{B}_0 \\ &= \left(\alpha + \beta\hat{\boldsymbol{K}}\right)(\mathbf{E}_0 + I\mathbf{B}_0)\end{aligned} \tag{5.26}$$

The representation $(\alpha + I\beta)(\mathbf{E}_0 + I\mathbf{B}_0)$ in which $(\alpha + I\beta)$ rotates vector into bivector is now transformed into $\left(\alpha + \beta\hat{\boldsymbol{K}}\right)(\mathbf{E}_0 + I\mathbf{B}_0)$ where $\hat{\boldsymbol{K}} = I\hat{\mathbf{k}}$ is the unit bivector perpendicular to the direction of propagation and $\left(\alpha + \beta\hat{\boldsymbol{K}}\right)$ rotates any vector $\mathbf{u}$ lying in the $\hat{\boldsymbol{K}}$ plane by the angle $\pm(\omega t - \mathbf{k}\cdot\mathbf{r})$. The sign here is given by the sign of L, which in turn determines the sign of β (see p. 64). This therefore represents an actual geometric rotation as opposed to just an analogy. The role that this scalar plus bivector form plays in rotations was touched on in Exercise 4.8.4 and is discussed in more detail in Section 9.3.

This discussion has a significant practical implication because it means that the plane wave represented by Equation (5.26) is inherently circularly polarized. Taking $\mathbf{E}_0$ as lying in the $\hat{\boldsymbol{K}}$ plane so as to satisfy $\mathbf{k}\cdot\mathbf{E}_0 = 0$, then at any fixed point $\mathbf{r}$, the vectors $\mathbf{E}$ and $\mathbf{B}$ simply rotate in quadrature about the $\mathbf{k}$ axis with frequency ω and the sense of rotation given by the sign of L. Figure 5.1 provides an illustration of this for one sense of polarization. The magnetic field is shown in its native bivector form with $\boldsymbol{B}$ equal to $\hat{\mathbf{k}}\mathbf{E}$ as per Equation (5.21) (b). With the wave traveling along $\mathbf{x}$, that is to say $\hat{\mathbf{k}} = \mathbf{x}$, we would then have $\hat{\boldsymbol{K}} = I\mathbf{x} = \mathbf{yz}$. The only degrees of freedom left are then the magnitude and initial polarization of $\mathbf{E}_0$, and the sense of circular polarization, clockwise or anticlockwise according to the sign of L. If at $t = 0$ the direction of $\mathbf{E}$ happens to be along $\mathbf{z}$, then $\left(\alpha + \beta\hat{\boldsymbol{K}}\right)\mathbf{E}_0 = E_0(\alpha + \beta\mathbf{yz})\mathbf{z} = E_0(\alpha\mathbf{z} + \beta\mathbf{y})$. It is readily seen that $\mathbf{E} = E_0(\alpha\mathbf{z} + \beta\mathbf{y})$ is a vector of constant magnitude E_0 lying in the $\mathbf{yz}$ plane and rotated by an angle $\arcsin\beta = \pm\omega t$ with respect to $\mathbf{z}$. Apart from an initial starting position along $\mathbf{y}$, the same behavior then applies to $\mathbf{B}$ so that it is always 90° out of phase with $\mathbf{E}$.

These circularly polarized solutions come out naturally, whereas following the traditional method of solution, they have to be constructed from linearly polarized

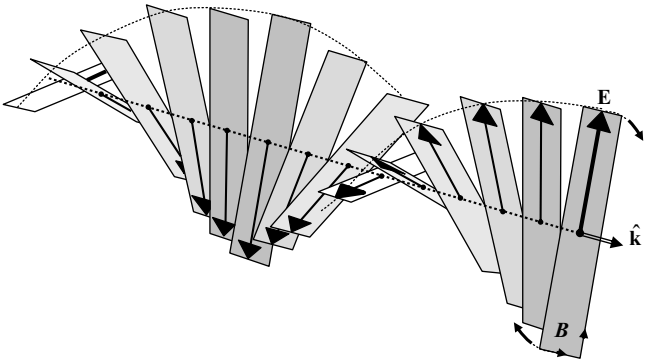

Figure 5.1 Circularly polarized electromagnetic plane wave. The figure shows a circularly polarized plane electromagnetic wave that has been sampled at phase intervals of 22.5° along the propagation direction. The shaded rectangular strips represent the magnetic field bivector $\boldsymbol{B}$. The sense of each of these bivectors is the same as that of the leading section. The arrow lying in the plane of each bivector shows the electric field vector $\mathbf{E}$, while the forward arrow $\hat{\mathbf{k}}$ is the unit vector in the propagation direction. In agreement with Equation (5.21) (b), $\boldsymbol{B} = \hat{\mathbf{k}}\mathbf{E}$. Because it rotates anticlockwise as it travels, the wave is right-hand polarized (the opposite of the right-hand screw rule).

solutions. It is possible to construct a field quantity that is superficially similar to $\boldsymbol{F}$ using a complex linear combination of $\mathbf{E}$ and $\mathbf{B}$. Stratton, for example, defined a complex electromagnetic field vector given by $\mathbf{B}+\mathrm{j}\sqrt{\mu\varepsilon}\mathbf{E}$, but in doing so he stated that, while this procedure offers compactness, it "has no apparent physical significance" [35, section 1.12, p. 32]. Here we had no need to construct anything, it came about naturally, and the physical significance has been obvious.

5.6 CHARGE CONSERVATION

As suggested earlier, it may be of interest to try applying geometric algebra to the continuity equation (Equation 3.6) that expresses the conservation of charge. The wave equation (Equation 5.22) provides us with a starting point.

Bearing in mind that $\boldsymbol{\nabla}^2-\partial_t^2$ is a scalar, $\left(\nabla^2-\partial_t^2\right)\boldsymbol{F}=(\boldsymbol{\nabla}-\partial_t)\boldsymbol{J}$ may be expanded to reveal

$$\begin{aligned}\left(\nabla^2-\partial_t^2\right)(\mathbf{E}+\boldsymbol{B}) &= (\boldsymbol{\nabla}-\partial_t)(\rho-\mathbf{J})\\ \Leftrightarrow \underbrace{\left(\nabla^2-\partial_t^2\right)\mathbf{E}}_{\text{vector}}+\underbrace{\left(\nabla^2-\partial_t^2\right)\boldsymbol{B}}_{\text{bivector}} &= \underbrace{-(\partial_t\rho+\boldsymbol{\nabla}\cdot\mathbf{J})}_{\text{scalar}}+\underbrace{\partial_t\mathbf{J}+\boldsymbol{\nabla}\rho}_{\text{vector}}-\underbrace{\boldsymbol{\nabla}\wedge\mathbf{J}}_{\text{bivector}}\end{aligned} \tag{5.27}$$

Upon applying the usual procedure of matching the terms of like grade on both sides we find three separate equations:

$$\begin{aligned}&\partial_t\rho+\boldsymbol{\nabla}\cdot\mathbf{J}=0\\ &\left(\nabla^2-\partial_t^2\right)\mathbf{E}=\partial_t\mathbf{J}+\boldsymbol{\nabla}\rho\\ &\left(\nabla^2-\partial_t^2\right)\boldsymbol{B}=-\boldsymbol{\nabla}\wedge\mathbf{J}\end{aligned} \tag{5.28}$$

The scalar equation expresses the conservation of charge, while the vector and bivector equations simply represent the wave equations for both $\boldsymbol{E}$ and $\boldsymbol{B}$ as usually derived from Maxwell's equations (the inhomogeneous Helmholtz equations). The bivector equation is perhaps more familiar in dual form as $\left(\nabla^2-\partial_t^2\right)\mathbf{B}=\boldsymbol{\nabla}\times\mathbf{J}$. Recalling the now recurrent theme, the three separate results in Equation (5.28) are all encoded in the single multivector equation $\left(\nabla^2-\partial_t^2\right)\boldsymbol{F}=(\boldsymbol{\nabla}-\partial_t)\boldsymbol{J}$. Given that we now have $\partial_t\rho+\boldsymbol{\nabla}\cdot\mathbf{J}=0$, it can be seen that the wave equation for $\boldsymbol{F}$ reduces to $\left(\nabla^2-\partial_t^2\right)\boldsymbol{F}=(\partial_t\mathbf{J}+\boldsymbol{\nabla}\rho)-\boldsymbol{\nabla}\wedge\mathbf{J}$, which readily splits into separate vector and bivector equations for $\boldsymbol{E}$ and $\boldsymbol{B}$, respectively.

As this example clearly illustrates, geometric algebra provides a means of systematizing related equations into a compact mathematical form that may be quickly unraveled into separate equations, often with the bonus that an extra equation, that might not have been considered to be part of the original set, may be revealed. The existence of a "bonus equation" is interesting in its own right and may well be of direct relevance to the underlying physics since it raises the question of exactly how it happens to be implied by the other equations.

The above method of expanding an equation into terms of separate grades and then examining each of the individual equations that result, provides a routine but very useful technique.

5.7 MULTIVECTOR POTENTIAL

We are accustomed to two types of electromagnetic potential: one scalar and the other vector. It now seems a certainty that in a geometric algebra over Newtonian space, these may be combined into a single multivector potential $\boldsymbol{A}$. Recalling that the scalar and vector potentials Φ and $\mathbf{A}$ give rise to scalar wave equations with ρ and $\mathbf{J}$ respectively as sources, we should now find a single wave equation relating $\boldsymbol{A}$ to $\boldsymbol{J}$. In fact, we should have

$$\begin{aligned} &\left(\nabla^2 - \partial_t^2\right)\boldsymbol{A} = \boldsymbol{J} \\ \Leftrightarrow\ &(\boldsymbol{\nabla} + \partial_t)(\boldsymbol{\nabla} - \partial_t)\boldsymbol{A} = (\boldsymbol{\nabla} + \partial_t)\boldsymbol{F} \end{aligned} \tag{5.29}$$

In asserting this, we are anticipating that there is some suitable linear combination of Φ and $\mathbf{A}$ that will form the requisite multivector potential $\boldsymbol{A}$ and that $\boldsymbol{F}$ must be derivable from it by the process of differentiation. It is clear that $\boldsymbol{\nabla} + \partial_t$ factors out of Equation (5.29) leaving us with

$$\boldsymbol{F} = (\boldsymbol{\nabla} - \partial_t)\boldsymbol{A} + \boldsymbol{F}' \tag{5.30}$$

where $\boldsymbol{F}'$ is any solution of the homogeneous (source-free) Maxwell's equation, $(\boldsymbol{\nabla} + \partial_t)\boldsymbol{F}' = 0$. We can therefore view $\boldsymbol{F}'$ as being an externally applied vector + bivector field satisfying some given boundary condition. Taking initially the simple condition $\boldsymbol{F}' = 0$ and recalling that $\left(\nabla^2 - \partial_t^2\right)$ is a scalar operator, $\boldsymbol{A}$ must have the same form as $\boldsymbol{J}$, that is to say, a scalar plus a vector, a particular form of multivector that is known as a paravector. By writing $\boldsymbol{A}$ in the form $-\Phi + [c]\mathbf{A}$ we therefore find

$$\begin{aligned} \boldsymbol{F} &= (\boldsymbol{\nabla} - \partial_t)\boldsymbol{A} \\ &= (\boldsymbol{\nabla} - \partial_t)(-\Phi + [c]\mathbf{A}) \\ \Leftrightarrow\ \underbrace{\mathbf{E}}_{\text{vector}} + \underbrace{[c]\boldsymbol{B}}_{\text{bivector}} &= \underbrace{\partial_t\Phi + [c]\boldsymbol{\nabla}\cdot\mathbf{A}}_{\text{scalar}} + \underbrace{\left(-\boldsymbol{\nabla}\Phi - [c]\partial_t\mathbf{A}\right)}_{\text{vector}} + \underbrace{[c]\boldsymbol{\nabla}\wedge\mathbf{A}}_{\text{bivector}} \end{aligned} \tag{5.31}$$

Using the now familiar rules where $[c]$ implies that the constant c is to be suppressed, and that objects of the same grade must match on each side of the equation, we find the instantly recognizable results

$$\begin{aligned} 0 &= \partial_t\Phi + \boldsymbol{\nabla}\cdot\mathbf{A} \\ \mathbf{E} &= -\boldsymbol{\nabla}\Phi - \partial_t\mathbf{A} \\ \boldsymbol{B} = \boldsymbol{\nabla}\wedge\mathbf{A} \quad &\Leftrightarrow \quad \mathbf{B} = \boldsymbol{\nabla}\times\mathbf{A} \end{aligned} \tag{5.32}$$

Here we recover the forms of the two equations that separately relate Φ to $\mathbf{E}$ and $\mathbf{A}$ to $\mathbf{B}$. The first equation, however, is the "bonus equation." Referred to as the Lorenz condition[2] [36, 37, sections 6.2–6.3], it arises because (1) $\boldsymbol{F}$ can have no scalar part and (2) it is required to support the initial conjecture that the multivector potential should satisfy the wave equation as given in Equation (5.29). This condition is referred to as a gauge, of which a different example is the Coulomb gauge, $\boldsymbol{\nabla} \cdot \mathbf{A} = 0$, which allows the scalar potential to be solved through Poisson's equation, $\nabla^2 \Phi = -\rho$, after which the vector potential may be solved separately through $\left(\nabla^2 - \partial_t^2\right)\mathbf{A} = -\mathbf{J} + \partial_t(\boldsymbol{\nabla}\Phi)$. However, since they do not satisfy the Lorenz condition, these solutions cannot be brought together as a multivector potential that obeys Equation (5.29). Another interesting comparison between the Lorenz condition and the Coulomb gauge is that the former eliminates the displacement current $\partial_t \mathbf{E}$ from the wave equation, whereas the latter retains it in the form $-\partial_t(\boldsymbol{\nabla}\Phi)$. Maxwell adopted the Coulomb gauge, a subject on which there has been much discussion. For further information on these issues, including the connection with the subject of retardation, see Reference 38 and the references cited therein.

Finally, returning to the assumption that $\boldsymbol{F}' = 0$, the presence of an external field cannot alter the scalar part of Equation (5.31) and so it has no effect on the Lorenz condition. Furthermore, adding an external field $\boldsymbol{F}'$ is equivalent to adding a potential $\boldsymbol{A}'$ that satisfies $\boldsymbol{F}' = (\boldsymbol{\nabla} - \partial_t)\mathbf{A}'$. Since $(\boldsymbol{\nabla} + \partial_t)\boldsymbol{F}' = 0$, this then implies $\left(\nabla^2 - \partial_t^2\right)\boldsymbol{A}' = 0$ and so it is only necessary to choose $\boldsymbol{A}'$ so that the given boundary conditions are satisfied.

Not only is the concept of a multivector potential very simple, it fits very nicely within the framework of (3+1)D geometric algebra. We have in Equations (5.30) and (5.32) another extremely compact statement of the key results.

Following up on the issue of complex field vectors mentioned by Stratton, complex potentials, referred to as $\mathbf{L}$ and Φ, respectively, also exist [35, p. 33], but since traditional complex analysis provides no way of combining vectors and scalars, these do not bear comparison with the multivector potential $\boldsymbol{A}$. Moreover, while our field vector $\boldsymbol{F} = \mathbf{E} + I\mathbf{B}$ may look like a complex vector, no similar analogy attaches to $\boldsymbol{A} = -\Phi + \mathbf{A}$. Taken as a whole therefore, even if complex vector fields and potentials share some points of similarity with their multivector counterparts, they are neither consistent with them nor do they have the same physical interpretation. Finally, they provide no equation that is comparable in simplicity with $\boldsymbol{F} = (\boldsymbol{\nabla} - \partial_t)\boldsymbol{A}$.

5.7.1 The Potential of a Moving Charge

In Section 5.4, the discussion of the electromagnetic field of point charges was limited to the quasistatic case in which time derivatives can be ignored. We now wish to discuss the classical approach to the problem of finding the potential of a moving point charge when the time derivative does have to be taken into account. Since the (3+1)D approach lacks the methodology of spacetime, the solution to this

[2] This is often wrongly attributed to H.A. Lorentz rather than L.V. Lorenz.

dynamical problem must be based, at least implicitly, on the solution of a wave equation. On the other hand, the spacetime approach neatly avoids the need for wave equations. The two approaches, however, must be compatible, and so our objective will be to compare them. The discussion here will commence with the multivector potential of a point charge undergoing arbitrary motion.

The starting point is the classical (3+1)D multivector form of the wave equation for the electromagnetic potential, $\left(\boldsymbol{\nabla}^2 - \partial_t^2\right)\boldsymbol{A} = \boldsymbol{J}$, which assumes the Lorenz gauge as discussed in Section 5.7. With some minor adaptation of Equation (5.11) to suit the situation, the source density for a single point charge q at position $\mathbf{r}_q$ and velocity $\mathbf{v}_q$ is given by $\boldsymbol{J}(t+\mathbf{r}) = q\delta\big(\mathbf{r}-\mathbf{r}_q(t)\big)\big(1-\mathbf{v}_q(t)\big)$ where $\delta(\mathbf{r})$ is the 3D Dirac delta function. This is a scalar function that vanishes everywhere except at $\mathbf{r}=0$ but nevertheless yields $\int\delta(\mathbf{r})d^3r = 1$ provided the volume of integration includes the origin [19, chapter 6, pp. 221–259 and section 14.7, pp. 621]. The solution for $\boldsymbol{A}$ is then found via a scalar Green's function taken from the scalar potential of a unit point charge at rest:

$$g\big((t-t')+(\mathbf{r}-\mathbf{r}')\big) = \frac{-1}{4\pi[\varepsilon_0]|\mathbf{r}-\mathbf{r}'|}\delta\left((t-t')-\frac{|\mathbf{r}-\mathbf{r}'|}{[c]}\right) \tag{5.33}$$

This Green's function may be obtained from the scalar wave equation with $q=1$ for the source, that is to say

$$\left(\boldsymbol{\nabla}^2 - \partial_t^2\right)g(t-t'+\mathbf{r}-\mathbf{r}') = \delta(\mathbf{r}-\mathbf{r}')\delta(t-t') \tag{5.34}$$

It helps to simplify matters if we express $g(t-t'+\mathbf{r}-\mathbf{r}')$ as $g(T+\mathbf{R})$ where $T = t-t'$ and $\mathbf{R} = \mathbf{r}-\mathbf{r}'$. Noting that $\boldsymbol{\nabla}_{\mathbf{R}}^2 - \partial_T^2 = \boldsymbol{\nabla}^2 - \partial_t^2$, this gives

$$\left(\boldsymbol{\nabla}_{\mathbf{R}}^2 - \partial_T^2\right)g(T+\mathbf{R}) = \delta(\mathbf{R})\delta(T) \tag{5.35}$$

which may be solved by the standard technique of expressing the functions on each side of the wave equation in terms of their Fourier transforms [19, section 14.7, pp. 619–624]. For example,

$$\begin{aligned} g(T+\mathbf{R}) &= \frac{1}{(2\pi)^4}\iint_{\mathbf{k},\omega}\tilde{g}(\omega+\mathbf{k})e^{-I(\mathbf{k}\cdot\mathbf{R}-\omega T)}d^3k\,d\omega \\ \delta(T)\delta(\mathbf{R}) &= \delta(T+\mathbf{R}) \\ &= \frac{1}{(2\pi)^4}\iint_{\mathbf{k},\omega}e^{-I(\mathbf{k}\cdot\mathbf{R}-\omega T)}d^3k\,d\omega \end{aligned} \tag{5.36}$$

where $\tilde{g}(\mathbf{k},\omega)$ denotes the Fourier transform of $g(T+\mathbf{R})$, which, incidentally, effectively expresses it as a linear superposition of plane waves similar to those we discussed in Section 5.5. This is effectively solving the wave equation that results from a point stimulus at $T+\mathbf{R}=0$ in terms of a spectrum of plane waves. In these terms, Equation (5.35) becomes

$$\begin{aligned}&\left(\nabla_{\mathbf{R}}^2-\partial_T^2\right)g(T+\mathbf{R})=\delta(T)\delta(\mathbf{R})\\ \Leftrightarrow&\left(\omega^2-\mathbf{k}^2\right)\tilde{g}(\omega+\mathbf{k})=1\\ \Leftrightarrow&\ \tilde{g}(\omega+\mathbf{k})=\frac{1}{\omega^2-\mathbf{k}^2}\end{aligned}\tag{5.37}$$

Substituting this result for $\tilde{g}(\mathbf{k},\omega)$ into Equation (5.36) reveals

$$g(T+\mathbf{R})=\begin{cases}-\dfrac{1}{4\pi}\dfrac{\delta(T+\mathbf{R})}{|\mathbf{R}|} & 0\le T\\ 0 & T<0\end{cases}\tag{5.38}$$

from which $g\left((t-t')+(\mathbf{r}-\mathbf{r}')\right)$ directly follows. Since the inverse Fourier transform of $\left(\omega^2-\mathbf{k}^2\right)^{-1}$ is a standard result, its evaluation is of little importance. The key point, however, is that we have obtained the required Green's function as a solution of the scalar wave equation. But, clearly this Green's function is the link with the spacetime approach that we will come to in Sections 11.7 and 11.8, for the delta function constrains the observation time t and the source time t' to be separated by the light travel time between $\mathbf{r}'$ and $\mathbf{r}$. Despite the fact that the Green's function has been obtained from a scalar wave equation, it also works with a vector wave equation such as $\left(\nabla^2-\partial_t^2\right)\boldsymbol{A}=\boldsymbol{J}$, the solution of which is obtained by integrating over the contributions from the entire source distribution over all space and time

$$\begin{aligned}\boldsymbol{A}(t+\mathbf{r})&=\frac{-1}{4\pi}\cdot\iint_{t'V}\frac{\boldsymbol{J}(t'+\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\delta\left((t-t')-|\mathbf{r}-\mathbf{r}'|\right)d^3r'dt'\\ &=\frac{-q}{4\pi}\cdot\iint_{t'V}\frac{\delta\left(\mathbf{r}'-\mathbf{r}_q(t')\right)\left(1-\mathbf{v}_q(t')\right)}{|\mathbf{r}-\mathbf{r}'(t')|}\delta\left((t-t')-|\mathbf{r}-\mathbf{r}'|\right)d^3r'dt'\\ &=\frac{q}{4\pi}\cdot\int_{t'}\frac{\mathbf{v}_q(t')-1}{|\mathbf{r}-\mathbf{r}_q(t')|}\delta\left((t-t')-|\mathbf{r}-\mathbf{r}_q(t')|\right)dt'\end{aligned}\tag{5.39}$$

Here the integral over space was achieved by applying one of the basic rules for integrals involving delta functions, namely rule (i) below:

$$\begin{aligned}\int f(t')\delta(t-t')dt'&=f(t')\big|_{t-t'=0}\\ &=f(t)\end{aligned}\qquad\text{(i)}$$

$$\begin{aligned}\int f(t')\delta(t-h(t'))dt'&=\int f(t')\delta(t-h(t'))\frac{dt'}{dh}dh\\ &=\frac{f(t')}{\partial_{t'}h}\bigg|_{t-h(t')=0}\end{aligned}\qquad\text{(ii)}\tag{5.40}$$

It was therefore only necessary to take the integrand evaluated at $\mathbf{r}' = \mathbf{r}_q(t')$. The integration over the remaining delta function, however, requires the use of rule (ii) because its argument, $(t-t') - |\mathbf{r} - \mathbf{r}_q(t')|$, is a function of t'. The constraint that this argument must vanish imposes the condition $t' = t - |\mathbf{r} - \mathbf{r}_q(t')|$ and $h(t')$ takes the form $t' + |\mathbf{r} - \mathbf{r}_q(t')|$. The chain rule in the form of $\partial_{t'} f(\mathbf{r}) = \boldsymbol{\nabla} f \cdot \partial_{t'} \mathbf{r}$ allows us to evaluate $\partial_{t'} h$ as follows:

$$\begin{aligned}
\partial_{t'} h &= 1 + \partial_{t'} |\mathbf{r} - \mathbf{r}_q(t')| \\
&= 1 + \mathring{\boldsymbol{\nabla}} \left| \mathbf{r} - \mathring{\mathbf{r}}_q(t') \right| \cdot \partial_{t'} (\mathbf{r} - \mathbf{r}_q(t')) \\
&= 1 - \frac{\mathbf{r} - \mathbf{r}_q(t')}{|\mathbf{r} - \mathbf{r}_q(t')|} \cdot \mathbf{v}_q(t') \\
&= 1 - \frac{\mathbf{R}(t')}{R(t')} \cdot \mathbf{v}_q(t')
\end{aligned} \tag{5.41}$$

where $\mathbf{R}(t') = \mathbf{r} - \mathbf{r}_q(t')$ and $R(t') = |\mathbf{R}(t')|$. The potential is thereby found to be

$$\begin{aligned}
\boldsymbol{A}(t + \mathbf{r}) &= \frac{q}{4\pi} \cdot \left[\frac{\mathbf{v}_q(t') - 1}{R(t')} \cdot \frac{1}{1 - \dfrac{\mathbf{v}(t') \cdot \mathbf{R}(t')}{R(t')}} \right]_{t' = t - |\mathbf{r} - \mathbf{r}_q(t')|} \\
&= \frac{q}{4\pi} \cdot \left[\frac{\mathbf{v}_q(t') - 1}{R(t') - \mathbf{v}(t') \cdot \mathbf{R}(t')} \right]_{t' = t - |\mathbf{r} - \mathbf{r}_q(t')|}
\end{aligned} \tag{5.42}$$

It is clear that the relationship between t' and t is such that $t - t'$ is exactly the time that it takes for any electromagnetic effect originating at $\mathbf{r}_q$, the location of the charge, to propagate to $\mathbf{r}$, the location where it is observed. The convention here is to write $t' = t^*$, where t^* is the so-called retarded time, that is to say, the time of observation less the propagation delay. In fact, it is clearly an earlier time rather than a later one. This leads to two other ways of expressing Equation (5.42). The first is to dispense with the square brackets and mark each variable that depends on t or t' with an asterisk. The second is simply to substitute t for t' and use the abbreviation *ret* in place of the constraint $t - t' - |\mathbf{r} - \mathbf{r}_q(t')| = 0$. This is understood as meaning that the value of t requires to be replaced with the value that t^* takes at t. The retarded potential of a point charge may therefore be written in either of the following equivalent forms:

$$\boldsymbol{A}(t + \mathbf{r}) = \frac{q}{4\pi} \cdot \frac{\mathbf{v}_q^* - 1}{R^* - \mathbf{v}_q^* \cdot \mathbf{R}^*} \quad \text{or} \quad \boldsymbol{A}(t + \mathbf{r}) = \frac{q}{4\pi} \cdot \left[\frac{\mathbf{v}_q - 1}{R - \mathbf{v}_q \cdot \mathbf{R}} \right]_{ret} \tag{5.43}$$

Although it is given here in compact multivector form, this result is referred to as the Liénard–Wiechert potentials [35, section 8.17, p. 475; 37, section 14.1, pp. 464–465; 39, 40]. Although the general idea of retardation was due to Lorenz

[36], the Liénard–Wiechert potential is probably the best known example of its use. In comparison with the vector and scalar potentials, it is worthwhile to note that $\mathbf{A}(t+\mathbf{r})$ and $\Phi(t+\mathbf{r})$ are far from independent since the numerator of Equation (5.43) tells us that the former must be simply $\mathbf{v}_q^*$ times the latter, that is to say, $\mathbf{A}(t+\mathbf{r}) = \Phi(t+\mathbf{r})\mathbf{v}_q^*$. This echoes the connection between the quasistatic electric and magnetic fields that may be inferred from Equation (5.16).

If the source moves only slowly compared with the speed of light or, more specifically, if the magnitude of its velocity in the direction of the observer is much less than the speed of light, then the amount of retardation is approximated by the instantaneous value of $[c^{-1}]|\mathbf{r}-\mathbf{r}_q|$ so that $t_q = t - \frac{1}{c}|\mathbf{r}-\mathbf{r}_q|$. Although an exact solution for the retarded time may be worked out for the case of uniform motion, a closed-form result may not be achievable for arbitrary charge trajectories, and so the use of retarded variables provides a convenient way of getting round the problem. As will be seen in Chapters 11 and 12, spacetime provides a much more suitable framework of dealing with retardation through the concept of null vectors.

The solution of Equation (5.43) for the case of uniform motion may be evaluated from Figure 5.2 by expressing the term $R^* - \mathbf{v}_q^* \cdot \mathbf{R}^*$ in the denominator as $R^* - vR^*\cos\theta^*$, where θ^* is the angle that $\mathbf{R}^*$ makes with $\mathbf{v}$ and the sign of v is taken as positive when the charge is approaching the observer. We may refer to this term as the effective distance. The figure reveals the essential relationships between

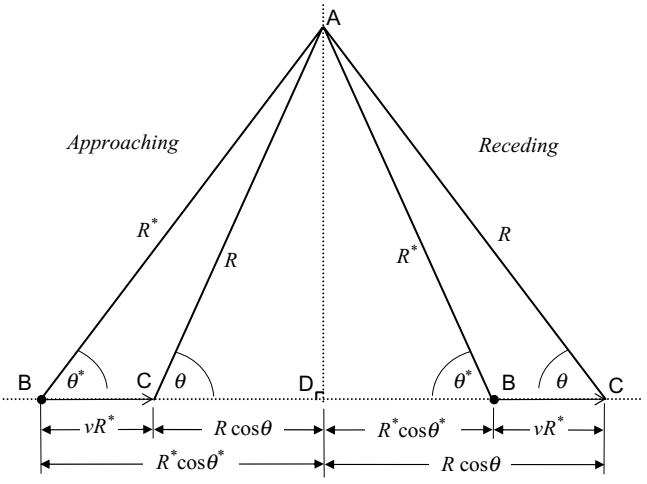

Figure 5.2 Retarded and unretarded measurements for a uniformly moving point charge. The left half of the figure applies to a charge q approaching the observer, whereas the right half applies when the charge is receding. The charge is instantaneously at **B** at time t^* and moves with constant velocity v along **BC**. An observer at **A** detects the field of the charge arising at this instant after a delay equal to $R^*/[c]$, by which time the distance to the observer is R and the time is t and the charge has traveled a further distance vR^* to reach **C**. From the information available, the retarded values R^* and $\cos\theta^*$ may be determined based on the values R and $\cos\theta$ *at the time of observation*, t. The left half of the figure corresponds to Jackson's figure 14.2 [37, section 14.1, p. 468].

the retarded and unretarded distances and angles given that the charge moves from **B** to **C** during, $[c^{-1}]R^*$, the time that it takes for the source information to travel the distance R^* from **B** to the observer at **A**. Basic trigonometry applied to $\mathbf{BC}+\mathbf{CA}=\mathbf{BA}$ and $\mathbf{BD}=\mathbf{BC}+\mathbf{CD}$ produces a pair of equations:

$$\begin{cases}\left(1-v^2\right)R^{*2}-2\left(vR\cos\theta\right)R^*-R^2=0\\ R^*\cos\theta^*=vR^*+R\cos\theta\end{cases}$$

These allow the retarded variables in $R^*-vR^*\cos\theta^*$ to be replaced with unretarded expressions to yield $R^*-\mathbf{v}_q^*\cdot\mathbf{R}^*=R\left(1-v^2\sin^2\theta\right)^{1/2}$ irrespective of whether v is positive (approaching the observer) or negative (receding from the observer). In addition, the effective distance $R\left(1-v^2\sin^2\theta\right)^{1/2}$ is always smaller or equal to the instantaneous distance R at the time of observation. For motion transverse to the observer, θ is equal to 90° (at $t=0$) so that the effective distance is $\gamma^{-1}R$ where $\gamma=\left(1-v^2\right)^{-1/2}$, but when the motion of the charge is directed in a straight line either toward or away from the observer, θ is 0 and the effective distance becomes R, just as if the entire principle of retardation had been ignored. While these facts may seem counterintuitive and contrary to the principles of retardation, their origin is more to do with the fact that the denominator in Equation (5.43) contains $R^*-vR^*\cos\theta^*$ rather than R^*. This difference arises in a subtle way that is best understood by following how the integration over t' takes us from Equation (5.39) to Equation (5.42).

Finally, there are some points worth noting about this solution for the Liénard–Wiechert potential. First, in geometric algebra, a multivector potential obeying Equation (5.30) automatically embodies the Lorenz condition and obeys a scalar wave equation. Since the Liénard–Wiechert potential has been derived on the basis of this electromagnetic wave equation, it should be relativistically correct, that it is to say, it embodies the principle that information travels at a finite speed, the speed of light. It is therefore not actually necessary to take an essentially relativistic approach to the problem. The Coulomb gauge, in which $\boldsymbol{\nabla}\cdot\mathbf{A}=0$, does not fit in with a *multivector* potential such as we have here. Rather, in the Coulomb gauge, the vector and scalar potentials obey characteristically different equations [36, section 6.5, pp. 181–183] and in particular the scalar potential does not obey a wave equation at all.

The second point is that if we wished to extend Equation (5.43) to find an equivalent result for the potential of an arbitrary source density, it would be necessary to know not only the net charge and current distributions but also the velocities of each charge distribution involved. But in the case of suitably smooth distributions, we may make the approximation $R(t')-\mathbf{v}(t')\cdot\mathbf{R}(t')\approx R(t)$. This holds because $\mathbf{v}\cdot\mathbf{R}=v_RR$ where v_R is the velocity of an element of source current directed toward the observer, so that v_RR is equal to $R(t')-R(t)$, the distance traveled toward the observer in the time $R(t)/[c]$. This approximation therefore enables us to replace $R(t')-\mathbf{v}(t')\cdot\mathbf{R}(t')$ in the denominator of Equation (5.42) with $R(t)$, a simplification that entirely removes the need to consider either charge velocities or the effects of

retardation on the distance R. As to the numerator of Equation (5.42), we have to replace the source density for an individual charge with the integral over the entire source distribution so that in total

$$\boldsymbol{A}(t+\mathbf{r}) = \frac{-1}{4\pi} \cdot \int_V \frac{\boldsymbol{J}^*}{R} d^3 r' \tag{5.44}$$

This is effectively the same as the results given by various authors [19, section 14.7, p. 624; 35, sections 8.1–8.2, pp. 424–428, equations 25 and 26; 37, section 6.6, p. 186, equation 6.66]. As with discrete charges, however, the approximation for $R(t') - \mathbf{v}(t') \cdot \mathbf{R}(t')$ may not hold for discontinuous distributions and highly dynamic situations, for example, a burst of charge accelerated to high velocities.

Finally, note that while the principles of retardation apply directly to the electromagnetic potential, this is not the case for the electromagnetic field. Based on Equation (5.44), it would be tempting to take the analogous route of applying retardation directly to the expression for the electromagnetic field as given in Equation (5.13). But the electromagnetic field is to be found from Equation (5.30) with $\boldsymbol{F}' = 0$, specifically $\boldsymbol{F} = (\boldsymbol{\nabla} - \partial_t)\boldsymbol{A}$. Whereas all the other contributions to $\boldsymbol{F}$ have a denominator of R^2, we find that $-\partial_t \boldsymbol{A}$ introduces a term involving $\partial_t \boldsymbol{J}/R$, the origin of electromagnetic radiation. Curious though it may seem, introducing retardation to Equation (5.13) bypasses the possibility of any such term and would therefore give an incomplete result. Turning to the case of a point charge undergoing acceleration, however, we find some insight into the nature of its electromagnetic field. We obtain the electromagnetic field from $(\boldsymbol{\nabla} - \partial_t)\boldsymbol{A}$ and apply this to Equation (5.43). We find that $\boldsymbol{\nabla}\boldsymbol{A}$ is bound to result in terms that have a denominator of order R^2 whereas $-\partial_t \boldsymbol{A}$ leaves the denominator as R, which is a particular hallmark of a radiation field. The main contribution to the radiation field will therefore be related to $q\mathbf{a}/4\pi R$ where $\mathbf{a}$ is the acceleration of the charge. This is not an exact result because we have neglected both retardation and additional factors arising from the differentiation, yet it serves to demonstrate that such a field exists. Since an acceleration will still be measurable in the rest frame of the charge, an observer there *will see not only the charge's usual Coulomb field but also this radiation field*. A more precise account of the electromagnetic field of accelerating charges is undertaken in Chapter 12.

5.8 ENERGY AND MOMENTUM

In this section, it will be convenient for practical reasons to refer to $\boldsymbol{B}$ in its dual form $I\mathbf{B}$. We know that in free space, the squared magnitudes of the vectors $\mathbf{E}$ and $\mathbf{B}$ are directly related to the electromagnetic energy density $\mathcal{E}$ through $\mathcal{E} = \frac{1}{2}[\varepsilon_0]\left(E^2 + \left[c^2\right]B^2\right)$ where $E^2 = \mathbf{E}^2$ and $B^2 = \mathbf{B}^2 = -\boldsymbol{B}^2$. We may enquire about a counterpart of this relationship in terms of $\boldsymbol{F}$. It would be reasonable to start by considering $\boldsymbol{F}^2$

$$\boldsymbol{F}^2 = (\mathbf{E} + I\mathbf{B})^2 = \mathbf{E}^2 - \mathbf{B}^2 + 2I\mathbf{E}\cdot\mathbf{B} \tag{5.45}$$

First of all, it can be seen that the result is not a simple scalar, rather a scalar plus pseudoscalar, but on further investigation it is clear that $\boldsymbol{F}^2$ must in any case vanish for a plane electromagnetic wave since, from Equation (5.21),[3] we have the usual result $E = [c]B$ and $\mathbf{E}\cdot\mathbf{B} = 0$.

Let us now turn to $\boldsymbol{F}\boldsymbol{F}^\dagger$ where, as already defined, $\boldsymbol{F}^\dagger$ means the reverse of $\boldsymbol{F}$. For vectors and bivectors alike, $\boldsymbol{F}\boldsymbol{F}^\dagger$ is a scalar, but since $\boldsymbol{F}$ is actually a multivector combining both vector and bivector parts, a simple scalar result cannot be guaranteed. To evaluate $\boldsymbol{F}\boldsymbol{F}^\dagger$, we first need to recall from Section 4.3 that $\boldsymbol{I}^\dagger = -\boldsymbol{I}$, from which we can determine $\boldsymbol{F}^\dagger = (\mathbf{E} + I\mathbf{B})^\dagger = \mathbf{E} + \mathbf{B}I^\dagger = \mathbf{E} - I\mathbf{B}$. Applying this to $\boldsymbol{F}\boldsymbol{F}^\dagger$ we find

$$\tfrac{1}{2}\boldsymbol{F}\boldsymbol{F}^\dagger = \tfrac{1}{2}(\mathbf{E} + I\mathbf{B})(\mathbf{E} - I\mathbf{B}) = \underbrace{\tfrac{1}{2}\mathbf{E}^2 + \tfrac{1}{2}\mathbf{B}^2}_{\text{scalar}} - \underbrace{I\mathbf{E}\wedge\mathbf{B}}_{\text{vector}} \tag{5.46}$$

Although the result here is not a simple scalar, its scalar part resembles the electromagnetic energy density while the vector term in the form $\mathbf{E}\times\mathbf{B}$ resembles the Poynting vector [35, section 2.19, pp. 131–134]. But in order to evaluate it, we cannot just make the assumption $\mathbf{E}^2 = \mathbf{E}_0^2$ and $\mathbf{B}^2 = \mathbf{B}_0^2$ for, even in the simple case of plane waves, both $\mathbf{E}$ and $\mathbf{B}$ are linear combinations of $\mathbf{E}_0$ and $\mathbf{B}_0$ that depend on time and position as a result of the waves being circularly polarized as given in Equation (5.24). We could indeed use Equation (5.24), but it is easier to evaluate $\frac{1}{2}\boldsymbol{F}\boldsymbol{F}^\dagger$ directly from the form of a general plane wave, $\boldsymbol{F} = \boldsymbol{F}_0 e^{\pm I(\omega t - \mathbf{k}\cdot\mathbf{r})}$. As previously, we need to remember that pseudoscalars change sign under reversal but vectors do not:

$$\begin{aligned}\tfrac{1}{2}\boldsymbol{F}\boldsymbol{F}^\dagger &= \tfrac{1}{2}(\mathbf{E}_0 + I\mathbf{B}_0)e^{\pm I(\omega t-\mathbf{k}\cdot\mathbf{r})}e^{\mp I(\omega t-\mathbf{k}\cdot\mathbf{r})}(\mathbf{E}_0 - I\mathbf{B}_0)\\ &= \tfrac{1}{2}(\mathbf{E}_0 + I\mathbf{B}_0)(\mathbf{E}_0 - I\mathbf{B}_0)\\ &= \tfrac{1}{2}\mathbf{E}_0^2 + \tfrac{1}{2}\mathbf{B}_0^2 - I\mathbf{E}_0\wedge\mathbf{B}_0\end{aligned} \tag{5.47}$$

In other words, the pseudoscalar phase factors simply cancel and, in spite of $\mathbf{E}$ and $\mathbf{B}$ themselves being dependent on time and position, the result is identical to the general form given in Equation (5.46). We may restate (5.47) as

$$\begin{aligned}\tfrac{1}{2}\left\langle\boldsymbol{F}\boldsymbol{F}^\dagger\right\rangle_0 &= \tfrac{1}{4}\left(\boldsymbol{F}\boldsymbol{F}^\dagger + \boldsymbol{F}^\dagger\boldsymbol{F}\right) = \tfrac{1}{2}\mathbf{E}_0^2 + \tfrac{1}{2}\mathbf{B}_0^2\\ \tfrac{1}{2}\left\langle\boldsymbol{F}\boldsymbol{F}^\dagger\right\rangle_1 &= \tfrac{1}{4}\left(\boldsymbol{F}\boldsymbol{F}^\dagger - \boldsymbol{F}^\dagger\boldsymbol{F}\right) = -I\mathbf{E}_0\wedge\mathbf{B}_0 = E_0B_0\hat{\mathbf{k}}\end{aligned} \tag{5.48}$$

Note that we cannot just take the inner and outer products of $\boldsymbol{F}$ and $\boldsymbol{F}^\dagger$ as being the requisite scalar and vector parts of $\boldsymbol{F}\boldsymbol{F}^\dagger$ since they are multivectors comprising

[3] This result can also be seen to come directly from the homogeneous Maxwell's equation $\boldsymbol{\nabla}\wedge\mathbf{E} + I\partial_t\mathbf{B} = 0$.

two separate grades (see Exercise 5.10.4). When we return the suppressed constants and include the factor of ε_0 to convert to the appropriate units, the scalar takes the familiar form of the electromagnetic energy density:

$$\mathcal{E} = \tfrac{1}{2}[\varepsilon_0]\mathbf{E}_0^2 + \tfrac{1}{2}[\mu_0^{-1}]\mathbf{B}_0^2 \tag{5.49}$$

The vector part is then equivalent to

$$\mathbf{g} = [\varepsilon_0]\mathbf{E}_0 \times \mathbf{B}_0 \tag{5.50}$$

where $\mathbf{g}$ is the electromagnetic momentum density vector [35, section 2.6, p.103]. We can combine both results as a multivector in two possible ways,

$$\tfrac{1}{2}\boldsymbol{F}\boldsymbol{F}^\dagger = \mathcal{E} + \mathbf{g} \quad \text{or} \quad \tfrac{1}{2}\boldsymbol{F}^\dagger\boldsymbol{F} = \mathcal{E} - \mathbf{g} \tag{5.51}$$

but it is clear we need to consider only either the one or the other.

It is clear we could have chosen to use the Poynting vector $\mathbf{S}$ rather than the momentum density vector in Equation (5.51) since $\mathbf{S} = c^2\mathbf{g}$ and only a change of dimensions is required. We then have the slightly different form $\frac{1}{2}[\varepsilon_0]\boldsymbol{F}\boldsymbol{F}^\dagger = \mathcal{E} + \left[\frac{1}{c^2}\right]\mathbf{S}$. Looking at the alternative form $\frac{1}{2}\boldsymbol{F}^\dagger\boldsymbol{F} = \mathcal{E} - \mathbf{S}$, note that the right-hand side is similar in appearance to $\rho - \mathbf{J}$. Given the conservation of charge is expressed as the continuity equation $\partial_t\rho + \boldsymbol{\nabla}\cdot\mathbf{J} = 0$, there would appear to be a parallel in $\partial_t\mathcal{E} + \boldsymbol{\nabla}\cdot\mathbf{S} = 0$ as a conservation equation for electromagnetic energy. This supports the interpretation of $\mathbf{S}$ as flow of energy. In contrast to charge, however, electromagnetic energy can be created or absorbed, for example, by moving charges around or transferring energy from one charge to another. The conservation law of course applies to the total energy, not just the electromagnetic contribution, and so for completeness we need to include $\mathcal{U}$, which we take as effectively being the energy per unit volume that is expended in changing the electromagnetic field, resulting in $\partial_t\mathcal{E} + \boldsymbol{\nabla}\cdot\mathbf{S} = \partial_t\mathcal{U}$. Jackson gave a version of this [37, section 6.9, p. 191, equation 6.85] that applies to the total energy.

Once again, we may find it refreshing to see how simply equations may be cast, manipulated, and analyzed with geometric algebra.

5.9 MAXWELL'S EQUATIONS IN POLARIZABLE MEDIA

In Section 5.2, we dealt with Maxwell's equations in their fundamental form, that is to say, as they apply to free space. Notwithstanding the reference to free space, these equations also apply to real media so long as all bound source distributions are explicitly accounted for. In dealing with media, therefore, it is more relevant to call them the *microscopic* equations. In contrast, however, even from their inception the traditional form of Maxwell's equations has been *macroscopic*. Bringing $\mathbf{D}$ and $\mathbf{H}$ into the equations allows the polarizable material to be taken into account without

explicitly involving these so-called bound sources. Here we use the term polarizable in the general sense including both magnetic and electric polarization so that the bound source distribution includes both charge and magnetization current. Permanent sources of polarization apart, these bound sources generally depend on the applied electromagnetic field, and when they change in response to it they consequently modify the field prevailing within the material or "medium". Such bound sources are therefore only implicitly involved in the governing equations, and consequently, the motivation for the macroscopic form of Maxwell's equations is to eliminate them. In a typical situation where the material characteristics are such that **D** and **H** are respectively proportional to **E** and **B**, this has the simplifying result that the macroscopic equations are of the same form as the free space equation and only the constants involved differ. However typical this situation might be, it is not the general case. What is missing is the step through which **D** and **H** are related to the prevailing electromagnetic field and polarization in the medium. This is summarized in the constitutive relations:

$$\begin{aligned} \mathbf{D} &= [\varepsilon_0]\mathbf{E} + \mathbf{P} \\ \mathbf{H} &= \frac{\mathbf{B}}{[\mu_0]} - \mathbf{M} \end{aligned} \tag{5.52}$$

The use of square brackets here means that ε_0 and μ_0 will subsequently be suppressed. Electromagnetic force is the physical basis of the definition of **E** and **B**, whereas **P**, the electric polarization, and **M**, the magnetic polarization or "magnetization", derive from the distribution of bound sources within the medium (note that at this stage, we are using the traditional vector forms, **B** and **M**). Both represent the net dipole moment of the appropriate type per unit volume. It is clear that the constitutive equations therefore *define* **D** and **H**. From a fundamental perspective, therefore, **D** and **H** are actually redundant. Their role is purely auxiliary and as such may be referred to as the auxiliary electric and magnetic fields respectively.

The static bound sources may be evaluated from the polarization by considering the medium to be made up of small cuboids, each of which may be treated as being uniformly polarized. Any uniformly polarized region still maintains a net zero distribution of molecular charge and magnetic current and only at its surface is any net source density to be found. Summing up the contributions to the bound source density from all the surfaces leads to the well-known results $\rho_{\text{bound}} = -\nabla \cdot \mathbf{P}$ and $\mathbf{J}_{\text{bound}} = \nabla \times \mathbf{M}$. But in addition to these static sources, there is also a current, the polarization current, due to the microscopic motion of the charges within polarized molecules. This may be inferred from applying Equation (5.11) to the charges involved in the polarization to gives us the complete bound source density due to polarization:

$$\begin{aligned} \boldsymbol{J}_P(\mathbf{r}) &= \sum_{i=+,-} \rho_i(\mathbf{r})(1 - \partial_t \mathbf{u}_i) \\ &= \sum_{i=+,-} \rho_i(\mathbf{r}) - \partial_t \sum_{i=+,-} \rho_i(\mathbf{r})\mathbf{u}_i \end{aligned} \tag{5.53}$$

Here $\mathbf{u}_i(t+\mathbf{r})$ represents the displacement of a small volume element of the charge distribution i from its mean location $\mathbf{r}$ at the time t. Now the leading term, $\sum_{i=+,-}\rho_i(\mathbf{r})$, gives us the net charge, which we must associate with $-\boldsymbol{\nabla}\cdot\mathbf{P}$, whereas the second term $\sum_{i=+,-}\rho_i(\mathbf{r})\mathbf{u}_i$ may be recognized as the dipole moment density, $\mathbf{P}$, that is to say, the polarization. Note that it is implied that only $\mathbf{u}_i$ changes with time, whereas ρ_i is taken as fixed. We may conclude, therefore, that the total electromagnetic source density, $\boldsymbol{J}_P$, due to both the spatial and temporal variation of the polarization is

$$\boldsymbol{J}_P = -\boldsymbol{\nabla}\cdot\mathbf{P} - \partial_t\mathbf{P} \tag{5.54}$$

Putting this together with the magnetic contribution to the bound source density, which has no counterpart to $\partial_t\boldsymbol{P}$, we find the total bound source density to be

$$\boldsymbol{J}_{\text{bound}} = \underbrace{-\boldsymbol{\nabla}\cdot\mathbf{P}}_{\rho_{\text{bound}}} - \underbrace{\partial_t\mathbf{P}+\boldsymbol{\nabla}\times\mathbf{M}}_{\mathbf{J}_{\text{bound}}} \tag{5.55}$$

This result allows the bound sources to be eliminated from the fundamental microscopic equations (Equation 5.3) by writing the total source densities, ρ and $\mathbf{J}$, as

$$\begin{aligned}\rho &= \rho_{\text{free}} - \boldsymbol{\nabla}\cdot\mathbf{P}\\ \mathbf{J} &= \mathbf{J}_{\text{free}} + \partial_t\mathbf{P} + \boldsymbol{\nabla}\times\mathbf{M}\end{aligned} \tag{5.56}$$

where the free sources, ρ_{free} and $\mathbf{J}_{\text{free}}$, represent all those other sources that do not originate from the polarization of the medium. Thereafter, it is only necessary to bring in the constitutive relations (Equation 5.52) in order to introduce the variables $\mathbf{D}$ and $\mathbf{H}$ in place of $\mathbf{P}$ and $\mathbf{M}$. Recalling that we are now going to drop the constants ε_0 and μ_0 from these relations, we find the standard form of Maxwell's equations:

$$\begin{aligned}\boldsymbol{\nabla}\cdot\mathbf{D} &= \rho_{\text{free}}\\ \boldsymbol{\nabla}\times\mathbf{E} &= -\partial_t\mathbf{B}\\ \boldsymbol{\nabla}\cdot\mathbf{B} &= 0\\ \boldsymbol{\nabla}\times\mathbf{H} &= \mathbf{J}_{\text{free}} + \partial_t\mathbf{D}\end{aligned} \tag{5.57}$$

The free sources appear as the independent variables within these macroscopic equations, whereas the bound sources are generally to be treated as being involved only implicitly. This is most usefully achieved when linear approximations such as $\mathbf{D}=\varepsilon\mathbf{E}$ and $\mu\mathbf{H}=\mathbf{B}$ are valid. The constants ε and μ are characteristic of the medium concerned, and any difference from ε_0 and μ_0 allows the effects of induced polarization to be accounted for.

With this familiar approximation, which only applies to media that are linear, homogeneous and isotropic, it is clear that Maxwell's equations take the same form as the free space equations with ε_0 and μ_0 being simply replaced by ε and μ, so that Equation (5.6) becomes

$$\left(\boldsymbol{\nabla}+\frac{1}{c'}\partial_t\right)(\mathbf{E}+c'\boldsymbol{B}) = \frac{\rho}{\varepsilon}-Z'\mathbf{J} \tag{5.58}$$

where $c'=(\varepsilon\mu)^{-1/2}$ and $Z'=(\mu/\varepsilon)^{1/2}$.

Let us now move on to consider how these macroscopic field equations are to be encoded in terms of geometric algebra. Starting with the source density, we have from Equation (5.55),

$$\begin{aligned}
\boldsymbol{J}_{\text{bound}} &= [\varepsilon_0^{-1}]\rho_{\text{bound}}-[Z_0]\mathbf{J}_{\text{bound}}\\
&= -[\varepsilon_0^{-1}]\boldsymbol{\nabla}\cdot\mathbf{P}-[Z_0](\partial_t\mathbf{P}+\boldsymbol{\nabla}\times\mathbf{M})\\
&= -\boldsymbol{\nabla}\cdot\mathbf{P}+I\boldsymbol{\nabla}\wedge\mathbf{M}-\partial_t\mathbf{P}\\
&= -\boldsymbol{\nabla}\cdot\mathbf{P}+\boldsymbol{\nabla}\cdot(I\mathbf{M})-\partial_t\mathbf{P}\\
&= -\boldsymbol{\nabla}\cdot([\mathbf{P}-I\mathbf{M})-\partial_t\mathbf{P}\\
&= -\boldsymbol{\nabla}\cdot\left([\varepsilon_0^{-1}]\mathbf{P}-[Z_0]\boldsymbol{M}\right)-[c^{-1}]\partial_t[\varepsilon_0^{-1}]\mathbf{P}\\
&= -\boldsymbol{\nabla}\cdot\boldsymbol{Q}-\partial_t\mathbf{P}
\end{aligned} \tag{5.59}$$

where we make use of the identity $\boldsymbol{\nabla}\cdot(I\mathbf{u})=I\boldsymbol{\nabla}\wedge\mathbf{u}$ from Exercise 4.8.3 and introduce $\boldsymbol{Q}=[\varepsilon_0^{-1}]\mathbf{P}-[Z_0]\boldsymbol{M}$ in which the suppressed constants are seen to take exactly the same form as in the definition of $\boldsymbol{J}$. We may refer to $\boldsymbol{Q}$ as the electromagnetic polarization multivector, but there clearly is a difficulty here for it would appear that we cannot derive a time-dependent bound electromagnetic source density from $\boldsymbol{Q}$ alone, we have to retain $\boldsymbol{P}$. However, noting that only grades 2 and 3 are involved in $\boldsymbol{\nabla}\wedge\boldsymbol{Q}$ and $\partial_t\boldsymbol{M}$, we could solve the problem by writing

$$\begin{aligned}
\boldsymbol{J}_{\text{bound}} &= \langle-\boldsymbol{\nabla}\cdot\boldsymbol{Q}-\partial_t\mathbf{P}-\boldsymbol{\nabla}\wedge\boldsymbol{Q}+\partial_t\mathbf{M}\rangle_{0,1}\\
&= -\langle(\boldsymbol{\nabla}+\partial_t)\boldsymbol{Q}\rangle_{0,1}
\end{aligned} \tag{5.60}$$

By separating $\boldsymbol{J}$ into $\boldsymbol{J}_{\text{free}}+\boldsymbol{J}_{\text{bound}}$, we find that Maxwell's equation (Equation 5.7) now becomes

$$(\boldsymbol{\nabla}+\partial_t)\boldsymbol{F}=\boldsymbol{J}_{\text{free}}-\langle(\boldsymbol{\nabla}+\partial_t)\boldsymbol{Q}\rangle_{0,1} \tag{5.61}$$

Noting that the electromagnetic source density is always a paravector, that is to say it comprises a scalar (grade 0) plus a vector (grade 1), Maxwell's equation may be split into a homogeneous equation plus an inhomogeneous equation as follows:

$$\begin{aligned}
&\langle(\boldsymbol{\nabla}+\partial_t)\boldsymbol{F}\rangle_{0,1}=\boldsymbol{J}_{\text{free}}-\langle(\boldsymbol{\nabla}+\partial_t)\boldsymbol{Q}\rangle_{0,1}\\
&\langle(\boldsymbol{\nabla}+\partial_t)\boldsymbol{F}\rangle_{2,3}=0\\
\Leftrightarrow&\begin{cases}\langle(\boldsymbol{\nabla}+\partial_t)(\boldsymbol{F}+\boldsymbol{Q})\rangle_{0,1}=\boldsymbol{J}_{\text{free}}\\ \langle(\boldsymbol{\nabla}+\partial_t)\boldsymbol{F}\rangle_{2,3}=0\end{cases}
\end{aligned} \tag{5.62}$$

In the traditional form of the macroscopic equations, $\mathbf{P}$ and $\boldsymbol{M}$ may be incorporated with $\mathbf{E}$ and $\boldsymbol{B}$ to form the auxiliary fields $\mathbf{D}$ and $\boldsymbol{H}$. It is clear that we may similarly incorporate $\boldsymbol{Q}$ with $\boldsymbol{F}$ to form an auxiliary field $\boldsymbol{G}$, where

$$\begin{aligned}\boldsymbol{G} &= \boldsymbol{F} + \boldsymbol{Q} \\ &= \left(\mathbf{E} + [\mathrm{c}]\boldsymbol{B}\right) + \left([\varepsilon_0^{-1}]\mathbf{P} - [Z_0]\boldsymbol{M}\right) \\ &= \left(\mathbf{E} + [\varepsilon_0^{-1}]\mathbf{P}\right) + [Z_0]\left([\mu_0^{-1}]\boldsymbol{B} - \boldsymbol{M}\right) \\ &= [\varepsilon_0^{-1}]\mathbf{D} + [Z_0]\boldsymbol{H}\end{aligned} \tag{5.63}$$

Note that the suppressed constants in the multivectors $\boldsymbol{G}$, $\boldsymbol{Q}$, and $\boldsymbol{J}$ present little difficulty because they are exactly the same in each case. We may refer to $\boldsymbol{G}$ as the *auxiliary* electromagnetic field multivector and restate Equation (5.62) as

$$\begin{aligned}\langle(\boldsymbol{\nabla} + \partial_t)\boldsymbol{G}\rangle_{0,1} &= \boldsymbol{J}_{\text{free}} \\ \langle(\boldsymbol{\nabla} + \partial_t)\boldsymbol{F}\rangle_{2,3} &= 0\end{aligned} \tag{5.64}$$

If we wish to express this as a single equation, we need to fall back on Equation (5.61) or, by adding $(\boldsymbol{\nabla} + \partial_t)\boldsymbol{Q}$ to each side, its alternative form

$$\begin{aligned}&(\boldsymbol{\nabla} + \partial_t)\boldsymbol{F} = \boldsymbol{J}_{\text{free}} - \langle(\boldsymbol{\nabla} + \partial_t)\boldsymbol{Q}\rangle_{0,1} \\ \Leftrightarrow\quad &(\boldsymbol{\nabla} + \partial_t)\boldsymbol{G} = \boldsymbol{J}_{\text{free}} + \langle(\boldsymbol{\nabla} + \partial_t)\boldsymbol{Q}\rangle_{2,3}\end{aligned} \tag{5.65}$$

While the alternative form of equation changes the bound source from $-\boldsymbol{\nabla}\cdot\boldsymbol{Q} - \partial_t\mathbf{P}$ to $\boldsymbol{\nabla}\wedge\boldsymbol{Q} - \partial_t\boldsymbol{M}$, this hardly amounts to any simplification. In fact, $\boldsymbol{\nabla}\wedge\boldsymbol{Q} - \partial_t\boldsymbol{M}$ is not even a proper source density. Rather than being a paravector, it is actually the *dual* of a paravector. Nevertheless, the result still has considerable significance. This is easier to see in the static case so that with $\partial_t = 0$, the two forms of the equation become

$$\begin{aligned}\boldsymbol{\nabla}\boldsymbol{F} &= \boldsymbol{J}_{\text{free}} - \boldsymbol{\nabla}\cdot\boldsymbol{Q} \\ \boldsymbol{\nabla}\boldsymbol{G} &= \boldsymbol{J}_{\text{free}} + \boldsymbol{\nabla}\wedge\boldsymbol{Q}\end{aligned} \tag{5.66}$$

wherein the bound source terms $-\boldsymbol{\nabla}\cdot\boldsymbol{Q}$ and $\boldsymbol{\nabla}\wedge\boldsymbol{Q}$ have the roles of charges and currents actually *reversed*. If we write $\boldsymbol{G} = \mathbf{D} + I\mathbf{H}$ and $\boldsymbol{Q} = \mathbf{P} - I\mathbf{M}$, that is to say we use the dual forms $I\mathbf{H}$ and $I\mathbf{M}$ for $\boldsymbol{H}$ and $\boldsymbol{M}$, we find

$$\begin{aligned}-\boldsymbol{\nabla}\cdot\boldsymbol{Q} &= -\boldsymbol{\nabla}\cdot(\mathbf{P} - I\mathbf{M}) \\ &= -\boldsymbol{\nabla}\cdot\mathbf{P} + I\boldsymbol{\nabla}\wedge\mathbf{M} \\ &= -\boldsymbol{\nabla}\cdot\mathbf{P} - \boldsymbol{\nabla}\times\mathbf{M} \qquad \text{(i)} \\ \boldsymbol{\nabla}\wedge\boldsymbol{Q} &= \boldsymbol{\nabla}\wedge(\mathbf{P} - I\mathbf{M}) \\ &= -I\boldsymbol{\nabla}\cdot\mathbf{M} + \boldsymbol{\nabla}\wedge\mathbf{P} \\ &= I\left(-\boldsymbol{\nabla}\cdot\mathbf{M} + \boldsymbol{\nabla}\times\mathbf{P}\right) \qquad \text{(ii)}\end{aligned} \tag{5.67}$$

In case (i), the electric source arises from the divergence of $\mathbf{P}$ while the magnetic source comes from the curl of $\mathbf{M}$. This is normally how we see things but, contrarily, in (ii) the familiar roles are reversed for, as we can see, the source involves the divergence of $\mathbf{M}$ and the curl of $\mathbf{P}$. Stratton pointed out [35, section 4.3, p. 228] that any magnetostatic problem may be reduced to an electrostatic one, that is to say we may replace current dipoles with magnetic charges, *poles*. Returning to Equation (5.66), we find that (ii) leads us to

$$\begin{aligned} \nabla \boldsymbol{G} &= \boldsymbol{J}_{\text{free}} + \nabla \wedge \boldsymbol{Q} \\ \Leftrightarrow \nabla(\mathbf{D} + I\mathbf{H}) &= \boldsymbol{J}_{\text{free}} + I(-\nabla \cdot \mathbf{M} + \nabla \times \mathbf{P}) \\ \Leftrightarrow \nabla(\mathbf{H} - I\mathbf{D}) &= -I\boldsymbol{J}_{\text{free}} - \nabla \cdot \mathbf{M} + \nabla \times \mathbf{P} \end{aligned} \tag{5.68}$$

The field that results from the magnetization alone therefore reduces to $\nabla \mathbf{H} = -\nabla \cdot \mathbf{M}$, so that the claim is indeed valid provided (1) we replace the bivector $\boldsymbol{B}$ with the *vector* $\mathbf{H}$ and (2), we do not consider what is happening on the atomic scale. This mathematical equivalence therefore allows us to think of the magnetic field as originating from poles provided that we take it as being given by $\mathbf{H}$ rather than $\boldsymbol{B}$. Though mathematically sound and very convenient for computational purposes, it is nevertheless diametrically opposed to the true physical situation, first because, if we consider Equation (5.68) as a whole, it clearly implies treating the free source density as $-I\boldsymbol{J}_{\text{free}}$ rather than $\boldsymbol{J}_{\text{free}}$ so that, for example, charge needs to be treated as a pseudoscalar. Second, and the critical issue, the magnetic force is specified in terms of $\boldsymbol{B}$ acting on currents rather than $\mathbf{H}$ acting on poles [2].

The main reason for employing $\mathbf{D}$ and $\mathbf{H}$ in the first and last of Maxwell's equations is that this eliminates the bound charges and magnetic currents from the source terms. While the wisdom of this is generally unchallenged, the resulting equations are incomplete without some form of constitutive relation such as the general form given in Equation (5.52), or the much simpler linear forms $\mathbf{D} = \varepsilon \mathbf{E}$ and $\mu \mathbf{H} = \mathbf{B}$. Only in the latter case is there any great benefit over simply retaining $\mathbf{P}$ and $\mathbf{M}$. Geometric algebra demonstrates the point here inasmuch as if we take the same approach these macroscopic equations turn out to be:

$$\begin{aligned} \nabla \wedge \boldsymbol{F} + \partial_t \boldsymbol{B} &= 0 \\ \nabla \cdot \boldsymbol{G} + \partial_t \mathbf{D} &= \boldsymbol{J}_{\text{free}} \end{aligned} \tag{5.69}$$

However we may try, we cannot make these into complete algebraic equations with $\boldsymbol{F}$ and $\boldsymbol{G}$ alone as field variables. There is always some contribution from the individual fields, in this case in the form of the time derivatives of $\boldsymbol{B}$ and $\mathbf{D}$. Finally, it is easy to be persuaded by the neatness of Equation (5.69) that it is somehow more useful than Equation (5.61), which by comparison may seem more clumsy. However, the fact that it involves four separate field variables rather than just two detracts from any advantage suggested by its appearance.

5.9.1 Boundary Conditions at an Interface

An interface between two different media usually involves a change in one or the other of the constants ε and μ. Stated in their most general form [35, section 1.3, p. 37; 37, section 4.4, p. 110 and section 5.9, p. 155], the boundary conditions at the interface are

$$\begin{aligned} \mathbf{n}\cdot(\mathbf{E}_2-\mathbf{E}_1)&=\sigma \\ \mathbf{n}\times(\mathbf{E}_2-\mathbf{E}_1)&=0 \\ \mathbf{n}\cdot(\mathbf{B}_2-\mathbf{B}_1)&=0 \\ \mathbf{n}\times(\mathbf{B}_2-\mathbf{B}_1)&=\mathbf{K} \end{aligned} \tag{5.70}$$

The regions on either side of the interface are denoted by the subscripts 1 and 2 and at any given point on the interface:

- $\mathbf{n}$ is the unit normal directed from side 1 to side 2.
- The fields involved are measured immediately adjacent to the point but on opposite sides of the interface.
- The total (free plus bound) surface charge and current densities are σ and $\mathbf{K}$ respectively.

Let us now put these four separate equations into the context of a geometric algebra. First of all, we introduce the simple scalar difference operator, Δ, such that $\Delta\boldsymbol{U}=\boldsymbol{U}_2-\boldsymbol{U}_1$. For example, we may express the boundary condition on $\boldsymbol{B}$ as $\mathbf{n}\cdot\Delta\mathbf{B}=0$. It is relatively straightforward to manipulate expressions involving Δ, for example, $\Delta(\mathbf{nu})$ is the same thing as $\mathbf{n}\Delta\mathbf{u}$ since $\mathbf{n}$ can be treated as being constant across the interface. However, take care that, as is easily shown, the general result is $\Delta(\mathbf{nu})=(\Delta\mathbf{n})\mathbf{u}_1+\mathbf{n}_1\Delta\mathbf{u}$.

By means of the identity $\mathbf{n}\times\Delta\mathbf{u}=-I\mathbf{n}\wedge\Delta\mathbf{u}$, we now replace the cross products and in order to restore the bivector form of the magnetic field, we make use of the identities $\mathbf{n}\cdot\Delta\mathbf{B}=-I\mathbf{n}\wedge(I\Delta\mathbf{B})=-I\mathbf{n}\wedge\Delta\boldsymbol{B}$ and $I\mathbf{n}\wedge\Delta\mathbf{B}=\mathbf{n}\cdot(I\Delta\mathbf{B})=\mathbf{n}\cdot\Delta\boldsymbol{B}$, both of which follow directly from Equation (4.6). The end result turns out simply enough as

$$\begin{aligned} \mathbf{n}\cdot\Delta\mathbf{E}&=\sigma \\ \mathbf{n}\wedge\Delta\mathbf{E}&=0 \\ \mathbf{n}\wedge\Delta\boldsymbol{B}&=0 \\ \mathbf{n}\cdot\Delta\boldsymbol{B}&=-\mathbf{K} \end{aligned} \tag{5.71}$$

However, it is immediately evident that these four equations may be combined into one by simply adding them all together. Doing this in two stages, we achieve

$$\begin{aligned}
&\mathbf{n}\cdot\Delta\mathbf{E}+\mathbf{n}\wedge\Delta\mathbf{E}=\sigma \\
\Leftrightarrow\quad &\mathbf{n}\Delta\mathbf{E}=\sigma \qquad \text{(i)} \\
&\mathbf{n}\cdot\Delta\boldsymbol{B}+\mathbf{n}\wedge\Delta\boldsymbol{B}=-\mathbf{K} \\
\Leftrightarrow\quad &\mathbf{n}\Delta\boldsymbol{B}=-\mathbf{K} \qquad \text{(ii)} \\
\Leftrightarrow\quad &\mathbf{n}\left(\Delta\mathbf{E}+\Delta\boldsymbol{B}\right)=\sigma-\mathbf{K} \\
\Leftrightarrow\quad &\mathbf{n}\Delta\boldsymbol{F}=\boldsymbol{K} \qquad \text{(iii)}
\end{aligned} \tag{5.72}$$

where $\boldsymbol{K}$ is the surface density of electromagnetic sources, analogous to the volume density multivector $\boldsymbol{J}$. Expressed in this way, the boundary conditions at a smooth interface are as simple as they could be. While it is usually necessary to refer to a textbook to recall the traditional forms, as in Equation (5.70), $\mathbf{n}\Delta\boldsymbol{F}=\boldsymbol{K}$ is easily remembered. What is also really useful is that, since $\mathbf{n}^2=1$ by definition, we may also write this as $\Delta\boldsymbol{F}=\mathbf{n}\boldsymbol{K}$, which gives us the discontinuity in the fields directly!

The form $\mathbf{n}\Delta\boldsymbol{F}=\boldsymbol{K}$, however, appears somewhat similar to $\boldsymbol{\nabla F}=\boldsymbol{J}$. Starting with Maxwell's equation $\left(\boldsymbol{\nabla}+\partial_t\right)\boldsymbol{F}=\boldsymbol{J}$, if we then

- drop the time derivative,
- replace $\boldsymbol{\nabla}$ with the *vector* difference operator $\mathbf{n}\Delta$ and
- replace the volume source density $\boldsymbol{J}$ with the surface source density $\boldsymbol{K}$,

we recover $\mathbf{n}\Delta\boldsymbol{F}=\boldsymbol{K}$. This correspondence is still apparent if we replace $\mathbf{E}$ and $\mathbf{B}$ in the two inhomogeneous boundary equations with $\mathbf{D}$ and $\mathbf{H}$, in which case σ and $\mathbf{K}$ represent the free surface charge and current densities alone, so that

$$\begin{aligned}
\mathbf{n}\cdot\Delta\mathbf{D}&=\sigma_{\text{free}} \\
\mathbf{n}\wedge\Delta\mathbf{E}&=0 \\
\mathbf{n}\wedge\Delta\boldsymbol{B}&=0 \\
\mathbf{n}\cdot\Delta\boldsymbol{H}&=-\mathbf{K}_{\text{free}} \\
\Leftrightarrow\ \mathbf{n}\wedge\Delta\boldsymbol{F}&=0 \\
\mathbf{n}\cdot\Delta\boldsymbol{G}&=\boldsymbol{K}_{\text{free}}
\end{aligned} \tag{5.73}$$

In comparison with Equation (5.69), the time derivatives have gone, and $\boldsymbol{\nabla}$ is once again replaced with $\mathbf{n}\Delta$ provided it is understood that we apply Δ before either the "dot" or the "wedge."

Because of this correspondence, we may form the idea that it is possible to derive the boundary conditions at a smooth interface directly from Maxwell's equations, that is to say, without bringing in the integral equations to a small volume or area bisected by the interface. It seems appropriate to begin by considering how a quantity such as $\boldsymbol{\nabla F}$ behaves when there is a discontinuity in $\boldsymbol{F}$. Effectively, we have to find some way of making the resulting infinities disappear by taking suitable limits.

To simplify matters initially, suppose mediums 1 and 2 have some scalar property, say s, for example, mass density or dielectric constant, that varies smoothly except in passing across the interface itself. The vector derivative $\boldsymbol{\nabla} s$ will be well behaved everywhere except at this surface where it will be dominated by the discontinuous change in s. However, just as for any vector, the vector operator $\boldsymbol{\nabla}$ may be expressed in any basis we like. Choosing an arbitrary orthonormal basis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$, we have

$$\boldsymbol{\nabla} = \mathbf{e}_1(\mathbf{e}_1 \cdot \boldsymbol{\nabla}) + \mathbf{e}_2(\mathbf{e}_2 \cdot \boldsymbol{\nabla}) + \mathbf{e}_3(\mathbf{e}_3 \cdot \boldsymbol{\nabla}) \tag{5.74}$$

In particular, choosing $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 = \mathbf{x}, \mathbf{y}, \mathbf{z}$ restores the usual form $\boldsymbol{\nabla} = \mathbf{x}\partial_x + \mathbf{y}\partial_y + \mathbf{z}\partial_z$. All Equation (5.74) represents is a change to a new set of basis vectors. Clearly, if at some point on the interface we choose $\mathbf{e}_1 = \mathbf{n}$ so that $\mathbf{e}_2$ and $\mathbf{e}_3$ lie in the plane that is tangent to the interface, then

$$\boldsymbol{\nabla} = \mathbf{n}(\mathbf{n} \cdot \boldsymbol{\nabla}) + \mathbf{e}_2(\mathbf{e}_2 \cdot \boldsymbol{\nabla}) + \mathbf{e}_3(\mathbf{e}_3 \cdot \boldsymbol{\nabla}) \tag{5.75}$$

Focusing on the leading term $\mathbf{n}(\mathbf{n} \cdot \boldsymbol{\nabla})$, just in the same way that $\mathbf{x}(\mathbf{x} \cdot \boldsymbol{\nabla}) = \mathbf{x}\partial_x$ represents the derivative in the $\mathbf{x}$ direction, we see this must simply represent the derivative in the direction of $\mathbf{n}$. This is therefore called the directional derivative along $\mathbf{n}$. But from a basic perspective, this must be dominated at the interface itself by any discontinuous change that occurs. Making the simplifying assumption that the change from one side of the interface to the other occurs linearly over a very small yet finite distance δ, the derivative at the surface may be made to appear well behaved

$$\mathbf{n}(\mathbf{n} \cdot \boldsymbol{\nabla}) s = \mathbf{n}\partial_n s = \mathbf{n}\frac{s_2 - s_1}{\delta} = \mathbf{n}\frac{\Delta s}{\delta} \tag{5.76}$$

Here n is the independent variable representing a distance along the direction $\mathbf{n}$. Of course, as $\delta \to 0$, $\mathbf{n}(\Delta s/\delta)$ will become very large indeed, and since, by assumption, there are no discontinuities parallel to the surface along $\mathbf{e}_2$ and $\mathbf{e}_3$, we must have $\mathbf{e}_2(\mathbf{e}_2 \cdot \boldsymbol{\nabla}) s + \mathbf{e}_3(\mathbf{e}_3 \cdot \boldsymbol{\nabla}) s \ll \mathbf{n}(\mathbf{n} \cdot \boldsymbol{\nabla}) s$ so that, provided we make δ small enough, the vector derivative of s at the interface itself reduces to

$$\boldsymbol{\nabla} s = \mathbf{n}(\mathbf{n} \cdot \boldsymbol{\nabla}) s = \mathbf{n}\frac{\Delta s}{\delta} \tag{5.77}$$

Given its simple linear form, this result must apply equally well when the scalar function s is replaced by any multivector function $\boldsymbol{U}$ provided that it too varies smoothly everywhere except across the interface itself. In general, therefore,

$$\boldsymbol{\nabla} \boldsymbol{U} = \mathbf{n}(\mathbf{n} \cdot \boldsymbol{\nabla}) \boldsymbol{U} = \mathbf{n}\frac{\Delta \boldsymbol{U}}{\delta} \tag{5.78}$$

But, given the presence of the indeterminate parameter δ, how does this help us? Let us return to Maxwell's equation in free space (Equation 5.6) in the form

$(\boldsymbol{\nabla}+\partial_t)\boldsymbol{F}=\rho-\mathbf{J}$. Clearly, at the interface between two media, $\boldsymbol{\nabla F}$ will be dominated by the large term contributed by $\mathbf{n}(\mathbf{n}\cdot\boldsymbol{\nabla})\boldsymbol{F}$ so that

$$\begin{aligned}(\nabla+\partial_t)\boldsymbol{F} &= \rho-\mathbf{J}\\ \Leftrightarrow \lim_{\delta\to 0}\left(\mathbf{n}\frac{\Delta\boldsymbol{F}}{\delta}\right) &= \rho-\mathbf{J}-\partial_t\boldsymbol{F}\\ \Leftrightarrow \qquad \mathbf{n}\Delta\boldsymbol{F} &= \lim_{\delta\to 0}\left(\delta(\rho-\mathbf{J})-\delta\partial_t\boldsymbol{F}\right)\\ &= \lim_{\delta\to 0}\left(\delta(\rho-\mathbf{J})\right)\end{aligned} \tag{5.79}$$

While $\lim_{\delta\to 0}\delta\partial_t\boldsymbol{F}$ simply vanishes, $\lim_{\delta\to 0}\delta(\rho-\mathbf{J})$ will not vanish when some surface density of charge or current is present. A finite surface source density corresponds to an infinite volume source density. The infinite volume source density is avoided by considering a volume of thickness δ at the surface. Since this will enclose an amount of source equal to $\delta(\rho-\mathbf{J})$ per unit area of the surface, $\delta(\rho-\mathbf{J})$ is the same as the surface source density $\sigma-\mathbf{K}$ so that we arrive at the neat result

$$\begin{aligned}\mathbf{n}\Delta\boldsymbol{F} &= \lim_{\delta\to 0}\left(\delta\cdot\frac{\sigma-\mathbf{K}}{\delta}\right)\\ &= \sigma-\mathbf{K}\\ &= \boldsymbol{K}\end{aligned} \tag{5.80}$$

which is none other than the final form of Equation (5.72). It has therefore been shown not only how the boundary conditions for the electromagnetic field at a plane interface follow directly from Maxwell's equation, as indeed they must, but also that geometric algebra casts these in an extremely simple form. Moreover, it has the versatility to blend seamlessly with other mathematical concepts such as scalar and vector difference operators.

Returning to Equations (5.61) and (5.65), which represent the alternative ways of including the electromagnetic polarization $\boldsymbol{Q}$ in Maxwell's equation, we could recover the boundary conditions for the auxiliary electromagnetic field $\mathbf{G}$, that is to say, $\mathbf{D}+\boldsymbol{H}$. But prior to doing so, it is instructive to try to find the bound sources that result from the discontinuous change in polarization $\Delta\boldsymbol{Q}$ that arises at an interface between different media. From Equation (5.78), we find that $\boldsymbol{\nabla Q}=(\mathbf{n}\Delta\boldsymbol{Q})/\delta$ at the interface, so that following the usual rules for vector multiplication we may write $\boldsymbol{\nabla}\cdot\boldsymbol{Q}=\Delta(\mathbf{n}\cdot\boldsymbol{Q})/\delta$ and $\boldsymbol{\nabla}\wedge\boldsymbol{Q}=\Delta(\mathbf{n}\wedge\boldsymbol{Q})/\delta$. Using Equation (5.61), we find

$$\begin{aligned}(\boldsymbol{\nabla}+\partial_t)\boldsymbol{F} &= \boldsymbol{J}_{\text{free}}-\boldsymbol{\nabla}\cdot\boldsymbol{Q}-\partial_t\mathbf{P}\\ \Leftrightarrow \qquad \mathbf{n}\Delta\boldsymbol{F} &= \boldsymbol{K}_{\text{free}}-\Delta(\mathbf{n}\cdot\boldsymbol{Q})\end{aligned} \tag{5.81}$$

The factor $1/\delta$ was eliminated from both sides of the equation by recalling from above that $\delta\boldsymbol{J}$ can be replaced with $\boldsymbol{K}$, and this applies equally well to the free

current as to the total current. But since $\mathbf{n}\Delta\boldsymbol{F} = \boldsymbol{K}$ where $\boldsymbol{K}$ is the *total* surface source density, $\Delta(-\mathbf{n}\cdot\boldsymbol{Q})$ must represent the bound surface source density, $\boldsymbol{K}_{\text{bound}}$ so that

$$\boldsymbol{K}_{\text{bound}} = -\mathbf{n}\cdot\Delta\boldsymbol{Q} \tag{5.82}$$

If however, we use Equation (5.65) as a starting point,

$$\begin{aligned}
&(\nabla+\partial_t)\boldsymbol{G} = \boldsymbol{J}_{\text{free}} + \boldsymbol{\nabla}\wedge\boldsymbol{Q} - \partial_t\boldsymbol{M} \\
\Leftrightarrow\quad &\Delta\boldsymbol{G} = \mathbf{n}\boldsymbol{K}_{\text{free}} + \mathbf{n}(\mathbf{n}\wedge\Delta\boldsymbol{Q}) \\
\Leftrightarrow\quad &\Delta\boldsymbol{G} = \mathbf{n}\boldsymbol{K}_{\text{free}} + \mathbf{n}(\mathbf{n}\wedge\Delta\mathbf{P} - \mathbf{n}\wedge\Delta\boldsymbol{M}) \\
\Leftrightarrow\ &\Delta\mathbf{D} + \Delta\boldsymbol{H} = \underbrace{\sigma_{\text{free}}\mathbf{n} + \Delta(\mathbf{n}\cdot(\mathbf{n}\wedge\mathbf{P}))}_{\text{vector}} - \underbrace{\mathbf{n}\wedge\mathbf{K}_{\text{free}} + I\Delta((\mathbf{n}\cdot\mathbf{M})\mathbf{n})}_{\text{bivector}}
\end{aligned} \tag{5.83}$$

Now, it is evident that $\Delta(\mathbf{n}\cdot(\mathbf{n}\wedge\mathbf{P}))$ simplifies to $\Delta\mathbf{P}^{//}$, whereas $\Delta((\mathbf{n}\cdot\mathbf{M})\mathbf{n})$ simplifies to $\Delta\mathbf{M}^{\perp}$ where $\mathbf{P}^{//}$ refers to the part of $\mathbf{P}$ that is parallel to $\mathbf{n}$ while $\mathbf{M}^{\perp}$ refers to the part of $\mathbf{M}$ that is perpendicular to $\mathbf{n}$ (we have avoided using the bivector $\boldsymbol{M}$ here so that these meanings are clear, e.g., $\boldsymbol{M}^{//}$ is parallel to the tangent plane rather than $\mathbf{n}$). Substituting these and separating out the vector and bivector parts of the field, we find

$$\begin{aligned}
\Delta\mathbf{D} &= \sigma_{\text{free}}\mathbf{n} + \Delta\mathbf{P}^{//} \\
\Delta\mathbf{H} &= -\mathbf{n}\times\mathbf{K}_{\text{free}} - \Delta\mathbf{M}^{\perp}
\end{aligned} \tag{5.84}$$

If we were to identify the magnetic field with $\mathbf{H}$, $-\Delta\mathbf{M}^{\perp}$ would be equivalent to a surface density of poles. This is directly analogous to the situation in Equations (5.67) (ii) and (5.68) where it was shown that $-\nabla\cdot\mathbf{M}$ would be equivalent to a volume density of poles. Rarely, if ever, has anyone thought to propose the analogous thing for $\mathbf{D}$ by treating $\Delta\mathbf{P}^{//}$ as the equivalent of a surface current density! Equation (5.84) also dispels another fallacy that $\mathbf{D}$ and $\boldsymbol{H}$ depend only on the free sources alone, for that is clearly not the case. Maxwell's macroscopic equations conceal how both of these fields depend on bound charge. However, in the alternative form of Maxwell's equation based on $\boldsymbol{G}$ rather than $\boldsymbol{F}$, namely Equation (5.65), the source terms clearly includes the bound charge density in the form of $(\boldsymbol{\nabla}+\partial_t)\wedge\boldsymbol{Q}$.

5.10 EXERCISES

1. Restore the necessary suppressed constants in the following simplified expressions and equations for your preferred system of units: $\mathbf{E}+I\mathbf{B}$, $1-\mathbf{v}$, $1-\mathbf{v}^2$, $\rho-\mathbf{J}$, $\partial_t\rho+\boldsymbol{\nabla}\cdot\mathbf{J}$, $\partial_t\rho+\boldsymbol{\nabla}\cdot\mathbf{J}=0$, and, assuming energy density is required, $\frac{1}{2}\mathbf{E}^2$ and $\frac{1}{2}\mathbf{B}_0^2$. Why is care needed in this last case?
2. If $\boldsymbol{M}$ is the magnetization of a uniformly magnetized sphere suspended in free space, then what is the equivalent net distribution of the magnetic current sources? What would be the corresponding distribution of magnetic poles?

3. A given isotropic medium has weak electric and magnetic polarizabilities α and β such that $\mathbf{P} - \boldsymbol{M} = \alpha\mathbf{E} - \beta\boldsymbol{B}$.

(a) Does it continue to make sense to define $\boldsymbol{F}$ as being $\mathbf{E} + [\mathrm{c}]\boldsymbol{B}$ where c is the speed of light in free space?

(b) Work out in detail the derivation of Equation (5.58).

(c) Derive an equation for the propagation of plane electromagnetic waves in the medium.

4. While it is true that $\boldsymbol{F}\boldsymbol{F}^\dagger = \left\langle \boldsymbol{F}\boldsymbol{F}^\dagger \right\rangle_0 + \left\langle \boldsymbol{F}\boldsymbol{F}^\dagger \right\rangle_1$, show that

(a) the grade 1 part does not correspond to $\boldsymbol{F} \wedge \boldsymbol{F}^\dagger$ and

(b) both parts actually derive from $\boldsymbol{F} \cdot \boldsymbol{F}^\dagger$.

5. **(a)** Show that the Lorentz force, $\mathbf{f} = q(\mathbf{E} + \boldsymbol{B} \cdot \mathbf{v})$, may be written as $\mathbf{f} = q\langle \boldsymbol{F}(1 + \mathbf{v}/[c]) \rangle_1$.

(b) Show that in the case of a multivector electromagnetic source density $\boldsymbol{J}$ as defined in Equation (3.3), the Lorentz force density $\mathcal{F}$ that acts on it is given by $\mathcal{F} = \varepsilon_0 \langle \boldsymbol{J}\boldsymbol{F} \rangle_1$.

6. Show that the boundary condition at an interface $\Delta\boldsymbol{F} = \mathbf{n}\boldsymbol{K}$ is fully equivalent to all of the separate traditional conditions:

$$\begin{aligned}
\mathbf{E}_2^\perp - \mathbf{E}_1^\perp &= \mathbf{n}\sigma \\
\mathbf{E}_2^{/\!/} - \mathbf{E}_1^{/\!/} &= 0 \\
\mathbf{B}_2^\perp - \mathbf{B}_1^\perp &= 0 \\
\mathbf{B}_2^{/\!/} - \mathbf{B}_1^{/\!/} &= -\mathbf{n} \times \mathbf{K}
\end{aligned}$$

7. Find the boundary conditions at an interface between two media characterized by $\mathbf{D}_1 = \varepsilon_1\mathbf{E}_1$; $\boldsymbol{B}_1 = \mu_1\boldsymbol{H}_1$ on the one side and $\mathbf{D}_2 = \varepsilon_2\mathbf{E}_2$; $\boldsymbol{B}_2 = \mu_2\boldsymbol{H}_2$ on the other.

8. Two circularly polarised plane waves, F_+ and F_-, are identical apart from their wave vectors being $\mathbf{k}_+ = \mathbf{m} + \mathbf{n}$ and $\mathbf{k}_- = \mathbf{m} - \mathbf{n}$ where $\mathbf{m} \perp \mathbf{n}$.

(a) Write down suitable formulae for $\boldsymbol{F}_+$ and $\boldsymbol{F}_-$ on the assumption that the electric fields at some suitable reference point are both equal to $\mathbf{E}_0 e^{I\omega t}$.

(b) Discuss the variation of the combined field $\boldsymbol{F}_+ + \boldsymbol{F}_-$ in both the $\mathbf{m}$ and $\mathbf{n}$ directions.

(c) Compare this with the field given by $\boldsymbol{F}_+ - \boldsymbol{F}_-$.

Chapter 6

Review of (3+1)D

Until now, we have presented the foundations of a geometric algebra using the case of 3D to illustrate its main principles and its application to electromagnetic theory.

We have seen the ability of geometric algebra to represent objects of different grades as a single multivector with the effect that most of the fundamental equations are encoded in a very compact form. Conversely, the grade structure is the instrument by which we may "unpack" these multivector equations into a more traditional form, a method that has by now been routinely exploited.

As to the practical application to electromagnetic theory:

- Scalar time t and vector position $\mathbf{r}$ may be combined as a multivector $\boldsymbol{R} = t + \mathbf{r}$ that defines an event in terms of both its position and time.
- Likewise, the time derivative and vector derivative combine to form the multivector operator $\boldsymbol{\nabla} + \partial_t$.
- Accordingly, we have referred to the present description in terms of a 3D geometric algebra as being "(3+1)D."
- Charge and current densities ρ and $\mathbf{J}$ combine into the multivector electromagnetic source density $\boldsymbol{J} = \rho - \mathbf{J}$.
- The electric dipole moment $\mathbf{p}$, polarization $\mathbf{P}$ and electric field $\mathbf{E}$ are fundamentally vector in character.
- The magnetic dipole moment $\boldsymbol{m}$, magnetization $\boldsymbol{M}$ and magnetic field $\boldsymbol{B}$ are fundamentally bivector rather than vector in character.
- The electric and magnetic fields $\mathbf{E}$ and $\boldsymbol{B}$ combine to form a multivector electromagnetic field $\boldsymbol{F} = \mathbf{E} + \boldsymbol{B}$.
- The scalar and vector potentials unite into a multivector potential of the form $\boldsymbol{A} = -\Phi + \mathbf{A}$ obeying the Lorenz condition $\boldsymbol{\nabla} \cdot \mathbf{A} = -\partial_t \Phi$.
- $\boldsymbol{A}$ obeys the simple inhomogeneous scalar wave equation $\left(\boldsymbol{\nabla}^2 - \partial_t^2\right)\boldsymbol{A} = \boldsymbol{J}$.
- Electrostatic and magnetostatic problems can be solved by the same equation simply by replacing the scalar charge density ρ with the multivector

Understanding Geometric Algebra for Electromagnetic Theory, First Edition. John W. Arthur.

electromagnetic source density $\boldsymbol{J}$. The result is the multivector $\boldsymbol{F}$ that gives both fields together.

- Maxwell's equations in free space reduce to $(\boldsymbol{\nabla} + \partial_t)\boldsymbol{F} = \boldsymbol{J}$, just a single equation in a single field quantity $\boldsymbol{F}$ and a single source density $\boldsymbol{J}$.
- In a medium, the multivector $\boldsymbol{Q} = \mathbf{P} - \boldsymbol{M}$ represents the combined electric and magnetic polarizations.
- Maxwell's equation then becomes a pair of equations, $\langle(\boldsymbol{\nabla} + \partial_t)\boldsymbol{G}\rangle_{0,1} = \boldsymbol{J}_{\text{free}}$ and $\langle(\boldsymbol{\nabla} + \partial_t)\boldsymbol{F}\rangle_{2,3} = 0$, in which $\boldsymbol{G} = \mathbf{D} + \boldsymbol{H}$ represents the auxiliary electromagnetic field.
- The constitutive relations are now simply $\boldsymbol{G} = \boldsymbol{F} + \boldsymbol{Q}$.
- Plane wave solutions to Maxwell's equation are naturally left or right circularly polarized.
- The quantity $\frac{1}{2}\boldsymbol{F}\boldsymbol{F}^\dagger$ yields $\mathcal{E} + \mathbf{g}$, a multivector that may be interpreted as the densities of electromagnetic energy and momentum in a plane wave.
- On the other hand, $\frac{1}{2}\boldsymbol{F}^\dagger\boldsymbol{F}$ results in the form $\mathcal{E} - \mathbf{S}$ which obeys the continuity equation $\partial_t\mathcal{E} + \boldsymbol{\nabla}\cdot\mathbf{S} = \partial_t\mathcal{U}$, justifying the standard interpretation of $\mathbf{S}$ as a flow of electromagnetic energy.
- The four boundary conditions for the electromagnetic field at an interface are expressible simply as $\mathbf{n}\Delta\boldsymbol{F} = \boldsymbol{K}$.
- The key equations are summarized in Table 6.1.

But there also seem to be some exceptions to the neat way that the (3+1)D model deals with the electromagnetic equations in general:

- A simple algebraic equation for the Lorentz force cannot be found using the unified field multivector $\boldsymbol{F}$. This can only be achieved using the grade selection operator, as in $\mathbf{f} = q\langle\boldsymbol{F}(1+\mathbf{v})\rangle_1$, otherwise the equation still involves $\mathbf{E}$ and $\boldsymbol{B}$ separately.
- A similar situation arises with the pair of Maxwell's equations in a polarizable media.
- We do not get a formal *time-dependent* solution to Maxwell's equation from $\boldsymbol{F} = \boldsymbol{\nabla}^{-1}\boldsymbol{J}$.

Useful though this (3+1)D regime may be, it seems that we come back to the issue that must have exercised Maxwell when he started looking at quaternions as a better way of setting down his equations—are we using the best toolset, that is to say, is this the best way of encoding the underlying physics into meaningful equations? Evidently, for us at least, the answer is nearly, but not quite. Notwithstanding the fact that we can combine ρ and $\mathbf{J}$ as a single multivector, the (3+1)D approach is essentially Newtonian in that stationary charge and moving charge are treated as independent types of sources, the one for $\mathbf{E}$ and the other for a different type of field $\boldsymbol{B}$. Maxwell's equation results in a wave equation, whereas Coulomb's law,

Table 6.1 Summary of the Key Results of (3+1)D Electromagnetic Theory

Lorentz force	$f = q(\mathbf{E} + \boldsymbol{B} \cdot \mathbf{v})$
Electromagnetic field	$\boldsymbol{F} = \mathbf{E} + \boldsymbol{B} = \mathbf{E} + I\mathbf{B}$
Electromagnetic polarization	$\boldsymbol{Q} = \mathbf{P} - \boldsymbol{M}$
Auxiliary electromagnetic field	$\boldsymbol{G} = \boldsymbol{F} + \boldsymbol{Q}$ $= \mathbf{D} + \boldsymbol{H} = \mathbf{D} + I\mathbf{H}$
Electromagnetic source density	$\left.\begin{aligned}\boldsymbol{J} &= \rho - \mathbf{J}\\ \boldsymbol{J}_{\text{bound}} &= -\boldsymbol{\nabla} \cdot \boldsymbol{Q}\end{aligned}\right\}$(Bulk) $\left.\begin{aligned}\boldsymbol{K} &= \sigma - \mathbf{K}\\ \boldsymbol{K}_{\text{bound}} &= -\mathbf{n} \cdot \Delta\boldsymbol{Q}\end{aligned}\right\}$(Surface)
Maxwell's equation in free space	$(\boldsymbol{\nabla} + \partial_t)\boldsymbol{F} = \boldsymbol{J}$
Maxwell's equation in a medium $\left[c' = (\varepsilon\mu)^{-\frac{1}{2}} \text{ and } Z' = (\mu/\varepsilon)^{\frac{1}{2}}\right]$	$(\boldsymbol{\nabla} + \partial_t)\boldsymbol{F} = \boldsymbol{J}_{\text{free}} - \boldsymbol{\nabla} \cdot \boldsymbol{Q}$ $\left(\boldsymbol{\nabla} + \frac{1}{c'}\partial_t\right)(\mathbf{E} + c'\boldsymbol{B}) = \frac{\rho_{\text{free}}}{\varepsilon} - Z'\mathbf{J}_{\text{free}}$
Electromagnetic field for an arbitrary distribution of quasistatic sources	$\boldsymbol{F}(\mathbf{r}) = \frac{1}{4\pi}\int_V \frac{(\mathbf{r} - \mathbf{r}')}{\lvert\mathbf{r} - \mathbf{r}'\rvert^3}\boldsymbol{J}(\mathbf{r}')dr'^3$
Electromagnetic field for circularly polarised plane waves	$\boldsymbol{F} = \boldsymbol{F}_0 e^{\pm I(\omega t - \mathbf{k}\cdot\mathbf{r})}$
Wave equation and conservation of charge	$\left(\nabla^2 - \partial_t^2\right)\boldsymbol{F} = (\boldsymbol{\nabla} - \partial_t)\boldsymbol{J}$
Multivector potential with Lorenz condition	$\boldsymbol{A} = -\Phi + \mathbf{A}$ $\boldsymbol{F} = (\boldsymbol{\nabla} - \partial_t)\boldsymbol{A}$ $\left(\nabla^2 - \partial_t^2\right)\boldsymbol{A} = \boldsymbol{J}$
Electromagnetic energy and momentum	$\frac{1}{2}\boldsymbol{F}\boldsymbol{F}^\dagger = \underbrace{\frac{1}{2}\left(\mathbf{E}_0^2 + \mathbf{B}_0^2\right)}_{\mathcal{E}} + \underbrace{(\mathbf{E}_0 \times \mathbf{B}_0)}_{\mathbf{g}}$
Boundary conditions at a smooth interface	$\Delta\boldsymbol{F} = \mathbf{n}\boldsymbol{K}$ $\mathbf{n} \cdot \Delta\boldsymbol{G} = \boldsymbol{K}_{\text{free}}$ $-\mathbf{n} \cdot \Delta\boldsymbol{Q} = \boldsymbol{K}_{\text{bound}}$

Note that for clarity, these equations are in terms of the modified variables referred to in Table 5.1. The same table may be used to recover the standard form.

in spite of being at the root of electromagnetic theory, appears to give little hint of this. The (3+1)D picture is therefore not the full story. It gives us a phenomenological view as opposed to a proper fundamental view. We will now go on to see how this shortfall can be overcome simply by starting from Coulomb's law and following a proper treatment. In fact, all electromagnetic phenomena can be understood in terms of this very simple concept. Spacetime geometric algebra, therefore, will be seen to provide better encoding and consequently makes the better toolset.

Geometric algebra does not always lend itself to making everything top level and simple, however. There are situations in which things can only be evaluated in terms of summations over products involving indexed components and basis elements in a manner very similar to tensor analysis. While geometric algebra is very adept at expressing reflections and rotations by means of bivectors (see, for example, Section 9.3 and References 6; 27, section 2.7 and 28, section 2.7) as will be discussed in Section 9.9, the simple matter of coping with a dilation turns out to be somewhat less elegant. Matos et al. [41] tackled this problem for the case of electromagnetic waves in an anisotropic medium. However impressive geometric algebra might appear in the light of the present discussion, it should not be seen as being detached from the rest of linear algebra over vector spaces, and it is by no means to be regarded as a complete toolset that fits all purposes with equal facility.

Before we move on to spacetime, however, let us digress a little to spare a thought for the reader who has come this far but feels that it may be too late to change from the traditional methods of vector analysis and learn a new system. While it is the purpose of this book to recommend the reader wholeheartedly to geometric algebra, it is nevertheless necessary to be sympathetic to the practicalities of such a major change, and so we propose that if the reader is in any difficulties of this sort, they should start out with the following "GA lite" regime:

- Concentrate purely on the familiar entities—scalars and vectors.
- Use the unit pseudoscalar I as though it were the same as the imaginary unit j.
- Work with multivector expressions of the form $\boldsymbol{U} = a + \mathrm{j}b + \mathbf{u} + I\mathbf{v}$.
- When a bivector form such as $I\mathbf{v}$ occurs, think of it as being an "imaginary vector".
- Expand the product $\mathbf{uv}$ as $\mathbf{u}\cdot\mathbf{v} + I\mathbf{u}\times\mathbf{v}$ rather than $\mathbf{u}\cdot\mathbf{v} + \mathbf{u}\wedge\mathbf{v}$.
- But always commute I to the left of any such product before expanding it; for example, $\mathbf{u}(I\mathbf{v}) = I\mathbf{uv} = I\mathbf{u}\cdot\mathbf{v} - \mathbf{u}\times\mathbf{v}$.
- The last two rules also apply to $\boldsymbol{\nabla}\mathbf{u}$ and $\boldsymbol{\nabla}(I\mathbf{u})$.

Most of the material at the level of Chapter 5 can be managed with this scheme, which is effectively an algebra of complex scalars and vectors. The key differences are the absence of true bivectors and the awkwardness in dealing with expressions such as $\boldsymbol{\nabla}(\mathbf{uv})$. As an example of how it works, Maxwell's equation $(\boldsymbol{\nabla} + \partial_t)\boldsymbol{F} = \boldsymbol{J}$ is dealt with by treating $\boldsymbol{F}$ as $\mathbf{E} + I\mathbf{B}$ upon which it may be manipulated as follows:

$$
\begin{aligned}
&(\nabla + \partial_t)(\mathbf{E} + I\mathbf{B}) = \rho - \mathbf{J} \\
\Leftrightarrow\ &\nabla\mathbf{E} + I\nabla\mathbf{B} + \partial_t(\mathbf{E} + I\mathbf{B}) = \rho - \mathbf{J} \\
\Leftrightarrow\ &\nabla\cdot\mathbf{E} + I\nabla\times\mathbf{E} + I(\nabla\cdot\mathbf{B} + I\nabla\times\mathbf{B}) + \partial_t(\mathbf{E} + I\mathbf{B}) = \rho - \mathbf{J} \\
\Leftrightarrow\ &\nabla\cdot\mathbf{E} + I\nabla\cdot\mathbf{B} - (\nabla\times\mathbf{B} - \partial_t\mathbf{E}) + I(\nabla\times\mathbf{E} + \partial_t\mathbf{B}) = \rho - \mathbf{J}
\end{aligned}
$$

By correlating real and "imaginary" scalar and vector terms on each side of the equation, we find

$$
(\nabla + \partial_t)\boldsymbol{F} = \boldsymbol{J} \Leftrightarrow \begin{cases} \nabla\cdot\mathbf{E} = \rho \\ I\nabla\cdot\mathbf{B} = 0 \\ -(\nabla\times\mathbf{B} - \partial_t\mathbf{E}) = -\mathbf{J} \\ I(\nabla\times\mathbf{E} + \partial_t\mathbf{B}) = 0 \end{cases}
$$

which has taken us back to Maxwell's microscopic equations in their usual form, albeit using simplified variables for clarity. Obviously, the complex field vectors and potentials mentioned by Stratton fit directly in with this scheme.

It is to be hoped that by experimenting with this cut-down version, the reader who has yet to be persuaded will find it less of a leap in abstraction. Then in time, by becoming increasingly familiar with it, they may be more ready to embrace the use of bivectors and advance to the 3D geometric algebra proper.

Chapter 7

Introducing Spacetime

This book opens with a quotation from James Clerk Maxwell: "it is a good thing to have two ways of looking at a subject . . .", an observation he made in commenting on two separate electromagnetic theories, one due to Faraday and the other to Weber. In broad terms, Faraday's theory was a model based on the concept of lines of force, while Weber had actually gone as far as to postulate that there were two types of electric charges, one positive and the other negative, and that the force law governing their interaction depended on velocity as well as distance. To Maxwell's mind, this was a more fundamental theory, yet he had objections concerning the force law because it did not fit in with Newtonian mechanics and went on to say: "There are [however] objections to making any ultimate forces in nature depend on the velocity of the bodies between which they act." At that time there was no complete theory of electromagnetism, and so both theories were not necessarily contrary to each other, rather two different ways of looking at the same thing. If there was a hint of special relativity in Weber's idea, it had already been there more or less a quarter of a century earlier in Ampere's law for the force between current elements. In essence, Newton's third law is broken by moving charges [2, pp. 31–32], and it would be another half century after Maxwell's musings on the subject before there would be yet another "way of looking" at it. Einstein called this a "new manner of expression" [42, p. 54]. Until then, in the "old manner of expression," electricity was due to charges and magnetism was due to currents, whereas in his new way of looking at it, there was only electricity. In short, a charged particle interacts only with the electric field that it sees in its own rest frame. Therein lay the answer to Maxwell's objection to a velocity-dependent noncentral force.

In everyday life, we ignore special relativity even though we encounter it every time we touch on something that involves any form of magnetism or electromagnetic effect: magnetic compass, ceramic fridge magnet, electric motor, generator, AC transformer, ignition coil, metal detector, CRT screen, electron microscope, security tag, magnetic storage disk drive, mobile phone, and finally MRI scanner, to name but a handful. Very few of the pioneers of these things, some of which involve an amazing amount of ingenuity and technical know-how, had any reason at all to take note of special relativity when they went about their business. It is necessary to

Understanding Geometric Algebra for Electromagnetic Theory, First Edition. John W. Arthur.

resort to special relativity only in very special circumstances; for the vast majority of situations, we have what appears to be a perfectly adequate phenomenological theory in the likes of Maxwell's equations, the constitutive relations and the Lorentz force. On the other hand, if we want to know how this phenomenological theory comes about, we must inevitably turn to special relativity. There are even "two ways of looking at" special relativity. Rather than starting from the Lorentz transformation itself, Minkowski's spacetime [42] provides an alternative, comprehensive way of dealing with the mathematics of relativity by incorporating a special metric from which the Lorentz transformation follows naturally as a type of orthogonal transformation that is inherently different from any combination of spatial reflections and rotations. Einstein himself took some time to concede that this was equivalent to his own way of looking at it. As we shall see, it works extremely well with geometric algebra as the toolset.

It is not necessary, however, to understand special relativity in order to make use of the spacetime geometric algebra. Spacetime geometric algebra may simply be used as an operational tool that fills in some of the gaps in the (3+1)D approach by treating time as a vector. Nevertheless, the basics of special relativity are inherent in the spacetime geometric algebra and so in Chapters 7–8 we cover only the essentials, whereas in Chapters 9–10 we develop the ideas a little further so that the interested reader can have some account of the physical as well as the mathematical background that underpins it. On a first reading, this may be skimmed, or even skipped, and simply treated as supplementary reference material. Nevertheless, it is hoped that most readers will eventually be sufficiently encouraged to return to it out of pure interest. Because of the comparatively simple way in which the basic concepts work in practice, it may be found that this material is not only rewarding in its own right but much less difficult to get to grips with than is often feared.

7.1 BACKGROUND AND KEY CONCEPTS

One of the key concepts of spacetime is that time is no longer treated as a separate scalar entity; rather, it is incorporated as an independent vector that adds an extra dimension to our worldly space. The Newtonian world appears to be (3+1)D because we deal with space and time separately. Our view of the world corresponds to a 3D space punctuated here and there with matter. Evolution in time is taken to be a continuous sequence of such views, each corresponding to an increment in time, just as in the frames of a never-ending movie, with each view having a slightly different arrangement of matter. Time itself is simply the scalar that parameterizes this unidirectional evolution.[1] But how then do we deal with time as a vector? A major problem we face in engaging with a 4D world is trying to envisage all four dimensions at the same time. But on the other hand, no one finds difficulty in thinking about three dimensions when we can see only two displayed on a page or screen. We intuitively suppress one 3D dimension when we make a drawing and, equally intuitively, we mentally recreate the lost dimension when we view it. Although these

[1] Einstein is reputed to have said that time exists only to stop everything from happening at once.

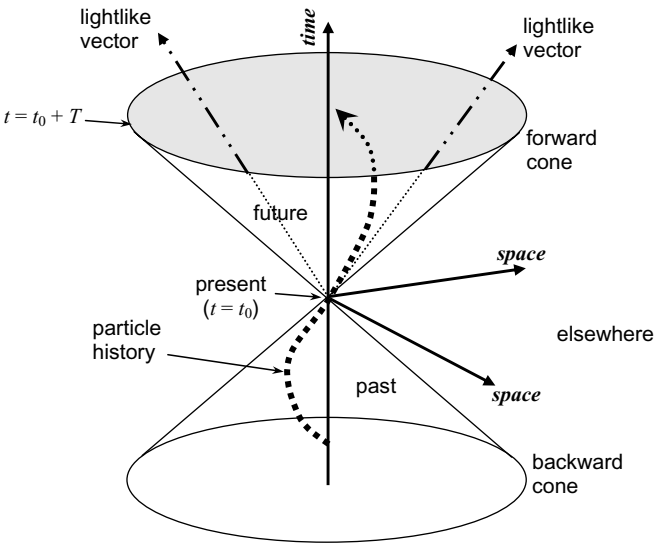

Figure 7.1 View of spacetime portrayed in 3D by suppressing one spatial dimension. Although we are restricted to using only two dimensions on the page, the brain is used to interpreting 2D forms as though they were actual 3D forms in space. The dotted curve represents the trajectory of a particle in spacetime—also called its history or world line. The particle appears to be spiraling as it travels upwards along time while looping in a spatial plane. If it emits a brief flash of light at $t=t_0$, the light will travel along every straight-line path in the surface of the upper cone. We see its history as well as its future, both of which must lie strictly within the double cone. Events that cannot cause an effect at the particle's present location are outside the cone are labeled as being "elsewhere". Representations such as this are commonly used in relativity theory, but they may appear more familiar if we imagine them as in Figure 7.2 where we are looking down the time axis from above to reveal the motion of the particle in the spatial plane alone. It is clear that the full spacetime view gives us much more information.

processes are intuitive, we do understand how it all works through projections and cues such as apparent perspective. And so it must be when we think in spacetime, as in Figure 7.1 which represents the trajectory of an orbiting particle. Here time and two spatial directions are represented on the 2D plane of the page. This is as much as our senses can apparently cope with and the third spatial dimension is simply lost. The equivalent diagram for the (3+1)D world is shown for comparison in Figure 7.2. The two spatial directions are properly represented, but the passage of time can only be suggested, for example, by displaying the locus of the point as a dotted line, which we can visualize as an evolving trajectory by following the sequence of dashes.

Paraphrasing slightly from Einstein's original text [42, pp. 37–38], the two key principles of special relativity are the following:

- The laws of electrodynamics and optics will be valid for all frames of reference, which are related by a rotation, a translation, or by uniform motion.
- Light is always propagated in empty space with the specific velocity c irrespective of the state of motion of either the emitter or the observer.

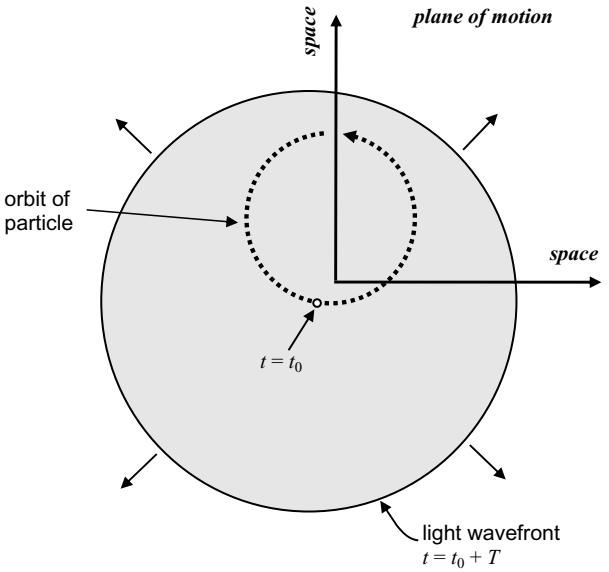

Figure 7.2 Conventional diagram with no time axis. We imagine the configuration of Figure 7.1 being redrawn so that time is now directed out of the page and we see only a spatial plane. The history of the particle has to be inferred by imagining it traveling around from dash to dash around the loop shown, or by using some form of real-time animation. The locus of the expanding beam front of Figure 7.1 is shown as the shaded disk with $t = t_0$ at the center and some future time at the boundary.

The first principle is just the extension of familiar Galilean or Newtonian relativity to include all of electromagnetic theory. It appears almost axiomatic that two situations related by such simple symmetries must be equivalent, and in fact, we can treat them just as different views of the same thing. Uniform motion is included because it has never been found to cause any difference to any other physical law. Previous aether-based theories differed on this point, and, in particular, they emphatically disagreed with the second principle, the invariance of the speed of light in free space.

Despite the simplicity of these underlying principles, the subject of special relativity is generally considered to be intellectually challenging. Although it underpins much of modern physics, including electromagnetic theory, for most of us, the very mention of the subject usually produces an adverse reaction. This may well stem from introductory undergraduate courses that dwell on explaining the many counterintuitive conclusions associated with the measurement of length and time, not to mention mass and force, in different inertial reference frames. Moreover, the process of transforming measurements between one inertial reference frame and another can be cumbersome and also depends on the sort of quantity involved. The approach usually runs out of steam when it comes to noninertial (accelerating) reference frames.

Spacetime, on the other hand, actually simplifies working with special relativity through a much more systematic approach to measurement. Although two key subtleties are involved, once the underlying rationale of spacetime is accepted there should be little further cause for concern. In any case, spacetime can also be used without any regard to special relativity when all we really want to do is to take advantage of the 4D approach to a physical problem. It will then be of little importance to us that the results obtained turn out to be consistent with relativity. But as far as the fundamental physical applications of the spacetime geometric algebra are concerned, it is more than just a case of convenience or expediency—it provides a solution to the gaps left by a (3+1)D approach. In the case of electromagnetic theory, the Lorentz force is expressed in terms of the electromagnetic field *as a whole* rather than as the sum of separate contributions from the electric and magnetic fields. A similar sort of improvement applies to Maxwell's equations in an electrically or magnetically polarizable medium. Furthermore, a vector derivative that now includes time provides an even neater way of encoding the basic equations. This last point does not just refer to the appearance of the equations, for in the case of Maxwell's equation, it reveals that the time-dependent solutions must belong to a particular class of function *that is a direct extension of the analytic functions from* 2*D into* 4*D*!

The two subtleties of spacetime that were alluded to above are simply things that we will have to become familiar with. The first is the distinctly non-Euclidean nature of the spacetime metric, or measure of "distance". The second is the need for some method of mapping between the 4D world of spacetime and the (3+1)D space that represents what we observe in the everyday world.

Let us now briefly consider the origin of spacetime. The concept was the brainchild of Herman Minkowski who presented it in a lecture entitled "Raum und Zeit" [42, pp. 75–96; 43], just a few years after Einstein published his theory of special relativity and, unfortunately, only a few months before his own untimely death. However, "raum und zeit" literally means space and time, whereas the concatenated forms space-time and spacetime originated much later, presumably as a result of the concept becoming universally adopted. Declaring boldly in his introduction "[the idea of] space by itself and time by itself are doomed to fade away . . .", he introduced it as a mathematical concept that would bring clarity and insight to the new physics of special relativity. In one sense, this would seem to be just a manipulation of the old (3+1)D system into a 4D space where time has been set on an equivalent vector footing to the spatial vectors and a mixed metric has been introduced so as to accommodate the needs of the physical model. But Minkowski also realized that it meant that it was possible to deal with relativity without having to constantly transform from one reference frame to another. He described this idea as "absolute spacetime", which does not mean that he was proposing that spacetime would provide some absolute frame of reference, a concept that is in fact diametrically opposed to special relativity; indeed, it was quite the opposite. Absolute spacetime simply means that simple spacetime vectors are invariant, that is to say they do not depend on any reference frame. Rather, it is the imposition of a reference frame that removes the property of invariance as a result of reducing these vectors *to terms that*

are relative to the chosen frame. This is a central point because, with few exceptions, (3+1)D vectors turn out to be essentially *frame-dependent* quantities.

Spacetime, however, does not require geometric algebra to be its mathematical framework. In the traditional mathematical approach, vectors have been commonly represented in component form and referred to as "four-vectors" (not to be confused with the 4-vector of geometric algebra, rather these are simply vectors with four components, three for space and one for time). The use of components implies the adoption of a specific set of basis vectors and generally leads to tensor analysis as the means of representing and manipulating the physical equations, whereas with geometric algebra, we can often represent equations in simple algebraic form without even mentioning basis vectors. Although we do not hesitate to use basis vectors as an expedient, on many occasions it is useful to avoid doing so, particularly when we want to achieve results that are as general as possible. This is often referred to as the "coordinate-free" or "basis-free" approach.

7.2 TIME AS A VECTOR

Time is now to be treated as a vector on an equal footing with space. We will use $\boldsymbol{t},\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}$ to represent our default choice of orthonormal basis vectors, distinguishing them from their (3+1)D counterparts, the scalar 1 and the vectors $\mathbf{x},\mathbf{y},\mathbf{z}$, by using bold italics throughout. We can refer to a set of basis vectors as a frame, and in particular, we will refer to this generic choice as the $\boldsymbol{t}$-frame. The vector $\boldsymbol{R} = x\boldsymbol{x} + y\boldsymbol{y} + z\boldsymbol{z} + [c]t\boldsymbol{t}$ that specifies a location $x\boldsymbol{x} + y\boldsymbol{y} + z\boldsymbol{z}$ at a given time t now defines an *event*. Here the constant $[c]$ ensures that all the dimensions are compatible, but according to our practice, it will now be suppressed by absorbing it within the variable t so that the symbol t now replaces the customary ct. As to how time should be included in the metric, it is clear that the simple Euclidean form that holds for 3D will not do. Given any two events $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ that take place at the same value of t, the separation vector $\boldsymbol{R}_{12} = \boldsymbol{R}_2 - \boldsymbol{R}_1$ is independent of t and should have the same measure as in 3D; that is, $\boldsymbol{R}_{12}^2 = (x_2 - x_1)^2\boldsymbol{x}^2 + (y_2 - y_1)^2\boldsymbol{y}^2 + (z_2 - z_1)^2\boldsymbol{z}^2$. Likewise, if we consider the two events as taking place at the same spatial coordinates, then it follows that $\boldsymbol{R}_{12} = (t_2 - t_1)\boldsymbol{t}$. In this case, it would therefore seem logical to have $\boldsymbol{R}_{12}^2 = (t_2 - t_1)^2\boldsymbol{t}^2$. Taking these facts together, the only possible simple metric appears to be $\boldsymbol{R}^2 = x^2\boldsymbol{x}^2 + y^2\boldsymbol{y}^2 + z^2\boldsymbol{z}^2 + t^2\boldsymbol{t}^2$, which on the face of it is still Euclidean. But this depends on the implicit assumption that normalization means $\boldsymbol{x}^2 = \boldsymbol{y}^2 = \boldsymbol{z}^2 = \boldsymbol{t}^2 = 1$. To take best advantage of geometric algebra and continue to avoid the use of complex numbers (with time being the customary imaginary quantity), it is possible to implement a different metric by defining the squares of the basis vectors so as to obey $\boldsymbol{x}^2 = \boldsymbol{y}^2 = \boldsymbol{z}^2 = -\boldsymbol{t}^2$. This new metric is no longer Euclidean, but nevertheless, it continues to be expressed as $|\boldsymbol{R}| = \left(x^2\boldsymbol{x}^2 + y^2\boldsymbol{y}^2 + z^2\boldsymbol{z}^2 + t^2\boldsymbol{t}^2\right)^{1/2}$!

This sort of metric is just a different way of framing Minkowski's spacetime metric, $|\boldsymbol{R}| = \left(x^2 + y^2 + z^2 - t^2\right)^{1/2}$, under which any point on an electromagnetic

wavefront expanding out in free space always has zero measure, or "separation", from its initial source point. For example, if the source is at the origin, a wavefront emitted at $t = 0$ travels out in any direction such that $\boldsymbol{R} = t\underset{\sim}{\boldsymbol{r}} + t\boldsymbol{t}$. Here $\boldsymbol{R}$ represents some time and position on the wavefront and $\underset{\sim}{\boldsymbol{r}}$ is the unit *spatial* vector along the given direction of travel, for example, $\boldsymbol{x}$. That is to say, the distance the wavefront travels along the direction $\boldsymbol{x}$ in the time t is $[c]t$. Now, with $\boldsymbol{x}^2 = -\boldsymbol{t}^2$, it is clear that at any time t, we have $\boldsymbol{R}^2 = t^2\boldsymbol{x}^2 + t^2\boldsymbol{t}^2 = 0$. This assumption is in effect one of the key tenets of special relativity, for it underpins the required invariance of the speed of light. Only by having $\boldsymbol{x}^2 = \boldsymbol{y}^2 = \boldsymbol{z}^2 = -\boldsymbol{t}^2$ do we get the required result $\boldsymbol{R}^2 = 0$. Any orthogonal transformation of the basis vectors, $\boldsymbol{t}$ included, leaves this result unchanged so that exactly the same view of a spherically expanding wavefront is maintained in all reference frames; that is, the speed of light is invariant. Under the chosen metric, we shall see that a Lorentz transformation [42, 44, 45] between reference frames is just such an orthogonal transformation.

But with a metric based on $\boldsymbol{x}^2 = \boldsymbol{y}^2 = \boldsymbol{z}^2 = -\boldsymbol{t}^2$, there remains a matter of choice as to whether to take $\boldsymbol{t}^2$ as being positive or negative. The normalization convention regarding the sign of the squares of the basis vectors is known as the metric signature and the usual choices are denoted by $(+---)$ and $(-+++)$ where the signs of $\boldsymbol{t}^2, \boldsymbol{x}^2, \boldsymbol{y}^2, \boldsymbol{z}^2$ are indicated in sequence. The former signature is quite common in the specialist literature, for example, in the works of authors such as Hestenes, Doran, Lasenby, and Gull, and has the benefit of making $\boldsymbol{u}^2$ positive for any timelike vector $\boldsymbol{u}$. The notion timelike vector will be more fully explained in Section 7.11, but for now we can take it to mean any vector that can represent time. The latter choice, $\boldsymbol{t}^2 = -1$, used for example by Lounesto [28, Section 8.10], is our choice because it corresponds to the familiar concept of spatial vectors having positive rather than negative squares, and it also corresponds to the usual choice in spacetime based on traditional four-vectors in which the time component is treated as an imaginary scalar.

Representing spacetime through geometric algebra, however, completely bypasses the need for imaginary quantities; all scalars are real. Since vectors can have a negative square, we must therefore remember that $|\boldsymbol{u}|$, the "length" or measure of a spacetime vector $\boldsymbol{u}$, needs to be expressed as $\left|\boldsymbol{u}^2\right|^{1/2}$ rather than $\left(\boldsymbol{u}^2\right)^{1/2}$. (Note that we use the single vertical bars $|\ldots|$ to denote the measure of vectors as well as the absolute value of scalars.) In the passing, however, it is interesting to observe that without the non-Euclidean spacetime metric, that is to say, if we simply wanted to create a conventional 4D Euclidean space with $\boldsymbol{x}^2 = \boldsymbol{y}^2 = \boldsymbol{z}^2 = \boldsymbol{t}^2 = 1$ simply in order to embody time as a vector, we would find $I^2 = +1$ (see Section 4.3). As a result, I would no longer be able to fulfill the same extremely useful role that it did in (3+1)D where complex scalars are superfluous. Complex scalars would have to be introduced. At this point, someone would suggest the bright idea of adopting $\boldsymbol{x}^2 = \boldsymbol{y}^2 = \boldsymbol{z}^2 = -\boldsymbol{t}^2$ simply to fix the problem. We would therefore be led back to the spacetime geometric algebra even without thinking about the potential relativistic benefits. Considerations of relativity apart, the spacetime norm is essential if we wish to derive any real benefit from a 4D geometric algebra.

Before we move on, however, there is a point of major importance to consider. We are well aware that when we choose $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$, we have an infinite number of possibilities, starting with how to orientate $\boldsymbol{x}$. It would be a mistake, therefore, to think that there is only one possibility for $\boldsymbol{t}$. In fact, using a single time vector is equivalent to imposing the usual nonrelativistic limit, $v \ll c$. This may be perfectly valid in many circumstances, but we will see in due course that the possibilities for $\boldsymbol{t}$ are actually limitless because any given time vector is always associated with some particular reference frame, and here the possibilities are indeed limitless. Usually $\boldsymbol{t}$ implies the time vector of our own local reference frame, but if we wish to choose a different reference frame it then turns out that we require a different time vector. When necessary, we will therefore treat $\boldsymbol{t}$ as being symbolic inasmuch as it could mean *any* time vector. It is common, however, to use the symbol $\boldsymbol{\upsilon}$ in connection with the time vector in the rest frame of some moving particle, and we will also use $\boldsymbol{\theta}$ when we need a more general symbol that is free of other possible associations. What these different choices actually mean will start to become clear in Sections 7.6 and 7.7, but for the benefit of those readers who do not wish to consider the relativistic implications at first reading, $\boldsymbol{t}$ should be regarded a being purely symbolic—just like $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$.

7.3 THE SPACETIME BASIS ELEMENTS

From the chosen orthonormal basis vectors $\boldsymbol{t},\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}$, we may readily generate all the other basis elements of spacetime as shown in Table 2.1(b). There will always be some choice involved in the order of products so that if we simply do this in an arbitrary way we will find that there are variations on a theme just due to differences of sign. To standardize the result, we can use a scheme to generate them based on the same pattern that was chosen for our 3D geometric algebra. Starting from $1,\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{yz},\boldsymbol{zx},\boldsymbol{xy},\boldsymbol{xyz}$, we incorporate exactly the same set of elements postmultiplied by $\boldsymbol{t}$, that is to say $\boldsymbol{t},\boldsymbol{xt},\boldsymbol{yt},\boldsymbol{zt},\boldsymbol{yzt},\boldsymbol{zxt},\boldsymbol{xyt},\boldsymbol{xyzt}$. The choice of postmultiplication rather than premultiplication here is somewhat arbitrary and, as we have just indicated, only affects the signs of some of the new elements. It is clear that all these new elements are mutually independent since multiplication of the original independent set by the independent factor $\boldsymbol{t}$, which cannot be expressed as any combination of $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$, can have no bearing on this property. The new set therefore simply inherits its independence from the original one. Likewise, they are all independent of those in the original set since none of these involves $\boldsymbol{t}$. We now have twice the original number of basis elements, exactly the number required for a 4D geometric algebra. Although we could have formed the basis elements in other ways, it will be useful to have them closely corresponding with the existing (3+1)D structure, as will soon become clear.

We now take the opportunity to comment on these elements and to illustrate other forms of notation that are likely to be encountered, as summarized in Table 7.1. The term $\boldsymbol{xyzt}$ provides us with our unit pseudoscalar, $\boldsymbol{I}$. Many of the authors who prefer the $(+---)$ signature usually represent it as $\gamma_0\gamma_1\gamma_2\gamma_3$, that is to say

Table 7.1 Other Forms of Representation of the Spacetime Basis Elements

Entity	This work	Hestenes, Doran, Lasenby, and Gull	Lounesto
Vectors	$\boldsymbol{t},\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}$	$\gamma_0,\gamma_1,\gamma_2,\gamma_3$	$\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3,\mathbf{e}_4$
Timelike bivectors	$\boldsymbol{xt},\boldsymbol{yt},\boldsymbol{zt}$	$\sigma_1,\sigma_2,\sigma_3$	$\mathbf{e}_{14},\mathbf{e}_{24},\mathbf{e}_{34}$
Spacelike bivectors	$\boldsymbol{yz},\boldsymbol{zx},\boldsymbol{xy}$	β_1,β_2,β_3 or $I\sigma_1,I\sigma_2,I\sigma_3$	$\mathbf{e}_{23},\mathbf{e}_{31},\mathbf{e}_{12}$
Trivectors	$-I\boldsymbol{t},-I\boldsymbol{x},-I\boldsymbol{y},-I\boldsymbol{z}$	$I\gamma_0,I\gamma_1,I\gamma_2,I\gamma_3$	$\mathbf{e}_{432},\mathbf{e}_{413},\mathbf{e}_{421},\mathbf{e}_{321}$
Pseudoscalar	$I=\boldsymbol{xyzt}$	$I=\gamma_0\gamma_1\gamma_2\gamma_3$	$\mathbf{e}_{1234}$
Vector derivative	$\boldsymbol{\nabla}$	∇	∂
Reversion	†	~	~

The entries should be compared with Table 2.2 and the differences noted. Other variations that may be seen include the use of $\mathbf{e}_0$ instead of $\mathbf{e}_4$ and a different order for some of the products.

$\boldsymbol{txyz}$ in our notation, which is the negative of I as we have defined it. This makes little or no difference to us, for it is easily verified, and worthwhile to remember, that irrespective of the chosen metric signature this particular unit pseudoscalar

- *anticommutes* with all vectors (see Exercise 4.8.9) and
- *obeys* $I^2=-1$.

There are now six bivectors. The three bivectors $\boldsymbol{xt},\boldsymbol{yt},\boldsymbol{zt}$ that were generated by multiplying the chosen basis vectors by $\boldsymbol{t}$ are called timelike bivectors, while the other three, $\boldsymbol{yz},\boldsymbol{zx},\boldsymbol{xy}$, which were carried over directly, involve only spatial vectors and are called spacelike bivectors. It is appropriate that there are now two separate classes of bivectors since the one sort turns out to be the dual of the other; for example, $\boldsymbol{yz}=I\boldsymbol{xt},\boldsymbol{zx}=I\boldsymbol{yt}\ldots$. Although the terms timelike and spacelike seem obvious enough in the present context, they also have a more specific meaning as will be discussed in Section 7.11.

The trivector $\boldsymbol{xyz}$, together with $\boldsymbol{yzt},\boldsymbol{zxt},\boldsymbol{xyt}$, that is, the bivectors $\boldsymbol{yz},\boldsymbol{zx},\boldsymbol{xy}$ postmultiplied by $\boldsymbol{t}$, form the four trivector basis elements. These turn out to be merely the duals of the vectors, albeit with a change of sign in each case; for example, $\boldsymbol{xyz}=-I\boldsymbol{t},\boldsymbol{yzt}=-I\boldsymbol{x}\ldots$ and so on.

We already know that the 3D geometric algebra splits into two sets of elements, one being the dual of the other. The same situation clearly exists in the spacetime geometric algebra where the first set comprises the scalars, vectors, and timelike bivectors, while the dual set comprises the pseudoscalars, the pseudovectors (trivectors), and the spacelike bivectors. The basis elements associated with each of these sets are $1,\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{xt},\boldsymbol{yt},\boldsymbol{xt}$ and $I,-I\boldsymbol{x},-I\boldsymbol{y},-I\boldsymbol{z},I\boldsymbol{xt},I\boldsymbol{yt},I\boldsymbol{zt}$ respectively. Note that the latter set is generated by *postmultiplying* each element of $1,\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{xt},\boldsymbol{yt},\boldsymbol{xt}$ by I. Remember here that the spacetime form of I is $\boldsymbol{xyzt}$ rather than $\boldsymbol{xyz}$ and it *anticommutes* with all the vectors. The latter point means that there are two ways in which the duals of the vectors may be formed—either by premultiplication or postmultiplication with I—and this gives rise to a choice of sign.

Table 7.2 The 4D Multiplication Table for Spacetime

4D	t	x	y	z	xt	yt	zt
t	−1	$-xt$	$-yt$	$-zt$	x	y	z
x	xt	1	$-Izt$	Iyt	t	$-Iz$	Iy
y	yt	Izt	1	$-Ixt$	Iz	t	$-Ix$
z	zt	$-Iyt$	Ixt	1	$-Iy$	Ix	t
xt	$-x$	$-t$	Iz	$-Iy$	1	Izt	$-Iyt$
yt	$-y$	$-Iz$	$-t$	Ix	$-Izt$	1	Ixt
zt	$-z$	Iy	$-Ix$	$-t$	Iyt	$-Ixt$	1

The table has been compacted by noting that spatial bivectors and trivectors can be written in terms of their duals and then multiplied, for example, $(xy)(zx)=(Izt)(Iyt)=-(zt)(yt)$, so that all of these elements may be omitted. Note however that in any product of duals, I commutes with bivectors but anticommutes with vectors so that if $I\boldsymbol{U}$ and $I\boldsymbol{V}$ are the duals of $\boldsymbol{U}$ and $\boldsymbol{V}$ respectively, then $I\boldsymbol{U}I\boldsymbol{V}=\pm\boldsymbol{U}\boldsymbol{V}$, the sign being positive when $\boldsymbol{U}$ is a vector and negative when it is a bivector.

To get some feeling for the new geometric algebra, a multiplication table can be worked out. The result is shown in Table 7.2 in which the overall picture has been simplified by exploiting the dual representation. The detailed workings are left as an exercise.

7.3.1 Spatial and Temporal Vectors

At this stage, we digress a little to introduce a device that will frequently prove useful in working out multivector expressions. As an illustration, take the case of multiplying any two vectors, say $\boldsymbol{u}=u_t\boldsymbol{t}+u_x\boldsymbol{x}+u_y\boldsymbol{y}+u_z\boldsymbol{z}$ and $\boldsymbol{v}=v_t\boldsymbol{t}+v_x\boldsymbol{x}+v_y\boldsymbol{y}+v_z\boldsymbol{z}$. Just as was the case in (3+1)D, the result $\boldsymbol{u}\boldsymbol{v}$ will be a scalar plus a bivector, but the bivector part will now generally be a sum of both timelike and spacelike sorts, as exemplified by $\boldsymbol{xt}$ and $\boldsymbol{yz}$ respectively. Because of the different properties of these two types of bivectors, it will be helpful to be able to distinguish between them. It will therefore be convenient to split each of $\boldsymbol{u}$ and $\boldsymbol{v}$ into two parts, one that includes the time vector $\boldsymbol{t}$ while the other does not. In the case of any vector that is expressed in terms of the basis vectors, such as $\boldsymbol{u}$ given above, it is clearly trivial to obtain the parts in question as being $u_t\boldsymbol{t}$ and $u_x\boldsymbol{x}+u_y\boldsymbol{y}+u_z\boldsymbol{z}$ respectively, and we may also do just the same for $\boldsymbol{v}$. Since $\boldsymbol{t}$ is orthogonal to each of $\boldsymbol{x},\boldsymbol{y}$, and $\boldsymbol{z}$, it is clear that these two parts must be orthogonal whichever vector we choose. This will therefore help to simplify extracting the scalar as well as the timelike and spacelike bivector parts from the product $\boldsymbol{u}\boldsymbol{v}$. Now we could proceed to demonstrate this with $\boldsymbol{u}$ and $\boldsymbol{v}$ represented as they are here in terms of all the basis vectors, but this would be clumsy as 16 separate products are involved. We are therefore going to demonstrate a neater method that requires us to know only the time vector concerned.

Given some time vector $\boldsymbol{t}$, we can write any vector $\boldsymbol{u}$ in the form $\boldsymbol{u} = \left(-\boldsymbol{t}^2\right)\boldsymbol{u} = -\boldsymbol{t}\left(\boldsymbol{t}\cdot\boldsymbol{u} + \boldsymbol{t}\wedge\boldsymbol{u}\right)$. Be wary of signs here as it is now necessary to get used to the fact that $\boldsymbol{t}^2 = \boldsymbol{t}\cdot\boldsymbol{t} = -1$. This gives $\boldsymbol{u}$ as the sum of two terms. The first term, $-\boldsymbol{t}\left(\boldsymbol{t}\cdot\boldsymbol{u}\right)$, is the projection of $\boldsymbol{u}$ onto the time vector $\boldsymbol{t}$, meaning that the second term, $-\boldsymbol{t}\left(\boldsymbol{t}\wedge\boldsymbol{u}\right)$, must be orthogonal to $\boldsymbol{t}$. To denote this, we therefore define an operator denoted by an under-tilde $\underset{\sim}{\ }$ such that

$$\underset{\sim}{\boldsymbol{u}} \equiv \boldsymbol{t}\left(\boldsymbol{u}\wedge\boldsymbol{t}\right) \tag{7.1}$$

from which it immediately follows

$$\boldsymbol{u} = \underset{\sim}{\boldsymbol{u}} - \left(\boldsymbol{u}\cdot\boldsymbol{t}\right)\boldsymbol{t} \tag{7.2}$$

Since $\underset{\sim}{\boldsymbol{u}}\cdot\boldsymbol{t} = \boldsymbol{u}\cdot\boldsymbol{t} + \left(\boldsymbol{u}\cdot\boldsymbol{t}\right)\boldsymbol{t}\cdot\boldsymbol{t} = \boldsymbol{u}\cdot\boldsymbol{t} - \left(\boldsymbol{u}\cdot\boldsymbol{t}\right) = 0$, we may readily confirm that $\underset{\sim}{\boldsymbol{u}}$ is always orthogonal to $\boldsymbol{t}$. A vector such as $\underset{\sim}{\boldsymbol{u}}$ that is orthogonal to the chosen time vector is then what we mean by a spatial vector as far as that particular time vector is concerned. It is clear that in terms of this definition, the standard basis vectors $\boldsymbol{x}, \boldsymbol{y}$, and $\boldsymbol{z}$ are all examples of vectors that are spatial with respect to the time vector $\boldsymbol{t}$. Any linear combination of $\boldsymbol{x}, \boldsymbol{y}$, and $\boldsymbol{z}$ alone is also a spatial vector in this context. When $\boldsymbol{u} = u_t\boldsymbol{t} + u_x\boldsymbol{x} + u_y\boldsymbol{y} + u_z\boldsymbol{z}$ we find $\underset{\sim}{\boldsymbol{u}} = u_x\boldsymbol{x} + u_y\boldsymbol{y} + u_z\boldsymbol{z}$, exactly as expected, but the key thing is that we can derive many important results without actually needing to know any of the components u_x, u_y, and u_z. However, in order to demonstrate that spatial vectors always have a positive square in our metric signature, that is to say $0 \le \underset{\sim}{\boldsymbol{u}}^2$ for any $\underset{\sim}{\boldsymbol{u}}$, it is still necessary to invoke some arbitrary $\boldsymbol{x}, \boldsymbol{y}$ and $\boldsymbol{z}$ that we can relate to each of the three + signs in $(-+++)$. This, however, is due only to a notation for metric signature that is more suited to working with components.

Any spatial vector such as $\underset{\sim}{\boldsymbol{u}}$ may be said to be in the "orthogonal space" of the time vector [45, Section 4.3, p. 41]. On the other hand, $-\left(\boldsymbol{u}\cdot\boldsymbol{t}\right)\boldsymbol{t}$ is clearly a pure time vector, which we may call the *temporal* part of $\boldsymbol{u}$. Spatial and temporal turn out to be much narrower in meaning than the terms timelike and spacelike, which will be more fully defined in Section 7.11, but basically, what is spatial and temporal depends on the chosen time vector, whereas what is timelike and spacelike does not.

Returning to the original question, we may apply our result to the product of the vectors $\boldsymbol{u}$ and $\boldsymbol{v}$. Taking advantage of the fact that both $\underset{\sim}{\boldsymbol{u}}$ and $\underset{\sim}{\boldsymbol{v}}$ are orthogonal to $\boldsymbol{t}$, we find

$$\boldsymbol{u}\boldsymbol{v} = \left(\underset{\sim}{\boldsymbol{u}} - \left(\boldsymbol{u}\cdot\boldsymbol{t}\right)\boldsymbol{t}\right)\left(\underset{\sim}{\boldsymbol{v}} - \left(\boldsymbol{v}\cdot\boldsymbol{t}\right)\boldsymbol{t}\right) = \underbrace{\left(\underset{\sim}{\boldsymbol{u}}\underset{\sim}{\boldsymbol{v}}\right)}_{\substack{\text{spacelike}\\ \text{bivector}}} + \underbrace{\left(\boldsymbol{u}\cdot\boldsymbol{t}\right)\underset{\sim}{\boldsymbol{v}}\boldsymbol{t} - \left(\boldsymbol{v}\cdot\boldsymbol{t}\right)\underset{\sim}{\boldsymbol{u}}\boldsymbol{t}}_{\substack{\text{timelike}\\ \text{bivector}}} - \underbrace{\left(\boldsymbol{u}\cdot\boldsymbol{t}\right)\left(\boldsymbol{v}\cdot\boldsymbol{t}\right)}_{\text{scalar}} \tag{7.3}$$

Here the timelike, spacelike, and scalar parts of the result are all separately identifiable from knowing the spatial and temporal parts of $\boldsymbol{u}$ and $\boldsymbol{v}$. It is therefore no longer necessary to express $\boldsymbol{u}$ and $\boldsymbol{v}$ in terms of a full set of basis vectors in order to achieve this.

Following from this, the simple identities below prove helpful in rearranging products of vectors and basis elements into a standard form. They are applicable in the $(-+++)$ signature, but the conversion to $(+---)$ is simple and generally affects only the signs within the results.

For any vector $\underset{\sim}{\boldsymbol{u}}$ that is spatial with respect to the time vector $\boldsymbol{t}$, we have

$$\boldsymbol{t}\underset{\sim}{\boldsymbol{u}} = -\underset{\sim}{\boldsymbol{u}}\boldsymbol{t} \quad \text{or} \quad \boldsymbol{t}\underset{\sim}{\boldsymbol{u}}\boldsymbol{t} = \underset{\sim}{\boldsymbol{u}} \tag{7.4}$$

As well as using the usual property that orthogonal vectors anticommute, this also makes use of the new rule $\boldsymbol{t}^2 = -1$.

Next, this generalizes to the case of any vector $\boldsymbol{u}$ as

$$\begin{aligned} \boldsymbol{t}\boldsymbol{u}\boldsymbol{t} &= \boldsymbol{t}(\underset{\sim}{\boldsymbol{u}} - (\boldsymbol{u}\cdot\boldsymbol{t})\boldsymbol{t})\boldsymbol{t} \\ &= \boldsymbol{t}\underset{\sim}{\boldsymbol{u}}\boldsymbol{t} - (\boldsymbol{u}\cdot\boldsymbol{t})\boldsymbol{t}^3 \\ \Leftrightarrow \boldsymbol{t}\boldsymbol{u}\boldsymbol{t} &= \underset{\sim}{\boldsymbol{u}} + (\boldsymbol{u}\cdot\boldsymbol{t})\boldsymbol{t} \end{aligned} \tag{7.5}$$

so that

$$\begin{aligned} \underset{\sim}{\boldsymbol{u}} &= \tfrac{1}{2}(\boldsymbol{u} + \boldsymbol{t}\boldsymbol{u}\boldsymbol{t}) \\ -(\boldsymbol{u}\cdot\boldsymbol{t}) &= \tfrac{1}{2}(\boldsymbol{u} - \boldsymbol{t}\boldsymbol{u}\boldsymbol{t}) \end{aligned} \tag{7.6}$$

We also have

$$\begin{aligned} \boldsymbol{u}\wedge\boldsymbol{t} &= \underset{\sim}{\boldsymbol{u}}\boldsymbol{t} \\ \boldsymbol{u}\cdot\boldsymbol{t} &= (\boldsymbol{u} - \underset{\sim}{\boldsymbol{u}})\boldsymbol{t} \end{aligned} \tag{7.7}$$

$$\begin{aligned} \underset{\sim}{\boldsymbol{u}} &= \boldsymbol{u} + (\boldsymbol{u}\cdot\boldsymbol{t})\boldsymbol{t} \\ &= \boldsymbol{t}(\boldsymbol{u}\wedge\boldsymbol{t}) \end{aligned} \tag{7.8}$$

and

$$\begin{aligned} 0 &\le \underset{\sim}{\boldsymbol{u}}^2 \\ \boldsymbol{u}^2 &= \underset{\sim}{\boldsymbol{u}}^2 - (\boldsymbol{u}\cdot\boldsymbol{t})^2 \end{aligned} \tag{7.9}$$

Since we have been able to exploit splitting $\boldsymbol{u}$ into the orthogonal spatial and temporal parts $\underset{\sim}{\boldsymbol{u}}$ and $-(\boldsymbol{u}\cdot\boldsymbol{t})\boldsymbol{t}$ in the case of an arbitrary time vector $\boldsymbol{t}$, it follows that we can do exactly the same for any given time vector. The identities given in Equations (7.4)–(7.9) consequently apply when $\boldsymbol{t}$ is replaced by any other time vector. It is therefore important to remember that the resulting spatial vector *always depends on the chosen time vector* since, by definition, the two must always be orthogonal to each other. If we chose $\boldsymbol{\theta}$, say, as an alternative time vector, then we would require to use $\boldsymbol{\theta}$ in Equation (7.1) and we would consequently get a different splitting of $\boldsymbol{u}$, this time into the spatial part $\boldsymbol{\theta}(\boldsymbol{u}\wedge\boldsymbol{\theta})$ and the temporal part $-(\boldsymbol{u}\cdot\boldsymbol{\theta})\boldsymbol{\theta}$.

Another simple but unrelated identity that may also be of use concerns the spacetime unit pseudoscalar:

$$I^\dagger = I \tag{7.10}$$

Because of the mixed metric signature, this is one of the exceptions to the general rule given in Section 4.3. The proof is simple enough and is one of the exercises below.

7.4 BASIC OPERATIONS

In Chapter 3 we saw that in (3+1)D, many useful expressions take the form of multivectors, for example, $\mathbf{E}+\boldsymbol{B}$, $\rho-\mathbf{J}$ and $t+\mathbf{r}$, while later on, we also discovered $\boldsymbol{\nabla}+\partial_t$. The last three examples all have the form of scalar plus vector that is known as a paravector, a concept that is clearly appropriate to the nature of (3+1)D. From our earlier discussion of the spacetime basis elements, it is clear that the third example, time and position, must correspond to a spacetime vector with a timelike part $t\boldsymbol{t}$, that is to say $t+\mathbf{r}$ now corresponds to $t\boldsymbol{t}+x\boldsymbol{x}+y\boldsymbol{y}+z\boldsymbol{z}$, or as we could now put it using the notation introduced in Section 7.3.1, $t\boldsymbol{t}+\underset{\sim}{\boldsymbol{r}}$. We may therefore speculate that something similar may also apply for the other paravectors $\rho-\mathbf{J}$ and $\boldsymbol{\nabla}+\partial_t$. In effect, this is where the term paravector originates, meaning similar to a vector. From the definition given in Equation (7.1), we may write any spacetime vector $\boldsymbol{u}$ in the form $u_t\boldsymbol{t}+\underset{\sim}{\boldsymbol{u}}$ by identifying u_t with $-(\boldsymbol{u}\cdot\boldsymbol{t})$. In general, we may form a paravector $\boldsymbol{U}=u_0+\mathbf{u}$ that corresponds to this spacetime vector by equating u_0 to u_t and equating the components of $\mathbf{u}$ with respect to the $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ basis with the components of $\underset{\sim}{\boldsymbol{u}}$ with respect to the $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$ basis. With this postulate in mind, let us examine the effects of the new time dimension on the multiplication process.

To begin with, let us compare multiplying two spacetime vectors, $\boldsymbol{u}=u_t\boldsymbol{t}+\underset{\sim}{\boldsymbol{u}}$ and $\boldsymbol{v}=v_t\boldsymbol{t}+\underset{\sim}{\boldsymbol{v}}$, against multiplying their (3+1)D counterparts, the paravectors $\boldsymbol{U}=u_0+\mathbf{u}$ and $\boldsymbol{V}=v_0+\mathbf{v}$. Note that it does not really matter whether or not the scalars u_t and v_t have anything to do with time, and while $\mathbf{u}$ and $\underset{\sim}{\boldsymbol{u}}$ may have identical component values, that is, $u_k=\underset{\sim}{u}_k$ for $k=x,y,z$, it must always be remembered that *they employ different sets of basis vectors*, $\mathbf{x},\mathbf{y},\mathbf{z}$, and $\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}$ respectively.

Multiplying the spacetime vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ results in (cf. Equation 7.3)

$$\begin{aligned}\boldsymbol{u}\boldsymbol{v}&=(u_t\boldsymbol{t}+\underset{\sim}{\boldsymbol{u}})(v_t\boldsymbol{t}+\underset{\sim}{\boldsymbol{v}})\\&=u_tv_t\boldsymbol{t}^2+u_t\boldsymbol{t}\underset{\sim}{\boldsymbol{v}}+v_t\underset{\sim}{\boldsymbol{u}}\boldsymbol{t}+\underset{\sim}{\boldsymbol{u}}\underset{\sim}{\boldsymbol{v}}\\&=-u_tv_t+v_t\underset{\sim}{\boldsymbol{u}}\boldsymbol{t}-u_t\underset{\sim}{\boldsymbol{v}}\boldsymbol{t}+\underset{\sim}{\boldsymbol{u}}\boldsymbol{t}\underset{\sim}{\boldsymbol{v}}\boldsymbol{t}\end{aligned} \tag{7.11}$$

while the multiplication of the corresponding (3+1)D paravectors $\boldsymbol{U}$ and $\boldsymbol{V}$ gives

$$\begin{aligned}\boldsymbol{U}\boldsymbol{V}&=(u_0+\mathbf{u})(v_0+\mathbf{v})\\&=u_0v_0+v_0\mathbf{u}+u_0\mathbf{v}+\mathbf{u}\mathbf{v}\end{aligned} \tag{7.12}$$

It is interesting that the resulting expressions are very similar in form. To make them fully equivalent, we would require $\mathbf{u}$, $\mathbf{v}$, and v_0 to be equal to $\underset{\sim}{\boldsymbol{ut}}$, $\underset{\sim}{\boldsymbol{vt}}$, and v_0, respectively, but, somewhat perversely, we must have $u_0 = -u_t$! Some other way of relating spacetime vectors to (3+1)D paravectors will therefore have to be found if we are to succeed in properly replicating multiplication.

Next, let us turn to the multiplication of a spacetime vector with a bivector. First, we will take the case of a timelike bivector $\boldsymbol{P}$, which, because of the definition of the basis elements, we can always write as $\underset{\sim}{\boldsymbol{pt}}$ where, as before, $\underset{\sim}{\boldsymbol{p}}$ is a purely spatial vector, for example, $\underset{\sim}{\boldsymbol{p}} = p_x\boldsymbol{x} + p_y\boldsymbol{y} + p_z\boldsymbol{z}$, which means that $\underset{\sim}{\boldsymbol{pt}}$ has the form of a general timelike bivector, $p_x\boldsymbol{xt} + p_y\boldsymbol{yt} + p_z\boldsymbol{zt}$. Clearly, however, $\underset{\sim}{\boldsymbol{pt}}$ is a much neater way of writing it. We then have

$$\begin{aligned}
\boldsymbol{uP} &= \left(u_t\boldsymbol{t} + \underset{\sim}{\boldsymbol{u}}\right)\underset{\sim}{\boldsymbol{pt}} \\
&= u_t\boldsymbol{t}\underset{\sim}{\boldsymbol{pt}} + \underset{\sim}{\boldsymbol{u}}\underset{\sim}{\boldsymbol{pt}} \\
&= \underbrace{u_t\underset{\sim}{\boldsymbol{p}} + \left(\underset{\sim}{\boldsymbol{u}}\cdot\underset{\sim}{\boldsymbol{p}}\right)\boldsymbol{t}}_{\text{vector}} + \underbrace{\left(\underset{\sim}{\boldsymbol{u}}\wedge\underset{\sim}{\boldsymbol{p}}\right)\boldsymbol{t}}_{\text{trivector}} \\
&= \quad \boldsymbol{u}\cdot\boldsymbol{P} \quad + \quad \boldsymbol{u}\wedge\boldsymbol{P}
\end{aligned} \tag{7.13}$$

Here we have used Equation (7.4) to rewrite $\boldsymbol{tpt}$ as $\underset{\sim}{\boldsymbol{p}}$. The result turns out to be a vector plus a trivector, just as we should expect for multiplying a vector with a bivector. The vector and trivector parts then simply represent the inner and outer products of $\boldsymbol{u}$ with $\boldsymbol{P}$.

Finally, on recalling that the spacelike bivectors are the duals of the timelike bivectors, for example, $\boldsymbol{yz} = \boldsymbol{Ixt}$, the product of vector with a spacelike bivector $\boldsymbol{Q}$ can be found in a similar manner by representing $\boldsymbol{Q}$ with $\boldsymbol{IP}$ where $\boldsymbol{P}$ is a timelike bivector as in Equation (7.13). We need to rearrange only the product $\boldsymbol{uIP}$ slightly to get the desired result:

$$\begin{aligned}
\boldsymbol{uQ} = \boldsymbol{uIP} &= -I\left(u_t\boldsymbol{t} + \underset{\sim}{\boldsymbol{u}}\right)\underset{\sim}{\boldsymbol{pt}} \\
&= -u_tI\underset{\sim}{\boldsymbol{p}} - \left(\underset{\sim}{\boldsymbol{u}}\cdot\underset{\sim}{\boldsymbol{p}}\right)I\boldsymbol{t} \;-\; \left(\underset{\sim}{\boldsymbol{u}}\wedge\underset{\sim}{\boldsymbol{p}}\right)I\boldsymbol{t} \\
&= \underbrace{-u_tI\underset{\sim}{\boldsymbol{p}} - \left(\underset{\sim}{\boldsymbol{u}}\cdot\underset{\sim}{\boldsymbol{p}}\right)I\boldsymbol{t}}_{\text{trivector}} + \underbrace{-\underset{\sim}{\boldsymbol{u}}\times\underset{\sim}{\boldsymbol{p}}}_{\text{vector}} \\
&= \quad \boldsymbol{u}\wedge\boldsymbol{Q} \quad + \quad \boldsymbol{u}\cdot\boldsymbol{Q}
\end{aligned} \tag{7.14}$$

Since I anticommutes with spacetime vectors, we must take care of the sign when we change its position within products. In addition, as an ad hoc device, we have used the same notation as for the 3D cross product (see Equation 2.8) to represent $\left(\underset{\sim}{\boldsymbol{u}}\wedge\underset{\sim}{\boldsymbol{p}}\right)I\boldsymbol{t}$. However, this is only to clarify that the result is a vector. Bear in mind that here $I\boldsymbol{t} = -\boldsymbol{xyz}$ and the basis vectors involved in it are of course $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$

rather than $\mathbf{x},\mathbf{y},\mathbf{z}$. The proper way to look at it, which will become easier with familiarity, is that $\underset{\sim}{\boldsymbol{u}} \wedge \underset{\sim}{\boldsymbol{p}}$ is a spacelike bivector and orthogonal to $\boldsymbol{t}$ so that $\left(\underset{\sim}{\boldsymbol{u}} \wedge \underset{\sim}{\boldsymbol{p}}\right)\boldsymbol{t}$ is a trivector and consequently its dual, $I\left(\underset{\sim}{\boldsymbol{u}} \wedge \underset{\sim}{\boldsymbol{p}}\right)\boldsymbol{t}$, must be a vector.

We could go on to work out all the other products, but these may be constructed along the lines demonstrated above. For example, to evaluate the product of two general bivectors that have both spacelike and timelike parts, we may write each bivector in similar forms, one as $\boldsymbol{P} = \left(\underset{\sim}{\boldsymbol{p}} + I\underset{\sim}{\boldsymbol{q}}\right)\boldsymbol{t}$ and the other as $\boldsymbol{U} = (\underset{\sim}{\boldsymbol{u}} + I\underset{\sim}{\boldsymbol{v}})\boldsymbol{t}$. In evaluating $\boldsymbol{PU}$, the expressions involved may be rearranged so that $\boldsymbol{t}$ drops out to yield the simple result $\boldsymbol{PU} = \left(\underset{\sim}{\boldsymbol{p}} + I\underset{\sim}{\boldsymbol{q}}\right)\boldsymbol{t}(-\boldsymbol{t})(\underset{\sim}{\boldsymbol{u}} - I\underset{\sim}{\boldsymbol{v}}) = \left(\underset{\sim}{\boldsymbol{p}} + I\underset{\sim}{\boldsymbol{q}}\right)(\underset{\sim}{\boldsymbol{u}} - I\underset{\sim}{\boldsymbol{v}})$. And so we may proceed with all the remaining forms that may be encountered.

7.5 VELOCITY

In (3+1)D, the paravector $t + \mathbf{r}$ specifies an event at position $\mathbf{r}$ and time t. As we saw in Section 7.2, in spacetime terms the same information is given by a vector like $\boldsymbol{r}$ that includes both position and time. If we can consider $\boldsymbol{r}$ to be a function of time, then $\boldsymbol{r}(t)$ defines the trajectory of some point, or particle, through spacetime. This being the case, it is natural to try to find the velocity of the particle. By defining the spacetime velocity in an analogous manner to its Newtonian counterpart, we find that $\mathbf{v} = \partial_t(t + \mathbf{r})$ simply leads us to $\boldsymbol{v} = \partial_t \boldsymbol{r}$. Note that velocity is what we may refer to as a derived vector, that is to say, it results from performing some operation on another vector, in this case differentiating a simple event vector with respect to time. Derived vectors may behave differently from simple vectors; for example, velocity clearly depends on the reference frame in which the time t is defined, whereas simple vectors do not depend on the choice of reference frame.

The trajectory of a particle at rest may be stated as $\boldsymbol{r}(t) = t\boldsymbol{t} + \underset{\sim}{\boldsymbol{r}}_0$. At any time t, its position is always given by the same purely spatial vector $\underset{\sim}{\boldsymbol{r}}_0$ (recall the definition of spatial in Section 7.3.1). In contrast, a particle moving with constant velocity will be described by $\boldsymbol{r}(t) = t\boldsymbol{t} + \underset{\sim}{\boldsymbol{r}}_0 + vt\underset{\sim}{\boldsymbol{s}}$ where $\underset{\sim}{\boldsymbol{s}}$ is a unit spatial vector along the direction of motion (so that $\underset{\sim}{\boldsymbol{s}}^2 = 1$). We could have written $x_0\boldsymbol{x} + y_0\boldsymbol{y} + z_0\boldsymbol{z}$ rather than $\underset{\sim}{\boldsymbol{r}}_0$ and $s_x\boldsymbol{x} + s_y\boldsymbol{y} + s_z\boldsymbol{z}$ rather than $\underset{\sim}{\boldsymbol{s}}$, but it is much more convenient to write simply $\underset{\sim}{\boldsymbol{r}}_0$ and $\underset{\sim}{\boldsymbol{s}}$. Now, since the particle moves through $v\underset{\sim}{\boldsymbol{s}}$ in a unit time interval and $\underset{\sim}{\boldsymbol{s}}$ is a unit spatial vector, it is clear that v is the same as the normal scalar velocity. Differentiating $\boldsymbol{r}(t)$ with respect to time, however, gives $\boldsymbol{v} = \boldsymbol{t} + v\underset{\sim}{\boldsymbol{s}}$. Although it seems novel that spacetime velocity should include the unit time vector, we have previously seen the analogous thing in (3+1)D where the time derivative of $t + \mathbf{r}$ is $1 + \mathbf{v}$ rather than just $\mathbf{v}$ (see Section 3.1).

If we write $v\underset{\sim}{\boldsymbol{s}}$ as $\underset{\sim}{\boldsymbol{v}}$, then $\boldsymbol{v} = \boldsymbol{t} + v\underset{\sim}{\boldsymbol{s}}$ may be more conveniently written as $\boldsymbol{v} = \boldsymbol{t} + \underset{\sim}{\boldsymbol{v}}$, and it is worthwhile remembering that this decomposition, as given in Equation (7.1), works with any time vector. Since by definition $\underset{\sim}{\boldsymbol{s}} \perp \boldsymbol{t}$ and $\underset{\sim}{\boldsymbol{s}}^2 = 1$, the magnitude of $\boldsymbol{v}$ is found from $\left|(\boldsymbol{t} + v\underset{\sim}{\boldsymbol{s}})^2\right|^{1/2}$ to be $\left(1 - v^2\right)^{1/2}$.

In summary therefore,

$$\begin{aligned} \boldsymbol{v} &\equiv \partial_t \boldsymbol{r} \\ &= \boldsymbol{t} + \underset{\sim}{\boldsymbol{v}} \end{aligned} \tag{7.15}$$

where

$$|\boldsymbol{v}| = \left(1 - v^2\right)^{\frac{1}{2}} \tag{7.16}$$

and t is the time parameter associated with the chosen time vector $\boldsymbol{t}$.

The trajectory of the particle from which the constant velocity $\boldsymbol{v}$ was derived is then simply expressed as $\boldsymbol{r}(t) = t\boldsymbol{v} + \underset{\sim}{\boldsymbol{r}}_0$. The spacetime velocity therefore allows us to write down the trajectory of a particle in uniform motion in just the same way that we would conventionally write it.

The fact that the spacetime velocity includes the time vector takes a little getting used to and is often troublesome to remember. Sometimes it is convenient to work with $\underset{\sim}{\boldsymbol{v}}$ and so, to avoid any confusion, we will refer to $\boldsymbol{v}$ as the velocity and $\underset{\sim}{\boldsymbol{v}}$ as the *spatial* velocity, that is to say, expressed in terms of basis vectors the velocity takes the form $\boldsymbol{t} + v_x\boldsymbol{x} + v_y\boldsymbol{y} + v_z\boldsymbol{z}$, whereas the spatial velocity takes the form $v_x\boldsymbol{x} + v_y\boldsymbol{y} + v_z\boldsymbol{z}$.

7.6 DIFFERENT BASIS VECTORS AND FRAMES

Let us now turn to the implications of transforming the basis vectors from one set to another equivalent set. The Galilean concept of relativity requires that this sort of transformation, that is, change of reference frame, does not alter any physical attributes such as size, shape, and mass. That is not to say that they necessarily *appear* to be the same. The particular transformations involved are spatial rotations and translations, including uniform motion, and so it is fairly clear here what we mean by being unchanged yet appearing different. Any change of appearance simply results from taking a different viewpoint. In (3+1)D, it is evident that the orthonormal vectors $\mathbf{x}, \mathbf{y}, \mathbf{z}$ are all equivalent. If we rotate them in the $\mathbf{xy}$ plane, that is, with $\mathbf{z}$ as axis, then we get another equivalent set $\mathbf{x}', \mathbf{y}', \mathbf{z}$ where the new $\mathbf{x}'$ and $\mathbf{y}'$ vectors are linear combinations of the original $\mathbf{x}$ and $\mathbf{y}$. But how about spacetime where we are now taking $\boldsymbol{t}$ to be on an equal footing with $\boldsymbol{x}, \boldsymbol{y}$, and $\boldsymbol{z}$? We have supposed in Section 7.1 that a simple spacetime vector will be invariant under change of basis. If we rotate $\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ in the $\boldsymbol{xy}$ plane as shown in Figure 7.3, then we get $\boldsymbol{t}, \boldsymbol{x}', \boldsymbol{y}', \boldsymbol{z}$ where $\boldsymbol{x}' = \alpha\boldsymbol{x} + \beta\boldsymbol{y}$ and $\boldsymbol{y}' = \alpha\boldsymbol{y} - \beta\boldsymbol{x}$ are normalized for some real scalars α and β that happen to be respectively the cosine and sine of θ, the angle of rotation, that is to say $\alpha^2 + \beta^2 = \cos^2\theta + \sin^2\theta = 1$. This represents an entirely equivalent choice of basis vectors to $\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$, and so everything seems the same as before apart from the fact that it is no longer possible to specify an axis of a rotation, just a plane. Any vector $\boldsymbol{u}$ that is entirely changed by the rotation and the vector $\boldsymbol{u}'$ into which it is rotated must lie in this plane, which may therefore be represented by the bivector $\boldsymbol{u} \wedge \boldsymbol{u}'$. For example, $\boldsymbol{x} \wedge \boldsymbol{x}' = \alpha\boldsymbol{x} \wedge \boldsymbol{x} + \boldsymbol{x} \wedge \beta\boldsymbol{y} = \beta\boldsymbol{xy}$ and $\boldsymbol{y} \wedge \boldsymbol{y}'$ gives exactly the same result.

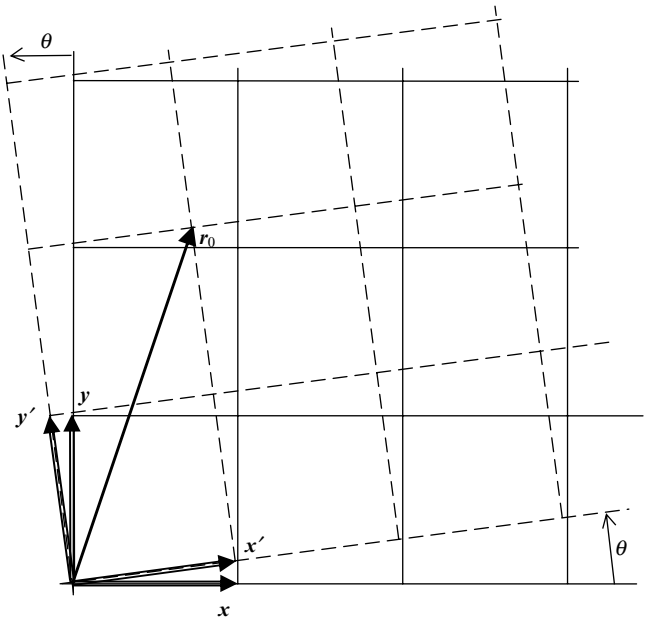

Figure 7.3 Change of basis vectors by rotation in the *xy* plane (spatial rotation). We are all familiar with this kind of basis transformation in which new basis vectors are obtained by rotating the spatial basis vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ through a given angle in the $\boldsymbol{xy}$ plane. The new basis vectors $\boldsymbol{x}', \boldsymbol{y}'$ are still orthonormal in our familiar Euclidean geometry. Two reference grids are shown. The one in solid lines is constructed from the basis vectors $\boldsymbol{x}, \boldsymbol{y}$, whereas the other, in dashed lines, is constructed from $\boldsymbol{x}', \boldsymbol{y}'$, which are rotated by an angle θ with respect to $\boldsymbol{x}, \boldsymbol{y}$. Each grid clearly forms an orthogonal, unit-spaced lattice. Since a vector is independent of the choice of the basis used, the components of any vector $\boldsymbol{r}_0$ representing some point in the plane may be read off in terms of the basis vectors of either grid. For the example shown $\boldsymbol{r}_0$ may be expressed as either $0.7\boldsymbol{x} + 2.1\boldsymbol{y}$ or $\boldsymbol{x}' + 2\boldsymbol{y}'$.

There is nothing to stop us from attempting a rotation in a plane that involves the time vector. If we try rotating the basis vectors in the $\boldsymbol{xt}$ plane, the result would be of the form $\boldsymbol{t}', \boldsymbol{x}', \boldsymbol{y}, \boldsymbol{z}$ where, for some other scalars α and β,

$$\begin{aligned} \boldsymbol{t}' &= \alpha\boldsymbol{t} + \beta\boldsymbol{x} \\ \boldsymbol{x}' &= \alpha\boldsymbol{x} + \beta\boldsymbol{t} \\ \text{with } \alpha^2 - \beta^2 &= 1 \end{aligned} \tag{7.17a}$$

The result is shown in Figure 7.4. The reader should ignore the additional detail in the figure at present and simply compare the new basis vectors and grid with those for a true rotation as shown in Figure 7.3. Although the transformed vectors look oblique rather than orthogonal, this is because we cannot represent them as they truly are. This is simply a conventional way of showing them with (α, β) and (β, α)

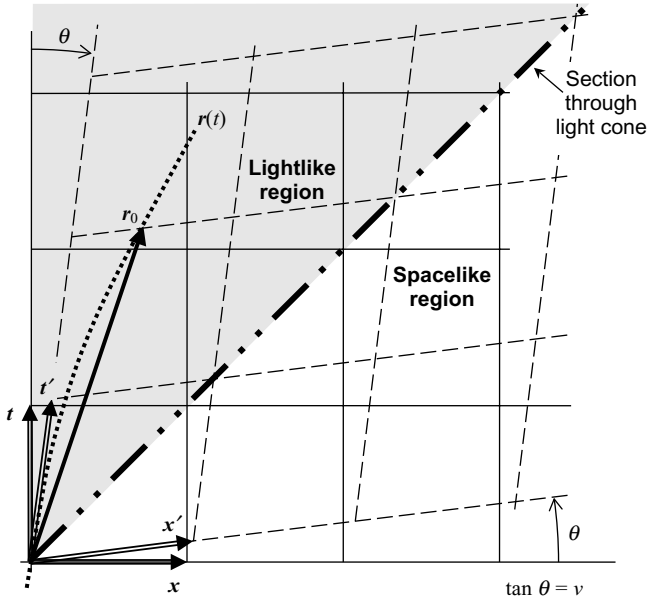

Figure 7.4 Change of basis vectors by rotation in the *xt* plane (Lorentz transformation). The $\boldsymbol{xt}$ plane of the spacetime $\boldsymbol{t}$-frame, in which the observer is at rest, is spanned by the basis vectors $\boldsymbol{t}$ and $\boldsymbol{x}$. The basis vectors $\boldsymbol{t}', \boldsymbol{x}'$ of a reference frame moving past the observer with relative velocity $v\boldsymbol{x}$, look neither normalized nor orthogonal, but this is because we think in terms of basic Euclidean geometry. Two reference grids are shown. The one in solid lines is for the $\boldsymbol{t}$-frame, and the other, in dashed lines, is for the $\boldsymbol{t}'$-frame. They are constructed so that each is an orthogonal, unit-spaced lattice in the non-Euclidean spacetime metric, where $\boldsymbol{t}^2 = -\boldsymbol{x}^2$ rather than $+\boldsymbol{x}^2$. Under this metric, the $\boldsymbol{t}'$ grid is equivalent to a "rotation" of the $\boldsymbol{t}$ grid through an angle $\theta = \arctan v$. Although it looks skewed, it nevertheless preserves both lengths and angles under the spacetime norm. The components of any event vector $\boldsymbol{r}_0$ may be read off in terms of the basis vectors of either grid. For the example shown, $\boldsymbol{r}_0$ may be expressed as either $2.1\boldsymbol{t} + 0.7\boldsymbol{x}$ or $2\boldsymbol{t}' + 0.4\boldsymbol{x}'$. The line $\boldsymbol{r}(t)$ represents the continuous sequence of events that makes up a trajectory, or history, leading to the event in question. The example shown could be the history of an accelerating particle that was at rest in the $\boldsymbol{t}'$-frame until $t = 0$. The section line represents part of the forward light cone from that point.

being treated as the coordinates of $\boldsymbol{t}'$ and $\boldsymbol{x}'$ in the Euclidean sense, that is to say, as though we had $\boldsymbol{t}^2 = 1$. However, with $\boldsymbol{t}^2 = -1$, the new basis vectors and grid are really orthogonal. The only significant change from a spatial rotation is that the condition for normalizing $\boldsymbol{t}'$ and $\boldsymbol{x}'$ is now $\alpha^2 - \beta^2 = 1$ rather than $\alpha^2 + \beta^2 = 1$. It therefore follows that there is still only one free parameter, θ, but it no longer corresponds to an actual angle of rotation. Although it is more properly described as an orthogonal transformation, the idea of a rotation in some sort of generalized sense still seems to be appropriate here, and since it has the benefit of being easy to visualize we will continue to use the term in this way.

Clearly, space and time are now mixed by Equation (7.17a), just as in Minkowski's bold idea. The key point, as alluded to at the end of Section 7.2, is that the time vector is not unique and, as can be seen here, nor is it totally isolated from

the spatial vectors. The change of basis implied by mixing space and time in this way is equivalent to a Lorentz transformation resulting in the time vector being transformed from $\boldsymbol{t}$ into $\boldsymbol{t}'$ where $\boldsymbol{t}' = \alpha\boldsymbol{t} + \beta\boldsymbol{x}$. In addition, $\boldsymbol{t}'$ must also have a new spatial partner, $\boldsymbol{x}' = \alpha\boldsymbol{x} + \beta\boldsymbol{t}$, so as to make $\boldsymbol{t}', \boldsymbol{x}', \boldsymbol{y}, \boldsymbol{z}$ an orthonormal basis. To distinguish the new set of basis vectors from the original $\boldsymbol{t}$-frame, we may simply call them the $\boldsymbol{t}'$-frame. It is easy enough to verify that the $\boldsymbol{t}'$-frame basis vectors are orthonormal since: we already know that $\boldsymbol{x}'$ and $\boldsymbol{t}'$ are normalized; second, it is clear that they are both still orthogonal to $\boldsymbol{y}$ and $\boldsymbol{z}$; and finally, $\boldsymbol{x}' \cdot \boldsymbol{t}' = (\alpha\boldsymbol{x} + \beta\boldsymbol{t}) \cdot (\alpha\boldsymbol{t} + \beta\boldsymbol{x}) = \alpha\beta\left(\boldsymbol{t}^2 + \boldsymbol{x}^2\right) = 0$.

In the (3+1)D view, time is excluded from the vectors with the consequence that a Lorentz transformation cannot be accommodated by some linear transformation that we can liken to a simple rotation. The move to spacetime rectifies this problem, and in doing so, what was previously *Galilean* relativity is now automatically extended to *special* relativity. It is the peculiarity of the metric signature that makes this possible simply because it reduces a Lorentz transformation to the same sort of operation as a spatial rotation—the choice of the plane of rotation being the only real difference. We will study this in some detail in Chapter 9.

7.7 EVENTS AND HISTORIES

7.7.1 Events

We introduced the idea in Section 7.2 that a constant spacetime vector defines an event; for example, $\boldsymbol{r}_0 = t_0\boldsymbol{t} + x_0\boldsymbol{x} + y_0\boldsymbol{y} + z_0\boldsymbol{z}$ occurs at the usual position coordinates (x_0, y_0, z_0) only at the specific scalar time t_0. Unlike a point in 3D space, which exists in perpetuity, an event only exists at the specified time. For example, an event could represent the collision of two particles or the emission of a flash of light.

7.7.2 Histories

In spacetime, a fixed point in space corresponds to a straight line parallel to the time axis. In the $\boldsymbol{t}$-frame, for instance, we must have $\boldsymbol{r}(t) = t\boldsymbol{t} + \underset{\sim}{\boldsymbol{r}}_0$ where $\underset{\sim}{\boldsymbol{r}}_0 = x_0\boldsymbol{x} + y_0\boldsymbol{y} + z_0\boldsymbol{z}$ is a constant spatial vector that gives the location of the point. The line begins at the time the point starts to exist and continues until it ceases to do so. More generally, we can say the vector $\boldsymbol{r}(t)$ is the history of some particle whose spatial location is given by $\underset{\sim}{\boldsymbol{r}}_0$. It is a continuum of events that tells us about the past, present, and future of the particle over some finite interval of time. In Section 7.5, we referred to this as a trajectory, and while the term world line is also fairly common, "history" immediately conveys the idea with little other explanation or qualification. Equations (7.28) and (7.29) and Figures 7.1, 7.4, 7.6, 10.3 and 11.1 all include further examples of spacetime histories, not all of which are simple straight lines, and those that are straight lines are not necessarily parallel to the time axis.

7.7.3 Straight-Line Histories and Their Time Vectors

For the purposes of this discussion, it will be useful to recall the properties of the spacetime velocity (Section 7.5) and to refer to Figure 7.5.

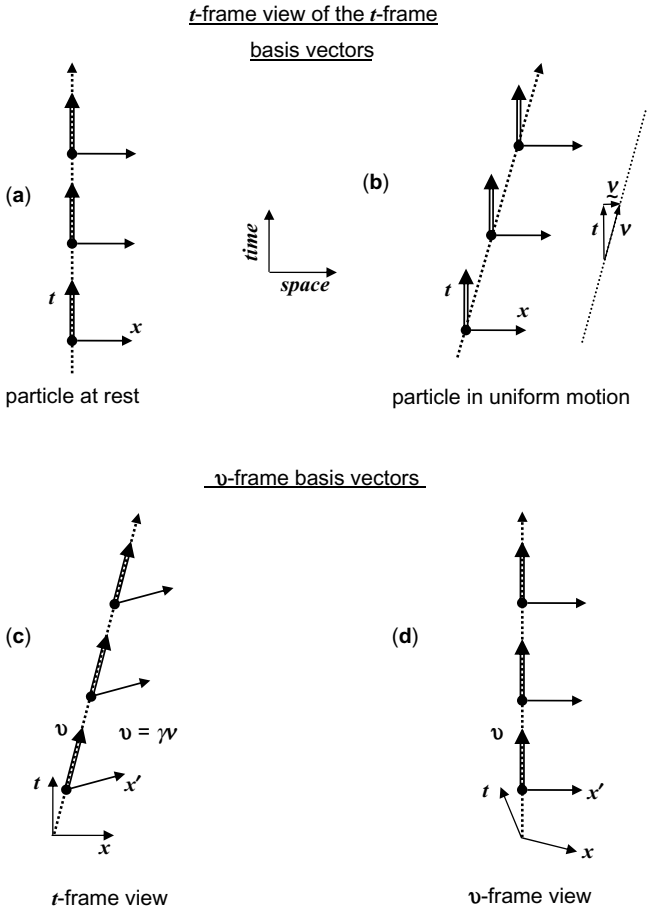

Figure 7.5 The time vector of a moving particle. In (a), we show three consecutive snapshots of a particle at rest taken at ~1.5 unit time intervals apart. In each case, the time vector, $\boldsymbol{t}$, and one representative spatial vector, $\boldsymbol{x}$, are shown. In (b), we see how this transfers to the case of a particle in uniform motion, with each snapshot being displaced to the right by the same distance on each occasion. Although the time vectors are still parallel to each other, they are no longer parallel to the straight line that represents the particle's history. But in the rest frame of the particle, the $\boldsymbol{\upsilon}$-frame, they are parallel to the history, as in (c). The spatial axis in the direction of motion must also be skewed so that both axes remain orthogonal under the spacetime metric. But in the particle's rest frame, the view must look just the same as the particle at rest situation in (a). This is provided for if we change basis vectors from the $\boldsymbol{t}$-frame to the $\boldsymbol{\upsilon}$-frame as shown in (d).

The history of some particle moving with constant velocity is a straight line that is oblique to the time axis of the frame from which we are observing the motion. For example, in the $\boldsymbol{t}$-frame, $\boldsymbol{r}(t) = t\boldsymbol{v} + \underset{\sim}{\boldsymbol{r}}_0$ is the history of a particle with velocity $\boldsymbol{v}$, as is readily confirmed from $\partial_t \boldsymbol{r}$. Since the magnitude of $\boldsymbol{v}$ is given by Equation (7.16) as $\left(1-v^2\right)^{1/2}$, it can be normalized by multiplying it by $\gamma = \left(1-v^2\right)^{-(1/2)}$. The resulting unit vector $\boldsymbol{\upsilon} = \gamma\boldsymbol{v}$ is referred to as the proper velocity of the particle, a concept which is more fully explained in Sections 10.4 and 10.5. It has a negative square and, more specifically, $\boldsymbol{\upsilon}^2 = -1$. Since in the $\boldsymbol{t}$-frame (that is to say using $\boldsymbol{t}$ as the time vector) $\boldsymbol{v}$ can be written as $\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}}$ where $\underset{\sim}{\boldsymbol{v}}$ is a purely spatial vector with magnitude v, it follows that $\boldsymbol{\upsilon}$ is more or less parallel to $\boldsymbol{t}$ for low velocities where $v \ll 1$, that is to say, $|\underset{\sim}{\boldsymbol{v}}| \ll |\boldsymbol{t}|$.[2] At much higher velocities, however, $\boldsymbol{\upsilon}$ and $\boldsymbol{t}$ may differ significantly. We can now make a relatively trivial rearrangement of the particle's history to obtain

$$\begin{aligned} \boldsymbol{r}(t) &= t\boldsymbol{v} + \underset{\sim}{\boldsymbol{r}}_0 \\ &= \gamma^{-1} t(\gamma\boldsymbol{v}) + \underset{\sim}{\boldsymbol{r}}_0 \\ &= \left(\gamma^{-1} t\right)\boldsymbol{\upsilon} + \underset{\sim}{\boldsymbol{r}}_0 \\ &= \tau\boldsymbol{\upsilon} + \underset{\sim}{\boldsymbol{r}}_0 \end{aligned} \tag{7.18}$$

where

$$\begin{aligned} \gamma &= \left(1-v^2\right)^{-\frac{1}{2}} \\ \boldsymbol{\upsilon} &= \gamma\boldsymbol{v} = \gamma\left(\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}}\right) \\ \tau &= \gamma^{-1} t \end{aligned} \tag{7.19}$$

Now,

- there is no time-dependent spatial vector in Equation (7.18);
- if $\boldsymbol{\upsilon}$ is taken to be the time vector and τ the time, $\boldsymbol{r}(\tau)$ has the same form as the history of a particle at rest; and
- together with the fact $\boldsymbol{\upsilon}^2 = -1$, this qualifies $\boldsymbol{\upsilon}$, *the particle's proper velocity*, as being *a time vector*.

We may therefore conclude that $\boldsymbol{r}(\tau) = \tau\boldsymbol{\upsilon} + \underset{\sim}{\boldsymbol{r}}_0$ *is indeed* the history of a particle at rest where $\boldsymbol{\upsilon}$ is the local time vector. But where the particle is at rest defines its own rest frame, so that

[2] Note, however, that if were to deduce this by evaluating $-\boldsymbol{\upsilon}\cdot\boldsymbol{t}$, the result γ represents the hyperbolic cosine of the angle between $\boldsymbol{\upsilon}$ and $\boldsymbol{t}$ rather than its cosine, as would be the case with two spatial vectors. Similarly, v represents a hyperbolic tangent rather than a tangent. The geometry of spacetime, therefore, is often described as hyperbolic rather than Euclidean.

- $\boldsymbol{\upsilon}$ is the time vector in the particle's rest frame;
- we may therefore call this frame the $\boldsymbol{\upsilon}$-frame; and
- here τ is the local time parameter, or *proper time*.

It will be noted, however, that although $\underset{\sim}{\boldsymbol{r}}_0$ is a constant vector, it is not actually orthogonal to $\boldsymbol{\upsilon}$. Orthogonality may be restored by replacing $\underset{\sim}{\boldsymbol{r}}_0$ with $\underset{\sim}{\boldsymbol{r}}_0' + \boldsymbol{\upsilon}\delta\tau$ and letting $\delta\tau$ be equal to $-\underset{\sim}{\boldsymbol{r}}_0 \cdot \boldsymbol{\upsilon}$ so that $\underset{\sim}{\boldsymbol{r}}_0' \cdot \boldsymbol{\upsilon} = \left(\underset{\sim}{\boldsymbol{r}}_0 - \boldsymbol{\upsilon}\delta\tau\right) \cdot \boldsymbol{\upsilon} = 0$. This adjustment alters Equation (7.18) to

$$\boldsymbol{r}(\tau) = (\tau + \delta\tau)\boldsymbol{\upsilon} + \underset{\sim}{\boldsymbol{r}}_0' \tag{7.20}$$

The net effect is only to introduce a difference $\delta\tau$ in the synchronization between the time parameters of the two frames. Although this issue is not so important to us, it will be familiar to many as one of the central points in any discussion of the implications of special relativity.

The symbols $\boldsymbol{\upsilon}$ for a local time vector and τ for the corresponding local time are common in the spacetime literature, usually with reference to a particle on some trajectory. There will be more about this in Sections 10.4 and 10.5 but for the moment, $\boldsymbol{\upsilon}$ and τ simply replace the labels $\boldsymbol{t}'$ and t' we have been applying in relation to some alternative frame to the $\boldsymbol{t}$-frame.

In Section 7.6, we took a geometric approach to changing basis vectors by rotating them in 4D. In particular, we argued that a rotation in the $\boldsymbol{xt}$ plane would generate a new time vector of the form $\boldsymbol{t}' = \alpha\boldsymbol{t} + \beta\boldsymbol{x}$ (Equation 7.17a). Having generated a new time vector $\boldsymbol{\upsilon} = \gamma\left(\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}}\right)$ on physical grounds, that is to say through a change of inertial reference frame, we can now look back on this from a different perspective. It must be possible to bring these two approaches together by identifying $\boldsymbol{t}' = \alpha\boldsymbol{t} + \beta\boldsymbol{x}$ with $\boldsymbol{\upsilon} = \gamma\left(\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}}\right)$ for some suitable choice of the real scalars α and β, that is to say we can identify

- $\boldsymbol{t}'$ with $\boldsymbol{\upsilon}$,
- α with γ, and
- $\beta\boldsymbol{x}$ with $\gamma\underset{\sim}{\boldsymbol{v}}$.

In which case, it follows that

- the $\boldsymbol{t}'$-frame is the same as the $\boldsymbol{\upsilon}$ frame;
- the $\boldsymbol{t}'$-frame therefore moves with spatial velocity $\underset{\sim}{\boldsymbol{v}} = v\boldsymbol{x}$ with respect to the $\boldsymbol{t}$-frame;
- the transformation of $\boldsymbol{t}$ into $\boldsymbol{t}'$ by a rotation in the $\boldsymbol{xt}$ plane is the same as changing the rest frame from the $\boldsymbol{t}$-frame to the $\boldsymbol{t}'$-frame;
- $\alpha = \gamma = \left(1 - v^2\right)^{-(1/2)}$ and $\beta = \gamma v$; and
- we may then restate Equation (7.17a) as

$$\begin{aligned} \boldsymbol{t}' &= \gamma \boldsymbol{t} + \gamma v \boldsymbol{x} \\ \boldsymbol{x}' &= \gamma \boldsymbol{x} + \gamma v \boldsymbol{t} \\ \gamma &= \left(1 - v^2\right)^{-\frac{1}{2}} \end{aligned} \tag{7.17b}$$

As will be confirmed in Section 9.4, the simple rearrangement of the history of a particle in uniform motion so as to appear the same as the history of a point at rest has quite remarkably divulged the Lorentz transformation that we simply associated with a change to an alternative set of spacetime basis vectors. The only real ingredients at work here are

- time being a vector,
- the spacetime metric,
- the equivalence between a particle seen as being in motion in one frame and a particle seen as being stationary in another, and
- the equivalence of a change of reference frame to a rotation of the basis vectors in a timelike plane.

While, as we have seen, we may alter the functional form of a simple spacetime vector by switching from one frame to another, by which we really mean transforming from the basis vectors of one frame to the basis vectors of the other, we must nevertheless continue to observe the principle that the vectors themselves do not depend on the choice of frame. Be aware, however, that this does not generally apply to a derived vector such as velocity.

7.7.4 Arbitrary Histories

The same principles may be extended to particles with arbitrary histories for it is only necessary to know how $\mathfrak{v}$ develops as a function of τ. In general, therefore, the history of any particle depends on, and only on, the evolution of *its local time vector*. Now, a particle's history is just the same thing as its spacetime trajectory, but this has a different interpretation from a 3D trajectory where at any instant we observe the particle heading along some direction in space. In a spacetime trajectory, the particle heads along its own time vector! Strange as it may at first seem, this is true irrespective of how the particle may move in any spatial direction. To see the basis of this, we only have to reflect once again that the particle is at rest in its own frame, the $\mathfrak{v}$-frame, and so here the spatial part of $\boldsymbol{r}(\tau)$, its history, must be constant. We can therefore focus entirely on the time part of its history. Taking the time vector to be $\mathfrak{v}$ at τ, then as long as the trajectory is smooth we can take it as being close to uniform, that is, a straight line, over a sufficiently small interval between τ and $\tau + \delta\tau$. Therefore, at $\tau + \varepsilon$ on this interval we have $\boldsymbol{r}(\tau + \varepsilon) = \varepsilon\mathfrak{v} + \boldsymbol{r}(\tau)$, giving $\delta\boldsymbol{r}$, the change in $\boldsymbol{r}$, as being $\boldsymbol{r}(\tau + \varepsilon) - \boldsymbol{r}(\tau) = \varepsilon\mathfrak{v}$. Since $\delta\boldsymbol{r} = \varepsilon\mathfrak{v}$, the direction of motion is clearly along $\mathfrak{v}$; in fact, taking limits, we can say $\partial_\tau \boldsymbol{r} = \lim_{\varepsilon \to 0} \delta\boldsymbol{r}/\varepsilon = \mathfrak{v}$.

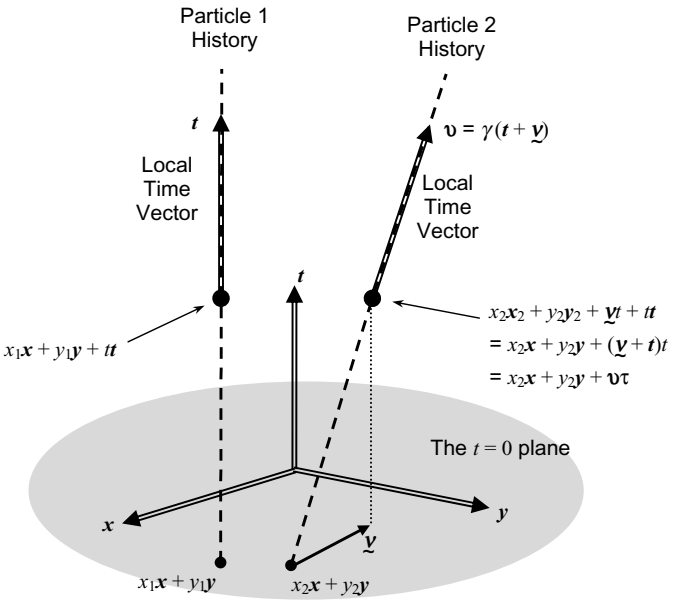

Figure 7.6 Particle histories and local time vectors. As in Figure 7.1, the time axis lies vertically in the page, and we see only a visual representation of the ***xy*** spatial plane. The basis vector for time is given as ***t*** and so we are in the ***t***-frame. Particle 1 is at rest in this frame so that its history is along an upward pointing vertical line parallel to ***t***. The event on its history at time t is given by $t\boldsymbol{t} + x_1\boldsymbol{x} + y_1\boldsymbol{y}$, and so all the time dependency of the history is contained in the term $t\boldsymbol{t}$. We can therefore say that the local time for this particle vector is ***t***. The second particle is clearly in motion because the spatial part of its history depends on time. The overall time-dependent part here is $t\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}}t = t(\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}})$. By normalizing $\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}}$, we arrive at particle 2's local time vector $\boldsymbol{\upsilon} = (\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}})/|\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}}|$ from which we can write its history in the form $\tau\boldsymbol{\upsilon} + x_2\boldsymbol{x} + y_2\boldsymbol{y}$. Here $\tau = |\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}}|t$ is the particle's local, or proper, time and $\boldsymbol{\upsilon}$ doubles as the particle's proper velocity. Note however that the basis vectors ***x*** and ***y*** need to be adjusted if we want all the basis vectors of the $\boldsymbol{\upsilon}$-frame, as we may now call it, to be orthonormal.

Figure 7.6 shows an example of how the histories of two different particles may be described, one at rest in the ***t***-frame and the other at rest in the $\boldsymbol{\upsilon}$-frame. Although the histories are shown as being more or less uniform, we can think of these as being small intervals during which the time vectors are effectively constant.

As seen from some frame, say the ***t***-frame which we generally take to be our own rest frame, a particle moving on an arbitrary trajectory will therefore have a local time vector that

- generally differs from the time vector of a particle at rest,
- generally includes a spatial part,

- but always remain normalized,
- is directly associated with its velocity, and
- always points in the direction in which the particle's spacetime trajectory is heading—that is to say it is the tangent vector to the particle's history.

We can imagine the particle carrying the $\mathfrak{v}$-frame basis vectors with it as it travels along, for as seen from the particle's perspective, these must remain fixed. This will be the case as long as the particle itself does not rotate, but since this aspect of particle dynamics is of no relevance to us, we can put it from our minds. The rest frame of a particle, that is to say its $\mathfrak{v}$-frame, is of particular interest to us because we know that here

- the particle interacts only with what is "sees" as being the *electric* field and
- the electromagnetic field of the particle itself may be determined from Coulomb's law.

It should be noted, however, that the electro-magnetic field of the particle is given directly by Coulomb's law only when it is in uniform motion. When it is accelerating, as we shall see in Chapter 12, it becomes necessary to evaluate the field indirectly via the electromagnetic potential. However, in either case, the field that we observe and what we see as the particle's interaction with the electromagnetic field are both crucially dependent on knowing $\mathfrak{v}$.

7.8 THE SPACETIME FORM OF ∇

The spacetime vector derivative is an essential element of Maxwell's equations, and it is also essential to the evaluation of the complete electromagnetic field of a point charge undergoing acceleration. However, it is necessary to discuss it in some detail because it is subject to the spacetime metric, a critical factor that introduces nuances not only into its form but also into the process of interepreting its function in terms of (3+1)D constructs such as $\boldsymbol{\nabla} + \partial_t$.

There is an established precedent defining the *vector* differential operator ∇ in a Euclidean space of any dimension, namely

$$\nabla \equiv \sum_k \boldsymbol{e}_k \partial_k \tag{7.21}$$

where the index k runs over all the orthonormal basis vectors $\boldsymbol{e}_k$. Any vector $\boldsymbol{u}$ may be written in terms of these vectors as $\sum_k u_k \boldsymbol{e}_k$, but there is the tacit assumption that the $\boldsymbol{e}_k$ forms an orthonormal basis in the strict Euclidean sense with $\boldsymbol{e}_k \cdot \boldsymbol{e}_j = \delta_{kj}$ where δ_{kj} takes the value 1 when $k = j$ and is otherwise 0. Now, we may choose some other sort of basis and system of measure, but this represents the simplest starting point. In Newtonian space, the standard basis $\mathbf{x},\mathbf{y},\mathbf{z}$ conforms to this Euclidean requirement in that $\mathbf{x}^2 = \mathbf{y}^2 = \mathbf{z}^2 = 1$ and $\mathbf{y}\cdot\mathbf{z} = \mathbf{z}\cdot\mathbf{x} = \mathbf{x}\cdot\mathbf{y} = 0$, but the

orthonormal spacetime basis $\boldsymbol{t},\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}$ does not conform since $\boldsymbol{t}^2=-\boldsymbol{x}^2$ breaks the Euclidean rule. No transformation of basis vectors can get round this problem and so the space is deemed to be non-Euclidean (but yet since it is so very similar, it is sometimes referred to as being quasi-Euclidean). The more general rule that is now required is $\boldsymbol{e}_k\cdot\boldsymbol{e}_j=g_{kj}$ where $g_{kj}=\delta_{kj}g_{kk}$ and some of the g_{kk} may be -1 rather than $+1$. This new rule still allows the basis to be orthogonal since $\boldsymbol{e}_k\cdot\boldsymbol{e}_j=0$ for $k\neq j$, but normalization now includes the possibility that $\boldsymbol{e}_k^2=g_{kk}=\pm1$. The g_{kk} therefore define the metric signature so that, for example, $(+---)$ simply represents the signs of the g_{kk} in row vector form. Since all the spatial vectors must have the same value of g_{kk}, there are only two possible choices of metric signature for spacetime. If we choose our usual spacetime metric signature to be $(-+++)$ where $\boldsymbol{t}^2=-1$, then $g_{tt}=-1$ with $g_{xx}=g_{yy}=g_{zz}=+1$.

Some care is also required with the interpretation of "length", or what we would more generally call the measure of an object. As we have already noted, when $\boldsymbol{u}$ is a spacetime vector we need to take care of a possible negative result for $\boldsymbol{u}^2$. In general, therefore, $|\boldsymbol{u}|=\left|\boldsymbol{u}^2\right|^{1/2}$. A similar problem arises in differentiation. A good test case in any dimension of space is differentiation of the position vector $\boldsymbol{r}=\sum_k x_k\boldsymbol{e}_k$ for which it is readily confirmed that $\nabla\boldsymbol{r}$ reduces to $\nabla\cdot\boldsymbol{r}=\sum_k\partial_k x_k$ so that $\nabla\boldsymbol{r}=N$, the dimension of the space. When we try this with our Euclidean (3+1)D space, the result is obviously upheld, but in spacetime with the definition of ∇ as in Equation (7.21), the result is 2 rather than 4! Looking in more detail at the problem, just as $\nabla(x\boldsymbol{x})=1$ returns the unit measure along $\boldsymbol{x}$, $\nabla(t\boldsymbol{t})$ should return the unit measure of time, but we find instead $\nabla(t\boldsymbol{t})=\boldsymbol{t}\partial_t(t\boldsymbol{t})=\boldsymbol{t}^2=-1$. The derivative of an increasing quantity should be positive, as in the case of all the spatial contributions to $\nabla\boldsymbol{r}$, such as $\boldsymbol{x}\partial_x(x\boldsymbol{x})=\boldsymbol{x}^2=+1$. To fix the problem, it is necessary to change the sign of the partial time derivative and so, in order to suit the non-Euclidean metric, we must modify the form of Equation (7.21) to

$$\boldsymbol{\nabla}=-\boldsymbol{t}\partial_t+\boldsymbol{x}\partial_x+\boldsymbol{y}\partial_y+\boldsymbol{z}\partial_z \tag{7.22}$$

This at once restores the desired property $\boldsymbol{\nabla}\boldsymbol{r}=N$. For those who are interested, this is explained in a somewhat more rigorous way in Appendix 14.7.

The vector operator $\boldsymbol{\nabla}$ is no different from any other vector in that it is independent of the choice of orthonormal basis, so that we could just as well express it as $\boldsymbol{\nabla}=-\boldsymbol{\upsilon}\partial_\upsilon+\boldsymbol{x}'\partial_{x'}+\boldsymbol{y}'\partial_{y'}+\boldsymbol{z}'\partial_{z'}$ in the $\boldsymbol{\upsilon}$-frame. In fact, we represent it here in terms of a basis only for simplicity. Hestenes offers a more general definition [46, chapter 2]. Also, as with other vectors, the dimension is implied and generally has to be taken from the context. Although the symbol $\Box$ has been used in the past for the spacetime form, no special symbol is actually required. While just now we did use the general form ∇, in keeping with our notation for the labels given to ordinary vectors, we use italic as in $\boldsymbol{\nabla}$ for the spacetime vector derivative and bold erect $\boldsymbol{\nabla}$ in (3+1)D. Although we may express them in different forms, symbolically, they all mean the same thing. In addition, just as for other vectors, we allow the use of the under-tilde notation to separately identify the spatial part of $\boldsymbol{\nabla}$ in a given frame. We

may therefore write $\boldsymbol{\nabla} = -\boldsymbol{t}\partial_t + \underset{\sim}{\nabla}$, which clearly has some potentially useful correspondence with the (3+1)D paravector form $\partial_t + \boldsymbol{\nabla}$.

7.9 WORKING WITH VECTOR DIFFERENTIATION

Here we illustrate some of the rules that allow the manipulation of expressions involving $\boldsymbol{\nabla}$. The general principles also apply to the 3D vector derivative $\boldsymbol{\nabla}$.

The key point to be borne in mind is that while the normal rules of differentiation apply, the order of the vectors in any product must be maintained. It is therefore necessary to allow for the possibility of $\boldsymbol{\nabla}$ being separated from its operand by some intervening term. In such cases, therefore, it is customary to over-mark the operator $\boldsymbol{\nabla}$ and the term it acts upon with some chosen symbol so as to keep track of the required association, for example, $\overset{\circ}{\boldsymbol{\nabla}}\boldsymbol{u}\overset{\circ}{\boldsymbol{v}}$ (over-marking with a simple dot would be confusing here as this is conventionally allotted to differentiation with respect to proper time, which we will come to in due course). When there is no such mark, however, $\boldsymbol{\nabla}$ is assumed to act on the term immediately to its right. For example, in $\overset{\circ}{\boldsymbol{\nabla}}\boldsymbol{u}\overset{\circ}{\boldsymbol{v}}$, it is $\boldsymbol{v}$ that is being differentiated, while in $\boldsymbol{\nabla}\boldsymbol{u}\boldsymbol{v}$, it is $\boldsymbol{u}$. In addition, in order to be absolutely clear, we will use brackets even though they may not be formally required, for example, as in $(\boldsymbol{\nabla}\boldsymbol{u})\boldsymbol{v}$, which is equivalent to $\boldsymbol{\nabla}\boldsymbol{u}\boldsymbol{v}$. We now give some useful identities involving $\boldsymbol{\nabla}$.

Given any differentiable vector functions $\boldsymbol{u}$, $\boldsymbol{v}$, and a vector function $\boldsymbol{f}(\lambda)$ that depends only a scalar parameter,

$$\boldsymbol{\nabla}(\boldsymbol{u}\boldsymbol{v}) = (\boldsymbol{\nabla}\boldsymbol{u})\boldsymbol{v} + \overset{\circ}{\boldsymbol{\nabla}}\boldsymbol{u}\overset{\circ}{\boldsymbol{v}} \quad \text{(product rule)} \tag{7.23}$$

$$\boldsymbol{\nabla}\boldsymbol{f}(\lambda) = (\boldsymbol{\nabla}\lambda)\partial_\lambda \boldsymbol{f} \quad \text{(chain rule)} \tag{7.24}$$

In addition,

$$\begin{aligned} \boldsymbol{\nabla}\boldsymbol{r}^2 &= (\boldsymbol{\nabla}\boldsymbol{r})\boldsymbol{r} + \overset{\circ}{\boldsymbol{\nabla}}\boldsymbol{r}\overset{\circ}{\boldsymbol{r}} \\ &= 4\boldsymbol{r} - 2\boldsymbol{r} \\ &= 2\boldsymbol{r} \end{aligned} \tag{7.25}$$

These three identities also hold in 3D, but as already mentioned, the following is different from the 3D result where $\boldsymbol{\nabla}\mathbf{r} = 3$:

$$\begin{aligned} \boldsymbol{\nabla}\boldsymbol{r} &= \left(-\boldsymbol{t}\partial_t + \boldsymbol{x}\partial_x + \boldsymbol{y}\partial_y + \boldsymbol{z}\partial_z\right)\left(t\boldsymbol{t} + x\boldsymbol{x} + y\boldsymbol{y} + z\boldsymbol{z}\right) \\ &= -\boldsymbol{t}^2 + \boldsymbol{x}^2 + \boldsymbol{y}^2 + \boldsymbol{z}^2 \\ &= 4 \end{aligned} \tag{7.26}$$

As an example of dealing with overdots, in order to evaluate $\overset{\circ}{\boldsymbol{\nabla}}\boldsymbol{u}\overset{\circ}{\boldsymbol{r}}$ it is only necessary to treat the intervening term $\boldsymbol{u}$ as though it were a constant:

$$\begin{aligned}\mathring{\boldsymbol{\nabla}}\boldsymbol{u}\,\mathring{\boldsymbol{r}} &= \left(-\boldsymbol{t}\partial_t + \boldsymbol{x}\partial_x + \boldsymbol{y}\partial_y + \boldsymbol{z}\partial_z\right)\boldsymbol{u}\left(t\boldsymbol{t} + x\boldsymbol{x} + y\boldsymbol{y} + z\boldsymbol{z}\right)\\ &= -\boldsymbol{tut} + \boldsymbol{xux} + \boldsymbol{yuy} + \boldsymbol{zuz}\\ &= -2\boldsymbol{u}\end{aligned} \tag{7.27}$$

This clearly illustrates a trap to be wary of in that $(\boldsymbol{\nabla r})\boldsymbol{r} \neq \mathring{\boldsymbol{\nabla}}\boldsymbol{r}\,\mathring{\boldsymbol{r}}$. Getting the final result here is a little tricky because the order of the vectors in any term such as $\boldsymbol{xux}$ or $-\boldsymbol{tut}$ can only be altered within the rules. While we may replace $\boldsymbol{xy}$ with $-\boldsymbol{yx}$, we cannot simply replace $\boldsymbol{ux}$ with $-\boldsymbol{xu}$. For example, on replacing $\boldsymbol{u}$ with $u_t\boldsymbol{t} + u_x\boldsymbol{x} + u_y\boldsymbol{y} + u_z\boldsymbol{z}$ in $\boldsymbol{xux}$, we find $\boldsymbol{x}\left(u_t\boldsymbol{t} + u_x\boldsymbol{x} + u_y\boldsymbol{y} + u_z\boldsymbol{z}\right)\boldsymbol{x} = -u_t\boldsymbol{t} + u_x\boldsymbol{x} - u_y\boldsymbol{y} - u_z\boldsymbol{z}$. The other three terms work out in similar fashion with the positive sign appearing against each basis vector in turn. All four such terms therefore add up to $\boldsymbol{u} - 3\boldsymbol{u} = -2\boldsymbol{u}$. In 3D, the corresponding result is $\boldsymbol{\nabla}\mathbf{u}\,\mathring{\mathbf{r}} = \mathbf{u} - 2\mathbf{u} = -\mathbf{u}$.

The basis of these and other useful identities are to be found in Reference 33 (chapter 2).

7.10 WORKING WITHOUT BASIS VECTORS

It was hinted at the end of Section 7.1 that it could be advantageous to be able to set down the equations we want to work with in a general form and to solve these in some equally general way without reference to a specific basis. We need to choose a basis only when it is convenient to do so, for example, when we want to see the results expressed in some chosen frame that defines a reference frame plus axis system. For example, our own rest frame along with our preferred spatial axes will be a frequent choice. Another choice could be the rest frame of some other body such as a moving point charge, for here the Lorentz force simply reduces to the force due to the electric field seen in that frame. We can also try this the other way round, because in the rest frame of a uniformly moving charge, its own electromagnetic field will just be its Coulomb field. If we can put this into an appropriately general spacetime form, then we should be able to derive the electromagnetic field that we will observe in any other frame simply by applying the appropriate basis. This turns out to be the same thing as choosing the appropriate time vector for each frame.

With vectors, it is not possible to determine absolute position; it is only possible to give position with respect to some stated origin and orientational framework. We are quite happy to work with vectors without this information, but we nevertheless take it as implied that there is an origin, from which we draw our vectors, and an **x**, **y**, **z** emanating from that origin as our preferred directions in space. We tend to be much better at abstraction when it comes to measure, for it bothers us little if we omit to mention that the units for x, y, and z are inches, meters, or whatever. The assumption of standard basis vectors is very useful for visualizing any sort of configuration that is a function of position, but it generally leads to seeing things in terms of **x**, **y**, **z** and x, y, z, which are simply constructs that we ourselves have imposed. We can be free of this constrained manner of thinking by writing vector

equations in an abstract way that takes care to avoid explicit references to either origin or orientation. For example, given that λ is some monotonic function of time, the parametric equation

$$\mathbf{r}(\lambda) = \mathbf{r}_0 + \lambda\mathbf{r}_1 + \lambda^2\mathbf{r}_2 \tag{7.28}$$

describes the trajectory of a particle through space without expressing the vectors involved in terms of any assumed basis. The origin and units of measurement are simply left to be fixed as and when we like. To emphasize the point, we have written λ instead of t and have not even specified a relationship between them.

As a more tangible example, we can express ordinary parabolic motion in a coordinate-free form as

$$\mathbf{r}(t) = \mathbf{r}_0 + \mathbf{v}t + \mathbf{h}t^2 \tag{7.29}$$

Although this equation is in the same form as Equation (7.28), it does refer to t explicitly. It will serve for any point mass undergoing uniform acceleration by choosing $\mathbf{r}_0$, $\mathbf{v}$, and $\mathbf{h}$ as appropriate to the situation. From here, if we have a particular situation in mind, for example, the motion of a projectile launched from the ground, we may substitute $-\frac{1}{2}g\mathbf{z}$ for $\mathbf{h}$ where $\mathbf{z}$ is orientated vertically. Depending on what suits our purpose, we may introduce a vector, say $\mathbf{w}$, which is horizontal and lies in the plane of motion, thereby reducing the problem to being 2D, as in

$$\mathbf{r}(t) = \mathbf{r}_0 + v_w t\,\mathbf{w} + \left(v_z t - \tfrac{1}{2}gt^2\right)\mathbf{z} \tag{7.30}$$

Clearly, even these choices are not the only possibilities. The very same considerations apply to equations involving spacetime vectors, *in particular, time*. As discussed in Section 7.6, the time vector is not unique, and time and space may be mixed. While $\boldsymbol{t}$ is the unit time vector *where we are*, according to Equation (7.19), $\boldsymbol{\upsilon} = \gamma\boldsymbol{v}$ is the unit time vector in the rest frame of a particle with spacetime velocity $\boldsymbol{v}$. It is therefore impossible to adhere to some arbitrarily chosen fixed basis vector for time. That would only be possible when the velocities involved are so small compared with the speed of light that to all intents $\gamma = 1$. It may also come as a bit of a surprise that even this condition is insufficient when it comes to the origin of something as commonplace as the magnetic field.

If spacetime is to serve any useful purpose, we therefore cannot expect $\boldsymbol{t}$, $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ to fit every conceivable situation. Rather than start out with such a basis set and deal with the subsequent problem of transforming to a different basis set whenever a different time vector is encountered, it will prove more fruitful to develop equations in a more general form that needs to be expressed in terms of a basis only when we find it convenient to do so. We will need to adopt this philosophy when it comes to the discussion of the electromagnetic field of a moving point charge.

7.11 CLASSIFICATION OF SPACETIME VECTORS AND BIVECTORS

We now return to the terms timelike and spacelike that we introduced in Sections 7.2 and 7.3. While the assignment of these terms to specific vectors and bivectors was done on an intuitive basis, that is to say, according to whether or not the time vector is involved, it is clear that the mixing of space and time means that a more general definition is required. For the vectors at least, we can use the fact that spatial vectors and the time vector have squares of opposite sign, but since it is possible to have a square equal to 0, we also need a category for null vectors, namely "light-like". It will be recalled from the discussion of the spacetime norm that null vectors are associated with things that propagate at the speed of light.

Any vector $\boldsymbol{u}$ is therefore classified as

- *timelike*, if $\boldsymbol{u}^2$ has the same sign as $\boldsymbol{t}^2$;
- *lightlike* (or *null*), if $\boldsymbol{u}^2 = 0$; and
- *spacelike*, if $\boldsymbol{u}^2$ has the opposite sign to $\boldsymbol{t}^2$.

In addition, any such timelike or lightlike vector can be said to be future pointing if $\boldsymbol{u} \cdot \boldsymbol{t}$ has the same sign as $\boldsymbol{t}^2$.

These definitions are independent of which time vector or metric signature we choose, and so reference to $\boldsymbol{t}$ is purely symbolic inasmuch as we can use whatever time vector we like. This implies that a vector will still retain its original timelike, lightlike, or spacelike character in any chosen frame. Even if a vector appears to comprise a mix of timelike and spacelike parts, it will still fall into one of the above categories.

Timelike vectors are of the real world since we travel along this kind of vector at a speed less than the speed of light. Lightlike or null vectors lie along paths taken by light. All the possible lightlike vectors from a given point define a double cone of vectors that point either to the future or the past, as shown in Figure 7.1. They connect the *source event* where an electromagnetic disturbance originates with any event at which it is subsequently observed, that is, the *observation event*. Any such vector may point either from the source to observation event or vice versa. Vectors in the remaining category, spacelike, join events that are not physically connectable, that is to say, neither event can affect the other because that would require a speed of interaction greater than the speed of light.

The terms timelike and spacelike are not to be confused with the terms spatial and temporal that were introduced in Section 7.3.1. Although they may seem similar, the meanings of the latter are somewhat different in that they depend on the chosen time vector. Spatial vectors belong to the orthogonal space of the time vector, whereas a temporal vector is, somewhat trivially, in the space spanned by the time vector. To make the distinction perfectly clear, take $\boldsymbol{t}$ as being some arbitrarily chosen time vector. We can then assert that any vector $\boldsymbol{u}$ is

- *Spatial*, if $\boldsymbol{u}\cdot\boldsymbol{t}=0$.
  - In the $\boldsymbol{t}$-frame, any linear combination of $\boldsymbol{x},\boldsymbol{y}$, and $\boldsymbol{z}$ qualifies.
- *Temporal*, if $\boldsymbol{u}\wedge\boldsymbol{t}=0$, that is to say it has no spatial part.
  - In the $\boldsymbol{t}$-frame, any vector of the form $\lambda\boldsymbol{t}$ qualifies.

In Section 7.3.1, we also introduced a convenient notation for representing a vector $\boldsymbol{u}$ in terms of temporal and spatial parts $u_0\boldsymbol{t}$ and $\underset{\sim}{\boldsymbol{u}}$ respectively. Equation (7.1) provides a useful means of splitting any vector into these parts and the result clearly depends on the chosen frame. On the other hand, given a vector $\boldsymbol{u}$, a change of basis vectors leaves $\boldsymbol{u}$ itself unchanged so that $\boldsymbol{u}^2$, which determines whether the vector itself is timelike, spacelike, or lightlike, cannot depend on the chosen frame. This, therefore, is the basis of the distinction between these two separate classes of vector.

As to the bivectors, they can be referred to as spacelike or timelike by relating them to the vectors involved. A bivector $\boldsymbol{U}$ is therefore

- *timelike*, if $\boldsymbol{U}\boldsymbol{U}^\dagger$ has the same sign as $\boldsymbol{t}^2$; examples are $\boldsymbol{xt},\boldsymbol{yt},\boldsymbol{zt}$;
- *null*, if $\boldsymbol{U}\boldsymbol{U}^\dagger=0$;
- *spacelike*, if $\boldsymbol{U}\boldsymbol{U}^\dagger$ has the opposite sign to $\boldsymbol{t}^2$; examples are $\boldsymbol{xy},\boldsymbol{yz},\boldsymbol{zx}$.

With respect to a given frame, the term temporal may be applied to a bivector that has the time vector as a factor, whereas the term spatial applies to one that has only spatial vectors as factors. These definitions imply that a bivector will still retain its original timelike or spacelike character in any other frame even if it appears to be a mix of timelike and spacelike parts (see Section 9.5). In the same way that we are able to separate a vector into spatial and temporal parts through Equation (7.1), there is a relatively simple way of separating a bivector into purely timelike and spacelike parts in a given frame:

$$\boldsymbol{U}=\underbrace{-(\boldsymbol{U}\cdot\boldsymbol{t})\boldsymbol{t}}_{\substack{\text{timelike}\\\text{bivector}}}+\underbrace{-(\boldsymbol{U}\wedge\boldsymbol{t})\boldsymbol{t}}_{\substack{\text{spacelike}\\\text{bivector}}} \tag{7.31}$$

The arrangement of each of the expressions on the right makes them easy to evaluate with each of the usual spacetime basis bivectors. Since this useful identity works with any combination of these, it must work for any bivector. It is of particular significance in electromagnetics because, as will be seen in Section 11.5, it allows us to split an electromagnetic field into electric and magnetic parts in any given frame.

7.12 EXERCISES

All the following refer to the spacetime geometric algebra with orthonormal basis vectros $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$, $\boldsymbol{t}$ under the metric signature $(+++-)$.

1. **(a)** Simplify $\boldsymbol{zyxt}$; $\boldsymbol{xtytzt}$; $\boldsymbol{zI}$; $\boldsymbol{tyIx}$; $\boldsymbol{IxIyIzIt}$.

(b) Determine the measure (area) of the bivector of $a\boldsymbol{xy} + b\boldsymbol{zt}$.

(c) Find inverses for $1 + bI$ and $a\boldsymbol{xy} + b\boldsymbol{zt}$.

2. Create an equation for the history of a particle in a stationary circular orbit of radius a lying in the $\boldsymbol{xy}$ plane and centered on the point $\boldsymbol{r}_0$ and find its velocity.

3. Complete Table 7.2 so as to get the full multiplication table of the spacetime geometric algebra.

4. Construct expressions for $\boldsymbol{\nabla} \cdot \boldsymbol{u}$ and $\boldsymbol{\nabla} \wedge \boldsymbol{u}$ in terms of the usual basis elements.

5. Simplify $\boldsymbol{\nabla} \cdot (\boldsymbol{\nabla u})$, $\boldsymbol{\nabla} \wedge (\boldsymbol{\nabla u})$, and $\boldsymbol{\nabla}^2 \boldsymbol{u}$ in terms of the usual basis elements.

6. What would be meant by $\overset{\circ}{\boldsymbol{u}}\overset{\circ}{\boldsymbol{\nabla}}$? How would the result differ from $\boldsymbol{\nabla u}$?

7. Confirm Equation (7.25) as an identity by expressing $\boldsymbol{\nabla}$ and $\boldsymbol{r}$ in terms of the usual basis elements.

8. Use $1 + \boldsymbol{r} + \boldsymbol{r}^2 + \boldsymbol{r}^3 \cdots = (1 - \boldsymbol{r})^{-1}$ for $|\boldsymbol{r}| < 1$ to evaluate $\boldsymbol{\nabla}(1 - \boldsymbol{r})^{-1}$. What is the corresponding (3+1)D result for $\boldsymbol{\nabla}(1 - \mathbf{r})^{-1}$?

9. Confirm the validity of Equations (7.1) and (7.31) as identities when

(a) a full set of vector and bivector basis elements is given and

(b) only the time vector is given.

10. Discuss the relationship between the time vector in the rest frame of any given particle and the particle's history.

11. In what sense is a Lorentz transformation similar to a rotation?

12. If the velocity of a particle is uniform and given by $\boldsymbol{v} = \boldsymbol{t} + 0.1[c]\boldsymbol{x}$:

(a) Find v, $\underset{\sim}{\boldsymbol{v}}$ and $\mathfrak{v}$.

(b) Give an example of what the particle's history might be when expressed in the $\boldsymbol{t}$-frame.

13. In the rest frame of the particle discussed in Exercise 7.12.12: $\boldsymbol{t}'$ is the local time vector, $\boldsymbol{x}'$ is the spatial vector corresponding to $\boldsymbol{x}$, t' is the local time parameter, $\mathfrak{v}$ is the particle's proper velocity and $\boldsymbol{v}'$ is its usual velocity as defined by $\partial_{t'}\boldsymbol{r}$.

(a) Express all of these parameters in the $\boldsymbol{t}$-frame, that is, in terms of $\boldsymbol{t}$ and $\boldsymbol{x}$.

(b) Give an example of what the particle's history might be when expressed in the $\boldsymbol{t}'$-frame.

Chapter 8

Relating Spacetime to (3+1)D

Even if we now take the view that the spacetime geometric algebra provides a truer way of modeling the fundamental processes of the natural world, our original (3+1)D model served well enough for most purposes. The question arises as to how these structurally different models correspond to one another. The obvious answer simply suggests that it is matter of exchanging the vector $t\boldsymbol{t}$ with the scalar t and the basis vectors $\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}$ with $\mathbf{x},\mathbf{y},\mathbf{z}$, but crucially, this turns out not to be the case, for the relationship between spacetime and (3+1)D involves yet another delicate twist. Far from being just another annoying quirk, however, this provides the basis of an extremely useful and powerful tool. There are at least three reasons why the spacetime and (3+1)D geometric algebras do not have a simple read-across relationship.

First, as shown in Table 2.1(b), there are $2^4 = 16$ basis elements in 4D, twice as many as compared with a 3D geometric algebra, and so any mapping between the two cannot be unique (not 1:1). We therefore need to understand how to make the correct associations; for example, we have a similar problem in going between a full 3D rendition of an object and some given 2D projection of it.

The second point is that, as we have already seen from Equations (7.11) and (7.12), multiplication of two (3+1)D paravectors gives a different result from multiplication of what would on the face of it seem to be the corresponding spacetime vectors.

The final point is perhaps even more fundamental. The vectors of spacetime and (3+1)D mean different things. A simple spacetime vector is independent of any choice of basis or reference frame, whereas the (3+1)D sort must always be stated as being in some given reference frame, for example, the so-called lab frame, the spaceship frame, or whatever. To go back from spacetime to (3+1)D, we must eliminate the time vector, and the particular time vector that we choose to eliminate will determine the result.

8.1 THE CORRESPONDENCE BETWEEN THE ELEMENTS

To discuss the mechanics of going between spacetime and (3+1)D, it will often be easier to make use of their respective orthonormal basis elements generated from

Understanding Geometric Algebra for Electromagnetic Theory, First Edition. John W. Arthur.

$\mathbf{x},\mathbf{y},\mathbf{z}$ and $\boldsymbol{t},\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}$. Unless otherwise stated, we will assume that the representation of the time vector as $\boldsymbol{t}$ is purely symbolic. For the moment, we can take it to belong to our own rest frame, or "lab frame," which for the present will also be the chosen reference frame for our (3+1)D vectors.

8.1.1 The Even Elements of Spacetime

The even elements of spacetime $1,\boldsymbol{xt},\boldsymbol{yt},\boldsymbol{zt},\boldsymbol{yz},\boldsymbol{zx},\boldsymbol{xy},I$ form a subalgebra, that is to say, they form a geometric algebra in their own right. Although this subalgebra, referred to as the even subalgebra of spacetime or just the even subalgebra, belongs to spacetime, it nevertheless turns out to be isomorphic to the 3D geometric algebra. Its basis elements, $1,\boldsymbol{xt},\boldsymbol{yt},\boldsymbol{zt},\boldsymbol{yz},\boldsymbol{zx},\boldsymbol{xy},I$, therefore correspond exactly with the (3+1)D basis elements $1,\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{yz},\mathbf{zx},\mathbf{xy},I$ that we associate with the Newtonian world. Multiplication between the timelike bivectors corresponds exactly to multiplication between the vectors in (3+1)D. This may be confirmed by taking the timelike bivectors as basis vectors for the even subalgebra and then multiplying combinations of these until all the other basis elements are found. For example, $(\boldsymbol{xt})(\boldsymbol{yt})=(\boldsymbol{xy})$, $(\boldsymbol{xt})(\boldsymbol{yt})(\boldsymbol{zt})=\boldsymbol{xyzt}=I$, and $I(\boldsymbol{xt})=(\boldsymbol{yz})$. This, then, is the key to the process of "translating" between the objects in spacetime and those of (3+1)D. This confirms what we found in Section 7.4 (Equations 7.11 and 7.12), that multiplication between spacetime vectors *does not* actually *correspond* to multiplication between (3+1)D vectors or paravectors. Rather, we now see that it must be the multiplication of the *timelike bivectors* that corresponds to the multiplication of 3D vectors. Based on the even subalgebra, the overall correspondence therefore turns out to be fairly simple and straightforward:

$$\text{even subalgebra of spacetime}\left\{\begin{array}{l}1\\ \boldsymbol{xt}\\ \boldsymbol{yt}\\ \boldsymbol{zt}\\ \boldsymbol{yz}\\ \boldsymbol{zx}\\ \boldsymbol{xy}\\ I\end{array}\right\}\leftrightarrow\left\{\begin{array}{l}1\\ \mathbf{x}\\ \mathbf{y}\\ \mathbf{z}\\ \mathbf{yz}\\ \mathbf{zx}\\ \mathbf{xy}\\ I\end{array}\right\}\ (3{+}1)\text{D} \tag{8.1}$$

Here the even spacetime element occupying any given row on the left of the mapping corresponds exactly to the (3+1)D element in the same row on the right. Note that what is called a bivector in one space may be a vector in another as they are both members of vector spaces, which makes them vectors in the general sense (see Appendix 14.4.1). However, rather than get too concerned about this, it may be easier to think of the mapping simply as a relabeling exercise; for example, $\boldsymbol{xt}$ is relabeled as $\mathbf{x}$ and $\boldsymbol{yz}$ is relabeled as $\mathbf{yz}$. The important thing is that we need such an isomorphism if the equations in both spaces are to correspond properly.

From here on, the symbol $\leftrightarrow$ will generally be used to represent a mapping between spacetime and (3+1)D. However, since there are twice as many elements in spacetime, this does not imply that the mapping is 1:1. The analogy of translating words or symbols from one language to another may be helpful in this respect since we often find that a word in one language does not always correspond to just a single word in another. In fact, we will borrow the word translation used in this sense to represent the mapping process. Equation (8.1) is therefore only part of the story, and though this is useful, it still leaves us with the problem of how to handle the odd elements of spacetime.

8.1.2 The Odd Elements of Spacetime

The odd elements, namely $\boldsymbol{t},\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},-I\boldsymbol{t},-I\boldsymbol{x},-I\boldsymbol{y},-I\boldsymbol{z}$, do not correspond directly with (3+1)D, and there is no odd subalgebra to appeal to as the odd elements do not include essential elements such as the scalars and pseudoscalars. However, this problem is averted by mapping the odd elements onto the even elements before translating them. The simple device of premultiplying each element by $-\boldsymbol{t}$ turns out to provide just the sort of mapping required, that is, $\boldsymbol{t},\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},-I\boldsymbol{t},-I\boldsymbol{x},-I\boldsymbol{y},-I\boldsymbol{z} \mapsto 1,\boldsymbol{xt},\boldsymbol{yt},\boldsymbol{zt},I,I\boldsymbol{xt},I\boldsymbol{yt},I\boldsymbol{zt}$. After this initial mapping, we may then apply a second mapping in the form of the same mapping as for the even elements. The complete translation of the odd elements of spacetime to (3+1)D is therefore given by

$$\text{odd elements of spacetime} \left\{ \begin{array}{c} \boldsymbol{t} \\ \boldsymbol{x} \\ \boldsymbol{y} \\ \boldsymbol{z} \\ -I\boldsymbol{t} \\ -I\boldsymbol{x} \\ -I\boldsymbol{y} \\ -I\boldsymbol{z} \end{array} \right\} \leftrightarrow \left\{ \begin{array}{c} 1 \\ \mathbf{x} \\ \mathbf{y} \\ \mathbf{z} \\ I \\ I\mathbf{x} \\ I\mathbf{y} \\ I\mathbf{z} \end{array} \right\} \quad (3+1)\text{D} \tag{8.2}$$

Equations (8.1) and (8.2) imply that two different spacetime elements correspond to the same (3+1)D object. There is therefore a choice of representing each type of (3+1)D object as either an odd or an even quantity in spacetime. This choice is usually to be made on the grounds of making the *physical* relations work out correctly, and so it is necessary to consider each case carefully on that basis. In fact, other choices are available; for example, we can modify the signs so that $x \leftrightarrow \mathbf{x}$ rather than $x \leftrightarrow -\mathbf{x}$, but we deal with this simply by adopting Equations (8.1) and (8.2) as the standard forms.

Whereas for our choice of metric signature the odd-to-even mapping may be represented by premultiplication with $-\boldsymbol{t}$, it turns out that in the $(+---)$ metric signature, where $\boldsymbol{t}^2 = +1$, it is necessary to postmultiply by $\boldsymbol{t}$ in order to achieve the

same correspondence. This is only one of many nuances associated with metric signature. A far more significant point, however, is that the choice of $\boldsymbol{t}$ itself is quite arbitrary; *any* time vector will qualify. It is important to keep stressing that, only by default is $\boldsymbol{t}$ taken to be the time vector where we are, or at least where we have chosen for our particular convenience to set up our "lab frame". Relativity treats any such choice with disdain, and so here $\boldsymbol{t}$ could mean ***any*** time vector.

8.1.3 From (3+1)D to Spacetime

In the reverse direction from (3+1)D to spacetime, the mapping is 1:2 because a choice has to be made as to whether to go to an odd or even element. While the choice is to be made on physical grounds, a general pattern to this soon becomes apparent and the process becomes fairly automatic. When an even result is required, there is no change other than to express each element in its spacetime form by replacing the (3+1)D basis vectors $\mathbf{x},\mathbf{y},\mathbf{z}$ with the spacetime bivectors $\boldsymbol{xt},\boldsymbol{yt},\boldsymbol{zt}$. It then follows that $1,\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{yz},\mathbf{zx},\mathbf{xy},I \mapsto 1,\boldsymbol{xt},\boldsymbol{yt},\boldsymbol{zt},I\boldsymbol{xt},I\boldsymbol{yt},I\boldsymbol{zt},I$ so as to maintain consistency in multiplication between the elements on each side of the mapping; for example, $I=\mathbf{xyz}\mapsto(\boldsymbol{xt})(\boldsymbol{yt})(\boldsymbol{zt})=\boldsymbol{xyzt}=I$. When an odd element is required, it is only necessary to premultiply the result with $\boldsymbol{t}$, which simply reverses the odd to even mapping that was used when going in the other direction from spacetime to (3+1)D. For example, $\mathbf{x}\mapsto\boldsymbol{xt}$ when an even element is required, whereas $\mathbf{x}\mapsto\boldsymbol{t}(\boldsymbol{xt})=\boldsymbol{x}$ for an odd one. In this case, we therefore have in total $1,\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{yz},\mathbf{zx},\mathbf{xy},I\mapsto\boldsymbol{t},\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},-I\boldsymbol{x},-I\boldsymbol{y},-I\boldsymbol{z},-I\boldsymbol{t}$, so that the (3+1)D scalars and vectors all translate into spacetime vectors, while the bivectors and pseudoscalar become trivectors, which are more recognizable in pseudovector form just as we have shown them here.

Provided that we know whether the result is to be even or odd, the mapping between the (3+1)D and spacetime basis elements may be summarized as

$$\text{(3+1)D}\quad \left.\begin{matrix}1\\ \mathbf{x}\\ \mathbf{y}\\ \mathbf{z}\\ I\\ I\mathbf{x}\\ I\mathbf{y}\\ I\mathbf{z}\end{matrix}\right\}\leftrightarrow\left\{\begin{matrix}\overbrace{1}^{\text{even}} & \overbrace{\boldsymbol{t}}^{\text{odd}}\\ \boldsymbol{xt} & \boldsymbol{x}\\ \boldsymbol{yt} & \boldsymbol{y}\\ \boldsymbol{zt} & \boldsymbol{z}\\ I & -I\boldsymbol{t}\\ \boldsymbol{yz} & -I\boldsymbol{x}\\ \boldsymbol{zx} & -I\boldsymbol{y}\\ \boldsymbol{xy} & -I\boldsymbol{z}\end{matrix}\right.\quad\text{spacetime}\tag{8.3}$$

where any the element in the odd column is clearly the neighboring element in the even column premultiplied by $\boldsymbol{t}$.

8.2 TRANSLATIONS IN GENERAL

If we express some general multivector object $\boldsymbol{U}$ in terms of a set of basis elements $\boldsymbol{X}_k$, for example $\boldsymbol{U} = \sum_k U_k \boldsymbol{X}_k$, then in translating back and forward between spacetime and (3+1)D, it should be clear that the coefficients U_k will not change; *only the basis elements* $\boldsymbol{X}_k$ *are affected*. This process is not the same as a *transformation* of basis vectors; rather, it is a *translation* from the basis vectors of one set to the basis vectors of the other. Recall that the term translation also implies that it is not a matter of mere substitution; we may also have to select the appropriate result, that is to say, an odd or even one. Both forms of the translation procedures that we have just discussed are summarized in Figure 8.1, which also attempts to show how the even subalgebra of spacetime acts as an interface.

8.2.1 Vectors

Having defined the odd-to-even mapping for the basis elements, let us now examine how it applies to spacetime vectors in general. Recall the notation introduced in

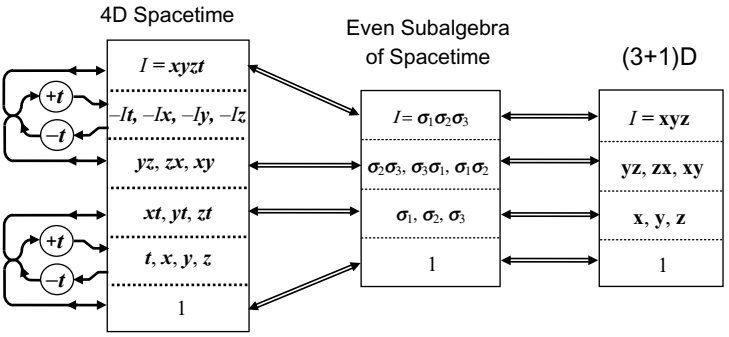

Figure 8.1 The connection between Newtonian space and 4D spacetime. The 4D spacetime geometric algebra relates to 3D through its even subalgebra. Because it is isomorphic to the 3D geometric algebra, the spacetime and 3D geometric algebras can share this subalgebra in common. The Newtonian vectors **x**, **y**, **z** equate to the basis vectors $\boldsymbol{\sigma}_1$, $\boldsymbol{\sigma}_2$, $\boldsymbol{\sigma}_3$ of the subalgebra, which are in turn identical to the timelike bivectors $\boldsymbol{xt}$, $\boldsymbol{yt}$, $\boldsymbol{zt}$ of spacetime. These and all other direct correspondences are shown by the tramline arrows. Alternatively, we may apply an additional step following the flow of solid arrows on the left. This allows us to switch between the even elements of spacetime and the odd ones by premultiplying by either $+\boldsymbol{t}$ or its inverse, $-\boldsymbol{t}$, as required; for example, $(-\boldsymbol{t})\boldsymbol{x} = \boldsymbol{xt}$ while in the other direction $(+\boldsymbol{t})\boldsymbol{xt} = \boldsymbol{x}$. As a further example, a spacetime vector $\boldsymbol{r}$ is an odd element, combining both time and spatial parts. To find its (3+1)D counterpart, we first premultiply by $-\boldsymbol{t}$ so as to form $-\boldsymbol{tr}$, which is then even. The multivector $-\boldsymbol{tr}$ comprises the scalar $r_t = -\boldsymbol{r}\cdot\boldsymbol{t}$ and the bivector $\boldsymbol{r}\wedge\boldsymbol{t}$. These then translate via the even subalgebra to the (3+1)D paravector $r_t + \mathbf{r}$ where $\mathbf{r} \equiv \boldsymbol{r}\wedge\boldsymbol{t}$. Going in the other direction, we simply reverse the process, premultiplying by $+\boldsymbol{t}$ in order to get back to an odd spacetime element. The choice of which route should be taken may seem arbitrary, but physical laws are at hand to determine what sort of spacetime element is required.

Section 7.3.1 that allows us to represent a general spacetime vector $\boldsymbol{u}$ in the form $u_t\boldsymbol{t} + \underset{\sim}{\boldsymbol{u}}$ where $u_t = -\boldsymbol{t}\cdot\boldsymbol{u}$. This splits the $\boldsymbol{u}$ into its separate temporal and spatial parts and provides a neat way of contracting $\boldsymbol{u} = u_t\boldsymbol{t} + u_x\boldsymbol{x} + u_y\boldsymbol{y} + u_z\boldsymbol{z}$. Since, by definition, $\underset{\sim}{\boldsymbol{u}} \perp \boldsymbol{t}$, we find that premultiplication by $-\boldsymbol{t}$ in order to effect the mapping into an even element gives

$$\begin{aligned} -\boldsymbol{t}\boldsymbol{u} &= -\boldsymbol{t}\left(u_t\boldsymbol{t} + \underset{\sim}{\boldsymbol{u}}\right) \\ &= -u_t\boldsymbol{t}^2 - \boldsymbol{t}\underset{\sim}{\boldsymbol{u}} \\ &= \underbrace{u_t}_{\text{scalar}} + \underbrace{\underset{\sim}{\boldsymbol{u}}\boldsymbol{t}}_{\substack{\text{timelike} \\ \text{bivector}}} \end{aligned} \tag{8.4}$$

In the same way that $-\boldsymbol{t}\boldsymbol{x} = \boldsymbol{x}\boldsymbol{t}$ is to be identified with $\mathbf{x}$, $-\boldsymbol{t}\underset{\sim}{\boldsymbol{u}} = \underset{\sim}{\boldsymbol{u}}\boldsymbol{t}$ is to be identified with the (3+1)D vector $\mathbf{u}$. If necessary, this may readily be verified by writing things out in component form. We may therefore conclude

$$\begin{aligned} \boldsymbol{u} &\leftrightarrow u_t + \mathbf{u} \\ \text{where} \quad u_t &= -\boldsymbol{t}\cdot\boldsymbol{u}, \\ \text{and} \quad \mathbf{u} &= -\boldsymbol{t}\wedge\boldsymbol{u} \end{aligned} \tag{8.5}$$

Provided that we are cautious about the implications of the equals sign, we may write this simply as

$$u_t + \mathbf{u} = -\boldsymbol{t}\boldsymbol{u} \tag{8.6}$$

To allow this, the mapping of Equation (8.1) is being treated as an equality. This is possible because the mapping is an equivalence inasmuch as it may treat things on both sides of the isomorphism as being interchangeable. However, this is not the case for Equation (8.2) since, for example, 1 is not interchangeable with $\boldsymbol{t}$. The relationship expressed by Equation (8.5) in the direction from spacetime to (3+1)D is an example of what is more generally called a spacetime split, a subject dealt with in more detail in Sections 8.3 and 9.5.

It will be recalled from Section 7.4 (Equations 7.11 and 7.12) that multiplying the spacetime vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ gives a different result from multiplying the corresponding (3+1)D paravectors $u_t + \mathbf{u}$ and $v_t + \mathbf{v}$. If, however, we apply the odd-to-even mapping so that we now multiply them as $-\boldsymbol{t}\boldsymbol{u}$ and $-\boldsymbol{t}\boldsymbol{v}$, we find

$$\begin{aligned} (-\boldsymbol{t}\boldsymbol{u})(-\boldsymbol{t}\boldsymbol{v}) &= \left(-\boldsymbol{t}\left(u_t\boldsymbol{t} + \underset{\sim}{\boldsymbol{u}}\right)\right)\left(-\boldsymbol{t}\left(v_t\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}}\right)\right) \\ &= \left(-u_t\boldsymbol{t}^2 - \boldsymbol{t}\underset{\sim}{\boldsymbol{u}}\right)\left(-v_t\boldsymbol{t}^2 - \boldsymbol{t}\underset{\sim}{\boldsymbol{v}}\right) \\ &= \left(u_t + \underset{\sim}{\boldsymbol{u}}\boldsymbol{t}\right)\left(v_t + \underset{\sim}{\boldsymbol{v}}\boldsymbol{t}\right) \\ &= \left(u_t + \mathbf{u}\right)\left(v_t + \mathbf{v}\right) \end{aligned} \tag{8.7}$$

This then is the logic behind the odd-to-even mapping—by creating an isomorphism, it preserves the correspondence with multiplication in (3+1)D that already exists for the even elements.

We will frequently need to make use of the translation between a spacetime vector and the corresponding (3+1)D paravector. Note the following:

- Equation (7.1) provides a convenient way of expressing the equivalence between $\boldsymbol{u}$ and $u_t + \mathbf{u}$, namely $u_t + \mathbf{u} = -\boldsymbol{t}\left(u_t \boldsymbol{t} + \underset{\sim}{\boldsymbol{u}}\right)$.
- A full set of basis vectors is therefore not required; only $\boldsymbol{t}$ needs to be known.
- $\boldsymbol{t}$ may be the time vector of *any* frame.
- The scalar u_t, likewise any component of $\underset{\sim}{\boldsymbol{u}}$ or $\mathbf{u}$, is not restricted to any particular meaning.
- It may assume various roles such as time and distance, charge and current, or even energy and momentum.
- But it does depend on the choice of $\boldsymbol{t}$.
- While the vectors $\underset{\sim}{\boldsymbol{u}}$ and $\mathbf{u}$ have the same components, remember that they have different basis vectors.
- We can go between $\underset{\sim}{\boldsymbol{u}}$ and $\mathbf{u}$ in either direction using $\mathbf{u} = \underset{\sim}{\boldsymbol{u}}\boldsymbol{t}$ or $\underset{\sim}{\boldsymbol{u}} = \boldsymbol{t}\mathbf{u}$ as required.

8.2.2 Bivectors

Given that they are even elements, the translation process for spacetime bivectors should be straightforward because no premultiplication by $-\boldsymbol{t}$ is required. However, the timelike and spacelike bivectors take different routes. As is evident from Equation (8.1), the one translates to a (3+1)D vector whereas the other translates to a bivector of a similar-looking form. Any timelike bivector $\boldsymbol{U}$ may be put in the form $\underset{\sim}{\boldsymbol{u}}\boldsymbol{t}$ that, as we have just discussed in Section 8.2.1, equates to the (3+1)D vector $\mathbf{u}$. Via the component route, however,

$$\boldsymbol{U} = U_{xt}\boldsymbol{xt} + U_{yt}\boldsymbol{yt} + U_{zt}\boldsymbol{zt} \quad \leftrightarrow \quad u_x\mathbf{x} + u_y\mathbf{y} + u_z\mathbf{z} \tag{8.8}$$

where u_k is the same thing as U_{kt} for $k = x, y, z$.

On the other hand, given that their basis elements have the same form in spacetime as in (3+1)D, the spacelike bivectors should appear unchanged. A spacelike bivector can always be written as the dual of a timelike one, and because of the way we generated the timelike bivector basis elements in Section 7.3, any timelike bivector may be expressed as $\underset{\sim}{\boldsymbol{v}}\boldsymbol{t}$ where $\underset{\sim}{\boldsymbol{v}}$ is some spatial vector. As a result, any spacelike bivector $\boldsymbol{V}$ may be put in the form $I\underset{\sim}{\boldsymbol{v}}\boldsymbol{t}$. Now, as an even element, the spacetime unit pseudoscalar $I = \boldsymbol{xyzt}$ translates directly into the (3+1)D unit pseudoscalar $I = \mathbf{xyz}$ and $\underset{\sim}{\boldsymbol{v}}$ translates into $\mathbf{v}$, so that $I\underset{\sim}{\boldsymbol{v}}\boldsymbol{t}$ should simply translate into (3+1)D as $I\mathbf{v}$, which of course is a bivector as anticipated. Otherwise, proceeding via the component route,

$$\boldsymbol{V} = V_{yz}\boldsymbol{yz} + V_{zx}\boldsymbol{zx} + V_{xy}\boldsymbol{xy} \quad \leftrightarrow \quad V_{yz}\mathbf{yz} + V_{zx}\mathbf{zx} + V_{xy}\mathbf{xy} \tag{8.9}$$

which is just a straight exchange of basis elements. In fact, we are able to refer to $\boldsymbol{V}$ in either spacetime or (3+1)D without referring to any translation process at all. We do have to be a little more careful translating timelike bivectors back to spacetime since a (3+1)D vector such as $\mathbf{r}$ can translate to spacetime either as a vector or a timelike bivector, and so we cannot use the same symbol in both cases. Whereas $\mathbf{r}$ may translate to the spacetime vector $\underset{\sim}{\boldsymbol{r}}$ (with no time part) and we may write $\mathbf{r} \leftrightarrow \underset{\sim}{\boldsymbol{r}}$, we cannot equate $\mathbf{r}$ to $\underset{\sim}{\boldsymbol{r}}$. Instead we must remember that it actually equates to the bivector form $\underset{\sim}{\boldsymbol{r}}\boldsymbol{t}$.

8.2.3 Trivectors

As we have already seen in Section 8.1.3, the spacetime unit pseudoscalar translates directly into (3+1)D without change, and it is clear that this must be the case for any pseudoscalar.

Now, as the name suggests, the spacetime trivectors are odd elements, and so they need premultiplication by $-\boldsymbol{t}$ before translation to (3+1)D. To be consistent with the way that we defined the basis elements, a trivector $\boldsymbol{U}$ that is the dual of some vector $\boldsymbol{u}$ is expressed as $-\boldsymbol{Iu}$ rather than $\boldsymbol{Iu}$. We may therefore conclude that in general, $\boldsymbol{U} \leftrightarrow -I\mathbf{u}$. The change of sign here may seem inconvenient, but if we tried to make it go away by defining the duals differently, then it would simply reappear in the relationship between the spacetime and (3+1)D spacelike bivectors. As we frequently work with bivectors, it would seem the lesser of two evils to avoid this. But luckily, we rarely need to use trivectors, and so in practice we are unlikely to have to remember about this awkward sign change.

If we stay with the dual form, the translation process is still expressed by

$$\begin{aligned} \boldsymbol{U} = -\boldsymbol{Iu} &\leftrightarrow -\boldsymbol{t}(\boldsymbol{Iu}) \\ &= -\boldsymbol{tI}\left(u_t\boldsymbol{t} + \underset{\sim}{\boldsymbol{u}}\right) \\ &= \boldsymbol{It}\left(u_t\boldsymbol{t} + \underset{\sim}{\boldsymbol{u}}\right) \\ &= -Iu_t - I\mathbf{u} \end{aligned} \tag{8.10}$$

For the sake of completeness, however, if we work in terms of basis elements using the actual trivector form, we get the much more cumbersome result

$$\begin{aligned} \boldsymbol{U} &\leftrightarrow -\boldsymbol{t}\left(U_{xyz}\boldsymbol{xyz} + U_{yzt}\boldsymbol{yzt} + U_{zxt}\boldsymbol{xzt} + U_{xyt}\boldsymbol{xyt}\right) \\ &= U_{xyz}(\boldsymbol{xt})(\boldsymbol{yz}) + U_{yzt}(\boldsymbol{yt})(\boldsymbol{zt}) + U_{zxt}(\boldsymbol{xt})(\boldsymbol{zt}) + U_{xyt}(\boldsymbol{xt})(\boldsymbol{yt}) \\ &= U_{xyz}\mathbf{xyz} + U_{yzt}\mathbf{yz} + U_{zxt}\mathbf{xz} + U_{xyt}\mathbf{xy} \\ &= U_{xyz}I + \left(U_{yzt}\mathbf{yz} + U_{zxt}\mathbf{xz} + U_{xyt}\mathbf{xy}\right) \end{aligned} \tag{8.11}$$

Note that from Equation (8.10), we also have $\boldsymbol{Iu} \leftrightarrow I\left(u_t t + \mathbf{u}\right)$, which supports the general result that in translations involving I, we may simply treat it as though

it were a scalar, that is to say for any spacetime multivector $\boldsymbol{V}$ and a (3+1)D multivector $\mathbf{V}$,

$$I\boldsymbol{V} \leftrightarrow I\mathbf{V} \quad \Leftrightarrow \quad \boldsymbol{V} \leftrightarrow \mathbf{V} \tag{8.12}$$

8.3 INTRODUCTION TO SPACETIME SPLITS

The translation processes discussed above are fairly simple examples of a more general concept known as a spacetime split. The spacetime split has the following characteristics:

- It does not require that the object concerned be expressed in terms of basis elements.
- Even if we have chosen basis elements, it allows different time vectors to be employed.
- Rather than being based on an algebraic correspondence between basis elements, it has the physical interpretation of being a *projection* from spacetime into (3+1)D.
- The frame chosen for the split, more specifically its time vector, controls the projection or "view" that is obtained.

The spacetime split therefore turns out to be an extremely useful tool. Although our present perception of it as a translation process will be quite adequate for many purposes, such as exploring the spacetime form of Maxwell's equations, for the sake of those readers who wish to go a bit more deeply into the subject, we will now attempt to give some idea of what a spacetime split means and why it is relevant.

While the time component of a spacetime vector need not actually represent time, for example it could be charge density, the time vector itself turns out to be of immense importance inasmuch as it is the key to the process of extracting a (3+1)D vector from its spacetime form in a coordinate-free manner. As we have seen, it is central to the odd-to-even mapping for vectors. What we have called translation is a simple algebraic process that "translates" between a quantity expressed in spacetime "language" and its (3+1)D equivalent form. The fundamental difference with the spacetime split [25; 27, pp. 134–135; 34; 47] is that it has a definite physical interpretation, and indeed it is not an idea that is at all specific to geometric algebra [45]. The case that we have been discussing up until now is only a particular example where the time vector belonging to our own rest frame also determines the reference frame for our (3+1)D vectors, or what we conveniently call the $\boldsymbol{t}$-frame. But as we have already tried to make clear, the point about spacetime is that there need be no such association with a particular time vector; *all time vectors are equally valid*.

We may regard spacetime as giving us a complete representation of what is happening in the universe, whereas (3+1)D gives only a particular view of it. The spacetime representation of an individual particle is its history. What we *observe*,

however, is only the (3+1)D rendition of this. Since we lose time as a separate dimension, the process of obtaining our (3+1)D view is in effect a projection, or what is called a spacetime split. Each different choice of unit time vector yields a different (3+1)D view, or spacetime split. For our particular view, we must use our time vector in the process. Likewise, each different observer must use *their* own local time vector. As we have seen in Section 7.7.3, the time vector is directly associated with motion. If the history of an observer is given as $\boldsymbol{r} = t\boldsymbol{v} + \underset{\sim}{\boldsymbol{r}}$, we are able to say that their spacetime velocity $\boldsymbol{v}$ automatically gives their local time vector as being $\gamma\boldsymbol{v}$, in which the parameter γ is simply a required normalization factor such that $(\gamma\boldsymbol{v})^2 = -1$. But we know that observers with different velocities are associated with different (3+1)D reference frames, and so a spacetime frame is in fact equivalent to a conventional (3+1)D reference frame. This then is the principle of the spacetime split; let us now sketch out the basics of its operation.

Our discussion here applies to vectors, but the same principle extends to all spacetime objects. Suppose we have a vector $\boldsymbol{u}$ that is given in the $\boldsymbol{t}$-frame as $\boldsymbol{u} = u_t\boldsymbol{t} + u_x\boldsymbol{x} + u_y\boldsymbol{y} + u_z\boldsymbol{z}$. Now *in the same frame*, we take the spacetime split of $\boldsymbol{u}$ as being the same thing as our translation process. For a vector, this is given by premultiplying it with $-\boldsymbol{t}$, that is to say, as in Equation (8.5). We can therefore write $\boldsymbol{u} \leftrightarrow -\boldsymbol{t}\boldsymbol{u} = u_t + \mathbf{u}$ where $u_t = -\boldsymbol{t}\cdot\boldsymbol{u}$ and $\mathbf{u} = -\boldsymbol{t}\wedge\boldsymbol{u}$. While the translation process applies only within a given frame, we can reason that if $\boldsymbol{u}$ were represented in some other frame, say the $\boldsymbol{t}'$-frame where we would have $\boldsymbol{u} = u_{t'}\boldsymbol{t}' + u_{x'}\boldsymbol{x}' + u_{y'}\boldsymbol{y}' + u_{z'}\boldsymbol{z}'$ (note that primes are also on the subscripts), then its spacetime split in that frame must be given in a like manner by $-\boldsymbol{t}'\boldsymbol{u}$. Therefore, as it does not matter what frame $\boldsymbol{u}$ happens to be represented in, or even whether there is any such representation at all, the spacetime split of $\boldsymbol{u}$ in any arbitrary frame $\boldsymbol{t}'$ is simply $\boldsymbol{u} \leftrightarrow -\boldsymbol{t}'\boldsymbol{u} = u_{t'} + \mathbf{u}'$ where $u_{t'} = -\boldsymbol{t}'\cdot\boldsymbol{u}$ and $\mathbf{u}' = -\boldsymbol{t}'\wedge\boldsymbol{u}$.

To see how this works in practice, let us find the spacetime split of $\boldsymbol{u}$ in the $\boldsymbol{t}'$-frame starting out from its $\boldsymbol{t}$-frame representation, $\boldsymbol{u} = u_t\boldsymbol{t} + u_x\boldsymbol{x} + u_y\boldsymbol{y} + u_z\boldsymbol{z}$. From Equation (7.17b), we have $\boldsymbol{t}' = \gamma(\boldsymbol{t} + v\boldsymbol{x})$ for the case where the $\boldsymbol{t}'$-frame is moving with velocity $v\boldsymbol{x}$ with respect to the $\boldsymbol{t}$-frame. This gives us $-\boldsymbol{t}'\boldsymbol{u}$ in a form we can work out as follows:

$$\begin{aligned}
-\boldsymbol{t}'\boldsymbol{u} &= -\gamma(\boldsymbol{t} + v\boldsymbol{x})(u_t\boldsymbol{t} + u_x\boldsymbol{x}) - \boldsymbol{t}'(u_y\boldsymbol{y} + u_z\boldsymbol{z}) \\
&= -\gamma\left(u_t\boldsymbol{t}^2 + u_x\boldsymbol{t}\boldsymbol{x} + vu_t\boldsymbol{x}\boldsymbol{t} + vu_x\boldsymbol{x}^2\right) - u_y\boldsymbol{t}'\boldsymbol{y} - u_z\boldsymbol{t}'\boldsymbol{z} \\
&= \gamma(u_t - vu_x) + \gamma(u_x - vu_t)\boldsymbol{x}\boldsymbol{t} + u_y\boldsymbol{y}\boldsymbol{t}' + u_z\boldsymbol{z}\boldsymbol{t}' \\
&= \underbrace{\gamma(u_t - vu_x)}_{\text{scalar}} + \underbrace{\gamma(u_x - vu_t)\mathbf{x} + u_y\mathbf{y} + u_z\mathbf{z}}_{\text{vector}}
\end{aligned} \tag{8.13}$$

so that on separating the scalar and vector parts, we have

$$\begin{aligned}
u_{t'} &= -\boldsymbol{t}'\cdot\boldsymbol{u} = \gamma(u_t - vu_x) \\
\mathbf{u}' &= \boldsymbol{t}'\wedge\boldsymbol{u} = \gamma(u_x - vu_t)\mathbf{x} + u_y\mathbf{y} + u_z\mathbf{z}
\end{aligned} \tag{8.14}$$

The only delicate point here is the justification of why, in the $\boldsymbol{t}'$-frame as we have here, $\mathbf{x}$ appears to be given by $\boldsymbol{xt}$ rather than $\boldsymbol{x}'\boldsymbol{t}'$, which would be in line with $\mathbf{y}$ and $\mathbf{z}$ being given by $\boldsymbol{yt}'$ and $\boldsymbol{zt}'$. But it is readily determined from Equation (7.17b) that $\boldsymbol{x}'\boldsymbol{t}' = \boldsymbol{xt}$, and so the point is resolved. Note that, as discussed in Section 10.6.2, the (3+1)D basis vectors are always represented by the same $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ irrespective of the choice of frame. Equation 8.14 clearly demonstrates how the spacetime split of a vector quantitatively depends on the chosen frame. The parameters γ and v together with the direction of the spatial part of its time vector are all involved in the resulting (3+1)D vector, and for this reason, it is called a *relative vector*.

The concept of a spacetime split is essential when moving charges come into play, in particular, it provides an ideal way to "project out" the electric and magnetic fields that we observe in different frames. In fact, for a charge in uniform motion, this amounts to simply projecting out from a charge's own Coulomb field. These, and other applications, are discussed in detail in Chapters 11 and 12. If readers wish to understand the spacetime split more fully, it is recommended that before attempting Section 9.5, they should read the preceding sections of Chapter 9 in order to get a fuller appreciation of how a change of reference frames affects basis vectors.

Provided the reader has clearly understood that the spacetime split is a physical concept in origin whereas the process of translation is mere algebraic manipulation, we can drop the formal distinction. Points to note, however, are as follows:

- Although the spacetime split is invertible inasmuch as $\boldsymbol{t}(-\boldsymbol{tu}) = \boldsymbol{u}$, the term itself is meant to apply in the direction from spacetime to (3+1)D.
- Because it is based on the substitution of basis elements, the simple translation process applies only in the frame in which an object is represented.
- However, if the spacetime split is invertible in a given context, then it may still use the symbol $\leftrightarrow$.
- The tie up between the two approaches derives from the fact that, as a principle of relativity, *any* time vector may be used in conjunction with Equation (8.5).
- If a frame is not specified for the spacetime split or it is not clear from the context, generally speaking, the $\boldsymbol{t}$-frame may be assumed.
- As with the translation process, there are other ways of implementing a spacetime split, such as by postmultiplication by $-\boldsymbol{t}$. For example, if $(-\boldsymbol{t})\boldsymbol{r} \leftrightarrow t + \mathbf{r}$ is the standard form of split for $\boldsymbol{r}$, then $\boldsymbol{r}(-\boldsymbol{t}) \leftrightarrow t - \mathbf{r}$. The only difference is in the sign of $\mathbf{r}$. As discussed in Section 3.1, there is no unique way of writing multivectors; the key issue is one of consistency. Premultiplication with $-\boldsymbol{t}$ will therefore be our standard spacetime split, and postmultiplication by $-\boldsymbol{t}$ will be used only as a special form if, for some particular reason, the change of sign turns out to be beneficial.

- The appearance of the standard spacetime split depends on the metric signature being used. Just as in the case of translation of basis vectors, in the $(-+++)$ metric signature its form changes to postmultiplication by the time vector. For example, the spacetime split of the vector $\boldsymbol{u}$ in the $\boldsymbol{\theta}$-frame is given by $\boldsymbol{u\theta}$ rather than $\boldsymbol{-\theta u}$.

We will return to the spacetime split in Section 10.6 where we discuss its geometric interpretation and its connection with the idea that (3+1)D vectors are in fact relative vectors.

8.4 SOME IMPORTANT SPACETIME SPLITS

8.4.1 Time

Equations (8.13) and (8.14) give alternative versions of the spacetime split of any simple vector in the case where the vector is given in the $\boldsymbol{t}$-frame, whereas the spacetime split in the $\boldsymbol{t}'$-frame is what we actually require. We could use this directly to find the spacetime split of $\boldsymbol{t}$ in some other frame, but clearly, we can also turn the question the other way round and ask, what is the spacetime split of any given time vector as seen in the $\boldsymbol{t}$-frame? Once again, from Equation (7.17b) we can take the time vector of the $\boldsymbol{t}'$-frame as having the form $\boldsymbol{t}' = \gamma(\boldsymbol{t} + v\boldsymbol{x})$. Since the choice of $\boldsymbol{x}$ is irrelevant, we may write this more generally as $\boldsymbol{t}' = \gamma(\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}})$, implying that the spatial origin of the $\boldsymbol{t}'$-frame has the spatial velocity $\underset{\sim}{\boldsymbol{v}}$ with respect to the spatial origin of the $\boldsymbol{t}$-frame. We then find

$$\begin{aligned} -\boldsymbol{t}\cdot\boldsymbol{t}' &= \gamma \\ -\boldsymbol{t}\wedge\boldsymbol{t}' &= \gamma\underset{\sim}{\boldsymbol{v}}\boldsymbol{t} = \gamma\mathbf{v} \end{aligned} \tag{8.15}$$

which we may put together as

$$-\boldsymbol{t}\boldsymbol{t}' = \gamma(1+\mathbf{v}) \tag{8.16a}$$

This tells us that the basic parameters γ and $\mathbf{v}$ we need to construct a Lorentz transformation between the $\boldsymbol{t}'$-frame and the $\boldsymbol{t}$-frame are entirely determined by the two time vectors alone. Furthermore, we can rearrange the order of the products to give

$$\begin{aligned} -\boldsymbol{t}'\boldsymbol{t} &= -\boldsymbol{t}'\cdot\boldsymbol{t} - \boldsymbol{t}'\wedge\boldsymbol{t} \\ &= -\boldsymbol{t}\cdot\boldsymbol{t}' + \boldsymbol{t}\wedge\boldsymbol{t}' \\ &= \gamma(1-\mathbf{v}) \end{aligned} \tag{8.16b}$$

This now represents the converse situation, the spacetime split of $\boldsymbol{t}$ in the $\boldsymbol{t}'$-frame. The result displays the essential symmetry with respect to Equation (8.16a) in that $\mathbf{v}$ is replaced by $-\mathbf{v}$.

8.4.2 Velocity

In this section, the main focus is on velocity, a vector that provides a piece of key information about the state of a particle. As discussed in Section 7.5, spacetime velocity is a frame-dependent *derived* vector. It is essentially frame dependent because it always depends on the time parameter of some chosen frame. For example, if $\boldsymbol{r}$ is the particle's history vector in the $\boldsymbol{t}$-frame, then we have $\boldsymbol{v} = \partial_t \boldsymbol{r}(t)$, while in the $\boldsymbol{t}'$-frame, we have $\boldsymbol{v}' = \partial_{t'} \boldsymbol{r}(t')$. In the $\boldsymbol{t}$-frame, $\boldsymbol{v}$ takes the convenient form $\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}}$ (Equation 7.15), where $|\underset{\sim}{\boldsymbol{v}}|$ is the usual scalar velocity, v, and $\hat{\underset{\sim}{\boldsymbol{v}}}$ is the direction of the motion in space.

From Equation (8.5), $\boldsymbol{v}$ translates to (3+1)D as

$$\begin{aligned} -\boldsymbol{t} \cdot \boldsymbol{v} &= -\boldsymbol{t} \cdot (\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}}) = 1 \\ -\boldsymbol{t} \wedge \boldsymbol{v} &= -\boldsymbol{t} \wedge (\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}}) = -\boldsymbol{t}\underset{\sim}{\boldsymbol{v}} = \underset{\sim}{\boldsymbol{v}}\boldsymbol{t} = \mathbf{v} \end{aligned} \tag{8.17}$$

Equation (8.17) illustrates how we may select either the vector or the scalar part of the result on its own, but we could equally well get the entire spacetime split in a single step simply by premultiplying $\boldsymbol{v}$ with $-\boldsymbol{t}$. Both methods inevitably lead to

$$\boldsymbol{v} \leftrightarrow 1 + \mathbf{v} \tag{8.18}$$

where $\mathbf{v}$ is just the usual (3+1)D velocity. We may therefore say that $1 + \mathbf{v}$ is the spacetime split of $\boldsymbol{v}$ in the $\boldsymbol{t}$-frame. But with derived vectors, there is often some form of complication, and here it is the fact that if we choose some other time vector for the spacetime split, say $\boldsymbol{t}'$ rather than $\boldsymbol{t}$, we also have to change the time parameter from t to t'. Applying the simple spacetime split rule, $-\boldsymbol{t}'\boldsymbol{v}$ gives $-\boldsymbol{t}'\partial_t \boldsymbol{r}$ rather than $-\boldsymbol{t}'\partial_{t'}\boldsymbol{r}$, and if we wish to find the latter we need to use the chain rule for differentiation, resulting in

$$\begin{aligned} -\boldsymbol{t}'\partial_{t'}\boldsymbol{r} &= -\boldsymbol{t}'\partial_t \boldsymbol{r}\partial_{t'}t \\ &= -\boldsymbol{t}'\boldsymbol{v}\partial_{t'}t \end{aligned} \tag{8.19}$$

To proceed any further we therefore need to know how to evaluate a total derivative such as $\partial_{t'}t$. For this, we need to know the relationship between the particle's spacetime coordinates in both frames. We will come to this in Chapter 9, but in the meantime, let us borrow Equation (9.29) in order to illustrate the point—many readers will in any case be familiar with this from the well-known Lorentz transformation for coordinates. Using $u\boldsymbol{x}$ rather than $v\boldsymbol{x}$ as the velocity of the $\boldsymbol{t}'$-frame with respect to the $\boldsymbol{t}$-frame, we may keep $\boldsymbol{v} = v\boldsymbol{x}$ for $\boldsymbol{t}$-frame velocity of the particle itself. This is sufficient to address the simplest case in which the velocities are collinear. Equation (9.29) then gives $t' = \gamma(t - ux)$ where $\gamma = (1 - u^2)^{-1/2}$. It is now possible to evaluate $\partial_{t'}t$ from $(\partial_t t')^{-1}$:

$$\begin{aligned}\partial_t t' &= \gamma\partial_t(t-ux)\\ &= \gamma(1-u\partial_t x)\\ &= \gamma(1-uv)\end{aligned} \tag{8.20}$$

$$\Leftrightarrow \partial_{t'} t = \frac{1}{\gamma(1-uv)}$$

This then allows us to complete the evaluation of Equation (8.19):

$$\begin{aligned}1+\mathbf{v}' &= -\boldsymbol{t}'\partial_{t'}\boldsymbol{r}\\ &= -\boldsymbol{t}'\boldsymbol{v}\partial_{t'}t\\ &= \frac{-\boldsymbol{t}'\boldsymbol{v}}{\gamma(1-uv)}\\ &= \frac{-\gamma(\boldsymbol{t}+u\boldsymbol{x})(\boldsymbol{t}+v\boldsymbol{x})}{\gamma(1-uv)}\\ &= \frac{1+v\boldsymbol{x}\boldsymbol{t}-u\boldsymbol{x}\boldsymbol{t}-uv}{(1-uv)}\\ &= 1+\frac{\mathbf{v}-\mathbf{u}}{(1-uv)}\end{aligned} \tag{8.21}$$

This result for the relative velocity of the particle as seen from $\boldsymbol{t}'$-frame is clearly more complicated than the usual straightforward form of spacetime split. There is no factor of γ, but instead, we have $(1-uv)^{-1}$, which is symmetric in u and v. In Section 10.6.4, however, we will find a much neater way of dealing with this but, for the moment, the present exercise is sufficient to illustrate some of the basic principles involved in dealing with the spacetime split of derived vectors and to underline the point that complications attach to them.

8.4.3 Vector Derivatives

The next part of our problem concerns translation between the (3+1)D *paravector* space-and-time derivative $\partial_t + \nabla$ and the spacetime *vector* derivative $\boldsymbol{\nabla}$. To establish the main principles, we will discuss derivatives of vectors before going on to consider derivatives of general multivectors.

The spacetime derivative of some vector $\boldsymbol{u}$ is represented by the product $\boldsymbol{\nabla u}$ where, for the purposes of multiplication, $\boldsymbol{\nabla}$ is treated just like any other vector. The result will therefore be of the form scalar plus bivector, and since both of these are even in character, the correspondence to (3+1)D through the even subalgebra will be direct. The derivative in question therefore needs only to be worked out in detail and then reassembled in (3+1)D form. First, we express both $\boldsymbol{\nabla}$ and $\boldsymbol{u}$ in a given frame, say the $\boldsymbol{t}$-frame. To facilitate this, we can use the now-familiar separation into temporal and spatial parts to write $\boldsymbol{u}$ as $u_t\boldsymbol{t}+\underset{\sim}{\boldsymbol{u}}$ where, from Section 8.3, the

corresponding (3+1)D vector part is $\mathbf{u} = -\boldsymbol{t}\underset{\sim}{\boldsymbol{u}}$. We may also use the same method to write $\boldsymbol{\nabla}$ as $-\partial_t \boldsymbol{t} + \underset{\sim}{\boldsymbol{\nabla}}$ where $\underset{\sim}{\boldsymbol{\nabla}} = \partial_x \boldsymbol{x} + \partial_y \boldsymbol{y} + \partial_z \boldsymbol{z}$ and $\boldsymbol{\nabla} = -\boldsymbol{t}\underset{\sim}{\boldsymbol{\nabla}} = \partial_x \mathbf{x} + \partial_y \mathbf{y} + \partial_z \mathbf{z}$ is the corresponding (3+1)D vector derivative. Following from this, $\boldsymbol{\nabla}\boldsymbol{u}$ can be translated into (3+1)D form through

$$\begin{aligned}
\boldsymbol{\nabla}\boldsymbol{u} &= -\boldsymbol{t}\partial_t(u_t\boldsymbol{t} + \underset{\sim}{\boldsymbol{u}}) + \underset{\sim}{\boldsymbol{\nabla}}(u_t\boldsymbol{t} + \underset{\sim}{\boldsymbol{u}}) \\
&= \partial_t u_t + \partial_t \underset{\sim}{\boldsymbol{u}}\boldsymbol{t} + \underset{\sim}{\boldsymbol{\nabla}}\boldsymbol{t}u_t + \underset{\sim}{\boldsymbol{\nabla}}(\boldsymbol{t}\underset{\sim}{\boldsymbol{u}}\boldsymbol{t}) \\
&= \partial_t u_t + \boldsymbol{\partial}_t(\underset{\sim}{\boldsymbol{u}}\boldsymbol{t}) + (\underset{\sim}{\boldsymbol{\nabla}}\boldsymbol{t})u_t + (\underset{\sim}{\boldsymbol{\nabla}}\boldsymbol{t})(\underset{\sim}{\boldsymbol{u}}\boldsymbol{t}) \\
&= \partial_t u_t + \partial_t \mathbf{u} + \boldsymbol{\nabla}u_t + \boldsymbol{\nabla}\mathbf{u} \\
&= (\partial_t + \boldsymbol{\nabla})(u_t + \mathbf{u})
\end{aligned} \tag{8.22}$$

In the second line of the working, we have used $\boldsymbol{t}\underset{\sim}{\boldsymbol{u}}\boldsymbol{t} = \underset{\sim}{\boldsymbol{u}}$ from Equation (7.4). It can be seen that the (3+1)D result corresponds directly with the spacetime expression in which $\boldsymbol{\nabla}$ is replaced by the paravector form $\partial_t + \boldsymbol{\nabla}$, and the vector being differentiated is replaced by its usual $\boldsymbol{t}$-frame spacetime split. We may express this as

$$\boldsymbol{\nabla}\boldsymbol{u} \leftrightarrow (\partial_t + \boldsymbol{\nabla})(u_t + \mathbf{u}) \tag{8.23}$$

But since we can write $\boldsymbol{\nabla}\boldsymbol{u}$ as $(\boldsymbol{\nabla}\boldsymbol{t})(-\boldsymbol{t}\boldsymbol{u})$ where $-\boldsymbol{t}\boldsymbol{u} = u_t + \mathbf{u}$ is the spacetime split of $\boldsymbol{u}$, this may be stated as

$$(\boldsymbol{\nabla}\boldsymbol{t})(u_t + \mathbf{u}) \leftrightarrow (\partial_t + \boldsymbol{\nabla})(u_t + \mathbf{u}) \tag{8.24}$$

suggesting that we may take $\boldsymbol{\nabla}\boldsymbol{t}$, that is to say $\boldsymbol{\nabla}$ *postmultiplied by* $+\boldsymbol{t}$, to be a *special spacetime split* of $\boldsymbol{\nabla}$ in its own right [25, Section 7, p. 27[1]].

In the case of the derivative of a vector, at least, the translation procedure turns out to be fairly simple:

$$\begin{aligned}
\boldsymbol{\nabla} &\leftrightarrow (-\boldsymbol{t}\partial_t + \underset{\sim}{\boldsymbol{\nabla}})\boldsymbol{t} \\
&= \partial_t + \boldsymbol{\nabla}
\end{aligned} \tag{8.25}$$

It is therefore not a true spacetime split in the normal sense; it behaves differently because the process of differentiation requires the metric signature to be taken into account. It is purely a manipulation that allows the spacetime split of $\boldsymbol{\nabla}\boldsymbol{u}$ to be written as the product of the two separate spacetime splits for $\boldsymbol{\nabla}$ and $\boldsymbol{u}$. Similarly, $(\partial_t + \boldsymbol{\nabla})(-\boldsymbol{t}) = -\boldsymbol{t}\partial_t + \underset{\sim}{\boldsymbol{\nabla}}\boldsymbol{t}(-\boldsymbol{t}) = -\boldsymbol{t}\partial_t + \underset{\sim}{\boldsymbol{\nabla}}\boldsymbol{t}$ takes us in the other direction.

Taking a more general spacetime split with respect to some other frame would seem to be complicated by the same sort of issue as the spacetime split involving a scalar derivative, but the form of $\boldsymbol{\nabla}$ does not depend on the choice of frame. As we might expect, in the $\boldsymbol{t}'$-frame, it is given by $\boldsymbol{\nabla} = -\partial_{t'}\boldsymbol{t}' + \partial_{x'}\boldsymbol{x}' + \ldots$ in much the same way as for an event vector $\boldsymbol{r}$ where we have $\boldsymbol{r} = t'\boldsymbol{t}' + x'\boldsymbol{x}' + \ldots$. Equation (8.25)

[1] But note the difference due to metric signature.

is therefore valid for any time vector; for example, $\boldsymbol{\nabla} \leftrightarrow \boldsymbol{\nabla t'} = \partial_{t'} + \boldsymbol{\nabla}'$ where $\boldsymbol{\nabla}' = \partial_{x'}\mathbf{x}' + \partial_{y'}\mathbf{y}' + \partial_{z'}\mathbf{z}'$.

8.4.4 Vector Derivatives of General Multivectors

The translation of $\boldsymbol{\nabla X}$ into its (3+1)D equivalent falls into two classes depending on whether the grade of $\boldsymbol{X}$ is even or odd. Clearly, we can split any general multivector $\boldsymbol{X}$ into odd and even parts prior to carrying out the translation and then recombine them afterwards, and so we need only to discuss the process for the odd and even cases separately.

For any odd grade of $\boldsymbol{X}$, the principle is the same as for a vector in that $\boldsymbol{\nabla X}$, being even, belongs to the even subalgebra so that it is then only necessary to write the result as $(\partial_t + \boldsymbol{\nabla})\mathbf{X}$. We have used $\mathbf{X}$ for the spacetime split of $\boldsymbol{X}$ irrespective of whether or not $\mathbf{X}$ is actually a vector purely for convenience. When $\boldsymbol{X}$ is even, however, the product $\boldsymbol{\nabla X}$ must be odd so that we have to transpose it to the even subalgebra by the usual means of premultiplying by $-\boldsymbol{t}$. As to $\boldsymbol{X}$ itself, no such premultiplication is necessary, so the result we require simply follows from reassembling $-\boldsymbol{t}\boldsymbol{\nabla X}$ in the appropriate (3+1)D form. But since $\boldsymbol{X} \leftrightarrow \mathbf{X}$, it can be seen that $-\boldsymbol{t}\boldsymbol{\nabla X} = (-\boldsymbol{t}\boldsymbol{\nabla})\boldsymbol{X} = (-\partial_t + \boldsymbol{\nabla})\mathbf{X}$ gives the desired result, so that contrary to the rule when $\boldsymbol{X}$ is odd, we now need to *premultiply* $\boldsymbol{\nabla}$ by $-\boldsymbol{t}$. The outcome can be summarized as

$$\boldsymbol{\nabla X} \leftrightarrow \begin{cases} (\partial_t + \boldsymbol{\nabla})\mathbf{X} & \boldsymbol{X} \text{ odd} \\ (-\partial_t + \boldsymbol{\nabla})\mathbf{X} & \boldsymbol{X} \text{ even} \end{cases} \tag{8.26}$$

This irregularity in the spacetime split of $\boldsymbol{\nabla}$ for odd and even multivectors has major implications for the spacetime form of the electromagnetic field bivector $\boldsymbol{F}$, as we will discuss below.

8.5 WHAT NEXT?

Chapter 7 provided an introduction to spacetime and, in particular, the role and interpretation of time as a vector. In Chapter 8, we tackled the problem of how spacetime and the (3+1)D world are related to each other. These foundations enable the reader to take up the ideas of special relativity that are essential to the understanding of the fundamental origins of electromagnetism, for example, why a moving charge produces a magnetic field, or how it is that the principles of retardation are actually embodied into spacetime.

But if the reader's main interest is simply to explore a new toolset for the phenomenological theory in which we take it for granted that a moving charge produces a magnetic field, they may have no interest in pursuing the special relativity. If this is the case, readers should go directly to Chapter 11.

In situations that do not involve either different reference frames or velocities that are relativistic, only one time vector, the symbolic time vector $\boldsymbol{t}$, needs to be considered. In this case, all spacetime splits are simple and equivalent to what we

have called translation. The process of translating between spacetime and (3+1)D, and vice versa, is summarized in Figure 8.1. It can be viewed as an algebraic process where basis elements in one space are simply substituted for those of another, the only nuance being that there is a choice of two spacetime basis elements for every one in (3+1)D.

In situations that involve a change of reference frame, for example, in Section 11.5 which deals with the transformation of the electromagnetic field, the basic ideas presented in Section 7.6 should be sufficient. It is when we come to deal with moving charges that it would be useful to have some appreciation of how to deal with different time vectors, such as $\boldsymbol{t}$ and $\boldsymbol{\upsilon}$, and how they affect the resultant (3+1)D view through a proper spacetime split. This has now been viewed as a geometric process, a projection, rather than a simple substitution. However, it is perhaps here that the reader may return to Chapters 9 and 10 for some assistance. A relatively easy starting point with moving charges is the spacetime form of the potential (Section 11.8.1). The arguments leading up to Equation (11.29) form the crucial step; thereafter, the problem is merely the mechanical one of evaluating the denominator $|\boldsymbol{R}\cdot\boldsymbol{\upsilon}|$. Nevertheless, the reader is encouraged to delve into Chapters 9 and 10, for not only are they of interest in their own right, they are also intended to help them to develop greater comprehension and facility when it comes to working with spacetime methods and concepts. This will doubtless repay the reader with a better appreciation of the fundamentals of electromagnetic theory through the medium, the spacetime geometric algebra.

8.6 EXERCISES

1. **(a)** Demonstrate that the subset of spacetime basis elements $1,\boldsymbol{xt},\boldsymbol{yt},\boldsymbol{zt},\boldsymbol{yz},\boldsymbol{zx},\boldsymbol{xy},\boldsymbol{xyzt}$ form the basis elements of a separate 3D geometric algebra.

(b) Identify the correspondence with the usual 3D basis elements $1,\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{yz},\mathbf{zx},\mathbf{xy},\mathbf{xyz}$.

(c) Why is there no odd subalgebra of spacetime?

(d) Identify and discuss the geometric subalgebras of 3D.

2. **(a)** Explain the circumstances under which a bivector may also be referred to as a vector.

(b) What determines whether an element of (3+1)D should map into an even or odd element of the spacetime geometric algebra?

3. Given $\boldsymbol{u}=(a\boldsymbol{t}+b\boldsymbol{x})$ and $\boldsymbol{v}=(p\boldsymbol{t}+q\boldsymbol{y})$, find the spacetime split of $\boldsymbol{uv}$ in the $\boldsymbol{t}$-frame by means of the translation process.

4. Concerning relating spacetime to (3+1)D, explain the difference in concept between the translation process and a spacetime split.

5. In the $\boldsymbol{t}$-frame, $\boldsymbol{t}$ and $\boldsymbol{x}$ are taken as two of the basis vectors, whereas in the $\boldsymbol{t}'$-frame, we have $\boldsymbol{t}'$ and $\boldsymbol{x}$. When these frames move relative to one another at low velocity, we may approximate $\boldsymbol{t}'$ and $\boldsymbol{x}'$ to first order by $\boldsymbol{t}+v\boldsymbol{x}$ and $v\boldsymbol{t}+\boldsymbol{x}$, respectively. Let $\boldsymbol{r}(\boldsymbol{t})=t(\boldsymbol{t}+u\boldsymbol{x})$ where $u\ll 1$.

(a) What is the spacetime split of $\boldsymbol{r}$ in the $\boldsymbol{t}$-frame?

(b) What is the corresponding result in the $\boldsymbol{t}'$-frame?

(c) Show that these results are consistent with Galilean relativity.

Chapter 9

Change of Basis Vectors

In Section 7.6, we discussed how basis vectors may be changed by rotating the set as a whole in some spacetime plane. If we choose a purely spatial plane, the result is a conventional rotation, but in Section 7.7.3 we were able to infer that when the plane includes the time vector, the result is a Lorentz transformation. We will use the word "rotation" to cover both cases, letting it be clear from the context whether we mean specifically the one sort or the other. Leaving aside reflections as only affecting right and left handedness or the direction of time, all such rotated sets represent equivalent choices. Given the symmetries of free space, shifts by any constant vector should be included in the discussion. However, while rotations and reflections are linear transformations that can be applied to basis vectors, shifts are not, since they actually leave the basis vectors unchanged. They need to be expressed by adding the shift vector $\boldsymbol{r}_0$, say, to the whole set of vectors, for example, $\boldsymbol{r} \mapsto \boldsymbol{r} + \boldsymbol{r}_0$. In Newtonian physics, this applies equally well to velocity shifts, $\mathbf{v} \mapsto \mathbf{v} + \mathbf{v}_0$, but spacetime is quite different—here the correct way to describe a velocity shift is by a linear transformation in the form of a rotation affecting the time vector. We will now address more fully how these linear transformations may be implemented using the mechanisms of geometric algebra.

9.1 LINEAR TRANSFORMATIONS

Linear transformations are frequently encountered in various sorts of relationships between sets of vectors belonging to any sort of vector space. They include scaling, dilation, rotation, reflection, projection, and any possible combinations thereof. But it is significant that in spite of being a quite different sort of object to the vectors they act on, linear transformations actually form a vector space in their own right. Under normal rules, the two separate spaces cannot be combined because addition between vectors and linear transformations is not allowed. On the other hand, a linear transformation acting on a vector can be thought of as a form of multiplication and, in this same spirit, the multiplication of transformations is allowed. However, this sort of "multiplication" is not, in general, to be to be associated with the different sort of multiplication that is defined for the geometric algebras that we have been

Understanding Geometric Algebra for Electromagnetic Theory, First Edition. John W. Arthur.

considering. We will therefore represent such objects in a *sans serif* typeface, for example, as T rather than $\mathbf{T}$, so that it will be clear that a different multiplication recipe is implied. The linear transformation resulting from the combined effect of the two linear transformations T_1 and T_2 can be written as a product $\mathsf{T}_2\mathsf{T}_1$ meaning that T_1 is applied first and then T_2, or in other words, $(\mathsf{T}_2\mathsf{T}_1)\boldsymbol{u} = \mathsf{T}_2(\mathsf{T}_1\boldsymbol{u})$ for any vector $\boldsymbol{u}$.

Some linear transformations may have inverses, but not all, and some classes of them may even form geometric algebras, but, as we shall soon see, not all. In the meantime, let us clarify how the way in which we express a linear transformation is affected by a change of basis vectors (a process that we know is actually a linear transformation in its own right).

In linear algebra, if the transformation R rotates the vector $\boldsymbol{u}$ into $\mathsf{R}\boldsymbol{u}$, we simply have

$$\boldsymbol{u} \mapsto \mathsf{R}\boldsymbol{u} = \boldsymbol{R}\boldsymbol{u} \tag{9.1}$$

Here a multiplication rule has been specifically defined so as to accommodate the fact that $\boldsymbol{R}$ and $\boldsymbol{u}$ are different sorts of objects in different vector spaces. The clearest example of this is in matrix algebra where $\boldsymbol{R}$ is represented by an orthogonal $N \times N$ matrix and $\boldsymbol{u}$ is an $N \times 1$ column vector. In the context of geometric algebra, the obvious question is whether it is possible for R and $\boldsymbol{u}$ to belong to the same algebra. While it turns out that this is not generally possible, it is nevertheless possible to implement rotations (and other sorts of linear transformation) in a geometric algebra by means of a different sort of rule from Equation (9.1). This being the case, the multivector that turns out to be associated with the rotation R will be denoted, in our usual notation, as $\boldsymbol{R}$. However, take care to note that this does not imply that $\mathsf{R}\boldsymbol{u} = \boldsymbol{R}\boldsymbol{u}$; rather, we now need to find out how $\boldsymbol{R}$ is to be determined and how it should act on the vector $\boldsymbol{u}$.

It is clear that any linear transformation such as $\boldsymbol{R}$, being a rotation, should have an inverse $\boldsymbol{R}^{-1}$ that will rotate $\boldsymbol{R}\boldsymbol{u}$ back to $\boldsymbol{u}$. In the case of matrix algebra, every orthogonal matrix has the property that its inverse is equal to its transpose, that is, $\boldsymbol{R}^{-1} = \boldsymbol{R}^t$, and so the inverse not only exists but is also rather easy to construct. In terms of a geometric algebra, this simply becomes $\boldsymbol{R}^{-1} = \boldsymbol{R}^\dagger$. If some other linear transformation T acts on $\boldsymbol{u}$, not necessarily a rotation but any transformation such that $\boldsymbol{u} \mapsto \mathsf{T}\boldsymbol{u} = \boldsymbol{T}\boldsymbol{u}$, then following this transformation by a rotation results in

$$\begin{aligned} \boldsymbol{T}\boldsymbol{u} &\mapsto \boldsymbol{R}\boldsymbol{T}\boldsymbol{u} \\ &= \left(\boldsymbol{R}\boldsymbol{T}\boldsymbol{R}^{-1}\right)(\boldsymbol{R}\boldsymbol{u}) \end{aligned} \tag{9.2}$$

If we were to apply the rotation to the whole space, that is to say, every vector in the space, each vector $\boldsymbol{u}$ would be substituted for by $\boldsymbol{R}\boldsymbol{u}$, and since $\boldsymbol{T}\boldsymbol{u}$ is itself a vector, it cannot escape the process and must be substituted for by $\boldsymbol{R}\boldsymbol{T}\boldsymbol{u}$. Then, the interpretation of Equation (9.2) is that after the rotation, we must also substitute $\boldsymbol{R}\boldsymbol{T}\boldsymbol{R}^{-1}$ for $\boldsymbol{T}$. With all vectors and linear transformations rotated in this way, everything will work as before, and in particular, existing equations will remain as equations.

In summary,

- a rotation acting on a linear transformation is expressed in a different form from the same rotation acting on a vector;
- for a vector, $\boldsymbol{u} \mapsto \boldsymbol{Ru}$; and
- for a linear transformation, $\boldsymbol{T} \mapsto \boldsymbol{RTR}^{-1}$.

9.2 RELATIONSHIP TO GEOMETRIC ALGEBRAS

As already mentioned, although we may not initially recognize them as being a class of vector, linear transformations are similar to members of a geometric algebra in that they obey the rules for a vector space and they have their own form of multiplication. In order for a geometric algebra to be formed, however, there is a constraint on the form of multiplication involved in that it must give rise to a graded structure. For example, we must require that $\mathbf{T}^0$ is a unit transformation that has no effect on the object that it acts on. Since it therefore acts like a unit scalar, we may simply refer to it as "1." Next, the square of any transformation $\mathbf{T}$ of grade 1 must obey $\mathbf{T}^2 = T^2\mathbf{T}^0$ where T is some scalar that we assume to be real and may be referred to as the magnitude of $\mathbf{T}$, and so on. This imposes quite a restriction on what sorts of transformations are allowed. However, 2D and 3D vector spaces do exist in which the elements are all linear transformations represented by matrices. For example, in 2D, a suitable basis is

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}; \quad \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad \equiv \quad 1; \boldsymbol{x}, \boldsymbol{y}; I \tag{9.3}$$

It is readily confirmed that this forms a simple geometric algebra, but note that the operations involve only the multiplication and addition of 2×2 matrices. The possibility of multiplying with 2×1 matrices is excluded. By introducing the imaginary scalar j into the arithmetic of this 2D space, we find that a 3D geometric algebra is formed from

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & \mathrm{j} \\ -\mathrm{j} & 0 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}; \quad \mathrm{j}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \mathrm{j}\begin{bmatrix} 0 & \mathrm{j} \\ -\mathrm{j} & 0 \end{bmatrix}, \mathrm{j}\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}; \quad \mathrm{j}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$
$$\equiv \quad 1; \quad \mathbf{x}, \mathbf{y}, \mathbf{z}; \quad I\mathbf{x}, I\mathbf{y}, I\mathbf{z}; \quad I \tag{9.4}$$

There need be no concern about using j here since its involvement arises only because some substitute for I is needed in order to make this set work. If, however, we allowed the scalars to be complex numbers, there would be no need for the four dual elements and we would be back to a 2D geometric algebra with complex scalars.

While this is all very interesting in its own right, the key points are the following:

- There are some well-known vector spaces that turn out to be geometric algebras.
- Vectors are not limited to simple things such as position vectors; they can include transformations, matrices, bivectors, and even general multivectors.
- The even subalgebra of spacetime is an example where the vectors are actually bivectors in some other space, that is, spacetime.
- Linear transformations form a vector space over some other vector space.
- There is, however, a clear distinction between linear transformations and ordinary vectors in that they transform as $\boldsymbol{T} \mapsto \boldsymbol{RTR}^{-1}$ rather than $\boldsymbol{T} \mapsto \boldsymbol{RT}$.
- *The same rule applies to the elements of a geometric algebra where there is an underlying isomorphism with linear transformations.*

9.3 IMPLEMENTING SPATIAL ROTATIONS AND THE LORENTZ TRANSFORMATION

Given that linear transformations can form a geometric algebra in their own right while the vectors they act on also belong to a geometric algebra, it is natural to ask whether they may both belong to the same geometric algebra. Indeed, this would provide a very powerful way of both representing and applying linear transformations.

In the 2D example above, the transformation $\boldsymbol{R} = \cos\theta - I\sin\theta$ acting on the algebraic basis elements $\boldsymbol{x}$ and $\boldsymbol{y}$ is equivalent to the orthogonal matrix $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ acting on $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$. The transformation $[\,\boldsymbol{\xi}\,] \mapsto [\,\boldsymbol{R}\,][\,\boldsymbol{\xi}\,][\,\boldsymbol{R}\,]^{-1}$ therefore represents a rotation of each basis element $[\,\boldsymbol{\xi}\,]$ by 2θ in the $\boldsymbol{xy}$ plane, as will readily be confirmed. By placing an object in square brackets here, we are simply highlighting the fact that we are referring to its matrix representation. Taking as an example $[\,\boldsymbol{R}\,][\,\boldsymbol{x}\,][\,\boldsymbol{R}\,]^{-1}$, we have

$$\begin{aligned} [\,\boldsymbol{R}\,][\,\boldsymbol{x}\,][\,\boldsymbol{R}\,]^{-1} &= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \\ &= \begin{bmatrix} -\sin 2\theta & \cos 2\theta \\ \cos 2\theta & \sin 2\theta \end{bmatrix} \\ &= \cos 2\theta\,[\,\boldsymbol{x}\,] + \sin 2\theta\,[\,\boldsymbol{y}\,] \end{aligned} \tag{9.5}$$

We must avoid confusing this result with the fact that the matrix $[\,\boldsymbol{R}\,]$ rotates a column vector $[\,\boldsymbol{u}\,]$ into $[\boldsymbol{Ru}]$ representing a rotation by an angle θ. Here we are applying the rotation to a matrix $[\,\boldsymbol{\xi}\,]$ by means of $[\,\boldsymbol{R}\,][\,\boldsymbol{\xi}\,][\,\boldsymbol{R}\,]^{-1}$ resulting in a rotation of 2θ. Now, treating Equation (9.5) algebraically by using the associations given in Equation (9.3) leads to

$$\begin{aligned} (\cos\theta - I\sin\theta)\boldsymbol{x}(\cos\theta + I\sin\theta) &= \left(\cos^2\theta - \sin^2\theta\right)\boldsymbol{x} + \cos\theta\sin\theta\,(\boldsymbol{x}I - I\boldsymbol{x}) \\ &= \cos 2\theta\boldsymbol{x} + \sin 2\theta\boldsymbol{y} \end{aligned} \tag{9.6}$$

In getting to this result, we have made use of $(\cos\theta - I\sin\theta)^{-1} = (\cos\theta + I\sin\theta)$ and $\boldsymbol{y} = -I\boldsymbol{x}$. It is exactly the same as Equation (9.5) without using matrices, but take note that here $\boldsymbol{x}$ and $\boldsymbol{y}$ belong to an apparently different space from the 3D position vectors $\mathbf{x}$ and $\mathbf{y}$.

In terms of our usual (3+1)D geometric algebra, however, we find $\cos\theta - I\mathbf{n}\sin\theta$ will do a similar job when $\mathbf{n}$ is a unit vector along the chosen axis of rotation:

$$\begin{aligned}\mathbf{u} &\mapsto (\cos\theta - I\mathbf{n}\sin\theta)\,\mathbf{u}\,(\cos\theta + I\mathbf{n}\sin\theta)\\ &= \mathbf{u}_{/\!/} + \cos 2\theta\,\mathbf{u}_{\perp} - \sin 2\theta\, I\mathbf{n}\wedge\mathbf{u}\\ &= \mathbf{u}_{/\!/} + \cos 2\theta\,\mathbf{u}_{\perp} + \sin 2\theta\,\mathbf{n}\times\mathbf{u}\end{aligned} \tag{9.7}$$

where $\mathbf{u}_{/\!/}$ and $\mathbf{u}_{\perp}$ are respectively the parts of $\mathbf{u}$ that are parallel and perpendicular to the axis of rotation, and the angle of rotation is 2θ.

In spacetime, we can clearly have the same thing, but the notion of an axis of rotation just does not exist (think about it!) and so we must replace it with the more general concept that a bivector identifies the *plane* of rotation. Note that in the last example (Equation 9.7), $I\mathbf{n}$ defines the requisite bivector $\boldsymbol{N}$ and it is simply coincidental that, in 3D, we can identify $\boldsymbol{N}$ with the dual of some vector that we can take as being an axis of rotation. In general, therefore, we have

$$\begin{aligned}\boldsymbol{u} \mapsto \mathsf{R}\boldsymbol{u} &= (a - b\boldsymbol{N})\,\boldsymbol{u}\,(a + b\boldsymbol{N})\\ &= \boldsymbol{R}\boldsymbol{u}\boldsymbol{R}^{\dagger}\end{aligned} \tag{9.8}$$

where $\boldsymbol{N}$ is the unit *bivector* defining the plane of rotation and the scalar parameters a and b are chosen so as to determine the required degree of rotation. A multivector such as $\boldsymbol{R}$ is referred to as a rotor. Note the following:

- Equation (9.8) is different from Equation (9.1) because it applies to elements that all belong to the same geometric algebra.
- It also exemplifies the difference between R, the linear transformation, and $\boldsymbol{R}$, the bivector needed to implement it.
- $\boldsymbol{R}^{\dagger}$ is the reverse of $\boldsymbol{R}$ whereas for a matrix we would write $\boldsymbol{R}^{t}$, its transpose. In both cases $\boldsymbol{R}^{\dagger}$ and $\boldsymbol{R}^{t}$ are the same thing as $\boldsymbol{R}^{-1}$.
- $\boldsymbol{N}^{\dagger} = -\boldsymbol{N}$ for any bivector $\boldsymbol{N}$.
- If $\boldsymbol{N}$ is a spacelike bivector, then $\boldsymbol{N}^2 = -1$, whereas if it is a timelike bivector, then $\boldsymbol{N}^2 = +1$.
- The bivector $\boldsymbol{u}\wedge\boldsymbol{v}$ rotates any vector lying in the $\boldsymbol{uv}$ plane by 90° in a direction *from* $\boldsymbol{v}$ *toward* $\boldsymbol{u}$, for example, $(\mathbf{xy})\mathbf{y} = \mathbf{x}$, so that the sense of rotation is from $\mathbf{y}$ to $\mathbf{x}$ rather than $\mathbf{x}$ to $\mathbf{y}$.
- See Exercises 4.8.1 and 4.8.4 for some further details and examples.

Now, in Section 7.6 we introduced the idea that rotations could be used as a way of changing from one set of spacetime basis vectors to another. There are two

choices: rotation in a spacelike plane and rotation in a timelike plane. The former leaves the time vector unchanged and so it is associated only with spatial rotations, whereas rotation in a timelike plane alters both the time vector and the spatial vector that makes up the bivector for the rotation plane. Equation (7.17b) gave us the general form of the transformed basis vectors that result from a rotation in the $\boldsymbol{xt}$ plane. In Section 7.7.3, however, we were able to identify such a rotation with a Lorentz transformation so that Equation (9.8) allows us to generalize our previous discussion to any plane of rotation. It applies to spatial rotations and Lorentz transformations alike, the only differences being

- whether or not the time vector is in the plane of rotation, and
- the parameters a and b are given by different expressions.

For a rotation in the spacelike plane $\boldsymbol{N}$ through an angle ψ,

$$a = \left(\frac{1+\cos\psi}{2}\right)^{\frac{1}{2}} = \cos\left(\frac{\psi}{2}\right); \quad b = \left(\frac{1-\cos\psi}{2}\right)^{\frac{1}{2}} = \sin\left(\frac{\psi}{2}\right) \tag{9.9}$$

Note that a and b are related to α and β in Equation (7.17a), but the latter correspond to $\cos\psi$ and $\sin\psi$, respectively.

On the other hand, for a rotation in a timelike plane $\boldsymbol{N}$ that is equivalent to a Lorentz transformation L with the (3+1)D velocity parameter $\underset{\sim}{\boldsymbol{v}}$,

$$\begin{aligned} \boldsymbol{u} \mapsto \mathsf{L}\boldsymbol{u} &= \boldsymbol{L}\boldsymbol{u}\boldsymbol{L}^{\dagger} \quad \text{where:} \\ \boldsymbol{L} &= a - b\boldsymbol{N} \\ \boldsymbol{N} &= \frac{1}{v}\underset{\sim}{\boldsymbol{v}}\boldsymbol{t} \\ \gamma &= \left(1-v^2\right)^{-\frac{1}{2}} \\ a &= \left(\frac{\gamma+1}{2}\right)^{\frac{1}{2}} \\ b &= \left(\frac{\gamma-1}{2}\right)^{\frac{1}{2}} \end{aligned} \tag{9.10}$$

Recall that spacetime velocity is of the form $\boldsymbol{v} = \boldsymbol{t} + \underset{\sim}{\boldsymbol{v}}$ where $\underset{\sim}{\boldsymbol{v}}$ represents only the spatial part so that we need to find $v\boldsymbol{N}$ from either $\underset{\sim}{\boldsymbol{v}}\boldsymbol{t}$ or $\boldsymbol{v} \wedge \boldsymbol{t}$. Furthermore, since $-\boldsymbol{t}\boldsymbol{v} = 1 + \underset{\sim}{\boldsymbol{v}}\boldsymbol{t} = 1 + \mathbf{v}$ is just the spacetime split of $\boldsymbol{v}$ in the $\boldsymbol{t}$-frame, we can equate $v\boldsymbol{N}$ to $\mathbf{v}$, that is to say, $\boldsymbol{N} = \hat{\mathbf{v}}$, the unit (3+1)D vector in the direction of the motion.

Lorentz transformations are orthogonal transformations and can be regarded as rotations in that they preserve length and angles, but only in accord with the specific non-Euclidean spacetime norm. Projected into 3D and depicted using Euclidean geometry as in Figure 7.4, however, they look anything but orthogonal, but that is simply a problem of our restricted ability to visualize such things.

9.4 LORENTZ TRANSFORMATION OF THE BASIS VECTORS

Many readers will be familiar with the idea of Lorentz transformations from one 3D reference to another. These sorts of transformations act on scalars such as time and mass, position coordinates, components of velocity, momentum and force, and even vector fields *under the assumption that the basis vectors themselves are unaffected* so that the **x**,**y**,**z** of one reference frame is just the same as that of any other. The spacetime approach is generally different. Simple spacetime vectors are independent of the chosen basis, but we may choose different frames in which to represent them, that is to say, a different choice of basis vectors where, in particular, the time vector is different. We may go from one such frame to another by means of a Lorentz transformation. At this stage, it is not essential to understand how the Lorentz transformation was derived; it is only necessary to be able to understand the principles of its application that we have just underlined. Since a frame is defined by its basis vectors, this simply means transforming each basis vector by means of Equations (9.8) and (9.10). We only need to know **v**, the relative velocity vector of the frame that we are transforming to, with respect to the frame we are starting from.

We will now consider a Lorentz transformation that, acting on the basis vectors $\boldsymbol{t},\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}$ of the $\boldsymbol{t}$-frame, results in the new basis set $\boldsymbol{t}',\boldsymbol{x}',\boldsymbol{y}',\boldsymbol{z}'$, or what we may call the $\boldsymbol{t}'$-frame. Without loss of generality, we can let the motion $\boldsymbol{v}$ be collinear with $\boldsymbol{x}$ so that the spatial origin of the $\boldsymbol{t}'$-frame has the history $\Omega'(t) = t\boldsymbol{t} + vt\boldsymbol{x}$ where $v\boldsymbol{x}$ is its velocity. For reference, the spacetime split of $v\boldsymbol{x}$ in the $\boldsymbol{t}$-frame is $v\mathbf{x} = \mathbf{v}$, the usual (3+1)D velocity. We then have the basis vectors in the $\boldsymbol{t}'$-frame given by Equations (9.8) and (9.10) as being

$$
\begin{aligned}
\boldsymbol{t} \mapsto \boldsymbol{t}' &= \boldsymbol{L}\boldsymbol{t}\boldsymbol{L}^{\dagger} \\
&= (a - b\boldsymbol{x}\boldsymbol{t})\,\boldsymbol{t}\,(a - b\boldsymbol{t}\boldsymbol{x}) = \left(a^2 + b^2\right)\boldsymbol{t} + 2ab\boldsymbol{x} \\
&= \gamma\left(\boldsymbol{t} + v\boldsymbol{x}\right) \\
\boldsymbol{x} \mapsto \boldsymbol{x}' &= \boldsymbol{L}\boldsymbol{x}\boldsymbol{L}^{\dagger} \\
&= (a - b\boldsymbol{x}\boldsymbol{t})\,\boldsymbol{x}\,(a - b\boldsymbol{t}\boldsymbol{x}) = \left(a^2 + b^2\right)\boldsymbol{x} + 2ab\boldsymbol{t} \\
&= \gamma\left(\boldsymbol{x} + v\boldsymbol{t}\right) \\
\boldsymbol{y} \mapsto \boldsymbol{y}' &= \boldsymbol{L}\boldsymbol{y}\boldsymbol{L}^{\dagger} \\
&= (a - b\boldsymbol{x}\boldsymbol{t})\,\boldsymbol{y}\,(a - b\boldsymbol{t}\boldsymbol{x}) = \left(a^2 - b^2\right)\boldsymbol{y} \\
&= \boldsymbol{y} \\
\boldsymbol{z} \mapsto \boldsymbol{z}' &= \boldsymbol{L}\boldsymbol{z}\boldsymbol{L}^{\dagger} \\
&= \boldsymbol{z}
\end{aligned}
\tag{9.11}
$$

Here $\boldsymbol{N}$, the spacetime plane of rotation, is represented by the bivector $\boldsymbol{x} \wedge \boldsymbol{t} = \boldsymbol{x}\boldsymbol{t}$, and we have also used $\boldsymbol{x}\boldsymbol{t} = -\boldsymbol{t}\boldsymbol{x}$ to give $\boldsymbol{L} = (a - b\boldsymbol{x}\boldsymbol{t})$ where, from

Equation (9.10), it is seen that $(a^2+b^2)=\gamma$, $(a^2-b^2)=1$, and $2ab=\gamma v$, with γ being the Lorentz factor $\left(1-v^2\right)^{-(1/2)}$. Equation (9.11) confirms in a more general way what we had already found by another route in Equation (7.17b).

Apart from the difference in sign for v, the two simple results $\boldsymbol{t}'=\gamma\left(\boldsymbol{t}+v\boldsymbol{x}\right)$ and $\boldsymbol{x}'=\gamma\left(\boldsymbol{x}+v\boldsymbol{t}\right)$ are very similar in appearance to, but not to be confused with, the standard Lorentz transformation for coordinates as given, for example, in References 2, section 5.1.2; 20, part V, section 1, p. 377; 45; and 48, chapter VII, and also Equation (9.29) later on in this chapter. The result for the transformation of the time vector is particularly useful, especially if we write it in the more general form

$$\boldsymbol{t}'=\gamma\left(\boldsymbol{t}+\underset{\sim}{\boldsymbol{v}}\right) \tag{9.12}$$

which means we do not need to specify the spatial part of the velocity as being along any basis vector such as $\boldsymbol{x}$; we simply refer to it as $\underset{\sim}{\boldsymbol{v}}$ by using the notation introduced in Section 7.3.1.

The transformation of the basis vectors $\boldsymbol{t}$ and $\boldsymbol{x}$ to $\boldsymbol{t}'$ and $\boldsymbol{x}'$ is depicted in Figure 7.4. Viewed in the $\boldsymbol{xt}$ plane, we can think of $\boldsymbol{t}'$ and $\boldsymbol{x}'$ as pointing in different directions to $\boldsymbol{t}$ and $\boldsymbol{x}$. The new basis vectors look neither normalized nor orthogonal, but this is because we are used to thinking in the way of normal Euclidean geometry. Both grids, one in solid lines and the other dashed, attempt to represent unit orthogonal grids in non-Euclidean spacetime where $\boldsymbol{t}^2=-\boldsymbol{x}^2$ rather than $+\boldsymbol{x}^2$. The components of any vector $\boldsymbol{r}_0$ (an event) may be read off in terms of the basis vectors of either grid, since *the vector itself is independent of the choice of basis*. For example, we have either $\boldsymbol{r}_0=2.1\boldsymbol{t}+0.7\boldsymbol{x}$ or $2\boldsymbol{t}'+0.45\boldsymbol{x}'$, but different though these two representations may seem, $\boldsymbol{r}_0$ itself is invariant. How these coordinates change as a result of choosing different basis vectors will now be discussed in a more general way.

The orthonormality of the new basis vectors $\boldsymbol{t}'$ and $\boldsymbol{x}'$ under the spacetime metric can be verified from

$$\begin{aligned}
\boldsymbol{t}'^2&=\gamma^2\left(\boldsymbol{t}^2+v\left(\boldsymbol{tx}+\boldsymbol{xt}\right)+v^2\boldsymbol{x}^2\right)\\
&=-\gamma^2\left(1-v^2\right)\\
&=-1\\
\boldsymbol{x}'^2&=\gamma^2\left(\boldsymbol{x}^2+v\left(\boldsymbol{xt}+\boldsymbol{tx}\right)+v^2\boldsymbol{t}^2\right)\\
&=\gamma^2\left(1-v^2\right)\\
&=1\\
\boldsymbol{t}'\cdot\boldsymbol{x}'&=\gamma^2\left(\boldsymbol{t}\cdot\boldsymbol{x}+v\left(\boldsymbol{t}^2+\boldsymbol{x}^2\right)+v^2\boldsymbol{x}\cdot\boldsymbol{t}\right)\\
&=0\\
\boldsymbol{x}'\cdot\boldsymbol{t}'&=\gamma^2\left(\boldsymbol{x}\cdot\boldsymbol{t}+v\left(\boldsymbol{t}^2+\boldsymbol{x}^2\right)+v^2\boldsymbol{t}\cdot\boldsymbol{x}\right)\\
&=0
\end{aligned} \tag{9.13}$$

In addition, we find that the area in the $\boldsymbol{xt}$ plane is preserved. We may take the bivector $\boldsymbol{x}\wedge\boldsymbol{t}$ as representing a unit directed area in the $\boldsymbol{t}$-frame, while for the $\boldsymbol{t}'$-frame, we have

$$\begin{aligned}\boldsymbol{x}'\wedge\boldsymbol{t}' &= \gamma(\boldsymbol{x}+v\boldsymbol{t})\wedge\gamma(\boldsymbol{t}+v\boldsymbol{x})\\ &= \gamma^2(\boldsymbol{x}\wedge\boldsymbol{t}+\boldsymbol{x}\wedge v\boldsymbol{x}+v\boldsymbol{t}\wedge\boldsymbol{t}+v\boldsymbol{t}\wedge v\boldsymbol{x})\\ &= \gamma^2\left(\boldsymbol{x}\wedge\boldsymbol{t}-v^2\boldsymbol{x}\wedge\boldsymbol{t}\right)\\ &= \gamma^2\left(1-v^2\right)\boldsymbol{x}\wedge\boldsymbol{t}\\ &= \boldsymbol{x}\wedge\boldsymbol{t}\end{aligned} \tag{9.14}$$

which demonstrates the point.

9.5 LORENTZ TRANSFORMATION OF THE BASIS BIVECTORS

In spacetime, we have two groups of basis bivectors: $\boldsymbol{xt},\boldsymbol{yt},\boldsymbol{zt}$ and $\boldsymbol{yz},\boldsymbol{zx},\boldsymbol{xy}$. We have seen how the first group, the timelike bivectors, corresponds to $\mathbf{x},\mathbf{y},\mathbf{z}$ in (3+1)D. The second group of spacelike bivectors directly correspond to their (3+1)D counterparts $\mathbf{yz},\mathbf{zx},\mathbf{zy}$ since, for example, $\mathbf{xy}\leftrightarrow\boldsymbol{xtyt}=-\boldsymbol{xyt}^2=\boldsymbol{xy}$. To see how these basis elements transform from one frame to another, the simplest route is to reformulate them from scratch using the transformed vectors so that $(\boldsymbol{xt})'=\boldsymbol{x}'\boldsymbol{t}'$ and so on. Given $\boldsymbol{L}^\dagger=\boldsymbol{L}^{-1}$, it is readily confirmed that this is indeed consistent with the transformation applied to the bivectors themselves since, for the product of any two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, we have $(\boldsymbol{uv})'=\boldsymbol{L}(\boldsymbol{uv})\boldsymbol{L}^\dagger=\left(\boldsymbol{LuL}^\dagger\right)\left(\boldsymbol{LvL}^\dagger\right)=\boldsymbol{u}'\boldsymbol{v}'$. Taking the velocity parameter for the transformation to the $\boldsymbol{t}'$-frame as being v along the $\boldsymbol{x}$ direction as before, we have first of all, for the transformation of the timelike bivectors,

$$\begin{aligned}(\boldsymbol{xt})' = \boldsymbol{x}'\boldsymbol{t}' &= \gamma(\boldsymbol{x}+v\boldsymbol{t})\gamma(\boldsymbol{t}+v\boldsymbol{x})\\ &= \gamma^2\left(\boldsymbol{xt}+v\boldsymbol{t}^2+v\boldsymbol{x}^2+v^2\boldsymbol{tx}\right)\\ &= \gamma^2\left(\boldsymbol{xt}+v^2\boldsymbol{tx}\right)\\ &= \left(1-v^2\right)^{-1}\left(1-v^2\right)\boldsymbol{xt}\\ &= \boldsymbol{xt}\\ (\boldsymbol{yt})' = \boldsymbol{y}'\boldsymbol{t}' &= \boldsymbol{yt}'\\ &= \boldsymbol{y}\gamma(\boldsymbol{t}+v\boldsymbol{x})\\ &= \gamma(\boldsymbol{yt}-v\boldsymbol{xy})\\ (\boldsymbol{zt})' = \boldsymbol{z}'\boldsymbol{t}' &= \boldsymbol{zt}'\\ &\quad\ \boldsymbol{z}\gamma(\boldsymbol{t}+v\boldsymbol{x})\\ &= \gamma(\boldsymbol{zt}+v\boldsymbol{zx})\end{aligned} \tag{9.15}$$

while for the spacelike bivectors

$$\begin{aligned}(yz)' &= y'z' = yz \\ (zx)' &= z'x' = z\gamma(x+vt) \\ &= \gamma(zx+vzt) \\ (xy)' &= x'y' = \gamma(x+vt)y \\ &= \gamma(xy-vyt)\end{aligned} \tag{9.16}$$

It is clear that the two classes of bivectors mix under the transformation of frame from t to t'. However, the timelike bivector parallel to the direction of motion and the spacelike bivector orthogonal to it are both unaffected by the transformation.

9.6 TRANSFORMATION OF THE UNIT SCALAR AND PSEUDOSCALAR

We would expect the spacetime unit scalar to be unaffected by a change of frame t to t'. However, since we know $x^2 = -t^2 = 1$, we can confirm this by applying the Lorentz transformation to either x or t and then evaluating x'^2 or t'^2 as appropriate. Indeed, we have $-t'^2 = x'^2 = 1$ from Equation (9.13) as proof of the point.

As to the unit pseudoscalar, $I' = x'y'z't' = x't'y'z' = (xt)'(yz)'$. From Equations (9.15) and (9.16) for the transformation of the basis bivectors, we have $(xt)'(yz)' = (xt)(yz) = I$, and so it too is unaffected by the transformation. This in turn implies that the remaining spacetime basis elements, the pseudovectors (the trivectors) $-It, -Ix, -Iy, -Iz$, transform in the same way as do the vectors.

It is necessary to remember, however, that (3+1)D scalars and pseudoscalars do not necessarily translate to spacetime scalars and pseudoscalars. For this reason, those that translate to a time vector or pseudovector, for example, charge density where we have $\rho t \leftrightarrow \rho$, are not invariant under a Lorentz transformation. In (3+1)D, we therefore have both invariant and noninvariant scalars and pseudoscalars. This is discussed further in Section 10.7. Finally, Table 9.1 summarizes those objects, or parts of objects, that change, or are left unchanged, by a Lorentz transformation.

9.7 REVERSE LORENTZ TRANSFORMATION

By a forward Lorentz transformation of a given frame, say the t-frame, we mean changing the basis vectors by a Lorentz transformation using velocity parameter v, say, to a new frame which we may call the t'-frame. The reverse transformation Lorentz simply takes us back from the new basis vectors to the original ones. In the forward direction, we have $\mathbf{e}_k \mapsto \mathbf{e}'_k = L\mathbf{e}_k L^\dagger$, while in the reverse direction, $\mathbf{e}'_k \mapsto \mathbf{e}_k = L^\dagger \mathbf{e}'_k L$. Since $L^\dagger = L^{-1}$, it follows that $L^\dagger \mathbf{e}'_k L = L^\dagger \left(L\mathbf{e}_k L^\dagger\right) L^\dagger = \mathbf{e}_k$, which confirms that for the reverse transformation, all we need to do is to exchange L and $L^\dagger$. We could also write $\mathbf{e}_k = \left(L\mathbf{e}'_k L^\dagger\right)^\dagger$, literally meaning "the

Table 9.1 Things that Change and Do Not Change under a Lorentz Transformation

Regime	Element	Affected parts
Spacetime	Scalars	Not affected
	Vectors	Parts parallel to $\boldsymbol{vt}$
	Timelike bivectors	Parts orthogonal to $\boldsymbol{vt}$
	Spacelike bivectors	Parts orthogonal to $\boldsymbol{Ivt}$
(3+1)D	Scalars (a)	Not affected
	Scalars (b)	Affected
	Basis vectors	Not affected
	Vectors (a)	Parts parallel to v
	Vectors (b) and bivectors	Parts orthogonal to v

Although the ideas of parallel and orthogonal intuitively apply to lines and planes, and hence to vectors and bivectors, two sorts of orthogonality apply to spacetime bivectors. In the first sort, the bivectors have some vector as a common factor, that is to say, their planes intersect in a line, for example, $\boldsymbol{xy}$ and $\boldsymbol{yz}$, whereas in the second sort, they share no vector factor at all and consequently do not intersect, for example, $\boldsymbol{xy}$ and $\boldsymbol{zt}$. This difference affects how such bivectors will change under a Lorentz transformation in a given transformation plane, and this also flows down into their various (3+1)D counterparts. For a Lorentz transformation in the $\underset{\sim}{\boldsymbol{vt}}$ transformation plane, the matter may be resolved by taking orthogonal to $\underset{\sim}{\boldsymbol{vt}}$ as the criterion for timelike bivectors but orthogonal to $\boldsymbol{I}\underset{\sim}{\boldsymbol{vt}}$ as the criterion for spacelike bivectors. The (3+1)D situation is much more obvious with the situation for bivectors being just the opposite of the case for the transformation of vectors. Objects that are tagged (a) correspond to a spacetime object of the same grade, whereas those tagged with (b) correspond to a spacetime object that is one grade higher. The main logical complication is that the (3+1)D basis vectors never change. For the purposes of the table, the Lorentz transformation may be regarded as applying either directly to the object concerned (active) or indirectly through a transformation of the spacetime basis vectors (passive).

reverse of the forward transformation." From Equation (9.8), however, we also have $\boldsymbol{L}(\boldsymbol{N}) = (a - b\boldsymbol{N})$ where we write $\boldsymbol{L}(\boldsymbol{N})$ to mean that $\boldsymbol{L}$ is a function of $\boldsymbol{N}$, the timelike bivector defined as the spacetime counterpart of $\hat{\mathbf{v}}$ through Equation (9.10). This implies that $\boldsymbol{L}^{\dagger}(\boldsymbol{N}) = \boldsymbol{L}\left(\boldsymbol{N}^{\dagger}\right) = \boldsymbol{L}(-\boldsymbol{N})$, which, if we now express $\boldsymbol{L}$ as a function of $\mathbf{v}$, leads us to $\boldsymbol{L}^{\dagger}(\mathbf{v}) = \boldsymbol{L}(-\mathbf{v})$. That is to say, the reverse transformation $\boldsymbol{L}^{\dagger}$ simply corresponds to the form of the original transformation, $\boldsymbol{L}$, with the velocity parameter reversed from $\mathbf{v}$ to $-\mathbf{v}$. This mathematical antisymmetry between the transformations in forward and reverse directions is inherent in the principle of relativity itself.

In terms of the reverse Lorentz transformation of the $\boldsymbol{t'}$-frame basis vectors, this asymmetry amounts to exchanging the primed vectors for unprimed ones and $\boldsymbol{v}$ with $-\boldsymbol{v}$ in Equation (9.11):

$$\begin{aligned} \boldsymbol{t} &= \gamma(\boldsymbol{t'} - v\boldsymbol{x'}) \\ \boldsymbol{x} &= \gamma(\boldsymbol{x'} - v\boldsymbol{t'}) \\ \boldsymbol{y} &= \boldsymbol{y'}; \quad \boldsymbol{z} = \boldsymbol{z'} \end{aligned} \tag{9.17}$$

Equations (9.11) and (9.17) represent the Lorentz transformations from the $\boldsymbol{t}$-frame to the $\boldsymbol{t}'$-frame and vice versa. It is important to note that both of these equations act on the basis vectors rather than the coordinates, which is one of the key differences in the spacetime approach, and both make abundantly clear the nonuniqueness of the time vector.

9.8 THE LORENTZ TRANSFORMATION WITH VECTORS IN COMPONENT FORM

Having a vector in component form can cause some confusion in the case of any sort of linear transformation because it is necessary to be quite clear about whether the transformation is to be applied to the vector itself or to the basis vectors used to express it in component form. We first of all discuss this from a general standpoint and then take a more detailed look at the case where the Lorentz transformation is applied to the basis vectors, which will generally be the way we will be applying it. We now consider an alternative approach that may help to clarify matters.

9.8.1 Transformation of a Vector versus a Transformation of Basis

It has already been emphasized that any given vector $\boldsymbol{r}$ is an entity in its own right, independent of the chosen basis. We therefore have to distinguish between a Lorentz transformation of the form $\boldsymbol{r} \mapsto \boldsymbol{r}' = \boldsymbol{L}\boldsymbol{r}\boldsymbol{L}^{\dagger}$, which *actually changes* $\boldsymbol{r}$ by rotating it in some spacetime plane, and one of the form $\mathbf{e}_k \mapsto \mathbf{e}'_k = \boldsymbol{L}\mathbf{e}_k\boldsymbol{L}^{\dagger}$, which *changes only the basis vectors*. Although there is also the possibility of transforming the vector *and* the basis vectors together as a whole, it is readily established that this case is in effect trivial and there is no effect on the components. For example, if we take basis vectors on the earth with one being along the axis of rotation and the others somewhere in the equatorial plane, the coordinates of a vector from the earth to the sun change due to the earth's rotation. On the other hand, the coordinates of any vector to a geostationary satellite, which of course by definition rotates along with the earth, clearly remain fixed so that, in effect, the rotation may be ignored. We therefore normally exclude this kind of situation.

In the case where only the basis vectors change, $\boldsymbol{r}$ itself is unaffected and the only thing to change is its representation, that is to say, from its representation in terms of the original basis vectors to a new representation in terms of the transformed ones. The first sort of transformation, sometimes referred to as an *active transformation*, is useful for boosting a particle's history to a new velocity; for example, a particle that is at rest in the $\boldsymbol{t}$-frame with a history $t\boldsymbol{t}$ would be changed by a Lorentz transformation with parameter $\underset{\sim}{\boldsymbol{v}}$ (see Equation 9.10 and preamble) to the history $\gamma t(\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}})$. The second sort, sometimes referred to as a *passive transformation*, allows us to represent the object's actual history in a different frame; for example, by the same transformation acting on the basis vectors alone, the same history has the new representation $t\boldsymbol{t} = \gamma t(\boldsymbol{t}' - \underset{\sim}{\boldsymbol{v}}')$.

As the conceptual difference is only slight, it may help to visualize these two different cases by thinking of the Lorentz transformation along the same lines as an actual rotation. The differences between the active and passive sorts of transformation processes are also easier to see when they are written out in matrix form. Taking the usual simple situation where the motion is along $\boldsymbol{x}$, only the $\boldsymbol{t}$ and $\boldsymbol{x}$ basis elements need to be considered, and the forward and reverse transformations of the basis elements may be expressed as

$$\begin{aligned}\begin{bmatrix}\boldsymbol{t}'\\ \boldsymbol{x}'\end{bmatrix}&=\begin{bmatrix}L^{+}_{tt} & L^{+}_{tx}\\ L^{+}_{xt} & L^{+}_{xx}\end{bmatrix}\begin{bmatrix}\boldsymbol{t}\\ \boldsymbol{x}\end{bmatrix}\\ \begin{bmatrix}\boldsymbol{t}\\ \boldsymbol{x}\end{bmatrix}&=\begin{bmatrix}L^{-}_{tt} & L^{-}_{tx}\\ L^{-}_{xt} & L^{-}_{xx}\end{bmatrix}\begin{bmatrix}\boldsymbol{t}'\\ \boldsymbol{x}'\end{bmatrix}\end{aligned} \tag{9.18}$$

where $\mathbf{L}^{+}$ and $\mathbf{L}^{-}$ are the matrices denoting the transformations in the forward ($+\mathbf{v}$) and reverse ($-\mathbf{v}$) directions, respectively. This may be confirmed by applying the results of Equation (9.11) to the primed basis vectors, from which we also find the actual components of the matrix $\mathbf{L}^{+}$. The components of $\mathbf{L}^{-}$ then follow by inversion (or from Equation 9.17).

$$\begin{aligned}\begin{bmatrix}\boldsymbol{t}'\\ \boldsymbol{x}'\end{bmatrix}&=\begin{bmatrix}\gamma & \gamma v\\ \gamma v & \gamma\end{bmatrix}\begin{bmatrix}\boldsymbol{t}\\ \boldsymbol{x}\end{bmatrix}\\ \begin{bmatrix}\boldsymbol{t}\\ \boldsymbol{x}\end{bmatrix}&=\begin{bmatrix}\gamma & -\gamma v\\ -\gamma v & \gamma\end{bmatrix}\begin{bmatrix}\boldsymbol{t}'\\ \boldsymbol{x}'\end{bmatrix}\end{aligned} \tag{9.19}$$

It is clear that, apart from taking $-v$ rather than $+v$ as the velocity parameter, $\mathbf{L}^{-}$ is identical to its inverse $\mathbf{L}^{+}$. It is worth noting that while the Lorentz transformation is represented algebraically as $\boldsymbol{r}\mapsto\boldsymbol{r}'=\boldsymbol{L}\boldsymbol{r}\boldsymbol{L}^{\dagger}$, here, for reasons that were discussed in Section 9.1, we find a single-sided transformation of the type $[\boldsymbol{r}]\mapsto[\boldsymbol{r}']=\mathbf{L}[\boldsymbol{r}]$ where $[\boldsymbol{r}]$ means the column vector of components representing $\boldsymbol{r}$ and $\mathbf{L}$ is either $\mathbf{L}^{+}$ or $\mathbf{L}^{-}$.

Taking the vector $\boldsymbol{r}$ in component form under the original basis as being $\boldsymbol{r}=r_t\boldsymbol{t}+r_x\boldsymbol{x}+r_y\boldsymbol{y}+r_z\boldsymbol{z}$, we need only to consider the t and x components as $\boldsymbol{y}$ and $\boldsymbol{z}$ are unaffected. Since they play no part, they may be omitted so as to keep things as simple as possible. We may therefore write

$$\boldsymbol{r}=\begin{bmatrix}\boldsymbol{t} & \boldsymbol{x}\end{bmatrix}\begin{bmatrix}r_t\\ r_x\end{bmatrix}=\begin{bmatrix}\boldsymbol{t}\\ \boldsymbol{x}\end{bmatrix}^{t}\begin{bmatrix}r_t\\ r_x\end{bmatrix} \tag{9.20}$$

Applying the Lorentz transformation *to the vector itself* therefore implies that $\boldsymbol{r}\mapsto\boldsymbol{r}'=\boldsymbol{L}\boldsymbol{r}\boldsymbol{L}^{\dagger}=r_t\boldsymbol{L}\boldsymbol{t}\boldsymbol{L}^{\dagger}+r_x\boldsymbol{L}\boldsymbol{x}\boldsymbol{L}^{\dagger}=r_t\boldsymbol{t}'+r_x\boldsymbol{x}'$, which may be put in a matrix form analogous to Equation (9.20) as

$$\boldsymbol{r}' = \begin{bmatrix} \boldsymbol{t}' \\ \boldsymbol{x}' \end{bmatrix}^t \begin{bmatrix} r_t \\ r_x \end{bmatrix} \tag{9.21}$$

From Equation (9.18), $\begin{bmatrix} \boldsymbol{t}' \\ \boldsymbol{x}' \end{bmatrix}$ is given by $\mathbf{L}^+ \begin{bmatrix} \boldsymbol{t} \\ \boldsymbol{x} \end{bmatrix}$ so that we find

$$\begin{aligned} \boldsymbol{r}' = \boldsymbol{L}\boldsymbol{r}\boldsymbol{L}^\dagger &= \left(\mathbf{L}^+ \begin{bmatrix} \boldsymbol{t} \\ \boldsymbol{x} \end{bmatrix} \right)^t \begin{bmatrix} r_t \\ r_x \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{t} \\ \boldsymbol{x} \end{bmatrix}^t \left[\mathbf{L}^+ \right]^t \begin{bmatrix} r_t \\ r_x \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{t} \\ \boldsymbol{x} \end{bmatrix}^t \begin{bmatrix} r_t^+ \\ r_x^+ \end{bmatrix} \end{aligned} \tag{9.22}$$

where

$$\begin{bmatrix} r_t^+ \\ r_x^+ \end{bmatrix} = \mathbf{L}^+ \begin{bmatrix} r_t \\ r_x \end{bmatrix} \tag{9.23}$$

Here we have used Equation (9.19), which shows that the matrix $\mathbf{L}^+$ is symmetric, to allow us to drop the transpose operation. Equation (9.23) therefore gives us the components r_t^+ and r_x^+ of the *transformed vector* $\boldsymbol{r}'$ *in terms of the original basis elements*. This procedure clearly applies to *any* vector that is expressed in terms of these basis elements, not just a position vector. As can be seen from Equation (9.21), however, had we also transformed the basis elements along with the vector, there would be *no change in the components*. This confirms the situation we have already excluded as being trivial, being analogous to the case where both the object and observer are rotated together such that the observer's view of the object remains the same throughout.

Now let us examine the case where the intention is *a change of basis* without any change to the vector itself:

$$\begin{aligned} \boldsymbol{r} &= \begin{bmatrix} \boldsymbol{t} \\ \boldsymbol{x} \end{bmatrix}^t \begin{bmatrix} r_t \\ r_x \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{t} \\ \boldsymbol{x} \end{bmatrix}^t \left(\mathbf{L}^+ \mathbf{L}^- \right) \begin{bmatrix} r_t \\ r_x \end{bmatrix} \\ &= \left(\mathbf{L}^+ \begin{bmatrix} \boldsymbol{t} \\ \boldsymbol{x} \end{bmatrix} \right)^t \mathbf{L}^- \begin{bmatrix} r_t \\ r_x \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{t}' \\ \boldsymbol{x}' \end{bmatrix}^t \begin{bmatrix} r_t^- \\ r_x^- \end{bmatrix} \end{aligned} \tag{9.24}$$

Here we have used the properties that $\mathbf{L}^-$ is the inverse of $\mathbf{L}^+$ and that $\mathbf{L}^+$ is symmetric. We can therefore represent this sort of transformation as $\boldsymbol{r} = r_t\boldsymbol{t} + r_x\boldsymbol{x} \mapsto \boldsymbol{r} = r_t^-\boldsymbol{t}' + r_x^-\boldsymbol{x}'$ where

$$\begin{bmatrix} r_t^- \\ r_x^- \end{bmatrix} = \mathbf{L}^- \begin{bmatrix} r_t \\ r_x \end{bmatrix} \tag{9.25}$$

Putting this together with Equations (9.21) and (9.19), the two different types of transformation may be summarized by

$$\begin{aligned} \boldsymbol{r}' &= \begin{bmatrix} \boldsymbol{t} \\ \boldsymbol{x} \end{bmatrix}^t \mathbf{L}^+ \begin{bmatrix} r_t \\ r_x \end{bmatrix} = \begin{bmatrix} \boldsymbol{t} \\ \boldsymbol{x} \end{bmatrix}^t \begin{bmatrix} r_t^+ \\ r_x^+ \end{bmatrix} \text{ where } \begin{bmatrix} r_t^+ \\ r_x^+ \end{bmatrix} = \begin{bmatrix} \gamma & \gamma v \\ \gamma v & \gamma \end{bmatrix} \begin{bmatrix} r_t \\ r_x \end{bmatrix} \quad \text{(a)} \\ \boldsymbol{r} &= \begin{bmatrix} \boldsymbol{t}' \\ \boldsymbol{x}' \end{bmatrix}^t \mathbf{L}^- \begin{bmatrix} r_t \\ r_x \end{bmatrix} = \begin{bmatrix} \boldsymbol{t}' \\ \boldsymbol{x}' \end{bmatrix}^t \begin{bmatrix} r_t^- \\ r_x^- \end{bmatrix} \text{ where } \begin{bmatrix} r_t^- \\ r_x^- \end{bmatrix} = \begin{bmatrix} \gamma & -\gamma v \\ -\gamma v & \gamma \end{bmatrix} \begin{bmatrix} r_t \\ r_x \end{bmatrix} \quad \text{(b)} \end{aligned} \tag{9.26}$$

The distinction between (a) and (b) is clear from the vectors on the left-hand side to which the primes attach. From this, it is plain that under a Lorentz transformation:

- Provided the same set of basis vectors is retained before and after the transformation, the components of any vector transform in the same way as the basis vectors themselves would transform (active transform).
- If the basis vectors are transformed along with the vector, then there is no difference in the components before and after the transformation. This is not surprising, for in the case of a rotation, nothing actually changes if we rotate *absolutely everything, including all observers*.
- If only the basis vectors are transformed while the vector itself remains unchanged (passive transformation), then the components of the vector transform in the opposite sense to the basis vectors, that is to say, with v replaced by $-v$.
- The reverse transformation of the basis elements (Equation 9.17) is particularly useful for expressing a vector in the new basis since it is only necessary to substitute the given expressions for each of the old basis vectors.

The same general principles apply to other objects in component form, for example, a bivector; we just have a different set of basis elements to consider. This is dealt with in Section 10.8.

As already mentioned, the Lorentz transformation is represented algebraically by the double-sided operation $\boldsymbol{r} \mapsto \mathbf{L}\boldsymbol{r} = \boldsymbol{L}\boldsymbol{r}\boldsymbol{L}^\dagger$, but when it is represented in matrix form, the more usual single-sided type of transformation results. This suggests

that the objects of a geometric algebra correspond to $n \times n$ matrices, which also transform by a double-sided operation, rather than $n \times 1$ column or row "vectors" which transform by a single-sided operation. This is supported by the examples given in Section 9.2. Note also the disadvantages of representing an abstract vector by a column vector of components. First, we cannot multiply column vectors in any way that is analogous to multiplying actual vectors, but we can multiply matrices! Second, in matrix algebra, there is ample opportunity for confusion because (contrary to the approach taken here in Equations 9.18–9.26) the basis vectors are rarely shown. For example, it is not possible to tell whether $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ means $\boldsymbol{x}$ or $\boldsymbol{x}'$.

9.8.2 Transformation of Basis for Any Given Vector

The preceding discussion highlighted the conflicting contexts in which the Lorentz transformation may be used and then compared their resulting forms for the basic case in which they act on vectors. It was assumed, however, that the vector was always given in terms of some set of basis vectors, or frame. We will now see how an arbitrary event vector or history may be evaluated in any given frame using a slightly different approach. In principle, the vector need not be an event vector and the transformation need not be a Lorentz transformation.

If we do not already know how the vector is represented in some chosen basis for which we know the transformation, it is necessary to find a standard means of generating such a representation. Given a vector, say $\boldsymbol{r}$, we may impose any basis we like by finding its projections along each normalized basis vector in turn. For the $\boldsymbol{t}$-frame, therefore

$$\boldsymbol{r} = (-\boldsymbol{t} \cdot \boldsymbol{r})\boldsymbol{t} + (\boldsymbol{x} \cdot \boldsymbol{r})\boldsymbol{x} + (\boldsymbol{y} \cdot \boldsymbol{r})\boldsymbol{y} + (\boldsymbol{z} \cdot \boldsymbol{r})\boldsymbol{z} \tag{9.27}$$

The reason for the negative sign in the first term is of course an effect of the metric signature, and it may be verified by taking $\boldsymbol{r} = \boldsymbol{t}$ that it gives the correct result since $(-\boldsymbol{t} \cdot \boldsymbol{t})\boldsymbol{t} = \boldsymbol{t}$. The inner products in brackets represent the scalars, which, as an ordered set, give the components of $\boldsymbol{r}$ with respect to the chosen basis $\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$. If we happen to know $\boldsymbol{r}$ in terms of the $\boldsymbol{t}$-frame basis vectors, for example, $\boldsymbol{r} = t\boldsymbol{t} + x\boldsymbol{x} + y\boldsymbol{y} + z\boldsymbol{z}$ where the components are just (t, x, y, z), we find that Equation (9.27) simply gives us back $\boldsymbol{r}$ in exactly the same form. However, we may equally well use Equation (9.27) for expressing $\boldsymbol{r}$ in the $\boldsymbol{t}'$-frame by simply replacing the unprimed basis vectors with the primed ones, that is to say $\boldsymbol{r} = (-\boldsymbol{t}' \cdot \boldsymbol{r})\boldsymbol{t}' + (\boldsymbol{x}' \cdot \boldsymbol{r})\boldsymbol{x}' + (\boldsymbol{y}' \cdot \boldsymbol{r})\boldsymbol{y}' + (\boldsymbol{z}' \cdot \boldsymbol{r})\boldsymbol{z}'$. Note that there is no prime here on $\boldsymbol{r}$ itself because it is unchanged; only the basis has been changed. Using Equation (9.11) to give the $\boldsymbol{t}'$-frame basis vectors from a *forward* Lorentz transformation of their $\boldsymbol{t}$-frame counterparts, we obtain

$$
\begin{aligned}
\boldsymbol{r} &= (-\boldsymbol{t}'\cdot\boldsymbol{r})\boldsymbol{t}' + (\boldsymbol{x}'\cdot\boldsymbol{r})\boldsymbol{x}' + (\boldsymbol{y}'\cdot\boldsymbol{r})\boldsymbol{y}' + (\boldsymbol{z}'\cdot\boldsymbol{r})\boldsymbol{z}' \\
&= -\boldsymbol{t}'\cdot(t\boldsymbol{t} + x\boldsymbol{x} + y\boldsymbol{y} + z\boldsymbol{z})\boldsymbol{t}' + \boldsymbol{x}'\cdot(t\boldsymbol{t} + x\boldsymbol{x} + y\boldsymbol{y} + z\boldsymbol{z})\boldsymbol{x}' + \\
&\quad \boldsymbol{y}'\cdot(t\boldsymbol{t} + x\boldsymbol{x} + y\boldsymbol{y} + z\boldsymbol{z})\boldsymbol{y}' + \boldsymbol{z}'\cdot(t\boldsymbol{t} + x\boldsymbol{x} + y\boldsymbol{y} + z\boldsymbol{z})\boldsymbol{z}' \\
&= \gamma[-(\boldsymbol{t} + v\boldsymbol{x})\cdot(t\boldsymbol{t} + x\boldsymbol{x} + y\boldsymbol{y} + z\boldsymbol{z})\boldsymbol{t}' + (\boldsymbol{x} + v\boldsymbol{t})\cdot(t\boldsymbol{t} + x\boldsymbol{x} + y\boldsymbol{y} + z\boldsymbol{z})\boldsymbol{x}'] + \\
&\quad \boldsymbol{y}\cdot(t\boldsymbol{t} + x\boldsymbol{x} + y\boldsymbol{y} + z\boldsymbol{z})\boldsymbol{y} + \boldsymbol{z}\cdot(t\boldsymbol{t} + x\boldsymbol{x} + y\boldsymbol{y} + z\boldsymbol{z})\boldsymbol{z} \\
&= \gamma[-(\boldsymbol{t}\cdot t\boldsymbol{t} + v\boldsymbol{x}\cdot x\boldsymbol{x})\boldsymbol{t}' + (\boldsymbol{x}\cdot x\boldsymbol{x} + v\boldsymbol{t}\cdot t\boldsymbol{t})\boldsymbol{x}'] + y\boldsymbol{y} + z\boldsymbol{z} \\
&= \gamma[(t - vx)\boldsymbol{t}' + (x - vt)\boldsymbol{x}'] + y\boldsymbol{y} + z\boldsymbol{z}
\end{aligned}
\tag{9.28}
$$

This therefore gives us exactly the same result as Equation (8.13), which was actually obtained by means of a spacetime split. Both of these methods therefore give us the components of $\boldsymbol{r}$ in the new basis. However, using yet another method that was discussed in Section 9.8.1 in connection with the derivation of Equation (9.19), we could start from $\boldsymbol{r} = t\boldsymbol{t} + x\boldsymbol{x} + y\boldsymbol{y} + z\boldsymbol{z}$ and use the *reverse Lorentz transformation* (Equation 9.17) to enable us to substitute $\boldsymbol{t}'$-frame basis vectors for those of the $\boldsymbol{t}$-frame. It therefore turns out that there are several different approaches to the problem and the one that will prove most fruitful will always depend on the information to hand. Basis vector substitution may not always be the neatest method, but it does have the advantage of being both readily understandable and simple to use.

As an illustration of the geometric interpretation of Equation (9.27), Figure 9.1(a) shows the time component of $\boldsymbol{r}$ in the $\boldsymbol{t}'$-frame in the form of its projection along (onto) $\boldsymbol{t}'$. Its $\boldsymbol{t}$-frame projection along $\boldsymbol{t}$ is also shown for comparison. While the same comparison applies to the $\boldsymbol{x}$ and $\boldsymbol{x}'$ components, those for $\boldsymbol{y}$ and $\boldsymbol{z}$ are unaffected because these basis vectors are unchanged. We could of course have used a more general form of Lorentz transformation with the velocity parameter along some arbitrary direction to bring all components into play, but this makes little difference to the principles involved.

The procedure we have just described illustrates one of the benefits of a coordinate-free approach. When the approach is coordinate based, we always think of a vector as being in terms of some particular basis. By default, we take the $\boldsymbol{t}$-frame as our basis so that we deal with $\boldsymbol{r}$ as though it needed to be identified with $t\boldsymbol{t} + x\boldsymbol{x} + y\boldsymbol{y} + z\boldsymbol{z}$. To express $\boldsymbol{r}$ in any other basis, we must then use this as our starting point. If, on the other hand, we take the more abstract approach of keeping $\boldsymbol{r}$ itself as the starting point, we then realize we can obtain its components in any frame from Equation (9.27) by using any set of basis vectors, not just $\boldsymbol{t}$,$\boldsymbol{x}$,$\boldsymbol{y}$,$\boldsymbol{z}$. All we have done here is to use this general idea to relate the components in one frame to the components in another. All that is required is the two sets of basis vectors, or at least one set expressed in terms of the other.

Equation (9.28), or its equivalent Equation (8.13), may be summarized by picking out the individual transformations that take us from each of the $\boldsymbol{t}$-frame components to its corresponding $\boldsymbol{t}'$-frame component as follows:

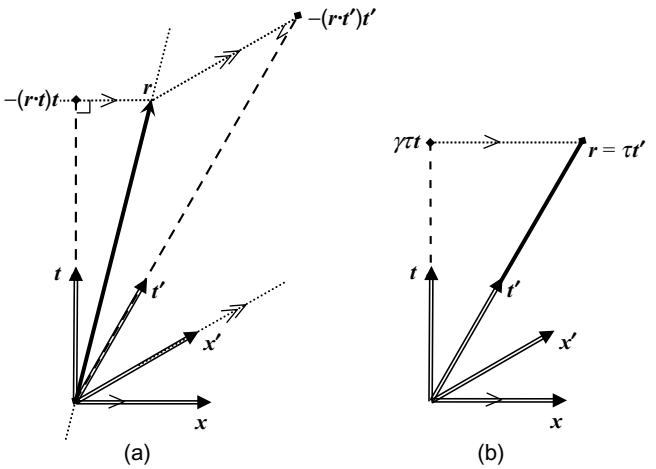

Figure 9.1 Projection of a spacetime vector onto the time axis in two different frames. The projection of $\boldsymbol{r}$ onto some local time vector $\boldsymbol{\theta}$ is given by $-(\boldsymbol{r}\cdot\boldsymbol{\theta})\boldsymbol{\theta}$. In (**a**), this is shown for the two cases $\boldsymbol{\theta}=\boldsymbol{t}$ and $\boldsymbol{\theta}=\boldsymbol{t}'$. The projection represents the temporal or time part of $\boldsymbol{r}$ expressed in terms of the time vector of the chosen frame, either $\boldsymbol{t}$ or $\boldsymbol{t}'$ as appropriate. The coefficient $-\boldsymbol{r}\cdot\boldsymbol{\theta}$ equates to the scalar part of the equivalent (3+1)D paravector for $\boldsymbol{r}$ that is observed in the $\boldsymbol{\theta}$-frame. A similar principle applies to the vector part using $\boldsymbol{r}\wedge\boldsymbol{\theta}$, but this is not so easily displayed. In (**b**), we take the specific example $\boldsymbol{r}=\tau\boldsymbol{t}'$, the spacetime vector representing the history of the spatial origin of the $\boldsymbol{t}'$-frame where the proper time is τ. The projection of $\boldsymbol{r}$ onto $\boldsymbol{t}'$ is $\boldsymbol{r}$ itself, whereas the projection back down $\boldsymbol{x}'$ onto $\boldsymbol{t}$ is $\gamma\tau\boldsymbol{t}$ (from Equation 9.12).

$$\begin{bmatrix} t \\ x \\ y \\ z \end{bmatrix} \mapsto \begin{bmatrix} t' \\ x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \gamma(t-vx) \\ \gamma(x-vt) \\ y \\ z \end{bmatrix} \tag{9.29}$$

It will be seen right away that this is exactly the same result as implied by Equation (9.26b) with $r_t = t$, $r_x = x$, $r_t^- = t'$, and $r_x^- = x'$, and it is just the familiar form of Lorentz transformation expressed in terms of components, the form that the reader is most likely to see in standard introductions to the special relativity [45; 48, chapter VII].

At this stage, it is worth recapitulating the key points:

- The difference between the Lorentz transformation "acting on the components" versus acting on basis elements amounts only to a minor but nevertheless critical detail, the sign of v.
- This is also a source of confusion in concept since the transformation does not actually act on the components.

- The components only exhibit consequential changes as a result of the transformation of the basis vectors (passive transformation).
- When the vector itself undergoes the transformation (active transformation), there is no change of basis vectors, but the components change in the same sense as for basis vectors (i.e., if the transformation had actually been applied to them).
- When only the basis is transformed, the components must change in the contrary sense to the basis vectors so that the net change to the vector itself is 0.
- Equation (9.27) is of fundamental importance. By treating $\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ as purely symbolic, it allows us to express a vector in terms of any given basis.

9.9 DILATIONS

Having dealt in depth with rotations in their spacetime form as a major class of linear transformation that fits in very well with geometric algebra, it is useful to touch on a different kind of linear transformation, which, although very simple indeed, proves not to be so cooperative when it comes to expressing it in a compact form, that is to say, it cannot generally be represented in such a simple manner as a rotation.

A dilation in one dimension may be thought of as a simple scaling operation applied only to that part of a vector that lies along a given direction. For a vector $\boldsymbol{u}$, dilation along the unit vector $\boldsymbol{a}$ would therefore result in

$$\boldsymbol{u} \mapsto \lambda_a \boldsymbol{u}_{\parallel} + \boldsymbol{u}_{\perp} \tag{9.30}$$

where $\boldsymbol{u}_{\parallel}$ and $\boldsymbol{u}_{\perp}$ are the parts of $\boldsymbol{u}$ that are respectively parallel and perpendicular to $\boldsymbol{a}$ and λ_a is the dilation factor. Note that a dilation factor of +1 results in no change, whereas a factor of −1 reflects the vector, and for the extreme case $\lambda_a = 0$, the vector is "flattened" by completely removing the part in the $\boldsymbol{a}$-direction. Ultimately, we can have an $\boldsymbol{n}$-dimensional dilation that is compounded from separate dilations along n independent axes. We could also include time dilation here, but for the moment, let us stay with purely spatial dilations. The simplest form of a general spatial dilation along three normalized axes $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ is

$$\boldsymbol{u} \mapsto \lambda_a (\boldsymbol{a} \cdot \boldsymbol{u}) \boldsymbol{a} + \lambda_b (\boldsymbol{b} \cdot \boldsymbol{u}) \boldsymbol{b} + \lambda_c (\boldsymbol{c} \cdot \boldsymbol{u}) \boldsymbol{c} \tag{9.31}$$

Or, if we subsume the dilation parameters into the axis vectors themselves by giving them magnitudes other than unity, we may replace the effect of λ_a with $\boldsymbol{a}^2$, λ_b with $\boldsymbol{b}^2$ and λ_c with $\boldsymbol{c}^2$, so as to express this a little more neatly as

$$\boldsymbol{u} \mapsto (\boldsymbol{a} \cdot \boldsymbol{u}) \boldsymbol{a} + (\boldsymbol{b} \cdot \boldsymbol{u}) \boldsymbol{b} + (\boldsymbol{c} \cdot \boldsymbol{u}) \boldsymbol{c} \tag{9.32}$$

From the similarity to Equation (9.27), it can be seen that these equations simply amount to a modification of the components of $\boldsymbol{u}$ along the chosen axes $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$, but

unless these axes are mutually orthogonal, there will be some vectors for which dilation along one direction will produce the same result as dilation in some other direction. A set of mutually orthogonal axes for which the dilations are uncoupled are called principal axes. If $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$ are the normalized principal axes of a dilation D, then $\boldsymbol{a} \mapsto \mathsf{D}\boldsymbol{a} = \lambda_a(\boldsymbol{a}\cdot\boldsymbol{a})\boldsymbol{a} + \lambda_b(\boldsymbol{b}\cdot\boldsymbol{a})\boldsymbol{b}\ldots = \lambda_a\boldsymbol{a}$ and similarly for $\boldsymbol{b}$ and $\boldsymbol{c}$, which demonstrates that the dilations of vectors along these axes are indeed uncoupled. Unfortunately, there is no more compact way of writing Equation (9.32). Geometric algebra simply does not offer a convenient way to express a general dilation along similar lines to a rotor.[1]

Dilations come into play in electromagnetic theory when we have some sort of anisotropic medium, for example, one of many types of natural crystals or structured synthetic media. For example, $\mathbf{D} = \varepsilon\mathbf{E}$ becomes $\mathbf{D} = \boldsymbol{\in}\mathbf{E}$ where, after the fashion of Equation (9.31), we can write $\boldsymbol{\in}\mathbf{E} = \varepsilon_a(\boldsymbol{a}\cdot\mathbf{E})\boldsymbol{a} + \varepsilon_b(\boldsymbol{b}\cdot\mathbf{E})\boldsymbol{b} + \varepsilon_c(\boldsymbol{c}\cdot\mathbf{E})\boldsymbol{c}$. The uniaxial case, however, is somewhat easier since this can be simplified to $\boldsymbol{\in}\mathbf{E} = (\varepsilon_a - \varepsilon_0)(\boldsymbol{a}\cdot\mathbf{E})\boldsymbol{a} + \varepsilon_0\mathbf{E}$ where $\boldsymbol{a}$ is the only axis of anisotropy. Even this is clearly more awkward to deal with than the simple isotropic case that we have been dealing with here [41].

9.10 EXERCISES

1. **(a)** If $\boldsymbol{a}$ and $\boldsymbol{b}$ are vectors, show that the rotor equation (Equation 9.8) applies to a bivector $\boldsymbol{ab}$ as well as the vector $\boldsymbol{u}$.

(b) Show that the same equation therefore applies to all objects in a geometric algebra.

2. **(a)** If $\boldsymbol{R}$ rotates $\boldsymbol{a}$ into $\boldsymbol{a}'$and $\boldsymbol{R}'$ rotates $\boldsymbol{a}'$ into $\boldsymbol{a}''$, what is the rotor that takes $\boldsymbol{a}$ into $\boldsymbol{a}''$?

(b) What is the rotor that takes $\boldsymbol{a}''$ into $\boldsymbol{a}$?

3. **(a)** Show that the Lorentz transformation of a null vector is itself a null vector.

(b) Show that timelike and spacelike vectors also retain their characters under a Lorentz transformation.

(c) Does future pointing stay the same under a Lorentz transformation?

(d) If $\underset{\sim}{\boldsymbol{u}}$ is a spatial vector in the $\boldsymbol{t}$-frame, that is, $\underset{\sim}{\boldsymbol{u}}\cdot\boldsymbol{t} = 0$, in what frame is $\boldsymbol{L}\underset{\sim}{\boldsymbol{u}}$ a spatial vector?

4. For any vector $\boldsymbol{u}$, let $\boldsymbol{u}' = \mathsf{L}\boldsymbol{u} = \boldsymbol{L}\boldsymbol{u}\boldsymbol{L}^\dagger$ where $\boldsymbol{L}$ takes the form $a - b\boldsymbol{N}$ with $\boldsymbol{N}$ being some unit bivector and a and b being constants as given in either Equations (9.9) or (9.10).

(a) Show $\boldsymbol{u}' = \boldsymbol{u} + 2ab(\boldsymbol{u}\cdot\boldsymbol{N}) + 2b^2(\boldsymbol{u}\cdot\boldsymbol{N})\cdot\boldsymbol{N}$.

(b) Interpret this result geometrically when $\boldsymbol{N}$ is (i) a spacelike bivector and (ii) a timelike bivector.

5. If $\boldsymbol{v}' = \mathsf{L}\boldsymbol{v} = \boldsymbol{L}\boldsymbol{v}\boldsymbol{L}^\dagger$ and $\boldsymbol{u}' = \mathsf{L}\boldsymbol{u} = \boldsymbol{L}\boldsymbol{u}\boldsymbol{L}^\dagger$ where $\boldsymbol{u}$ and $\boldsymbol{v}$ are both vectors, evaluate $\boldsymbol{u}'\boldsymbol{v}'$, $\boldsymbol{u}'\cdot\boldsymbol{v}'$, and $\boldsymbol{u}'\wedge\boldsymbol{v}'$.

6. Let $\boldsymbol{J}_0$ be the electromagnetic source density vector for a charge density ρ that is stationary in the $\boldsymbol{t}$-frame.

(a) What is the spacetime form taken by $\boldsymbol{J}_0$?

[1] Similar constructions may be used, but the result is not a pure dilation.

(b) Find the form of $\boldsymbol{J}_0'$, the electromagnetic source density vector obtained by applying a Lorentz transformation with velocity parameter $\mathbf{v}$ directly to $\boldsymbol{J}_0$.

(c) Show that the spacetime split of $\boldsymbol{J}_0'$ in the $\boldsymbol{t}$-frame represents a charge density ρ' and current density $\rho'\mathbf{v}$.

(d) Contrast this with the conventional definition of electromagnetic source density in (3+1)D and explain the difference.

(e) Express $\boldsymbol{J}_0$ and $\boldsymbol{J}_0'$ in the $\mathfrak{v}$-frame that moves with velocity $\mathbf{v}$ relative to the $\boldsymbol{t}$-frame.

7. (a) Provided $\boldsymbol{N}^2 = 1$, show that $\boldsymbol{L} = a - b\boldsymbol{N}$ may be written in the alternative form $\boldsymbol{L} = e^{-\psi \boldsymbol{N}}$ where $a = \cosh(\psi/2)$ and $b = \sinh(\psi/2)$.

(b) Two Lorentz transformations are given by $\boldsymbol{L}_1 = a_1 - b_1\boldsymbol{N}_1$ and $\boldsymbol{L}_2 = a_2 - b_2\boldsymbol{N}_2$. Compare the compounded transformations $\boldsymbol{L}_1\boldsymbol{L}_2$ and $\boldsymbol{L}_2\boldsymbol{L}_1$.

(c) Simplify the results for the case when $\boldsymbol{N}_1$ and $\boldsymbol{N}_2$ are in the same plane.

(d) Comment on how compounding the two transformations works using instead the exponential forms $\boldsymbol{L}_1 = e^{-\psi_1 \boldsymbol{N}_1}$ and $\boldsymbol{L}_2 = e^{-\psi_2 \boldsymbol{N}_2}$.

8. Differentiate between active and passive transformations and show how they are related.

9. Evaluate $\boldsymbol{\nabla} e^{\psi(t)\boldsymbol{N}}$ and $\boldsymbol{\nabla} e^{\boldsymbol{r}\cdot\boldsymbol{N}}$ where ψ is a scalar, $\boldsymbol{r}$ is the vector $x\boldsymbol{x} + y\boldsymbol{y} + z\boldsymbol{z} + t\boldsymbol{t}$ and $\boldsymbol{N}$ is some bivector constant. What grades are involved in each case?

Chapter 10

Further Spacetime Concepts

In Chapters 7 and 8, we introduced the basic ideas of spacetime, the spacetime geometric algebra, and the relationship between spacetime and (3+1)D. From these, we derive the main tools we will need in order to get to grips with spacetime electromagnetic theory. This is sufficient if all we wish to do is to find spacetime counterparts for the (3+1)D equations that were the subject of Chapter 5. The fundamentals of electromagnetic theory, however, originate in special relativity, and so in order to be able to explore these even at a basic level, we need to understand changing basis vectors through a Lorentz transformation, which we touched on briefly in Sections 7.6 and 7.7 and discussed at length in Chapter 9. Now, it is relatively easy to understand changing basis vectors as a process in its own right without having to take on board much of the theory of relativity itself, particularly when we make the analogy with other forms of transformation such as a rotation and let the peculiar spacetime metric take care of the actual details. In this section, however, we consolidate on these basic ideas by engaging more directly with the physical interpretation, that is to say, special relativity. If the reader is not inclined to go much further in this direction, it would nevertheless be worth reviewing the ideas concerning frames and time vectors, proper time, and proper velocity in the light of Sections 10.1, 10.4 and 10.5 respectively, and the idea of relative vectors in the main part of Section 10.6. That said, for those readers who wish to get the most benefit from the spacetime approach, it is hoped that this section will afford an understanding of the basic concepts that is at least sufficient to allow the electromagnetic field of a moving point charge to be appreciated for what it really is.

10.1 REVIEW OF FRAMES AND TIME VECTORS

Let us review our discussion of frames so far.

We are at liberty to choose $\boldsymbol{t},\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}$ as the spacetime basis vectors and call this "the $\boldsymbol{t}$-frame". In the nomenclature typical of Hestenes and Doran and Lasenby, the $\boldsymbol{t}$-frame is called the γ_0-frame, γ_0 being the counterpart of $\boldsymbol{t}$ (see Table 7.1). It may be "our rest frame", "the lab frame", and so on, in which the basis, or frame, stays fixed with respect to us, wherever we are. Alternatively, we may mean some arbitrary

Understanding Geometric Algebra for Electromagnetic Theory, First Edition. John W. Arthur.

frame for which $\boldsymbol{t}$ is merely a label. It is important, however, to realize that there is nothing special about a frame, and so this does not really matter. Apart from the fact that it is extremely convenient to have them orthonormal, we may also orientate $\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}$ in any way we like and measure time by our chosen clock, but unless we have some specific axes in mind these basis elements, and hence the frame itself, remain as abstractions.

Whether or not our $\boldsymbol{t}$-frame is real or abstract, if we place ourselves at its spatial origin Ω, we will have a history given by $t\boldsymbol{t}+0\boldsymbol{x}+0\boldsymbol{y}+0\boldsymbol{z}=t\boldsymbol{t}$. Anyone who is at rest with us shares this time vector and this frame so that their history will simply be $\boldsymbol{r}=t\boldsymbol{t}+\underset{\sim}{\boldsymbol{r}}$ where $\underset{\sim}{\boldsymbol{r}}$ is a fixed vector that indicates to us their location within our frame. As we have previously pointed out, while $\boldsymbol{r}$ is independent of any choice of basis, that is to say invariant, $\underset{\sim}{\boldsymbol{r}}$ is not, because it depends on the choice of $\boldsymbol{t}$ that goes with it. The $\boldsymbol{t}$-frame velocity associated with the evolution of any vector is obtained by differentiation with respect to t as discussed in Section 7.5, and in the case of $\boldsymbol{r}$, this gives $\boldsymbol{v}=\partial_t(t\boldsymbol{t}+\underset{\sim}{\boldsymbol{r}})=\boldsymbol{t}$. Given that the velocity translates to (3+1)D through Equation (8.17), that is to say $\boldsymbol{v}\leftrightarrow -\boldsymbol{t}\boldsymbol{v}=1+\mathbf{v}$, we then have $1+\mathbf{v}=-\boldsymbol{t}\boldsymbol{t}=1$ so that $\mathbf{v}=0$. In other words we observe the point specified $\boldsymbol{r}$ to be at rest. The (3+1)D view from our frame therefore agrees with the spacetime representation.

As discussed in Sections 7.7.3 and 9.4, another frame with its origin $\boldsymbol{\Omega}'$ moving with velocity $\boldsymbol{v}$ with respect to the $\boldsymbol{t}$-frame, may equally well be chosen, and we may label its basis vectors as $\boldsymbol{t}',\boldsymbol{x}',\boldsymbol{y}',\boldsymbol{z}'$. Given its time vector is $\boldsymbol{t}'$, we refer to it as the $\boldsymbol{t}'$-frame. Expressed in terms of the $\boldsymbol{t}$-frame, the history of its origin will be $\boldsymbol{\Omega}'(t)=t\boldsymbol{v}$, which, by doing the now familiar separation into temporal and spatial parts, we can put into the convenient form $t\boldsymbol{v}=t(\boldsymbol{t}+\underset{\sim}{\boldsymbol{v}})$. While any initial offset between the two origins has been ignored, the spatial part of the velocity vector $\underset{\sim}{\boldsymbol{v}}$ can be in any direction. From the $\boldsymbol{t}$-frame, the (3+1)D velocity observed for $\boldsymbol{\Omega}'$ is found from Equation (8.17) to be $\mathbf{v}=-\boldsymbol{t}\underset{\sim}{\boldsymbol{v}}=\underset{\sim}{\boldsymbol{v}}\boldsymbol{t}$. Now, turning things around to the $\boldsymbol{t}'$-frame's viewpoint, here $\boldsymbol{\Omega}'$ is at rest, exactly the same situation that applied to Ω in the $\boldsymbol{t}$-frame. If the history of Ω in terms of $\boldsymbol{t}$-frame parameters is $\boldsymbol{\Omega}(t)=t\boldsymbol{t}$, one of the key implications of relativity is that the history of $\boldsymbol{\Omega}'$ in terms of $\boldsymbol{t}'$-frame parameters must take exactly the same form, that is to say, we must have $\boldsymbol{\Omega}'(t')=t'\boldsymbol{t}'$. This asserts that the spatial part of $\boldsymbol{\Omega}'$ is stationary (in fact it is zero) in that frame and the only thing that actually changes with time is the local time parameter t'. It therefore follows that $t'\boldsymbol{t}'$ and $t(\boldsymbol{t}+\underset{\sim}{\boldsymbol{v}})$ are just different representations of the same spacetime vector $\boldsymbol{\Omega}'$ as expressed in the two different frames. This simple fact led, via Equation (7.19), to the identification of $\boldsymbol{t}'$ (or $\mathfrak{v}$) with $\gamma(\boldsymbol{t}+\underset{\sim}{\boldsymbol{v}})$ where γ is merely the parameter required to normalize $\boldsymbol{t}'$, and is therefore equal to $\left(1-\underset{\sim}{\boldsymbol{v}}^2\right)^{-1/2}$ or, more familiarly, $\left(1-v^2\right)^{-1/2}$. Given $\mathbf{v}=\underset{\sim}{\boldsymbol{v}}\boldsymbol{t}$, this is also identical to $\left(1-\mathbf{v}^2\right)^{-1/2}$. This is confirmed by Equation (9.11) to be the same thing as applying a Lorentz transformation to the $\boldsymbol{t}$-frame basis vectors in order to generate the corresponding set in the $\boldsymbol{t}'$-frame.

While it seems quite amazing that the relationship between $\boldsymbol{t}'$ and $\boldsymbol{t}$ is to be found so easily, it must be borne in mind that we have treated these frames correctly from a simple principle of relativity and the spacetime metric does the rest.

The time vector is clearly of particular significance to the concept and role of frames. We are generally only interested in the time vectors here because it is tacitly assumed that spatial rotations are not involved, that is to say, we have no rotating reference frames. Any change in the spatial basis vectors therefore arises only as a consequence of a change in time vector so that the orthonormality of the entire basis is maintained. Choosing the time vector is the same as choosing the reference frame velocity. The approach that most of us will be familiar with is choosing the latter without any consideration given to time vectors.

10.2 FRAMES IN GENERAL

Einstein's theory of special relativity takes the equivalence of all inertial, that is to say nonaccelerating, reference frames as a basic principle. Indeed, it is emphasized that there is no such thing as an absolute rest frame; all things are relative to one another. This principle does not, however, forbid us from adopting some reference frame as a standard frame to which we can refer all things, that is to say, what we have been calling the $\boldsymbol{t}$-frame. It does not matter where we choose the origin and, as long as it remains constant, the idea of taking velocity into consideration is pointless. We want our frame to be where we usually are and the fact that *we* may be traveling with respect to some other possible origin simply does not matter.

It would be ideal if we were able to choose a proper inertial reference frame, but then for practical purposes, it turns out that we are actually quite happy to use something set up on the surface of the earth, an object that is not flat and which rotates on its own axis while traveling in an elliptical orbit, as a convenient choice for most practical purposes! For any measurement we may make, all we need to be sure of is that either this does not matter or that we are duly correcting for it. In fact, any other choice would be downright inconvenient except in special circumstances such as space travel.

For thousands of years, the stars have provided a reference frame that people have used to navigate by when there are no other observable reference points. Latterly, we have used this for travel in space itself, and we have also created an artificial frame of reference by using satellite systems to provide a round-the-clock substitute for the stars. We observe these satellites by means of their microwave signals, which are encoded so as to give us the position and time at our location virtually anywhere in the world with unprecedented accuracy. Consequently, on earth itself, we have a limitless choice of suitable reference frames, and at worst, it requires a little computation to translate between them. There is consequently no absolute or completely standardized frame but, as we have already said, we only need to be able to identify some particular frame and by common consent agree on it as *our* frame of reference. It is therefore in this spirit that we generally use the terms $\boldsymbol{t}$-frame, our rest frame, the lab frame, and so on.

Choosing a frame is not only about choosing a suitable set of axes, but also about choosing a clock, which is where the local time vector comes into play. Our clock must stay with us, that is to say, with our frame, but the science of measuring

time is such that we have no problem in keeping track of it to an extraordinarily high degree of accuracy. We are also able to replicate our clock anywhere within our frame that we please by means of any number of convenient clocks that are all accurately synchronized. The real distinction we make is that when we set up a new reference frame traveling at some nonzero velocity with respect to our own, we cannot consider the clocks in the two different frames as being synchronized—even if this was the case at some previous time. The new frame has a different clock and therefore a different time vector.

Frames give some sense of reality to a particular physical problem, but, as previously pointed out, the ability to work without specific frames is of key importance and a spacetime vector $\boldsymbol{r}$ is an entity in its own right, independent of any frame. For most of us, however, it is difficult to work with abstract things, and frames and basis elements creep in at an early stage, particularly if we want to draw a diagram to represent our thoughts. This does not matter a great deal and creating a tangible set of basis vectors may help us to express a situation or formulate a problem. The reader, however, must always be aware that the notion of frames is merely an adjunct to spacetime itself, like different windows that allow us to express what is going on in spacetime from different viewpoints.

Given all the possible flexibility in choosing a frame, it will be of great benefit to introduce some sort of consistency between the available options. We have to make an initial choice of four basis vectors, say $\boldsymbol{t}$, $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$, which we once again simply refer to as the $\boldsymbol{t}$-frame. This frame generates an infinite number of possible $\boldsymbol{t}'$-frames in which the basis vectors $\boldsymbol{t}', \boldsymbol{x}', \boldsymbol{y}', \boldsymbol{z}'$ are related to $\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ by a pure Lorentz transformation. Since spatial rotations and reflections are excluded, the alignment of all such frames remains consistent in the sense that they always project out the same set of (3+1)D basis vectors $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$. For example, we note from Equation (9.14) that under such a transformation with velocity parameter along $\boldsymbol{x}$, we have $\boldsymbol{x}' \wedge \boldsymbol{t}' = \boldsymbol{x} \wedge \boldsymbol{t}$, that is to say $\mathbf{x}' = \mathbf{x}$, while $\mathbf{y}$ and $\mathbf{z}$ are also unaffected. There will be further discussion on this particular point in Section 10.6.2, but it can be safely assumed that all the frames that we discuss will belong to this important class of related frames that are referred to as *Lorentz frames*.

An observer can find no means to differentiate between the (3+1)D basis vectors $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ and those of any other corresponding Lorentz frame. All such observers see their basis vectors as orthonormal triples where any two corresponding basis vectors, for example, $\mathbf{z}$ and $\mathbf{z}'$, are collinear and of the same measured length. Indeed, it is embedded in special relativity that any two such sets should be equivalent. There is therefore no point in putting primes on the basis vectors of the primed frame for they are in fact just the same as the original basis vectors. Interestingly, the same idea also applies to time in as much as a time vector may be represented differently in different frames, whereas its (3+1)D counterpart, the unit scalar, is the same in *all* frames.

Ubiquitous though it may seem, it must always be remembered that

- the $\boldsymbol{t}$-frame is not absolute;
- by default, it generally means some unspecified, hypothetical frame;

- although we may often take it to imply our own rest frame, this is not essential;
- when we say "moving" or "at rest" in no particular context, we generally mean with respect to the ***t***-frame; and
- "local" refers to some particular frame given by the context.

Furthermore,

- any two Lorentz frames are equivalent, and
- they also share the same (3+1)D basis vectors.

10.3 MAPS AND GRIDS

Whatever the perception of time and distance within a given frame, an observer can always deduce what is happening in, or relative to, some other frame by means of observation (the radar principle, accredited to E.A. Milne [44, p. 58; 49]). Any observer may therefore use observations to construct some sort of map of their environment, for example a star map, that shows some selected reference points that have been taken as being fixed within the observer's adopted frame. The validity of any such map will therefore be with regard to this construct, which we may regard as being associated with a standard ***t***-frame; that is, we associate our spatial basis vectors with given orientations on the map and adopt a standard clock. Any observer may then measure their progress through space with reference to the map and associated clock. Given some suitable reference points and the standard clock, a 4D grid can be constructed by interpolation; for example, the 2D grid shown in solid lines in Figure 7.4 could be a slice through one such grid. It would be natural to use this procedure as a means of choosing the basis vectors by aligning them with the edges of any unit cell of the grid. This then provides the coordinate system (always now including a clock) for our map. All observers can agree on this "grid plus clock" as a way of locating any event, or trajectory, within our universe. The map itself need only to be a list of spatial co-ordinates for the known "fixed objects," including those that were used to determine the grid in the first place. Clearly, the time does not need to be recorded on the map itself as long as it records only fixed objects. In principle, if it is possible to predict movement with respect to the grid, that is to say a trajectory, the map could be made to update itself with the passage of time—but for this, it would need to have access to the grid clock to ensure that everything remains synchronized. By means of suitable observations of established reference points on the map, any observer is then able to determine their own trajectory with respect to the map and grid as a function of time.[1] This may seem a challenge, but

[1] We do not specify what time we mean here, but by knowing their trajectory according to their own clock, any observer would be able to work out the relationship between their local time vector and the grid's time vector.

such a system is effectively in place for all the major bodies in the solar system and some even beyond it.

A coordinate system for space may well be the province of astronomers, but for maps on earth we do have a well-known system determined by North, West, and altitude, and for our clock, we have GMT and date. The existence of various time zones demonstrates just how arbitrary the choice of clock is—the important thing is that we can always trace the time on our clocks back to GMT. We also have to choose the units in which to measure space and time, but again, this is arbitrary: meters, feet, nautical miles, light-years, seconds, calendar years, sidereal years, and so on. Although there is no such system that covers everything in the universe, this does not prevent us from referring to some standard, universal map and grid on a conceptual basis.

As we have previously discussed, special relativity tells us that there is no absolute reference frame, but it does not preclude us from using the idea of a common map and clock, even though in many cases it may be notional rather than actual. Any observer would know that a grid square in the ***xy*** plane of the map represents a square measuring 1×1 and that a tick of its clock represents a time interval of 1, *in the frame used to define the map*. Since everybody would be able to agree on it, it would be a very convenient choice for the ***t***-frame. But observers will also know that a map square will not measure 1×1 in their own frame unless they happen to deduce from their observations that they are actually at rest with respect to the frame of the map. The same discrepancy will also apply to their observations of the map's clock. But at least they will know that they are at rest on the map if the time is the only spacetime coordinate that they see changing with respect to the grid. On the other hand, if their position coordinates are also changing, they can readily work out their instantaneous velocity in the frame of the map.

The very idea of drawing spacetime vectors on a diagram incorporating a specific coordinate system or set of basis vectors seems totally contradictory to the basic principle that they are not to be fixed down in this way, yet this is in fact just the thing that we usually do. But all we are doing is to exercise our freedom to choose any convenient origin and set of axes to measure from, and the principle itself lies in this freedom of choice. We are not prevented from making a choice, but whatever choice we make is no different from any of the other possibilities.

Given the principle of a standard 4D grid map of the universe, irrespective of their state of motion, any observer would be able to use its four coordinates as a means of navigation. More generally, it provides a means of parameterizing events and histories in a way that we can all agree on. But on the other hand, we do not always have to refer things to a map. As discussed in Section 7.10, it is usually of benefit to solve physical problems in a general way by using the variables and parameters that are most convenient for the purpose. Not only does this give much more flexibility than being tied to a specific map, the results may always be used in conjunction with a specific grid or map when we wish to apply them to a specific situation.

10.4 PROPER TIME

Earlier we saw that a change of basis vectors from one frame to another resulted in a different time parameter, for example, in Equation (7.19) where τ replaces t. We will now discuss the meaning and significance of this change.

Equation (9.27) allows us to express the history of a particle $\boldsymbol{r}(\lambda)$ in terms of the basis vectors of any Lorentz frame. This automatically extends to being able to convert the resulting expression from one frame to another. While in applying this in the form of Equation (9.28) the algebra was simplified by taking the direction of motion to be along $\boldsymbol{x}$, $\boldsymbol{x}$ itself is just a convenient label for the unit vector along the direction of motion. Take, for example, $\boldsymbol{r}(t) = t\boldsymbol{t} + vt\boldsymbol{x}$, which is the familiar history of a particle moving with velocity v along $\boldsymbol{x}$ and passing through the spatial origin of the $\boldsymbol{t}$-frame at $t = 0$. According to Equation (9.28), this history is represented in the $\boldsymbol{t}'$-frame by

$$\begin{aligned}\boldsymbol{r}(t) &= \gamma\left[\left(t - v(vt)\right)\boldsymbol{t}' + \left(vt - vt\right)\boldsymbol{x}'\right] \\ &= \gamma\left(1 - v^2\right)t\boldsymbol{t}' + 0\boldsymbol{x}' \\ &= \left(\gamma^{-1}t\right)\boldsymbol{t}' + 0\boldsymbol{x}' \\ &= \left(\gamma^{-1}t\right)\boldsymbol{t}'\end{aligned} \tag{10.1}$$

which describes a particle at rest at the spatial origin of the $\boldsymbol{t}'$-frame, but with the time measured as $t' = \gamma^{-1}t$ along the basis vector $\boldsymbol{t}'$. Since $\gamma^{-1} = \left(1 - v\right)^2 < 1$, we find $t' < t$ so that time runs slower on the $\boldsymbol{t}'$-frame clock than on the $\boldsymbol{t}$-frame clock. Two events in the $\boldsymbol{t}$-frame that are seen to be separated in time by the differential dt will therefore be seen in the $\boldsymbol{t}'$-frame as being separated by $d\tau$ where

$$d\tau = \gamma^{-1}dt \tag{10.2}$$

A particle at rest in $\boldsymbol{t}'$-frame may be thought of as carrying the $\boldsymbol{t}'$-frame along with it. The passage of time as seen by this particle, that is to say where $\boldsymbol{t}'$ is the unit time vector, will therefore be

$$\tau_2 - \tau_1 = \int_{t_1}^{t_2} \gamma(t)^{-1}\, dt \tag{10.3}$$

This is called the proper time interval for the particle, meaning the passage of time according to the particle itself or, put in the usual terms, according to a clock at rest with the particle. The meaning of the word "proper" in this context needs clarifying. It is intended to convey the idea of being a property of the particle itself. Being free from any arbitrarily defined external reference frame, it is therefore unambiguous. It is not even necessary that the particle's velocity should be constant, Equation (10.3) always allows us to relate the particle's proper time

to the passage of time in some other frame using, say, the $\boldsymbol{t}$ clock. The relationship expressed in Equations (10.2) and (10.3) is crucial to the analysis of dynamical problems.

Now, before leaving this subject, it is important to clarify one of those typical paradoxes associated with relativity. Let us take two different frames, where the one that measures time as τ is traveling with velocity v with respect to the other that measures the time as t. As we have just seen, the rate at which their respective clocks are seen to run is given by $d\tau = \gamma^{-1}dt$. But we can equally well turn this situation round the other way so that the frame that measures time as t travels with velocity $-v$ with respect to the one that measures the time as τ. Because $\gamma(v) = \gamma(-v)$, we now find $dt = \gamma^{-1}d\tau$. Now $\gamma \neq 1$ unless the two frames are at rest with respect to each other; these two statements appear to be irreconcilable, yet they are nevertheless true. The problem is clarified by recalling that one of the two frames is taken as a rest frame and the other as a moving frame. Whichever way round we make the choice, the clock in the moving frame is always measured to be slower than the clock of the frame that is at rest. At the heart of this apparent conundrum is the fact that each frame sees the time in the other frame as being dependent on position. For example, in the case of the usual $\boldsymbol{t}$ and $\boldsymbol{t}'$ frames, from Equation (9.29) $t' = \gamma(t - vx)$ while, by symmetry, $t = \gamma(t' + vx')$. If we have $t' = \gamma^{-1}t$, then we must have $x = vt$ and $x' = 0$, whereas if we have $t = \gamma^{-1}t'$, then this requires $x' = -vt$ and $x = 0$. The relationships $t' = \gamma^{-1}t$ and $t = \gamma^{-1}t'$ therefore apply to two separate locations. From the $\boldsymbol{t}$-frame we see $t' = \gamma^{-1}t$ at the origin of the $\boldsymbol{t}'$-frame, while from the $\boldsymbol{t}'$-frame we see $t = \gamma^{-1}t'$ at the origin of the $\boldsymbol{t}$-frame.

10.5 PROPER VELOCITY

As discussed in Section 10.3, the moving observer may note their passage through space using the grid of some agreed map. Quite naturally, they may wish to use *their own clock* rather than the *map clock* to record their rate of progress, for example, so that they may readily work out how long it will take to go from $\boldsymbol{a}$ to $\boldsymbol{b}$ on the map without having to transform map distances to local distances or map time to their own local time. On the one hand, their velocity along some interval $d\boldsymbol{r}$ on the map (which of course includes the time on its clock as well as the usual spatial features) is given by $d\boldsymbol{r}/dt$ where dt is measured in map time. The result is the usual velocity that we have been denoting by $\boldsymbol{v}$. On the other hand, using *their own clock, on which τ is their proper time*, an observer measures their velocity as being $d\boldsymbol{r}/d\tau$, and the result may be quite different. This, then, gives rise to the notion of "proper" velocity $\boldsymbol{\upsilon}$, which was first introduced in Sections 7.7.3–7.7.4:

$$\boldsymbol{\upsilon} \equiv \partial_\tau \boldsymbol{r} = \frac{d\boldsymbol{r}}{d\tau} = \dot{\boldsymbol{r}} \tag{10.4}$$

The use of the small overdot to indicate differentiation dates back to Newton and the origins of calculus, but Minkowski used it with the specific meaning of differentiation with respect to proper time, now a convention of spacetime physics.

Recall that, in contrast, we employed an open overdot to identify which variable within a product is to be differentiated, for example, as in Equation (7.23).

If the observer's history is given in the $\boldsymbol{t}$-frame as $\boldsymbol{r}(t)$, we then have

$$\begin{aligned}\boldsymbol{\upsilon} &= \frac{d\boldsymbol{r}(t)}{dt}\cdot\frac{dt}{d\tau}\\ &= \boldsymbol{v}\gamma\\ &= \gamma(\boldsymbol{t}+\underset{\sim}{\boldsymbol{v}})\end{aligned} \tag{10.5}$$

As before, we have separated $\boldsymbol{v}$ into $\boldsymbol{t}+\underset{\sim}{\boldsymbol{v}}$ so that $\underset{\sim}{\boldsymbol{v}}$ gives the purely spatial part of the velocity associated with the magnitude and direction of the motion in space. It follows directly that $\boldsymbol{\upsilon}$ is normalized and timelike:

$$\begin{aligned}\boldsymbol{\upsilon}^2 &= \gamma^2\left(\boldsymbol{t}^2+(\boldsymbol{t}\underset{\sim}{\boldsymbol{v}}+\underset{\sim}{\boldsymbol{v}}\boldsymbol{t})+\underset{\sim}{\boldsymbol{v}}^2\right)\\ &= \gamma^2\left(-1+\underset{\sim}{\boldsymbol{v}}^2\right)\\ &= -1\end{aligned} \tag{10.6}$$

Now, the remarkable thing is that the proper velocity is identical to the time vector $\boldsymbol{\upsilon}$ in Equations (7.18) and (7.19) where $\boldsymbol{\upsilon}$ was determined simply by rearranging the history of a moving particle so as to make it take the form of a particle at rest. In Equation (9.11), we found the same thing through a Lorentz transformation acting on the time vector $\boldsymbol{t}$. Although we took $\underset{\sim}{\boldsymbol{v}}$ in the form of $v\boldsymbol{x}$, the transformed time vector $\boldsymbol{t}'$ is otherwise the same as the proper velocity $\boldsymbol{\upsilon}$ as we have defined it here. The proper velocity is therefore identical to the time vector in the rest frame of the moving particle or observer! The rest frame of a particle with proper velocity $\boldsymbol{\upsilon}$ must then be the $\boldsymbol{\upsilon}$-frame. Being intrisically normalized, the magnitude of the local time vector is consequently invariant under a change of frame. This is also consistent with the fact that the Lorentz transformation is an orthogonal transformation and so preserves measure.

It was pointed out in Section 7.7.3 that $\boldsymbol{\upsilon}$ also represents the unit tangent vector to the particle's history. In its own rest frame, where the local time is the proper time τ, and the time vector is the proper velocity $\boldsymbol{\upsilon}$, the particle's history must change by $\boldsymbol{\upsilon}d\tau$ during an instant $d\tau$. The particle's history may therefore be found by integrating $\boldsymbol{\upsilon}$ with respect to proper time:

$$\boldsymbol{r}(\tau) = \int_{\tau_1}^{\tau_2}\boldsymbol{\upsilon}(\tau)d\tau \tag{10.7}$$

Any vector $\boldsymbol{r}$ represents a particle history only if it can be put in the form of Equation (10.7). When the velocity is constant, this takes the simple form $\boldsymbol{r}(\tau)=\tau\boldsymbol{\upsilon}+\underset{\sim}{\boldsymbol{r}}$ as in Equation (7.18). Any equation of this form then qualifies as a history provided it can be adjusted so that $\boldsymbol{\upsilon}^2=-1$, $0<\underset{\sim}{\boldsymbol{r}}^2$, and the scalar τ is a monotonically increasing parameter. This means that $\boldsymbol{\upsilon}$ qualifies as a time vector, which in turn allows us to identify the $\boldsymbol{\upsilon}$-frame in which τ is the time and $\underset{\sim}{\boldsymbol{r}}$ is some

constant spatial vector equal to $\boldsymbol{r}(0)$. Proper velocity is consequently a concept of fundamental importance.

It is clear from Equation (10.5), however, that the representation of $\mathfrak{v}$ has a spatial part in any frame other than the $\mathfrak{v}$-frame itself. In the case of the $\boldsymbol{t}$-frame, the spatial part is $\gamma\underset{\sim}{\boldsymbol{v}}$. While this tells us the direction of the motion of the particle, its magnitude actually exceeds the speed of light when $1/\sqrt{2} < v$. While it may appear to observers in the $\mathfrak{v}$-frame that they are traveling across the map faster than the speed of light, in reaching this result, they have simply used the distances as shown *on the map* and divided them by *their own* measure of elapsed time. The map distances are not the same as the moving observer would measure them to be. This observer's time, their proper time, is γ^{-1} times the map time, and any distance the observer actually measures will be γ^{-1} times the distance given on the map. These factors cancel, so that the $\mathfrak{v}$-frame observer measures the same spatial velocity $\underset{\sim}{\boldsymbol{v}}$ as does the $\boldsymbol{t}$-frame observers.

10.6 RELATIVE VECTORS AND PARAVECTORS

Spacetime gives us a framework within which we can express the events and histories of interest to us without having to consider the view of individual observers. The observer's view of the world is the (3+1)D view that applies only in their own rest frame. How do we generate these views? In Chapter 8, we showed how we may go between spacetime vectors and the more familiar (3+1)D vectors. Provided that they are both associated with the same frame, for example the $\boldsymbol{t}$-frame, this is accomplished by means of a simple "translation" procedure that may be reduced to the algebraic form of Equation (8.5). However, we then went on to show that the idea could be readily extended to different frames, resulting in what is more generally known as a spacetime split. A spacetime split has the physical interpretation of a projection that eliminates the time vector in some chosen frame, leaving us with an observable (3+1)D rendition. To make Equation (8.5) completely general, it is only necessary to regard the choice of $\boldsymbol{t}$ as extending to the time vector in *any* observer's rest frame, that is to say, we can replace $\boldsymbol{t}$ with any other time vector, $\boldsymbol{t}'$. Given some spacetime vector $\boldsymbol{u}$, then the corresponding (3+1)D view of it as seen by any observer at rest in the $\boldsymbol{t}$-frame is given by $-\boldsymbol{t}\boldsymbol{u} = u_t + \mathbf{u}$. An observer at rest in the $\boldsymbol{t}'$-frame, however, gets a different (3+1)D view given by $-\boldsymbol{t}'\boldsymbol{u} = u'_t + \mathbf{u}'$. It is not too difficult to guess that $u_t + \mathbf{u}$ and $u'_t + \mathbf{u}'$ will be related by a Lorentz transformation. The (3+1)D view of any spacetime vector is therefore inherently frame dependent, and, as a result, we may describe (3+1)D paravectors as being "relative." Relative paravectors, and their relative scalar and vector parts, may be considered as being the observables of the physical (3+1)D world. They always require some frame to be specified or implied, and it will be convenient to use the same label, for example, $\boldsymbol{t}$-frame and $\boldsymbol{t}'$-frame, to refer to both a spacetime frame and its associated (3+1)D reference frame. When any form of the word "observe" is used in this context, it may be assumed that what is being observed will be "relative."

We will now proceed with the development of these ideas, but, before moving on, what about other relative objects? While in principle we could have a relative trivector, this is only the same thing as the dual of a relative vector. The timelike and spacelike bivectors are frame dependent since, as we have seen, they intermix under a Lorentz transformation. Being even multivectors, they also translate directly into (3+1)D without premultiplication by some chosen time vector. Since the term "relative" would apply just as well to the spacetime bivector as to its (3+1)D counterpart, there is no point in using it to make the kind of distinction between them that applies to vectors. Spacetime scalars and pseudoscalars also translate into (3+1)D without change, and so the term relative scalar (or pseudoscalar) applies only to the scalar part of a relative paravector (or its dual). In short, relative paravectors, and their relative vector and scalar constituents, are the only relative objects we need to consider.

10.6.1 Geometric Interpretation of the Spacetime Split

Finding the (3+1)D paravector $u_t + \mathbf{u}$ that corresponds to some spacetime vector $\boldsymbol{u}$ has been described as being equivalent to projecting spacetime onto a (3+1)D space. In fact, the projection into the chosen 3D space gives us $\mathbf{u}$, whereas the projection onto the corresponding time axis gives us t. However, this would be much easier to imagine starting from a 3D space rather than a 4D one since the vector part is then projected into a plane rather than a volume. The spacetime split process can therefore be understood by first studying the analogous 3D process in which an object is projected onto a 2D plane.

Take three unit vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ such that $\mathbf{e}_3$ is normal to both $\mathbf{e}_1$ and $\mathbf{e}_2$ and the bivector $\mathbf{e}_1 \wedge \mathbf{e}_2$ determines a plane. Following the basic principle behind Equation (9.27), any vector $\mathbf{u}$ may be expressed as $\mathbf{u} = (\mathbf{e}_1 \cdot \mathbf{u})\mathbf{e}_1 + (\mathbf{e}_2 \cdot \mathbf{u})\mathbf{e}_2 + (\mathbf{e}_3 \cdot \mathbf{u})\mathbf{e}_3$. The effect of the projection is that the part of $\mathbf{u}$ that is parallel to the projection axis $\mathbf{e}_3$ is simply lost. The vector $\mathbf{u}$ will therefore be projected onto the $\mathbf{e}_1 \wedge \mathbf{e}_2$ plane as $\mathbf{u}_{/\!/} = (\mathbf{e}_1 \cdot \mathbf{u})\mathbf{e}_1 + (\mathbf{e}_2 \cdot \mathbf{u})\mathbf{e}_2$. The projection therefore amounts to collapsing everything in the 3D space straight up or down the $\mathbf{e}_3$ axis onto the plane. The vector $\mathbf{u}_{/\!/}$ is simply the part of $\mathbf{u}$ that is parallel to the plane but, by the same token, perpendicular to the projection axis. Note that since $\mathbf{e}_3$ is normal to both $\mathbf{e}_1$ and $\mathbf{e}_2$, we can express the projection onto the plane in the form $\mathbf{e}_3 \wedge \mathbf{u} = (\mathbf{e}_1 \cdot \mathbf{u})\mathbf{e}_3\mathbf{e}_1 + (\mathbf{e}_2 \cdot \mathbf{u})\mathbf{e}_3\mathbf{e}_2 = \mathbf{e}_3\mathbf{u}_{/\!/}$. While the projection itself is given by $\mathbf{u}_{/\!/}$, it will be useful for the present to retain the bivector form $\mathbf{e}_3\mathbf{u}_{/\!/}$. On the other hand, the scalar $\mathbf{e}_3 \cdot \mathbf{u}$ gives us the component of $\mathbf{u}$ that is parallel to the projection axis. But this does not work for just $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$; we can choose any axis to project along. In doing so, the vector $\mathbf{u}$ remains the same throughout and it is only the view of it in the selected plane that changes. We can then put these two pieces of information together in the form

$$\mathbf{e}_3\mathbf{u} = \underbrace{\mathbf{e}_3 \cdot \mathbf{u}}_{\text{scalar}} + \underbrace{(\mathbf{e}_1 \cdot \mathbf{u})\mathbf{e}_3\mathbf{e}_1 + (\mathbf{e}_2 \cdot \mathbf{u})\mathbf{e}_3\mathbf{e}_2}_{\text{projected bivector}} \tag{10.8}$$

Note that by including the scalar part, this is now equivalent to projecting from 3D into a (2+1)D space, and, as no information is lost, the procedure becomes invertible.

Following this analogy, taking a spacetime split simply amounts to projecting out from a 4D space into a (3+1)D space by simultaneously

- projecting *along*, or collapsing, some chosen time axis to form a 3D space spanned by bivectors, and
- projecting *onto* the time axis to form 1D space of scalars, which preserves the information lost by collapsing the time axis.

The scalar part of the spacetime split of a vector in both the $\boldsymbol{t}$-frame and the $\boldsymbol{t}'$-frame was discussed in Section 9.8.2 and is depicted in Figure 9.1. The vector part of the $\boldsymbol{t}$-frame split is equivalent to suppressing the time axis as in going from Figures 7.1 to 7.2. A $\boldsymbol{t}'$-frame split implies that we are doing the same thing but by suppressing the $\boldsymbol{t}'$ axis. If we include the scalar terms projected onto the respective time axes, the resulting (3+1)D space is how the world would be seen from either frame. It is no coincidence that the form of Equation (10.8) is just the same in principle as writing

$$-\boldsymbol{tu} = -\boldsymbol{t}\cdot\boldsymbol{u} + (\boldsymbol{x}\cdot\boldsymbol{u})\boldsymbol{xt} + (\boldsymbol{y}\cdot\boldsymbol{u})\boldsymbol{yt} + (\boldsymbol{z}\cdot\boldsymbol{u})\boldsymbol{zt} \tag{10.9}$$

which is actually what we have been using as the basis of the spacetime split. The spacetime split therefore goes beyond the straight translation procedure that originated from the simple substitution of basis elements. Using the time vector for our rest frame, it amounts to the same thing, but by using a different time vector, we are actually projecting out into a different (3+1)D space or reference frame. Since we can do this for *any* reference frame we choose, it means no longer having to transform between one (3+1)D reference frame and another, which can often be not only tedious but also confusing, especially when there are more than two frames involved. It is only necessary to create a general spacetime description of the situation and then project out into whatever reference frame we please—merely by choosing the appropriate time vector as the projection axis for the spacetime split.

The rationale for this process as it applies to any frame is depicted in Figure 10.1 for the simple case that there are two frames, the usual $\boldsymbol{t}$-frame and the $\boldsymbol{t}'$-frame, where the latter is moving along the $\boldsymbol{x}$ direction with velocity $v\boldsymbol{x}$. While the figure also shows the $\boldsymbol{y}$-axis, the $\boldsymbol{z}$-axis is suppressed; that is, the z-coordinate is constant throughout the diagram. This is all just to make things simple enough to see what is going on. Following Figure 7.4, using the same origin, we can draw in the basis vectors for the $\boldsymbol{t}'$-frame, namely $\boldsymbol{t}'$, $\boldsymbol{x}'$, and $\boldsymbol{y}'$. As before, $\boldsymbol{y}' = \boldsymbol{y}$ and $\boldsymbol{z}' = \boldsymbol{z}$, the only difference being the additional spatial dimension so that we can envisage the 3D space that we are projecting the spacetime split onto as a plane. The choice of a position vector $\boldsymbol{r}$ for the projection is completely arbitrary—the procedure clearly applies to any spacetime vector.

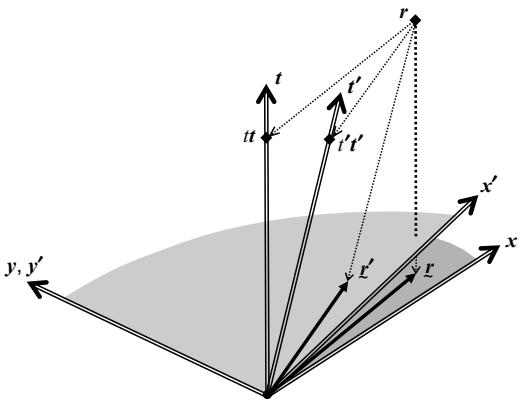

Figure 10.1 Projection of a spacetime vector onto a 3D space in two different frames. Here we show the projection of a spacetime vector onto a 3D space—but $\boldsymbol{z}$ has simply been suppressed since we can represent only time and two spatial dimensions on the page. Two frames are represented, $\boldsymbol{t}$ and $\boldsymbol{t}'$, with the spatial origin of the $\boldsymbol{t}'$-frame moving with velocity $v\boldsymbol{x}$ with respect to the $\boldsymbol{t}$-frame as in previous figures. The y-axis and unit $\boldsymbol{y}$ vector are therefore common to both frames. In the $\boldsymbol{t}$-frame, the event $\boldsymbol{r}$ projects back down $\boldsymbol{t}$ onto $\underset{\sim}{\boldsymbol{r}}$ in the $\boldsymbol{xy}$ plane. In exactly the same way, in the $\boldsymbol{t}'$-frame it projects back down $\boldsymbol{t}'$ onto $\underset{\sim}{\boldsymbol{r}}'$ in the $\boldsymbol{x}'\boldsymbol{y}'$ plane. Note that in a similar way, we can then get the time components of $\boldsymbol{r}$ in each frame by projecting back down either $\underset{\sim}{\boldsymbol{r}}$ or $\underset{\sim}{\boldsymbol{r}}'$ onto the relevant time axis, as is shown in Figure 9.1.

Let us take the case of the $\boldsymbol{t}$-frame first. The figure reveals only the $\boldsymbol{xy}$ plane of the 3D space onto which the projection is made. On projecting $\boldsymbol{r}$ down time dimension onto $\boldsymbol{xy}$, the resulting vector is $\underset{\sim}{\boldsymbol{r}}$, that is to say, the purely spatial part of $\boldsymbol{r}$ in the $\boldsymbol{t}$-frame. This projection is therefore the same as removing the temporal part from $\boldsymbol{r}$ in the given frame. We may refer to this collapsing of the space along a given time vector as being the geometric part of the process. The algebraic part is the formation of $\underset{\sim}{\boldsymbol{r}}\boldsymbol{t}$, the spacetime bivector that is equivalent to the required (3+1)D vector $\mathbf{r}$, as discussed in Section 8.1. Exactly the same process can be applied in the $\boldsymbol{t}'$-frame where we project $\boldsymbol{r}$ back along $\boldsymbol{t}'$, *this time to the* $\boldsymbol{x}'\boldsymbol{y}'$ *plane*, resulting in $\underset{\sim}{\boldsymbol{r}}'$. As we can see, this plane is not the same as the $\boldsymbol{xy}$ plane as it is tilted from it about the $\boldsymbol{y}$-axis (c.f. Figure 7.2 which shows how $\boldsymbol{x}'$ appears tilted as a result of the spacetime norm).

The relative vectors corresponding to the spacetime event $\boldsymbol{r}$ are then $\mathbf{r} = \underset{\sim}{\boldsymbol{r}}\boldsymbol{t}$ in the $\boldsymbol{t}$-frame and $\mathbf{r}' = \underset{\sim}{\boldsymbol{r}}'\boldsymbol{t}'$ in the $\boldsymbol{t}'$-frame. However, the (3+1)D basis vectors corresponding to the two sets of spacetime vectors $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{x}'$, $\boldsymbol{y}'$ are the same, simply $\mathbf{x}$ and $\mathbf{y}$ in each case—a subtlety that originally came up in Section 8.3 and will be fully resolved in Section 10.6.2. Although different, the $\boldsymbol{xy}$ and $\boldsymbol{x}'\boldsymbol{y}'$ planes must therefore translate into equivalent $\mathbf{xy}$ planes, which we may then merge as illustrated in Figure 10.2. Next, we relabel $\underset{\sim}{\boldsymbol{r}}\boldsymbol{t}$ as $\mathbf{r}$ and $\underset{\sim}{\boldsymbol{r}}'\boldsymbol{t}'$ as $\mathbf{r}'$, each of which is then the relative (3+1)D vector counterpart of $\boldsymbol{r}$ in their respective frames. It will often be expedient to keep the scalar part together with the relative vector, and so $t + \mathbf{r}$ and $t' + \mathbf{r}'$ each express the relevant projection of the event $\boldsymbol{r}$ onto (3+1)D as a *relative paravector*.

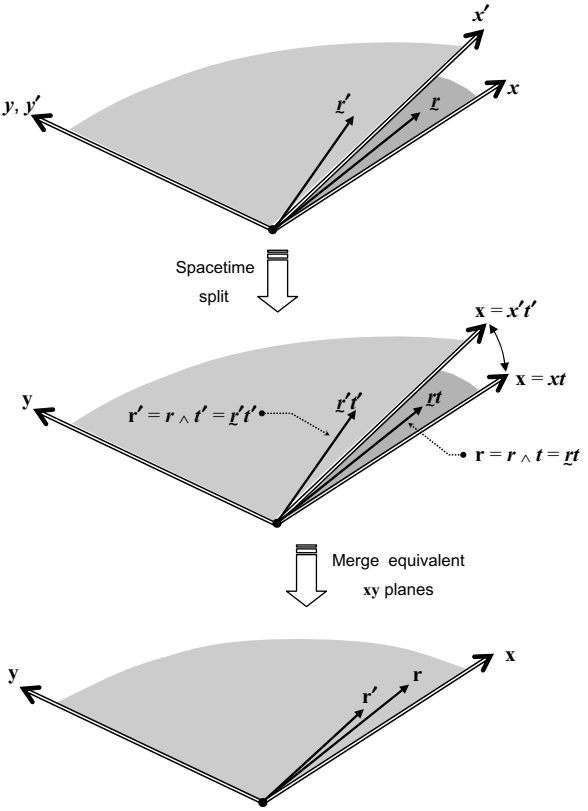

Figure 10.2 Merging the projections onto the ***xy*** and ***x′y′*** planes onto a single **xy** plane. Once again, the ***xy*** and ***x′y′*** planes here are meant to represent 3D subspaces of spacetime with the ***z*** dimension suppressed. Following on from Figure 10.1, we take the spacetime split of the vectors concerned. In (3+1)D, however, the basis vectors themselves do not change from frame to frame so that the ***xy*** and ***x′y′*** planes translate into equivalent **xy** planes, which we may then merge, then relabel $\underset{\sim}{\boldsymbol{rt}}$ as **r** and $\underset{\sim}{\boldsymbol{r}}'\boldsymbol{t}'$ as **r′**, each of which is then the (3+1)D vector counterpart of ***r*** in their respective frames. In algebraic terms, the process is simply $\mathbf{r} = \underset{\sim}{\boldsymbol{rt}}$ and $\mathbf{r}' = \underset{\sim}{\boldsymbol{r}}'\boldsymbol{t}'$.

As a further illustration, Figure 10.3 shows how the projection part of the spacetime split works with a particle history rather than just a single event. The concept is simple enough, as it only involves applying the procedure to every event on the particle's history. The parametric curves $\underset{\sim}{\boldsymbol{r}}(t)$ and $\underset{\sim}{\boldsymbol{r}}'(t')$ thereby projected onto each plane may then be converted to (3+1)D relative vector form by the process outlined above. The results correspond to the trajectories that would be observed for the particle from within each frame. The points on each curve may then be translated into (3+1)D vector form just as previously described so as to reveal the trajectories of the particle in ordinary space, but again, as they would be seen from each frame.

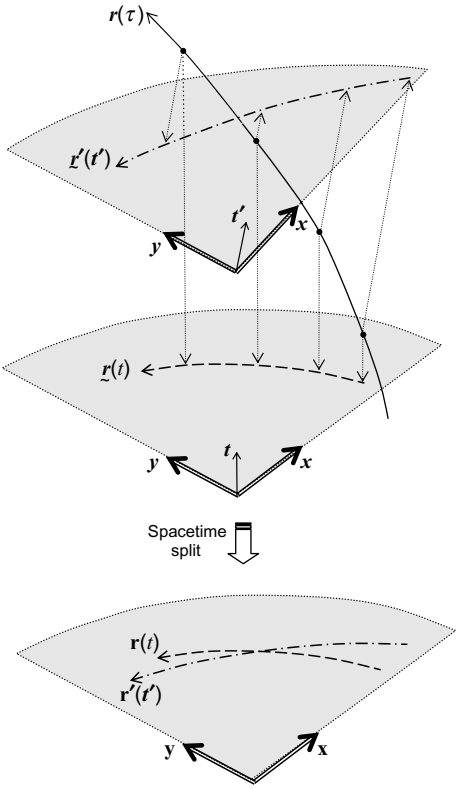

Figure 10.3 The history of a moving particle as projected onto two different (3+1)D frames. The particle has a history $\boldsymbol{r}(\tau)$ with its proper time τ as parameter. The figure shows it projected along the time axes of the $\boldsymbol{t}$ and $\boldsymbol{t}'$ frames onto two separate spacetime planes, $\boldsymbol{xy}$ and $\boldsymbol{x}'\boldsymbol{y}$ respectively, at regular intervals. These projections correspond to the trajectories observed in the two (3+1)D reference frames that are associated with $\boldsymbol{t}$ and $\boldsymbol{t}'$ respectively. The $\boldsymbol{xy}$ and $\boldsymbol{x}'\boldsymbol{y}$ planes have been placed apart for clarity, and the time axes have been compressed so as to fit the diagram within the page.

10.6.2 Relative Basis Vectors

An important point that was previously only touched on is the relationship between the (3+1)D basis vectors and the spacetime set. We now explore this relationship in more detail in order to justify why $\boldsymbol{t},\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}$ and $\boldsymbol{t}',\boldsymbol{x}',\boldsymbol{y}',\boldsymbol{z}'$ should both translate to the same set of (3+1)D basis vectors, $\mathbf{x},\mathbf{y},\mathbf{z}$.

All paravectors projected out from spacetime vectors are relative to their chosen frame, and a set of (3+1)D basis vectors for the frame may be derived from any three orthonormal spacetime vectors that are also orthogonal to the frame's time vector. For example, in Figure 10.2, $\boldsymbol{x}'$ is chosen to be orthogonal to $\boldsymbol{t}'$, and this then translates to the basis vector $\mathbf{x}'$ given by $\boldsymbol{x}'\boldsymbol{t}'$. As mentioned in Section 10.2, as long as we keep the discussion to Lorentz frames, that is to say spatial rotations and reflections

are excluded, $\mathbf{x}'$ is the same thing as $\mathbf{x}$, and in general, all such frames share the same (3+1)D basis vectors, for example, $\boldsymbol{xt}, \boldsymbol{yt}, \boldsymbol{zt} \equiv \boldsymbol{x't'}, \boldsymbol{y't'}, \boldsymbol{z't'} \equiv \boldsymbol{x''t''}, \boldsymbol{y''t''}, \boldsymbol{z''t''} \cdots$. All of these combinations are therefore conveniently labeled identically as $\mathbf{x}, \mathbf{y}, \mathbf{z}$.

But clearly, this apparent uniqueness of the relative basis vectors does not extend to their spacetime bivector equivalents since these are all differentiated by the time vector involved. Only in the case of a basis vector that lies parallel to the direction of motion do we have an equality, for example, $\boldsymbol{x't'} = \boldsymbol{xt}$, as in our example.

It will now be demonstrated that $\mathbf{x}, \mathbf{y}, \mathbf{z}$ do indeed form an orthonormal right-handed set irrespective of being expressed in terms of bivectors belonging to different frames. Let us start by noting that $\boldsymbol{t'}, \boldsymbol{x'}, \boldsymbol{y}, \boldsymbol{z}$, the spacetime basis vectors of the $\boldsymbol{t'}$-frame, are simply a Lorentz transformation of those of the $\boldsymbol{t}$-frame, and from Equation (9.13), these must then form an orthonormal set. Moreover, although $\boldsymbol{x'}$ may not be parallel to $\boldsymbol{x}$, it is nevertheless orthogonal to $\boldsymbol{y}$ and $\boldsymbol{z}$ (e.g., see $\boldsymbol{x'}$and $\boldsymbol{y}$ in Figure 10.1). There is no change in handedness as a result of a simple Lorentz transformation, and so this property is simply inherited. We now express $\mathbf{x}, \mathbf{y}, \mathbf{z}$ in terms of the $\boldsymbol{t'}$-frame basis vectors and check first of all the normalization. We find

$$\begin{aligned} |\mathbf{x}|^2 &= \boldsymbol{x't'x't'} \\ &= \boldsymbol{x'x'} = 1 \\ |\mathbf{y}|^2 &= \boldsymbol{yt'yt'} \\ &= \boldsymbol{yy} = 1 \\ |\mathbf{z}|^2 &= \boldsymbol{zt'zt'} \\ &= \boldsymbol{zz} = 1 \end{aligned} \tag{10.10}$$

so that $\mathbf{x}, \mathbf{y}, \mathbf{z}$ are indeed normalized. As to their orthogonality, we have by way of example for the typical case where the motion between the frames is along $\boldsymbol{x}$,

$$\begin{aligned} \mathbf{x} \cdot \mathbf{y} &= \tfrac{1}{2}(\mathbf{xy} + \mathbf{yx}) \\ &= \tfrac{1}{2}(\boldsymbol{x't'yt'} + \boldsymbol{yt'x't'}) \\ &= \tfrac{1}{2}(\boldsymbol{x't'yt'} + \boldsymbol{yt'x't'}) \\ &= \tfrac{1}{2}(\boldsymbol{x'y} + \boldsymbol{yx'}) \\ &= \boldsymbol{x'} \cdot \boldsymbol{y} \\ &= \gamma(\boldsymbol{t} + v\boldsymbol{x}) \cdot \boldsymbol{y} \\ &= 0 \end{aligned} \tag{10.11}$$

It is therefore clear that the time vector will simply cancel out of any inner product between two relative basis vectors so that, given $\boldsymbol{x'}, \boldsymbol{y}, \boldsymbol{z}$ are mutually orthogonal, the entire set of $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ must be mutually orthogonal.

It is extremely useful for practical purposes that our (3+1)D basis vectors $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ may be regarded as being the same irrespective of the choice of Lorentz frame; all the changes seen in going from one such frame to another are observed in the

components alone, as in Equation (9.29). In spacetime, we may change the basis vectors as a result of a Lorentz transformation, whereas in (3+1)D only the components can change.

The underlying assumption is that each inertial frame has an identical set of standards for length and time[2] specified *at rest in that frame*. Such standards are therefore *frame independent* provided they are always at rest in whichever frame they are used. Note that when we compare, say, a standard meter in some inertial frame (a) to an identical standard meter in a relatively moving inertial frame (b), the observer in frame (a) sees its own standard meter in the $\mathbf{x}$ direction as being $1\mathbf{x}$, whereas it is not hard to discover that he sees the identical thing in frame (b) as measuring $\gamma^{-1}\mathbf{x}$. But we could have done things differently here by saying that he sees it as $1\mathbf{x}'$, where the measurement stays fixed at 1 but the standard changes. This, however, is not the convention; each reference frame is deemed to use the same standard metersticks, which amounts to sharing the same orthonormal basis vectors $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$. Although each relatively moving observer would say that the other standard is shorter than their own, the difference here lies in the measurements (or coordinates), not the standards (or basis vectors). This observational difference is the stuff of the Lorentz–Fitzgerald contraction [20, part V, section 1, pp. 378–379; 48, chapter V, section B, pp. 36–43] of elementary special relativity.

10.6.3 Evaluating Relative Vectors

The principle of evaluating relative vectors through the spacetime split has already been discussed in Sections 8.3 and 8.4. There are, however, a number of potential pitfalls when it comes to the actual practice, for example, as discussed in Section 8.4.2. We will restrict the discussion here to the case of frames and particles at rest or in uniform motion.

Let us begin with a vector in component form. Consider the vector $\boldsymbol{r}$ that is represented by $t\boldsymbol{t}+x\boldsymbol{x}+y\boldsymbol{y}+z\boldsymbol{z}$ and $t'\boldsymbol{t}'+x'\boldsymbol{x}'+y'\boldsymbol{y}'+z'\boldsymbol{z}'$ in the $\boldsymbol{t}$-frame and $\boldsymbol{t}'$-frame respectively. To obtain the relative paravectors for $\boldsymbol{r}$ in each frame, it is necessary to premultiply each expression by the relevant time vector, namely $-\boldsymbol{t}$ in one case and $-\boldsymbol{t}'$ in the other, giving

$$\begin{aligned} -\boldsymbol{t}\boldsymbol{r} &= -\boldsymbol{t}\left(t\boldsymbol{t}+x\boldsymbol{x}+y\boldsymbol{y}+z\boldsymbol{z}\right) & -\boldsymbol{t}'\boldsymbol{r} &= -\boldsymbol{t}'\left(t'\boldsymbol{t}'+x'\boldsymbol{x}'+y'\boldsymbol{y}'+z'\boldsymbol{z}'\right) \\ &= t+x\boldsymbol{x}\boldsymbol{t}+y\boldsymbol{y}\boldsymbol{t}+z\boldsymbol{z}\boldsymbol{t} & &= t'+x'\boldsymbol{x}'\boldsymbol{t}'+y'\boldsymbol{y}'\boldsymbol{t}'+z'\boldsymbol{z}'\boldsymbol{t}' \\ &= t+x\mathbf{x}+y\mathbf{y}+z\mathbf{z} & &= t'+x'\mathbf{x}+y'\mathbf{y}+z'\mathbf{z} \end{aligned} \tag{10.12}$$

The crucial step here is that, as discussed in Section 10.6.2, the relative basis vectors must be the same for each frame. This demonstrates that in any given frame, the components of the spacetime vector and its relative paravector are identical. This

[2] The current internationally agreed standards are derived from (1) a specified atomic spectral line of Caesium 133 that corresponds to 9.19263177 GHz at absolute zero, and (2) the meter being defined such that the speed of light in vacuo corresponds to 2.99792458×10^8 m/s. A meter therefore corresponds to 30.663319 wavelengths of this microwave radiation.

means that one way to evaluate the relative paravectors would be to use Equation (9.29), which determines the components under a Lorentz transformation of the basis vectors, at least in the basic case where these frames are related by a velocity v along the $\boldsymbol{x}$ direction. But we may equally well obtain the required $\boldsymbol{t}'$-frame paravector simply by using the $\boldsymbol{t}$-frame representations of both $-\boldsymbol{t}'$ and $\boldsymbol{r}$ to evaluate $-\boldsymbol{t}'\boldsymbol{r}$. Now, Equation (7.17b) gives us $\boldsymbol{t}'$ and as usual $\boldsymbol{r} = t\boldsymbol{t} + x\boldsymbol{x} + y\boldsymbol{y} + z\boldsymbol{z}$, so that

$$\begin{aligned} t' + x'\mathbf{x} + y'\mathbf{y} + z'\mathbf{z} &= -\boldsymbol{t}'\boldsymbol{r} \\ &= -\boldsymbol{t}'(t\boldsymbol{t} + x\boldsymbol{x} + y\boldsymbol{y} + z\boldsymbol{z}) \\ &= -\gamma(\boldsymbol{t} + v\boldsymbol{x})(t\boldsymbol{t} + x\boldsymbol{x}) + y\boldsymbol{y}\boldsymbol{t}' + z\boldsymbol{z}\boldsymbol{t}' \\ &= \gamma(t + x\boldsymbol{x}\boldsymbol{t} - vt\boldsymbol{x}\boldsymbol{t} - vx) + y\mathbf{y} + z\mathbf{z} \\ &= \gamma(t - vx) + \gamma(x - vt)\mathbf{x} + y\mathbf{y} + z\mathbf{z} \end{aligned} \tag{10.13}$$

Even though it would seem natural to eliminate $\boldsymbol{t}'$ at the earliest stage by substituting $\gamma(\boldsymbol{t} + v\boldsymbol{x})$ throughout, it is crucial here to leave products such as $\boldsymbol{y}\boldsymbol{t}'$ and $\boldsymbol{z}\boldsymbol{t}'$ as they stand so that they can be replaced by $\mathbf{y}$ and $\mathbf{z}$, respectively. If we were indeed to substitute $\gamma(\boldsymbol{t} + v\boldsymbol{x})$ for $\boldsymbol{t}'$ in the bivector $\boldsymbol{y}\boldsymbol{t}'$, say, a spacelike bivector $\gamma v\boldsymbol{x}\boldsymbol{y}$ would result. Since there is no possibility of replacing $\boldsymbol{x}\boldsymbol{y}$ with a (3+1)D vector, this is a typical pitfall to be avoided. The end result, however, is exactly the same as we would have obtained by applying Equation (9.29) to find the components of $\boldsymbol{r}$ in the $\boldsymbol{t}$-frame and then equating these to the components of the relative paravector $t' + x'\mathbf{x} + y\mathbf{y} + z\mathbf{z}$.

While dealing with spacetime vectors in component form is relatively straightforward, as discussed in Secton 7.10, it is frequently more convenient to work without a full set of basis vectors and instead to have the vectors represented in a parametric form. For example, the history of some particle in uniform motion may be generally expressed as $\boldsymbol{r}(\lambda) = \boldsymbol{u}\lambda + \boldsymbol{w}$ where $\boldsymbol{w} = \boldsymbol{r}(0)$. The form taken in the $\boldsymbol{t}$-frame would then be $\boldsymbol{r}(t) = \boldsymbol{u}t + \underset{\sim}{\boldsymbol{w}}$ where $\boldsymbol{u}$ is the particle's spacetime velocity and $\underset{\sim}{\boldsymbol{w}}$ is a constant spatial vector. The velocity itself may then be split into temporal and spatial parts with respect to $\boldsymbol{t}$ as $\boldsymbol{u} = \boldsymbol{t} + \underset{\sim}{\boldsymbol{u}}$, with $\underset{\sim}{\boldsymbol{u}}$ giving us the velocity u as $|\underset{\sim}{\boldsymbol{u}}|$ and its direction as $\hat{\underset{\sim}{\boldsymbol{u}}}$, for example, $\underset{\sim}{\boldsymbol{u}} = u\boldsymbol{z}$. In the $\boldsymbol{t}$-frame, the spacetime vector $\boldsymbol{r}$ therefore gives rise to the relative paravector

$$\begin{aligned} t + \mathbf{r} &= -\boldsymbol{t}\boldsymbol{r} \\ &= -\boldsymbol{t}\left((\boldsymbol{t} + \underset{\sim}{\boldsymbol{u}})t + \underset{\sim}{\boldsymbol{w}}\boldsymbol{r}\right) \\ &= t(1 + \mathbf{u}) + \mathbf{w} \end{aligned} \tag{10.14}$$

This result involves the relative vectors $\mathbf{u} = \underset{\sim}{\boldsymbol{u}}\boldsymbol{t} = \boldsymbol{u} \wedge \boldsymbol{t}$ and $\mathbf{w} = \underset{\sim}{\boldsymbol{w}}\boldsymbol{t}$. The factor $1 + \mathbf{u}$ is the spacetime split of the velocity vector $\boldsymbol{u}$ in the $\boldsymbol{t}$-frame (Equation 8.18), and $\mathbf{u} = u\mathbf{z}$ for the simple example given above where $\underset{\sim}{\boldsymbol{u}} = u\boldsymbol{z}$. Now, since $\underset{\sim}{\boldsymbol{w}}$ and $\underset{\sim}{\boldsymbol{u}}$ are purely spatial vectors with respect to the $\boldsymbol{t}$-frame, it is clear that they correspond directly to the relative vectors $\mathbf{w}$ and $\mathbf{u}$ of the same frame. Since Equation (7.2) allows us to split any spacetime vector into its temporal and spatial parts in a given frame, finding its relative paravector for that frame vector follows directly from

evaluating its spatial part *in the same frame*. For example, given $\boldsymbol{s} = s_t\boldsymbol{t}' + \underset{\sim}{\boldsymbol{s}}'$ then $\mathbf{s}' = \underset{\sim}{\boldsymbol{s}}'\boldsymbol{t}'$. The vectors $\mathbf{s}'$ and $\underset{\sim}{\boldsymbol{s}}'$ are therefore directly related and frame dependent.

Let us now turn to the corresponding relative vectors for $\boldsymbol{r}$ in the $\boldsymbol{t}'$-frame for which $\boldsymbol{t}' = \gamma(\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}})$ independently of whether or not $\underset{\sim}{\boldsymbol{v}}$ happens to be along $\boldsymbol{x}$. First of all, let us consider the case where the particle is actually at rest in that frame, in which case, $\underset{\sim}{\boldsymbol{u}}$ and $\underset{\sim}{\boldsymbol{v}}$ are the same. The relative paravector for $\boldsymbol{r}$ in the $\boldsymbol{t}'$-frame is then found by evaluating $-\boldsymbol{t}'\boldsymbol{r}$. However, in assembling the result, we need to take care to avoid forming spacelike bivectors for, as we have just been discussing, only scalars and timelike bivectors relate to (3+1)D paravectors. This may be conveniently accomplished by splitting $\underset{\sim}{\boldsymbol{w}}$ into $\underset{\sim}{\boldsymbol{w}}_{/\!/} + \underset{\sim}{\boldsymbol{w}}_{\perp}$ where and $\underset{\sim}{\boldsymbol{w}}_{/\!/}$ and $\underset{\sim}{\boldsymbol{w}}_{\perp}$ are parallel and perpendicular to $\underset{\sim}{\boldsymbol{v}}$ respectively. Having done this, recall that it does not matter whether we form those relative vectors that are *perpendicular to the motion* using either $\boldsymbol{t}'$ or $\boldsymbol{t}$. Since in this particular case $\underset{\sim}{\boldsymbol{u}} = \underset{\sim}{\boldsymbol{v}}$, evaluating $-\boldsymbol{t}'\boldsymbol{r}$ in this way readily yields

$$\begin{aligned} t' + \mathbf{w}' &= -\boldsymbol{t}'\big((\boldsymbol{t} + \underset{\sim}{\boldsymbol{u}})t + \underset{\sim}{\boldsymbol{w}}\big) \\ &= -\gamma\big(\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}}\big)\big((\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}})t + \underset{\sim}{\boldsymbol{w}}_{/\!/}\big) + \underset{\sim}{\boldsymbol{w}}_{\perp}\boldsymbol{t}' \\ &= \gamma\Big(\big(1 - v^2\big)t + \underset{\sim}{\boldsymbol{w}}_{/\!/}\boldsymbol{t} - \underset{\sim}{\boldsymbol{w}}_{/\!/}\boldsymbol{t}\underset{\sim}{\boldsymbol{v}}\boldsymbol{t}\Big) + \mathbf{w}_{\perp} \\ &= \underbrace{\gamma^{-1}t - \gamma\mathbf{w}\cdot\mathbf{v}}_{t'} + \underbrace{\gamma\mathbf{w}_{/\!/} + \mathbf{w}_{\perp}}_{\mathbf{w}} \end{aligned} \tag{10.15}$$

Note that, as is easily shown, parallel and perpendicular are preserved when taking the relative vectors of spatial vectors. Furthermore, $\underset{\sim}{\boldsymbol{w}}_{/\!/}\underset{\sim}{\boldsymbol{v}}$ has been manipulated into $\underset{\sim}{\boldsymbol{w}}_{/\!/}\boldsymbol{t}\underset{\sim}{\boldsymbol{v}}\boldsymbol{t}$ so as to obtain $\mathbf{w}_{/\!/}\mathbf{v}$, which can then be written as $\mathbf{w}\cdot\mathbf{v}$.

Turn now to a different situation in which we have some arbitrary vector specified in the $\boldsymbol{t}$-frame, say $\boldsymbol{w} = w_t\boldsymbol{t} + \underset{\sim}{\boldsymbol{w}}$. Unlike a history such as $\boldsymbol{r}(\lambda)$, this is simply the spacetime representation of a paravector $w_t + \mathbf{w}$. Although w_t may well be time dependent, it no longer represents time. An example is the electromagnetic source density in which $w_t = \rho$ (this will be discussed in further detail in Section 11.2.1). Let us therefore split $\boldsymbol{w}$ into $w_t\boldsymbol{t} + \underset{\sim}{\boldsymbol{w}}_{/\!/} + \underset{\sim}{\boldsymbol{w}}_{\perp}$ where $w_t = -\boldsymbol{t}\cdot\boldsymbol{r}$, and $\underset{\sim}{\boldsymbol{w}}_{/\!/}$ and $\underset{\sim}{\boldsymbol{w}}_{\perp}$ are parallel and perpendicular to $\underset{\sim}{\boldsymbol{v}}$ respectively, that is to say, one lies along the direction of motion of the $\boldsymbol{t}'$-frame and while the other lies counter to it. By definition, however, both $\underset{\sim}{\boldsymbol{w}}_{/\!/}$ and $\underset{\sim}{\boldsymbol{w}}_{\perp}$ are orthogonal to $\boldsymbol{t}$. The single spatial vector $\underset{\sim}{\boldsymbol{w}}$ is therefore replaced by $\underset{\sim}{\boldsymbol{w}}_{/\!/} + \underset{\sim}{\boldsymbol{w}}_{\perp}$. On taking the $\boldsymbol{t}'$-frame spacetime split, we find

$$\begin{aligned} w_{t'} + \mathbf{w}' &= -\boldsymbol{t}'\boldsymbol{r} \\ &= -\gamma\big(\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}}\big)\big(w_t\boldsymbol{t} + \underset{\sim}{\boldsymbol{w}}_{/\!/}\big) - \boldsymbol{t}'\underset{\sim}{\boldsymbol{w}}_{\perp} \\ &= \gamma\big(w_t + \underset{\sim}{\boldsymbol{w}}_{/\!/}\boldsymbol{t} - w_t\underset{\sim}{\boldsymbol{v}}\boldsymbol{t} - \underset{\sim}{\boldsymbol{w}}_{/\!/}\boldsymbol{t}\underset{\sim}{\boldsymbol{v}}\boldsymbol{t}\big) + \mathbf{w}_{\perp} \\ &= \gamma\big(w_t + \mathbf{w}_{/\!/} - w_t\mathbf{v} - \mathbf{w}_{/\!/}\mathbf{v}\big) + \mathbf{w}_{\perp} \\ &= \underbrace{\gamma\big(w_t - \mathbf{w}\cdot\mathbf{v}\big)}_{w_{t'}} + \underbrace{\gamma\big(\mathbf{w}_{/\!/} - w_t\mathbf{v}\big) + \mathbf{w}_{\perp}}_{\mathbf{w}'} \end{aligned} \tag{10.16}$$

This result gives us $w_{t'} + \mathbf{w}'$, the relative paravector for $\boldsymbol{w}$ in $\boldsymbol{t}'$-frame, expressed in terms of the $\boldsymbol{t}'$-frame relative paravector $w_t + \mathbf{w}$ and the $\boldsymbol{t}'$-frame velocity $\mathbf{v}$. This

is comparable with Equation (8.13) if we put it in component form and let $\underset{\sim}{\boldsymbol{v}} = v\boldsymbol{x}$. Equation (10.16) looks different from Equation (10.15) due to the fact that we cannot associate $w_{t'}$ with a time parameter, specifically, the time parameter of the $\boldsymbol{t}'$-frame. When $\mathbf{w}_\perp = 0$ and $\mathbf{w}_{/\!/} = w_t\mathbf{v}$, however, the relative vector $\mathbf{w}'$ vanishes and the result reduces to a scalar $w_{t'} = \gamma(w_t - w_t\mathbf{v}\cdot\mathbf{v}) = \gamma^{-1}w_t$. If $\boldsymbol{w}$ did indeed represent an electromagnetic source density, then in this situation the charge density would be at rest in the $\boldsymbol{t}'$-frame in which there is correspondingly zero current density.

As a final case for examination, we now evaluate the relative paravector for $\boldsymbol{r}$ in the $\boldsymbol{t}'$-frame when the particle velocity $\boldsymbol{u}$, though still fixed, is of arbitrary magnitude and direction. While the evaluation is a little more complicated than the initial case in which we had $\boldsymbol{r}$ at rest in the $\boldsymbol{t}'$-frame (with the resulting simplification $\underset{\sim}{\boldsymbol{u}} = \underset{\sim}{\boldsymbol{v}}$), it nevertheless proceeds in the same general way. The main difference is that it is now also necessary to split $\boldsymbol{u}$ into parts, $\underset{\sim}{\boldsymbol{u}}_{/\!/}$ and $\underset{\sim}{\boldsymbol{u}}_\perp$, that are parallel and perpendicular to $\underset{\sim}{\boldsymbol{v}}$ respectively, so as to avoid once again the pitfall concerning unwanted bivectors appearing in the result. We now find

$$\begin{aligned}
t' + \mathbf{w}' &= -\boldsymbol{t}'\big((\boldsymbol{t} + \underset{\sim}{\boldsymbol{u}})t + \underset{\sim}{\boldsymbol{w}}\big) \\
&= -\gamma\big(\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}}\big)\big((\boldsymbol{t} + \underset{\sim}{\boldsymbol{u}}_{/\!/})t + \underset{\sim}{\boldsymbol{w}}_{/\!/}\big) + \big(\underset{\sim}{\boldsymbol{u}}_\perp t + \underset{\sim}{\boldsymbol{w}}_\perp \boldsymbol{t}'\big) \\
&= \gamma\big((1 - \underset{\sim}{\boldsymbol{u}}_{/\!/}\boldsymbol{t}\underset{\sim}{\boldsymbol{v}}\boldsymbol{t})t - \underset{\sim}{\boldsymbol{w}}_{/\!/}\boldsymbol{t}\underset{\sim}{\boldsymbol{v}}\boldsymbol{t} + \big(\underset{\sim}{\boldsymbol{w}}_{/\!/}\boldsymbol{t} + \underset{\sim}{\boldsymbol{u}}_{/\!/}\boldsymbol{t}\boldsymbol{t} - \underset{\sim}{\boldsymbol{v}}\boldsymbol{t}t\big)\big) + \big(\mathbf{u}_\perp t + \mathbf{w}_\perp\big) \\
&= \gamma\big(1 - \mathbf{u}_{/\!/}\mathbf{v}\big)t - \mathbf{w}_{/\!/}\mathbf{v} + \gamma\big(\mathbf{w}_{/\!/} + (\mathbf{u}_{/\!/} - \mathbf{v})t\big) + \big(\mathbf{w}_\perp + \mathbf{u}_\perp t\big) \\
&= \underbrace{\gamma\big(t - (\mathbf{w}_{/\!/} + \mathbf{u}_{/\!/}t)\mathbf{v}\big)}_{t'} + \underbrace{\gamma\big(\mathbf{w}_{/\!/} + (\mathbf{u}_{/\!/} - \mathbf{v})t\big) + \big(\mathbf{w}_\perp + \mathbf{u}_\perp t\big)}_{\mathbf{w}'}
\end{aligned} \tag{10.17}$$

This gives us the relative paravector $t' + \mathbf{w}'$ in terms of the time and relative vectors for $\boldsymbol{r}$, $\boldsymbol{u}$, and $\boldsymbol{v}$ as expressed in the $\boldsymbol{t}$-frame. Setting $\underset{\sim}{\boldsymbol{u}} = \underset{\sim}{\boldsymbol{v}}$, that is to say, with the particle at rest in the $\boldsymbol{t}'$-frame, $\mathbf{u}_{/\!/} = \mathbf{v}$ and $\mathbf{u}_\perp = 0$ so that, given $1 - \mathbf{v}^2 = 1 - v^2 = \gamma^{-2}$, Equation (10.15) is recovered. With $\mathbf{u} = 0$, the particle is at rest in the $\boldsymbol{t}$-frame, that is to say, at any time t, its position is $\mathbf{w}$, so that its relative time t' and position $\mathbf{r}'$ observed from the $\boldsymbol{t}'$-frame are $t' = \gamma(t - \mathbf{w}_{/\!/}\mathbf{v})$ and $\mathbf{w}' = \gamma(\mathbf{w}_{/\!/} - \mathbf{v}t) + \mathbf{w}_\perp$. Equation (10.17) therefore gives Equation (10.13) in a completely basis-free form at the expense of only a little extra complexity.

10.6.4 Relative Vectors Involving Parameters

In Section 8.4.2, we encountered a different sort of problem with the spacetime split of velocity. Velocity is an example of a derived vector, that is to say, a vector that is derived from some other object. In this case, velocity is the time derivative of a history vector. But, to be clear on this, there is no rule that states velocity must be the derived vector. For example, given the proper velocity of a particle, we can find its history by means of Equation (10.7). Here, then, is an example that is the other way around, where velocity is the given vector while history is the derived vector.

Spacetime vectors often involve one or more parameters, for example, a history $\boldsymbol{r}(t)$ involves the time parameter t. Using a spacetime split to project a relative vector onto a given frame does not necessarily produce a result in which the

parameters are relevant to that frame for, without modification, they will still be those of the original frame. Therefore, when a given parameter is frame dependent, it is necessary to choose which representation of it we want to see in the final result, for example, t or t'. In the case of $\boldsymbol{r}(t)$, t is conventionally the time vector of the $\boldsymbol{t}$-frame. Taking the spacetime split of $\boldsymbol{r}(t)$ in the $\boldsymbol{t}'$-frame by means of evaluating $-\boldsymbol{t}'\boldsymbol{r}(t)$ gives us $t'(t)+\mathbf{r}'(t)$, but it is clear that since premultiplying by $-\boldsymbol{t}'$ has no effect whatsoever on the independent parameter t, the end result still continues to be a function of the parameter t. To form an illustration, let us simplify the result of Equation (10.17) by setting $\mathbf{u}$ (and consequently its parallel and orthogonal parts) to 0. This then gives us $t'(t)+\mathbf{w}'(t)$ when $\boldsymbol{r}(t)$ represents the history of some particle at rest at $\mathbf{w}$ in the $\boldsymbol{t}$-frame. But it is clear that this equation would be of more use to an observer in the $\boldsymbol{t}'$-frame if $\mathbf{w}'$ were given as a function of t', the local time parameter, rather than t. But then the scalar part of Equation (10.17) gives us t' as a function of t in the form of $t'=\gamma(t-\mathbf{w}_{/\!/}\mathbf{v})=\gamma(t-\mathbf{w}\cdot\mathbf{v})$, which may readily be inverted to find

$$t=\gamma^{-1}(t'+\mathbf{v}\cdot\mathbf{w}) \tag{10.18}$$

Substituting this for t in the vector part of Equation (10.17) then gives $\mathbf{w}'(t')$, the trajectory of $\mathbf{r}$ seen from the $\boldsymbol{t}'$-frame observer's perspective, as

$$\begin{aligned}\mathbf{w}'(t') &= \gamma(\mathbf{w}_{/\!/}-\mathbf{v}t)+\mathbf{w}_{\perp}\\ &= \gamma\left(\mathbf{w}_{/\!/}-\mathbf{v}\gamma^{-1}(t'+\mathbf{v}\cdot\mathbf{w})\right)+\mathbf{w}_{\perp}\\ &= \gamma\mathbf{w}_{/\!/}+\mathbf{w}_{\perp}-(t'+\delta t)\mathbf{v}\end{aligned} \tag{10.19}$$

where $\delta t=\mathbf{v}\cdot\mathbf{w}$. This is now an equation that $\boldsymbol{t}'$-frame observers can use directly, and, since t' is their own local time, $\mathbf{w}$ and $\mathbf{v}$ are the only two pieces of information that the $\boldsymbol{t}'$-frame observers require to take from the $\boldsymbol{t}$-frame. There is therefore little point in using the paravector form $t'+\mathbf{w}'(t')$, and the relative vector can therefore stand on its own. The parameter t is not lost, however, because we may still recover it from Equation (10.18). Note that the appearance of the time offset δt here shows up the well-known and often intellectually challenging issue of clock synchronization between (3+1)D reference frames, but thankfully, spacetime averts such problems and any such synchronization offset simply emerges, as here, from the spacetime splits in each frame.

As discussed in Section 8.4.2, when the vector in question is a derived vector, the issue of parameters is further complicated by the need to change the variable of differentiation (or integration). We are now in a position to give a simpler method of finding the velocity vector for a moving particle in any frame. The key point now is that we will be working not from the velocity as seen in one arbitrary frame to the velocity as seen in another, but from the particle's proper velocity, as introduced in Section 10.5. For a particle history $\boldsymbol{r}(\tau)$ that is parameterized by τ, the particle's proper time, the particle's proper velocity is given by Equation (10.4) as being $\boldsymbol{\upsilon}(\tau)=\partial_{\tau}\boldsymbol{r}(\tau)$. On the basis that the form of the relative paravector for $\boldsymbol{r}$ in the $\boldsymbol{t}$-frame will be $-\boldsymbol{t}\boldsymbol{r}=t+\mathbf{r}$, we then have

$$\begin{aligned}\partial_\tau(t+\mathbf{r}) &= \partial_\tau(-\boldsymbol{t}\boldsymbol{r}) \\ &= -\boldsymbol{t}\,\mathfrak{v} \\ &= -\mathfrak{v}\cdot\boldsymbol{t} + \mathfrak{v}\wedge\boldsymbol{t} \\ \Leftrightarrow &\begin{cases}\partial_\tau t = -\mathfrak{v}\cdot\boldsymbol{t} \\ \partial_\tau \mathbf{r} = \mathfrak{v}\wedge\boldsymbol{t}\end{cases}\end{aligned} \tag{10.20}$$

The means of changing the variable of differentiation, $\partial_\tau t$, has neatly fallen out of this so that we may conclude

$$\begin{aligned}\mathbf{v} &= \partial_t \mathbf{r} \\ &= \partial_t \tau \partial_\tau \mathbf{r} \\ &= (\partial_\tau t)^{-1}\,\mathfrak{v}\wedge\boldsymbol{t} \\ &= \frac{\mathfrak{v}\wedge\boldsymbol{t}}{-\mathfrak{v}\cdot\boldsymbol{t}}\end{aligned} \tag{10.21}$$

Since there can be nothing special in the choice of $\boldsymbol{t}$-frame, and $\mathfrak{v}$ is independent of this choice, we may use any given frame here, say the $\boldsymbol{\theta}$-frame, so that we may generalize this result to

$$\mathbf{u}(\mathfrak{v},\boldsymbol{\theta}) = \frac{\mathfrak{v}\wedge\boldsymbol{\theta}}{-\mathfrak{v}\cdot\boldsymbol{\theta}} \tag{10.22}$$

where $\boldsymbol{\theta}$ is the proper velocity, that is, the time vector, of the $\boldsymbol{\theta}$-frame. The relative velocity $\mathbf{u}(\mathfrak{v},\boldsymbol{\theta})$ is therefore the relative velocity of the $\mathfrak{v}$-frame as seen from the $\boldsymbol{\theta}$-frame.

The relative vector for the velocity of a particle in another frame is worked out in detail in Section 10.9, using a slightly different route. The result (Equation 10.40) may also be found from Equation (10.22). Finally, evaluating the relative vector for any other sort of derived vector, such as acceleration, needs a similarly careful approach.

10.6.5 Transforming Relative Vectors and Paravectors to a Different Frame

Just as in the case of spacetime vectors, it is necessary to have the objective of any transformation clear from the outset. When we say that $\mathbf{u}$ is a relative vector in the $\boldsymbol{t}$-frame, what we mean is that $\mathbf{u}$ is the spacetime split of some vector $\boldsymbol{u}$ in this frame. The spacetime split, given as usual by $-\boldsymbol{t}\boldsymbol{u}$, actually yields more than just the vector $\mathbf{u}$; it in fact gives us a paravector $u_t + \mathbf{u}$ where $u_t = -\boldsymbol{t}\cdot\boldsymbol{u}$ and $-\boldsymbol{t}\wedge\boldsymbol{u} = \mathbf{u}$. As discussed in Section 8.3, by starting from the entire paravector rather than just its vector part, the spacetime split has an inverse, and we may find the original spacetime vector $\boldsymbol{u}$ simply by premultiplying $u_t + \mathbf{u}$ by the time vector that was originally used to make this spacetime split, in this case $\boldsymbol{t}$. For example, $\boldsymbol{t}(u_t + \mathbf{u}) = \boldsymbol{t}(-\boldsymbol{t}\boldsymbol{u}) = -\boldsymbol{t}^2\boldsymbol{u} = \boldsymbol{u}$. It is

then only necessary to form $-\boldsymbol{t}'\boldsymbol{u}$ in order to find the corresponding paravector $u'_t + \mathbf{u}'$ in the $\boldsymbol{t}'$-frame. The paravectors $u_t + \mathbf{u}$ and $u'_t + \mathbf{u}'$ are simply alternative projections of $\boldsymbol{u}$ into different (3+1)D frames. We have seen that different Lorentz frames share the same relative basis vectors, generally referred to as $\mathbf{x},\mathbf{y},\mathbf{z}$, so that when the relative vector $\mathbf{u}$ is given in terms of basis vectors, transforming it to a different frame affects only the components.

This leads to a version of the Lorentz transformation that can be applied directly to relative vectors, or paravectors, in any given frame. The transformation from $u_t + \mathbf{u}$ to $u'_t + \mathbf{u}'$ is readily revealed by

$$u'_t + \mathbf{u}' = -\boldsymbol{t}'\boldsymbol{u} = -\boldsymbol{t}'\boldsymbol{t}\left(u_t + \mathbf{u}\right) \tag{10.23}$$

that is to say, we need only multiply the paravector to be transformed by $-\boldsymbol{t}'\boldsymbol{t}$. Taking $\boldsymbol{t}' = \gamma\left(\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}}\right)$ as before, we find, as in Equation (8.16b), that $-\boldsymbol{t}'\boldsymbol{t} = \gamma\left(1 - \mathbf{v}\right)$ where $\mathbf{v} = \underset{\sim}{\boldsymbol{v}}\boldsymbol{t}$. The "relative" Lorentz transformation would then appear to be

$$u_t + \mathbf{u} \mapsto u'_t + \mathbf{u}' = \gamma\left(1 - \mathbf{v}\right)\left(u_t + \mathbf{u}\right) \tag{10.24}$$

This is directly equivalent to the Lorentz transformation when $\mathbf{u}$ and $\mathbf{v}$ are parallel, but when $\mathbf{u}$ is perpendicular to $\mathbf{v}$, a term $\gamma\mathbf{u} \wedge \mathbf{v}$ arises. For Equation (10.24) to be of any use, we therefore need to follow the now familiar theme of splitting $\mathbf{u}$ into $\mathbf{u} = \mathbf{u}_{/\!/} + \mathbf{u}_\perp$, where the subscripts identify the parts of a vector that are parallel and perpendicular to $\mathbf{v}$ respectively. Taking first $\mathbf{u}_\perp$, we have

$$\begin{aligned} \mathbf{u}'_\perp &= \gamma\left(1 - \mathbf{v}\right)\mathbf{u}_\perp \\ &= \gamma\mathbf{u}_\perp - \gamma\mathbf{v}\mathbf{u}_\perp \\ &= \gamma\boldsymbol{u}_\perp\boldsymbol{t} - \gamma\underset{\sim}{\boldsymbol{v}}\boldsymbol{t}\boldsymbol{u}_\perp\boldsymbol{t} \\ &= \gamma\boldsymbol{u}_\perp\boldsymbol{t} + \gamma\boldsymbol{u}_\perp\underset{\sim}{\boldsymbol{v}} \end{aligned} \tag{10.25}$$

Recalling $\boldsymbol{t}' = \gamma\left(\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}}\right)$, the troublesome spacelike bivector appearing in the form of $\boldsymbol{u}_\perp\underset{\sim}{\boldsymbol{v}}$ may be dealt with by substituting $\boldsymbol{t}' - \gamma\boldsymbol{t}$ for $\gamma\underset{\sim}{\boldsymbol{v}}$ so as to obtain

$$\begin{aligned} \mathbf{u}'_\perp &= \gamma\boldsymbol{u}_\perp\boldsymbol{t} + \boldsymbol{u}_\perp\left(\boldsymbol{t}' - \gamma\boldsymbol{t}\right) \\ &= \boldsymbol{u}_\perp\boldsymbol{t}' \end{aligned} \tag{10.26}$$

The problem has therefore been resolved by expressing $\mathbf{u}'_\perp$ in terms of a bivector composed using *the transformed time vector* rather than *the original one*. From this result, it can be seen that the relative vector for $\boldsymbol{u}_\perp$ in the $\boldsymbol{t}'$-frame is exactly the same as in the $\boldsymbol{t}$-frame except that the new time vector replaces the old. Just as we did in the case of the (3+1)D basis vectors, we may therefore say $\mathbf{u}'_\perp = \mathbf{u}_\perp$.

The problem of getting round the spacelike bivector does not arise with $u_t + \mathbf{u}_{/\!/}$ for which we simply find

$$\begin{aligned} u'_t + \mathbf{u}'_{/\!/} &= \gamma\left(1 - \mathbf{v}\right)\left(u_t + \mathbf{u}_{/\!/}\right) \\ &= \gamma\left(u_t - \mathbf{v}\cdot\mathbf{u}\right) + \gamma\left(\mathbf{u}_{/\!/} - u_t\mathbf{v}\right) \end{aligned} \tag{10.27}$$

Putting these last two results together, we finally obtain

$$\begin{aligned} u_t' + \mathbf{u}' &= \gamma(1-\mathbf{v})(u_t + \mathbf{u}_{//}) + \mathbf{u}_\perp \\ &= \underbrace{\gamma(u_t - \mathbf{v}\cdot\mathbf{u})}_{u_t'} + \underbrace{\gamma(\mathbf{u}_{//} - u_t\mathbf{v}) + \mathbf{u}_\perp}_{\mathbf{u}'} \end{aligned} \tag{10.28}$$

which is exactly the same form of result as Equation (10.16), except that we started from the spacetime vector $\boldsymbol{w}$ rather than the paravector $u_t + \mathbf{u}$.

This example clearly demonstrates the following:

- The process of transforming relative paravectors from one frame to another amounts to a Lorentz transformation.
- The Lorentz transformation parameter is the relative vector $\mathbf{v}$ that arises in the product of the time vectors of the two frames involved.
- This may be put in the form of Equation (10.22) to give $\mathbf{v}(\boldsymbol{t}', \boldsymbol{t}) = \dfrac{\boldsymbol{t}' \wedge \boldsymbol{t}}{-\boldsymbol{t}' \cdot \boldsymbol{t}}$ for the case where the transformation is from the $\boldsymbol{t}$-frame to the $\boldsymbol{t}'$-frame.
- As expected, the parts of relative vectors that are perpendicular to $\mathbf{v}$ are unaffected by the transformation.
- Scalars and the parallel parts of relative vectors may be transformed using Equation (10.27).
- A relative vector on its own will be transformed as though it were a paravector with scalar part set to zero.
- In the case of a position vector $\mathbf{r}$, the result produced by Equation (10.28) would therefore only be valid at $t = 0$.
- It is therefore safer to apply this transformation to a complete paravector, for example, $t + \mathbf{r}$.

10.7 FRAME-DEPENDENT VERSUS FRAME-INDEPENDENT SCALARS

According to Equation (8.5), the scalars of (3+1)D may be translated to timelike vectors in spacetime, for example, $a \leftrightarrow a\boldsymbol{t}$, but since there is always the choice of going into an odd or even spacetime element, it is also possible to translate them directly as spacetime scalars. It was said at an early stage that the underlying physics should dictate which choice should be made. As an example, we have scalar quantities m for a particle's mass and q for its charge. Now special relativity holds that the observed mass is frame dependent while charge is invariant. This dictates that the spacetime representation of mass–energy must be a timelike vector, whereas charge, being invariant, must simply be a spacetime scalar.

The fact that mass–energy is represented by a spacetime vector is not the whole story. If we start with a particle at rest in the $\boldsymbol{t}$-frame where its mass is

represented by the vector $m\boldsymbol{t}$, then following the procedure of Equation (9.28) (effectively just by replacing t with m and setting x, y, and z all to 0), we will observe $\gamma m\boldsymbol{t}' - \gamma mv\boldsymbol{x}'$ in the $\boldsymbol{t}'$-frame where the particle is seen to have velocity $-v\boldsymbol{x}$. The time part must continue to be the particle's observed mass, while the spatial part making its appearance can be clearly recognized as the momentum! The usual Lorentz factor γ attaches to both, in agreement with the usual textbook derivations. What we are calling the particle's mass here actually equates to its rest mass plus energy, and so the mass does not stand alone, it is part of an overall energy–momentum vector, which, in the particle's rest frame, of course, represents just the mass.

For convenience, let us turn this around to the more usual way where the particle is at rest in the $\boldsymbol{t}'$-frame, while we, the observer, are in the $\boldsymbol{t}$-frame. This is only a matter of exchanging $\boldsymbol{t}$ for $\boldsymbol{t}'$ and v for $-v$ so as to keep the labeling consistent with our previous discussions, giving us $\gamma m\boldsymbol{t} + \gamma mv\boldsymbol{x}$ as the observed energy–momentum. We previously identified the particle's proper velocity $\boldsymbol{\upsilon}$ with its local time vector, and so we may write $\boldsymbol{\upsilon} = \boldsymbol{t}' = \gamma(\boldsymbol{t} + v\boldsymbol{x})$. This being the case

$$\begin{aligned} \gamma m\boldsymbol{t} + \gamma mv\boldsymbol{x} &= \gamma m(\boldsymbol{t} + v\boldsymbol{x}) \\ &= m\boldsymbol{t}' \\ &= m\boldsymbol{\upsilon} \end{aligned} \tag{10.29}$$

This, therefore, is the *proper* energy–momentum vector for our particle, which we could now think of as being in its native form. The spacetime split of this vector in the particle's own frame will show only its rest mass m, whereas for the observer in the $\boldsymbol{t}$-frame, the split gives $\gamma m + \gamma m\mathbf{v}$. Note that there are two schools of thought as to whether mass should be considered as being velocity dependent by taking on board the factor γ, or as being constant by keeping the factor γ separate. This is merely a matter of semantics as to what we call mass. Whether we use $m\gamma$ or m is a matter of choice; both imply the total of rest mass plus kinetic energy that we have been calling mass–energy. To get round any potential ambiguity, the symbol m_0 is often used to specifically indicate rest mass.

Frame-dependent scalars that translate to a spacetime vector rather than a scalar, such as time, particle mass and charge density, are inevitably associated with some form of paravector, for example, $t + \mathbf{r}$, $\gamma m(1 + \mathbf{v})$ and $\gamma\rho(1 - \mathbf{v})$ respectively. While the scalars themselves relate to the temporal parts of a spacetime vector, in some different frame, they will also have spatial parts, and as a consequence, these will project into relative paravectors rather than simple scalars in that frame. Once again, it is usually better to consider the relative paravector as a whole when it comes to the question of transformation to another frame. That being the case, Equation (10.28) provides the appropriate transformation. For example, if we transform the energy–momentum paravector $\gamma m(1 + \mathbf{v})$ to the rest frame of the particle, we find $m' + \mathbf{p}' = \gamma(1 - \mathbf{v})\gamma m(1 + \mathbf{v}) = m$, which is exactly what we should expect, but had we started from the scalar part γm alone, the result would have been $\gamma^2 m(1 - \mathbf{v})$, which is clearly erroneous.

10.8 CHANGE OF BASIS FOR ANY OBJECT IN COMPONENT FORM

In Section 9.8, we have seen that a vector such as $\boldsymbol{u} = u_t\boldsymbol{t} + u_x\boldsymbol{x} + \ldots$ in component form may be expressed in terms of a new set of Lorentz transformed basis vectors. We evaluated the "transformed components" in a rather systematic way, with the result for the vector $\boldsymbol{r}$ given in Equation (9.26). We now do the same thing in a more obvious way by the procedure of replacing the old basis vectors $\boldsymbol{t},\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}$ in $u_t\boldsymbol{t} + u_x\boldsymbol{x} + \ldots$ with the corresponding expressions in Equation (9.17) for what we refer to as the reverse transformation of the basis vectors, that is to say, $\boldsymbol{t} = \gamma(\boldsymbol{t}' - v\boldsymbol{x}')$, $\boldsymbol{x} = \gamma(\boldsymbol{x}' - v\boldsymbol{t}')$, and so on. We obtain the coefficients in the new basis by rearranging the result to be in the form $u_{t'}\boldsymbol{t}' + u_{x'}\boldsymbol{x}' + u_{y'}\boldsymbol{y}' + u_{z'}\boldsymbol{z}'$. We therefore find

$$\begin{aligned}\boldsymbol{u} &= u_t\boldsymbol{t} + u_x\boldsymbol{x} + u_y\boldsymbol{y} + u_z\boldsymbol{z}\\ &= u_t\gamma(\boldsymbol{t}' - v\boldsymbol{x}') + u_x\gamma(\boldsymbol{x}' - v\boldsymbol{t}') + u_y\boldsymbol{y} + u_z\boldsymbol{z}\\ &= \gamma(u_t - vu_x)\boldsymbol{t}' + \gamma(u_x - vu_t)\boldsymbol{x}' + u_y\boldsymbol{y} + u_z\boldsymbol{z}\\ &= \quad u_{t'}\boldsymbol{t}' \quad + \quad u_{x'}\boldsymbol{x}' \quad + u_{y'}\boldsymbol{y}' + u_{z'}\boldsymbol{z}'\end{aligned} \tag{10.30}$$

which is the same result as Equation (9.28) but found by a slightly different route. Here we continue to take the basis vector $\boldsymbol{x}$ as being parallel to the transformation parameter $\boldsymbol{v}$ so that $\boldsymbol{y}' = \boldsymbol{y}$ and $\boldsymbol{z}' = \boldsymbol{z}$. The coefficients of the original basis vectors have simply been replaced by the transformed ones, whereas the new coefficients appear as though the Lorentz transformation had been applied to them. While we may write Equation (10.30) in a more succinct way by using the same prescription as Equation (9.27), it is generally difficult to evaluate the coefficients without effectively going through the procedure just described. For example, see Equation (9.28).

Now, provided the basis elements within each grade are chosen to form an orthonormal set, exactly the same process applies for any sort of basis element. For any basis element $\boldsymbol{X}_j$ belonging to $\boldsymbol{X}_1 \ldots \boldsymbol{X}_N$, the entire basis of the original frame, we have $\boldsymbol{X}_j = \sum_k \alpha_{jk}\boldsymbol{X}'_k$ where $\boldsymbol{X}'_1 \ldots \boldsymbol{X}'_N$ are the corresponding basis elements within the new frame (there is no reason to exclude the vectors here). Since vectors transform into vectors, and likewise with objects of all grades, the sum can be restricted to just those elements $\boldsymbol{X}'_k$ that are of the same grade as $\boldsymbol{X}_j$, that is to say, j. Provided that we do so, the required coefficients are given by $\alpha_{jk} = h_{kk}\boldsymbol{X}_j \cdot \boldsymbol{X}'_k$ where the factor $h_{kk} = \boldsymbol{X}_k \cdot \boldsymbol{X}_k = \boldsymbol{X}'_k \cdot \boldsymbol{X}'_k$ deals with those cases where $\boldsymbol{X}_k^2 = -1$ (see Section 7.8 for a similar discussion in relation to the vector derivative). It is also useful to note that since $\boldsymbol{X}_j \cdot \boldsymbol{X}'_k = \boldsymbol{X}'_k \cdot \boldsymbol{X}_j$, it is only necessary to replace each of the α_{jk} with $h_{jj}\alpha_{jk}h_{kk}$ to get the coefficients for the transformation in the other direction, that is to say, they are just the same apart from a change of sign when $j \neq k$.

So far, this part of the discussion has been completely general, but reverting to the simple case the transformation parameter $\boldsymbol{v}$ is along the basis vector $\boldsymbol{x}$, the coefficients α_{jk} for the vectors can be written down from Equation (9.17) as being

$$\begin{aligned}\alpha_{tt} &= \alpha_{xx} = \gamma \\ \alpha_{xt} &= \alpha_{tx} = -\gamma v \\ \alpha_{yy} &= \alpha_{zz} = 1\end{aligned} \tag{10.31}$$

where all the others are zero. Provided we change the sign of v, the coefficients for the bivectors can be obtained from Section 9.5 and, since the pseudoscalars are invariant, the trivectors behave exactly like the vectors. In practice, therefore, the bivectors are as high as we need to go for spacetime. If we express the basis vectors as $\boldsymbol{e}_i = \sum_m \alpha_{im} \boldsymbol{e}'_m$ where i and m range over t, x, y, z, then the bivectors are given by $\boldsymbol{e}_i \boldsymbol{e}_l = \sum_{m,n} \alpha_{im} \alpha_{ln} \boldsymbol{e}'_m \boldsymbol{e}'_n$ for $i \neq l$. Since $\boldsymbol{e}'_m \boldsymbol{e}'_n = -\boldsymbol{e}'_n \boldsymbol{e}'_m$, it must be the case that the contribution arising from the bivector $\boldsymbol{X}_{mn}$ that equates to $\boldsymbol{e}'_m \boldsymbol{e}'_n$ is given by $(\alpha_{in}\alpha_{lm} - \alpha_{im}\alpha_{ln})\boldsymbol{X}'_{mn}$. By defining $\beta_{il,mn} \equiv \alpha_{im}\alpha_{ln} - \alpha_{in}\alpha_{lm}$, we then arrive at $\boldsymbol{X}_{il} = \sum_{mn} \beta_{il,mn} \boldsymbol{X}'_{mn}$ where the sum is now restricted to the standard bivectors, which have the paired indices il and mn in the correct order, that is to say, xt, yt, zt, yz, zx, xy. Using the coefficients given in Equation (10.31), we find

$$\begin{aligned}\beta_{yt,yt} &= \beta_{zt,zt} = \beta_{zx,zx} = \beta_{xy,xy} = \gamma \\ \beta_{yt,xy} &= \beta_{zt,zx} = \beta_{xy,yt} = \beta_{zx,zt} = -\gamma v \\ \beta_{xt,xt} &= \beta_{yz,yz} = 1\end{aligned} \tag{10.32}$$

and again, all the others vanish. Nevertheless, this procedure may by applied to any Lorentz transformation once the values of the α_{jk} for the transformation of the basis vectors have been established. As with the vectors, we can interpret the $\beta_{il,mn}$ as projections, in this case, the projections of the bivectors $\boldsymbol{X}_{il}$ onto each of the basis bivectors $\boldsymbol{X}'_{mn}$ of the $\boldsymbol{t}'$-frame.

The "transformed" components of any bivector $\boldsymbol{U}$ may then be worked out from

$$\begin{aligned}\boldsymbol{U} &= \sum_{M=xt,yt\ldots} h_{MM}(\boldsymbol{U}\cdot\boldsymbol{X}_M)\boldsymbol{X}_M \\ &= (\boldsymbol{U}\cdot(\boldsymbol{xt})')(\boldsymbol{xt})' + (\boldsymbol{U}\cdot(\boldsymbol{yt})')(\boldsymbol{yt})' + (\boldsymbol{U}\cdot(\boldsymbol{zt})')(\boldsymbol{zt})' \\ &\quad -(\boldsymbol{U}\cdot(\boldsymbol{yz})')(\boldsymbol{yz})' - (\boldsymbol{U}\cdot(\boldsymbol{zx})')(\boldsymbol{zx})' - (\boldsymbol{U}\cdot(\boldsymbol{xy})')(\boldsymbol{xy})'\end{aligned} \tag{10.33}$$

Expanding this in an analogous way to Equation (10.30) where we found the components of vectors in a new basis for the usual simple test case in which the transformation parameter is $v\boldsymbol{x}$, we find

$$\begin{aligned}\boldsymbol{U} &= U_{xt}\boldsymbol{xt} + U_{yt}\boldsymbol{yt} + U_{zt}\boldsymbol{zt} + U_{yz}\boldsymbol{yz} + U_{zx}\boldsymbol{zx} + U_{xy}\boldsymbol{xy} \\ &= U_{xt}(\boldsymbol{xt})' + U_{yt}\gamma\big((\boldsymbol{yt})' + v(\boldsymbol{xy})'\big) + U_{zt}\gamma\big((\boldsymbol{zt})' - v(\boldsymbol{zx})'\big) \\ &\quad + U_{yz}(\boldsymbol{yz})' + U_{zx}\gamma\big((\boldsymbol{zx})' - v(\boldsymbol{zt})'\big) + U_{xy}\gamma\big((\boldsymbol{xy})' + v(\boldsymbol{yt})'\big) \\ &= U_{xt}(\boldsymbol{xt})' + \gamma\big(U_{yt} + vU_{xy}\big)(\boldsymbol{yt})' + \gamma\big(U_{zt} - vU_{zx}\big)(\boldsymbol{zt})' \\ &\quad + U_{yz}(\boldsymbol{yz})' + \gamma\big(U_{zx} - U_{zt}v\big)(\boldsymbol{zx})' + \gamma\big(U_{xy} + vU_{yt}\big)(\boldsymbol{xy})'\end{aligned} \tag{10.34}$$

In the particular case that in the original basis $\boldsymbol{U}$ is a temporal bivector with no spatial part (see Section 7.11), this takes the form

$$\begin{aligned}\boldsymbol{U} = U_{xt}(\boldsymbol{xt})' + \gamma U_{yt}(\boldsymbol{yt})' + \gamma U_{zt}(\boldsymbol{zt})' \\ - \gamma v U_{zt}(\boldsymbol{zx})' + \gamma v U_{yt}(\boldsymbol{xy})'\end{aligned} \tag{10.35}$$

It is clear that, in the new basis, a spatial part appears and the converse holds if $\boldsymbol{U}$ is initially a spacelike bivector with no temporal part.

These sort of techniques will be applied to bivectors in Section 11.5.3 as a means of finding the Lorentz transformation for the electromagnetic field.

10.9 VELOCITY AS SEEN IN DIFFERENT FRAMES

Here we explore a problem that puts to the test several of the techniques we have been developing for manipulating spacetime expressions and transforming from one frame to another—given the velocity a particle with arbitrary velocity $\boldsymbol{v}$ in one frame, how do we find $\boldsymbol{v}'$, the form it takes in some other frame? Understanding how to tackle an apparently simple problem of this sort will be of benefit when it comes to dealing with electromagnetic fields of moving charges as discussed in the latter part of Chapter 11 and in Chapter 12.

Referring once more to Figure 7.4, we see the arbitrary history of some particle. We may express this history as either $\boldsymbol{r}(t)$ in the $\boldsymbol{t}$-frame (solid grid) or $\boldsymbol{r}(t')$ in the $\boldsymbol{t}'$-frame (dashed grid). As already noted, the history itself is a spacetime vector that is independent of which frame we choose, it is only the expression of $\boldsymbol{r}$ in terms of the chosen basis vectors and time parameter that changes. In fact, if we restrict ourselves to Lorentz frames, that is to say no reflections or spatial rotations are involved, we need to concern ourselves only with the change in the time vector. Given $\boldsymbol{r}(t)$, finding $\boldsymbol{r}(t')$ leads to the allied question of how the instantaneous velocity $\boldsymbol{v}'$ in one frame is related to $\boldsymbol{v}$ in the other. For example, in the simple case of uniform motion, the two histories are related by $\boldsymbol{r}(t) = \boldsymbol{v}t + \underset{\sim}{\boldsymbol{r}}_0$ and $\boldsymbol{r}(t') = \boldsymbol{v}'t' + \underset{\sim}{\boldsymbol{r}}'_0$. While we can readily find t' and $\underset{\sim}{\boldsymbol{r}}'_0$ as a result of a change of frame, $\boldsymbol{v}'$ cannot be found just by expressing the $\boldsymbol{t}$-frame form of $\boldsymbol{v}$ in terms of the $\boldsymbol{t}'$-frame basis vectors because velocity is frame dependent. In fact, in Section 8.4.2 we pointed out that $\boldsymbol{v}' = \partial_{t'}\boldsymbol{r}$ whereas $\boldsymbol{v} = \partial_t \boldsymbol{r}$. It was necessary to take heed of this point when finding the relative velocity vector for a particle in motion. A more general yet much neater method of finding relative velocity vectors in any frame was shown in Section 10.6.4, and here we now use a similar approach for spacetime velocities.

Starting from the particle's proper velocity, $\mathfrak{v}$, the required procedure is similar to a spacetime split except that instead of projecting onto (3+1)D, we project onto the spacetime $\boldsymbol{t}'$-frame, that is to say, we project separately onto $\boldsymbol{t}'$ and its orthogonal space. Equation (9.27) is an example of projection of this sort when all the basis vectors are given, and we may use any basis vectors, not just $\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$. Here the difference is that only the time vector of each frame is available and all spatial vectors are basis free. We begin by stating the known vectors in terms of the $\boldsymbol{t}$-frame. The particle's velocity, which need not be constant, is taken as $\boldsymbol{v}$ so that its proper veloc-

ity $\mathfrak{v}$ is given by $\mathfrak{v} = \hat{\boldsymbol{v}} = \gamma_v \boldsymbol{v}$. As usual, we may write $\boldsymbol{v} = \boldsymbol{t} + \underset{\sim}{\boldsymbol{v}}$, while velocity of the origin of the $\boldsymbol{t}'$-frame is taken as $\boldsymbol{u} = \boldsymbol{t} + \underset{\sim}{\boldsymbol{u}}$. Normalizing $\boldsymbol{u}$ then gives the time vector of the $\boldsymbol{t}'$-frame as $\boldsymbol{t}' = \gamma_u \boldsymbol{u} = \gamma_u (\boldsymbol{t} + \underset{\sim}{\boldsymbol{u}})$. Recall that both $\underset{\sim}{\boldsymbol{v}}$ and $\underset{\sim}{\boldsymbol{u}}$ are implicitly orthogonal to $\boldsymbol{t}$, but since we now have two such spatial velocities, we have the two separate normalization factors $\gamma_v = \left(1 - v^2\right)^{-1/2}$ and $\gamma_u = \left(1 - u^2\right)^{-1/2}$ where v and u are the magnitudes of $\underset{\sim}{\boldsymbol{v}}$ and $\underset{\sim}{\boldsymbol{u}}$, respectively.

The first step is to form our projection of $\mathfrak{v}$ onto the $\boldsymbol{t}'$-frame. Whatever the result, it must take the form $\mathfrak{v} = \gamma_{v'} \boldsymbol{v}' = \gamma_{v'} (\boldsymbol{t}' + \underset{\sim}{\boldsymbol{v}}')$ where $\boldsymbol{v}' = \partial_{t'} \boldsymbol{r}$ is the velocity of the particle in the $\boldsymbol{t}'$-frame, and the required normalization is provided by $\gamma_{v'} = \left(1 - v'^2\right)^{-1/2}$ where $v' = |\underset{\sim}{\boldsymbol{v}}'|$. The projection is formed by constructing the identity

$$\begin{aligned} \mathfrak{v} &= -\mathfrak{v} \boldsymbol{t}' \boldsymbol{t}' \\ &= -(\mathfrak{v} \cdot \boldsymbol{t}' + \mathfrak{v} \wedge \boldsymbol{t}') \boldsymbol{t}' \\ &= -(\mathfrak{v} \cdot \boldsymbol{t}') \boldsymbol{t}' - (\mathfrak{v} \wedge \boldsymbol{t}') \boldsymbol{t}' \\ &= (-\mathfrak{v} \cdot \boldsymbol{t}') \left(\boldsymbol{t}' + \left(\frac{\mathfrak{v} \wedge \boldsymbol{t}'}{\mathfrak{v} \cdot \boldsymbol{t}'} \right) \boldsymbol{t}' \right) \end{aligned} \tag{10.36}$$

Equation (9.27) would give the projection as $\mathfrak{v} = (-\mathfrak{v} \cdot \boldsymbol{t}') \boldsymbol{t}' + (\mathfrak{v} \cdot \boldsymbol{x}') \boldsymbol{x}' + (\mathfrak{v} \cdot \boldsymbol{y}') \boldsymbol{y}' + (\mathfrak{v} \cdot \boldsymbol{z}') \boldsymbol{z}'$ so that by comparison with the third line of Equation (10.36), we see that $-(\mathfrak{v} \wedge \boldsymbol{t}') \boldsymbol{t}'$ replaces all of $(\mathfrak{v} \cdot \boldsymbol{x}') \boldsymbol{x}' + (\mathfrak{v} \cdot \boldsymbol{y}') \boldsymbol{y}' + (\mathfrak{v} \cdot \boldsymbol{z}') \boldsymbol{z}'$.

By the very nature of the projection we have constructed, $\left(\dfrac{\mathfrak{v} \wedge \boldsymbol{t}'}{\mathfrak{v} \cdot \boldsymbol{t}'} \right) \boldsymbol{t}'$ should be a vector that is orthogonal to $\boldsymbol{t}'$. We may verify this as follows since $\mathfrak{v} \wedge \boldsymbol{t}'$ is a bivector and $(\mathfrak{v} \wedge \boldsymbol{t}') \boldsymbol{t}' = (\mathfrak{v} \wedge \boldsymbol{t}') \cdot \boldsymbol{t}'$. The inner product of a vector $\boldsymbol{t}'$ with a bivector $\boldsymbol{U}$ automatically results in a vector that is orthogonal to $\boldsymbol{t}'$ (see for example Figure 2.1h). Since the denominator $\mathfrak{v} \cdot \boldsymbol{t}'$ is a scalar, these properties are simply carried over into $\left(\dfrac{\mathfrak{v} \wedge \boldsymbol{t}'}{\mathfrak{v} \cdot \boldsymbol{t}'} \right) \boldsymbol{t}'$. Given that $\mathfrak{v}$ is projected into the $\boldsymbol{t}'$-frame as $\mathfrak{v} = \gamma_{v'} \boldsymbol{v}' = \gamma_{v'} (\boldsymbol{t}' + \underset{\sim}{\boldsymbol{v}}')$ where $\underset{\sim}{\boldsymbol{v}}$ is in the orthogonal space of $\boldsymbol{t}'$, by identifying the bottom line of Equation (10.36) with $\gamma_{v'} (\boldsymbol{t}' + \underset{\sim}{\boldsymbol{v}}')$, we may directly conclude

$$\begin{aligned} \gamma_{v'} &= -\mathfrak{v} \cdot \boldsymbol{t}' \\ \underset{\sim}{\boldsymbol{v}}' &= \left(\frac{\mathfrak{v} \wedge \boldsymbol{t}'}{\mathfrak{v} \cdot \boldsymbol{t}'} \right) \boldsymbol{t}' \end{aligned} \tag{10.37}$$

In the discussion leading to Equation (10.21), it was pointed out that $1/(-\mathfrak{v} \cdot \boldsymbol{t}')$ is precisely the factor required to take into account the change of the variable of differentiation in going from $\partial_t \boldsymbol{r}$ to $\partial_{t'} \boldsymbol{r}$. The surprising thing is that here there has been no need to consider this issue explicitly, it has been subsumed into the analysis simply by imposing the required form, $\gamma_{v'} (\boldsymbol{t}' + \underset{\sim}{\boldsymbol{v}}')$, on $\mathfrak{v}$.

The evaluation of $-\mathfrak{v} \cdot \boldsymbol{t}'$ and $-(\mathfrak{v} \wedge \boldsymbol{t}') \boldsymbol{t}'$ subsequently produces

$$\begin{aligned} -\mathfrak{v} \cdot \boldsymbol{t}' &= -\gamma_v \boldsymbol{v} \cdot \gamma_u \boldsymbol{u} \\ &= -\gamma_v \gamma_u \boldsymbol{v} \cdot \boldsymbol{u} \\ &= \gamma_v \gamma_u (1 - \underset{\sim}{\boldsymbol{v}} \cdot \underset{\sim}{\boldsymbol{u}}) \end{aligned} \tag{10.38}$$

and

$$\begin{aligned}
-(\boldsymbol{\upsilon}\wedge\boldsymbol{t}')\boldsymbol{t}' &= -\left(\gamma_v(\boldsymbol{t}+\boldsymbol{v}_{/\!/}+\boldsymbol{v}_\perp)\wedge\boldsymbol{t}'\right)\boldsymbol{t}' \\
&= -\left(\gamma_v(\boldsymbol{t}+\boldsymbol{v}_{/\!/})\wedge\gamma_u(\boldsymbol{t}+\boldsymbol{v}_{/\!/})\right)\boldsymbol{t}' - (\gamma_v\boldsymbol{v}_\perp\wedge\boldsymbol{t}')\boldsymbol{t}' \\
&= -\gamma_v\gamma_u(\boldsymbol{v}_{/\!/}\boldsymbol{t}-\underset{\sim}{\boldsymbol{u}}\boldsymbol{t})\boldsymbol{t}' - (\gamma_v\boldsymbol{v}_\perp\boldsymbol{t}')\boldsymbol{t}' \\
&= -\gamma_v\gamma_u(\boldsymbol{v}_{/\!/}-\underset{\sim}{\boldsymbol{u}})\boldsymbol{t}\boldsymbol{t}' + \gamma_v\boldsymbol{v}_\perp \\
&= \gamma_v\gamma_u(\boldsymbol{v}_{/\!/}-\underset{\sim}{\boldsymbol{u}})\gamma_u(1+\underset{\sim}{\boldsymbol{u}}\boldsymbol{t}) + \gamma_v\boldsymbol{v}_\perp \\
&= \gamma_v\gamma_u^2\left(\boldsymbol{v}_{/\!/}-\underset{\sim}{\boldsymbol{u}}+\left(\boldsymbol{v}_{/\!/}\underset{\sim}{\boldsymbol{u}}-\underset{\sim}{\boldsymbol{u}}^2\right)\boldsymbol{t}\right) + \gamma_v\boldsymbol{v}_\perp \\
&= \gamma_v\gamma_u^2\left(\boldsymbol{v}_{/\!/}-\underset{\sim}{\boldsymbol{u}}+\left(\underset{\sim}{\boldsymbol{v}}\cdot\underset{\sim}{\boldsymbol{u}}-\underset{\sim}{\boldsymbol{u}}^2\right)\boldsymbol{t}\right) + \gamma_v\boldsymbol{v}_\perp
\end{aligned} \tag{10.39}$$

Here we have used $\boldsymbol{t}\boldsymbol{t}' = \gamma_u\boldsymbol{t}(\boldsymbol{t}+\underset{\sim}{\boldsymbol{u}}) = -\gamma_u(1+\underset{\sim}{\boldsymbol{u}}\boldsymbol{t})$ and, once again, we have resorted to the device of splitting up a vector into parts that are parallel and perpendicular to a given direction. In writing $\underset{\sim}{\boldsymbol{v}} = \boldsymbol{v}_{/\!/} + \boldsymbol{v}_\perp$, we have dropped the under-tildes on the right-hand side for the sake of readability, the subscripts // and $\perp$ are with respect to $\underset{\sim}{\boldsymbol{u}}$, and it is evident that $\boldsymbol{v}_{/\!/}\underset{\sim}{\boldsymbol{u}} = \underset{\sim}{\boldsymbol{v}}\cdot\underset{\sim}{\boldsymbol{u}}$. This done, the difficulty of resolving $(\underset{\sim}{\boldsymbol{u}}\wedge\underset{\sim}{\boldsymbol{v}})\cdot\boldsymbol{t}'$ into a useable form is neatly avoided. Returning to Equation (10.37), we then find $\gamma_{v'}$ and $\boldsymbol{v}'$

$$\begin{aligned}
\gamma_{v'} &= -\boldsymbol{\upsilon}\cdot\boldsymbol{t}' \\
&= \gamma_v\gamma_u(1-\underset{\sim}{\boldsymbol{v}}\cdot\underset{\sim}{\boldsymbol{u}}) \\
\boldsymbol{v}' &= \boldsymbol{t}' + \underset{\sim}{\boldsymbol{v}}' \\
&= \boldsymbol{t}' + \left(\frac{\boldsymbol{\upsilon}\wedge\boldsymbol{t}'}{\boldsymbol{\upsilon}\cdot\boldsymbol{t}'}\right)\boldsymbol{t}' \\
&= \boldsymbol{t}' + \frac{\gamma_u\left(\boldsymbol{v}_{/\!/}-\underset{\sim}{\boldsymbol{u}}+\left(\underset{\sim}{\boldsymbol{v}}\cdot\underset{\sim}{\boldsymbol{u}}-\underset{\sim}{\boldsymbol{u}}^2\right)\boldsymbol{t}\right)}{(1-\underset{\sim}{\boldsymbol{v}}\cdot\underset{\sim}{\boldsymbol{u}})} + \frac{\boldsymbol{v}_\perp}{\gamma_u(1-\underset{\sim}{\boldsymbol{v}}\cdot\underset{\sim}{\boldsymbol{u}})}
\end{aligned} \tag{10.40}$$

Given that we already know that $\boldsymbol{t}' = \gamma_u\boldsymbol{u}$, this result gives us $\boldsymbol{v}'$, the particle's velocity in the $\boldsymbol{t}'$-frame based only on the information available to us in the $\boldsymbol{t}$-frame. Despite the fact that $\boldsymbol{v}'$ is given from $\boldsymbol{t}'+\underset{\sim}{\boldsymbol{v}}'$, $\underset{\sim}{\boldsymbol{v}}'$ itself is expressed in terms of vectors that are specific to the $\boldsymbol{t}$-frame. Note, however, that the term $1-\underset{\sim}{\boldsymbol{v}}\cdot\underset{\sim}{\boldsymbol{u}}$ may be expressed in any frame as just $-\boldsymbol{v}\cdot\boldsymbol{u}$, and it is also available from the (3+1)D velocities as $1-\mathbf{v}\cdot\mathbf{u}$. The result begins to become clear if we express $\boldsymbol{v}$ and $\boldsymbol{u}$ in terms of the $\boldsymbol{t}$-frame basis $\boldsymbol{t},\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}$. Assigning $\boldsymbol{x}$ to be along the direction of $\underset{\sim}{\boldsymbol{u}}$ gives us $\underset{\sim}{\boldsymbol{u}} = u\boldsymbol{x}$, $\boldsymbol{v}_{/\!/} = v_x\boldsymbol{x}$, and $\underset{\sim}{\boldsymbol{v}}\cdot\underset{\sim}{\boldsymbol{u}} = v_x u$, leading to

$$\begin{aligned}
\boldsymbol{v}' &= \boldsymbol{t}' + \underset{\sim}{\boldsymbol{v}}' \\
&= \boldsymbol{t}' + \frac{\gamma_u\left((v_x-u)\boldsymbol{x}+\left(v_x u-u^2\right)\boldsymbol{t}\right)}{1-v_x u} + \frac{\gamma_u^{-1}}{1-v_x u}\left(v_y\boldsymbol{y}+v_z\boldsymbol{z}\right) \\
&= \boldsymbol{t}' + \frac{(v_x-u)\gamma_u(\boldsymbol{x}+u\boldsymbol{t})}{1-v_x u} + \frac{\gamma_u^{-1}}{1-v_x u}\left(v_y\boldsymbol{y}+v_z\boldsymbol{z}\right) \\
&= \boldsymbol{t}' + \frac{v_x-u}{1-v_x u}\boldsymbol{x}' + \frac{\gamma_u^{-1}}{1-v_x u}\left(v_y\boldsymbol{y}'+v_z\boldsymbol{z}'\right)
\end{aligned} \tag{10.41}$$

which has a much more familiar look. The difference is that all the vectors are now in terms of the $\boldsymbol{t}'$-frame basis, $\boldsymbol{t}', \boldsymbol{x}', \boldsymbol{y}', \boldsymbol{z}'$, whereas all the *components* are expressions belonging to the $\boldsymbol{t}$-frame. This is exactly the situation that applies when we transform basis vectors.

Equation (10.41) readily gives us $\mathbf{v}'$, the particle's relative velocity in the $\boldsymbol{t}'$-frame, from the simple spacetime split $1+\mathbf{v}' = -\boldsymbol{t}'\boldsymbol{v}'$:

$$\begin{aligned} 1+\mathbf{v}' &= -\boldsymbol{t}'\left(\boldsymbol{t}' + \frac{v_x - u}{1-v_x u}\boldsymbol{x}' + \frac{\gamma_u^{-1}}{1-v_x u}\left(v_y \boldsymbol{y}' + v_z \boldsymbol{z}'\right)\right) \\ &= 1 + \frac{v_x - u}{1-v_x u}\boldsymbol{x}'\boldsymbol{t}' + \frac{\gamma_u^{-1}}{1-v_x u}\left(v_y \boldsymbol{y}'\boldsymbol{t}' + v_z \boldsymbol{z}'\boldsymbol{t}'\right) \\ &= 1 + \frac{v_x - u}{1-v_x u}\mathbf{x} + \frac{\gamma_u^{-1}}{1-v_x u}\left(v_y \mathbf{y} + v_z \mathbf{z}\right) \\ \Leftrightarrow \mathbf{v}' &= \frac{v_x - u}{1-v_x u}\mathbf{x} + \frac{\gamma_u^{-1}}{1-v_x u}\left(v_y \mathbf{y} + v_z \mathbf{z}\right) \end{aligned} \tag{10.42}$$

Because the motion between the two frames was chosen arbitrarily to be along $\boldsymbol{x}$, this may be put in the more general but simpler form,

$$\mathbf{v}' = \frac{\mathbf{v}_{/\!/} - \mathbf{u}}{1-\mathbf{v}_{/\!/}\mathbf{u}} + \frac{\mathbf{v}_\perp}{\gamma_u\left(1-\mathbf{v}_{/\!/}\mathbf{u}\right)} \tag{10.43}$$

where $\mathbf{u}$ is the relative velocity of the $\boldsymbol{t}'$-frame with respect to the $\boldsymbol{t}$-frame and $\mathbf{v}_{/\!/}$ and $\mathbf{v}_\perp$ give the particle's velocity parallel and transverse to $\mathbf{u}$ respectively.

Finding relative velocity vectors in a new frame may also be tackled by changing basis vectors and allowing for the change of the variable of differentiation. However, this step is neatly eliminated by projecting from the particle's proper velocity onto the frame in question. Equation (10.22) and its counterpart Equation (10.36) may look simple enough, but they will both produce an undesirable spacelike bivector unless the precaution of splitting $\boldsymbol{v}$ into $\boldsymbol{v}_{/\!/}$ is $\boldsymbol{v}_\perp$ is taken beforehand.

Equation (10.43) encapsulates the Lorentz transformation for velocities. Note that it is another example of the Lorentz transformation of a derived vector and also that we have not insisted that $\boldsymbol{u}(t)$ is constant, as is usually the case when this result is found by compounding two Lorentz transformations. We must remember, however, that the result applies at a time t in the $\boldsymbol{t}'$-frame so that t' has to be found from $\gamma(t - \boldsymbol{r}\cdot\boldsymbol{v})$. This is not possible to solve without knowing $\boldsymbol{r}$, the particle's history. Whereas in the simple case of constant velocity we do not need to worry about this minor point, in principle, we are able to relate the timing of the two observed velocities if we know enough of the particle's history.

10.10 FRAME-FREE FORM OF THE LORENTZ TRANSFORMATION

The simple representations of the Lorentz transformation in Sections 9.4–9.8 relied on some nominal frame, the $\boldsymbol{t}$-frame, and in addition, the direction of the transformation parameter $\underset{\sim}{\boldsymbol{v}}$ was usually associated with the basis vector $\boldsymbol{x}$ and written as $v\boldsymbol{x}$. For practical purposes, it would be useful to have the full power of Equation (9.8) in a simple form that is nevertheless still frame free.

In Equation (10.24), we found a fairly simple version for the Lorentz transformation that did not involve basis vectors. Unfortunately, it applies to relative vectors rather than spacetime vectors, and we also had the problem of ensuring the outcome was in terms of timelike bivectors alone, for these are the only sort that equate to relative vectors. Here we derive an equivalent result for spacetime vectors and, in so doing, we manage to avoid the problem by bringing in the idea of splitting the vector to be transformed into appropriately defined parallel and perpendicular parts from the outset.

Equation (9.8) defines the Lorentz transformation in a perfectly frame-free manner as a rotation in a spacetime plane. The plane is specified by the unit timelike bivector $\boldsymbol{N}$, while the magnitude of the rotation is related to the velocity parameter v. However, the spacetime split of $v\boldsymbol{N}$ equates to a relative velocity $\mathbf{v}$ (Equation 9.10) that must be related to the time vector used for the split. We could write this split in any frame, say the $\boldsymbol{\theta}$-frame, as $-(v\boldsymbol{N}\cdot\boldsymbol{\theta})\boldsymbol{\theta} = \underset{\sim}{\boldsymbol{v}}\boldsymbol{\theta} \leftrightarrow \mathbf{v}$ where $|\mathbf{v}| = |\underset{\sim}{\boldsymbol{v}}| = v$. This is just another way of saying that any given timelike bivector may be expressed as the product of any time vector $\boldsymbol{\theta}$ and some corresponding spatial vector, taken here to be $\underset{\sim}{\boldsymbol{v}}$. The vector $\underset{\sim}{\boldsymbol{v}}$, being projected out of $v\boldsymbol{N}$ by the inner product $-v\boldsymbol{N}\cdot\boldsymbol{\theta}$, must be orthogonal to $\boldsymbol{\theta}$ and so we may say that it is spatial in the $\boldsymbol{\theta}$-frame. Now, while relative vectors are generally associated with a frame, that is to say the frame from which they were projected into (3+1)D, there is no way of distinguishing which relative vectors came from which frame unless we kept that information. We may therefore attach them to any frame, and consequently we may identify $v\boldsymbol{N}$, which has no associated frame, with a relative vector $\mathbf{v}$. To express $v\boldsymbol{N}$ in some frame, we only need to equate the timelike basis bivectors of that frame to $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$. We may then transform from any given frame to some new frame by using $\mathbf{v}$ as the relative velocity between the frames. Note that although $|\boldsymbol{v}| = \left(1-v^2\right)^{1/2}$, the magnitudes of $\mathbf{v}$, $v\boldsymbol{N}$, and $\underset{\sim}{\boldsymbol{v}}$ are all given by v.

It will by now be no surprise that it proves easiest to split the vector undergoing the transformation into parts that are parallel and perpendicular to the plane of the transformation, $\boldsymbol{N}$. We therefore write $\boldsymbol{u} = \boldsymbol{u}_{/\!/} + \boldsymbol{u}_{\perp}$ where $\boldsymbol{u}_{\perp}$ is the perpendicular part and $\boldsymbol{u}_{/\!/}$ is the part lying in the plane. Now, it will be key to evaluation of the transformation acting on $\boldsymbol{u}_{/\!/} + \boldsymbol{u}_{\perp}$ that $\boldsymbol{N}$ must commute with $\boldsymbol{u}_{\perp}$ but anticommute with $\boldsymbol{u}_{/\!/}$. To see this, let us first of all deal with $\boldsymbol{u}_{\perp}$. The orthogonality between any two objects may be established if their inner product vanishes, that is to say, in our case $\boldsymbol{u}_{\perp}\cdot\boldsymbol{N} = 0$. But from the rules given in Equation (4.6) for the inner product of a vector with a bivector, we must then have $\frac{1}{2}(\boldsymbol{u}_{\perp}\boldsymbol{N} - \boldsymbol{N}\boldsymbol{u}_{\perp}) = 0$, so that $\boldsymbol{N}\boldsymbol{u}_{\perp} = \boldsymbol{u}_{\perp}\boldsymbol{N}$

as required. Conversely, if two objects are parallel we must have $\boldsymbol{u}_{//} \wedge \boldsymbol{N} = 0$ so that $\frac{1}{2}(\boldsymbol{u}_\perp \boldsymbol{N} + \boldsymbol{N}\boldsymbol{u}_\perp) = 0$ and $\boldsymbol{N}\boldsymbol{u}_\perp = -\boldsymbol{u}_\perp \boldsymbol{N}$.

We may now write out Equation (9.8) in terms of $\boldsymbol{u}_{//} + \boldsymbol{u}_\perp$ and apply these commutation properties to find

$$\begin{aligned}
\boldsymbol{u} \mapsto \boldsymbol{u}' &= (a - b\boldsymbol{N})(\boldsymbol{u}_{//} + \boldsymbol{u}_\perp)(a + b\boldsymbol{N}) \\
&= (a - b\boldsymbol{N})^2 \boldsymbol{u}_{//} + (a - b\boldsymbol{N})(a + b\boldsymbol{N})\boldsymbol{u}_\perp \\
&= \left(a^2 - 2ab\boldsymbol{N} + b^2\boldsymbol{N}^2\right)\boldsymbol{u}_{//} + \left(a^2 - b^2\boldsymbol{N}^2\right)\boldsymbol{u}_\perp \\
&= \left(a^2 - 2ab\boldsymbol{N} + b^2\right)\boldsymbol{u}_{//} + \left(a^2 - b^2\right)\boldsymbol{u}_\perp \\
&= (\gamma - \gamma v\boldsymbol{N})\boldsymbol{u}_{//} + \boldsymbol{u}_\perp \\
&= \gamma(1 - \mathbf{v})\boldsymbol{u}_{//} + \boldsymbol{u}_\perp
\end{aligned} \tag{10.44}$$

In reaching this result, we have made use of some further properties:

- $\boldsymbol{N}^2 = 1$, since any timelike bivector has a positive square and, by definition, $|\boldsymbol{N}| = 1$.
- From Equation (9.10),
  - $a^2 + b^2 = \left(1 - v^2\right)^{-1/2} = \gamma$
  - $2ab = \left(\gamma^2 - 1\right)^{1/2} = \gamma v$
  - $a^2 - b^2 = 1$.

It is confirmed that $\boldsymbol{u}_\perp$ is unaffected by the transformation while $\boldsymbol{u}_{//}$ appears to be altered in a very simple way. As discussed above, we may still consider this form to be frame free as we do not need to say beforehand what frame $\mathbf{v}$ is relative to. Clearly, Equation (10.44) is a much simpler result than we could have found by expressing an arbitrary vector $\boldsymbol{u}$ and the velocity $\mathbf{v}$ in terms of some set of basis such as $\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ and then evaluating the transformation from, say, Equations (9.8) or (9.11).

Let us therefore see how Equation (10.44) relates to the basic transformation given in Equation (9.11). Resorting to the $\boldsymbol{t}$-frame, we will take the transformation plane to be $\boldsymbol{N} = \mathbf{v} = v\boldsymbol{x}\boldsymbol{t}$. Any linear combination of $\boldsymbol{y}$ and $\boldsymbol{z}$ will therefore qualify as $\boldsymbol{u}_\perp$ whereas any linear combination of $\boldsymbol{t}$ and $\boldsymbol{x}$ will qualify as $\boldsymbol{u}_{//}$, so that we may take it as being either $\boldsymbol{t}$ or $\boldsymbol{x}$. This should give us back the simple form of the Lorentz transformation acting on the basis vectors that we seek. Indeed, we find, as expected, $\boldsymbol{t}' = \gamma(1 - v\boldsymbol{x}\boldsymbol{t})\boldsymbol{t} = \gamma(\boldsymbol{t} + v\boldsymbol{x})$ and $\boldsymbol{x}' = \gamma(1 - v\boldsymbol{x}\boldsymbol{t})\boldsymbol{x} = \gamma(\boldsymbol{x} + v\boldsymbol{t})$, while both $\boldsymbol{y}$ and $\boldsymbol{z}$ come under $\boldsymbol{u}_\perp$ and therefore are unchanged.

For a slightly more general case, we can take $\boldsymbol{u}_{//} = \boldsymbol{t}$ without defining the motion to be along any basis vector. By simply specifying the spatial part of the velocity to be $\underset{\sim}{\boldsymbol{v}}$ where $\mathbf{v} = \underset{\sim}{\boldsymbol{v}}\boldsymbol{t}$, we find $\boldsymbol{t}' = \gamma(1 - \underset{\sim}{\boldsymbol{v}}\boldsymbol{t})\boldsymbol{t} = \gamma(\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}})$. This is a useful relationship that allows us to find the new time vector quickly. In fact, we see this is just the same as the proper velocity $\mathfrak{v} = \gamma(\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}})$ for a particle at rest in the $\boldsymbol{t}'$-frame (Equation 10.5), which is no surprise.

There seems to be no disadvantage to the frame-free form of the Lorentz transformation as we have it here in Equation (10.44). It is no more difficult to use than the basic $\boldsymbol{t}$-frame form of Equation (9.11), yet it is far less constrained. We do have to find $\boldsymbol{u}_{/\!/}$ and $\boldsymbol{u}_{\perp}$ from $\boldsymbol{u}$, but at least $\boldsymbol{u}_{\perp}$ can be found from $\boldsymbol{u}_{\perp} = (\boldsymbol{u} \wedge \boldsymbol{N}) \cdot \boldsymbol{u}$ and it is also helpful to note that, relative to the same frame, the vectors $\mathbf{u}_{/\!/}$ and $\mathbf{u}_{\perp}$ may be taken parallel and to perpendicular $\mathbf{v}$, respectively, such that $\mathbf{u} = \mathbf{u}_{/\!/} + \mathbf{u}_{\perp}$ is the relative vector for $\boldsymbol{u}$.

10.11 EXERCISES

1. The pair of individual transformations $\boldsymbol{t} \mapsto \boldsymbol{t}' = \gamma(\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}})$ and $\underset{\sim}{\boldsymbol{v}} \mapsto \underset{\sim}{\boldsymbol{v}}' = \gamma\left(\underset{\sim}{\boldsymbol{v}} + v^2\boldsymbol{t}\right)$ defines a transformation in 2D that acts on the time vector $\boldsymbol{t}$ and the mutually orthogonal spatial vector $\underset{\sim}{\boldsymbol{v}}$.

(a) What is the inverse transformation back to $\boldsymbol{t}$ and $\underset{\sim}{\boldsymbol{v}}$?

(b) Is the transformation orthogonal?

(c) How does it compare with the transformation given by Equation (9.11)?

(d) How may the Lorentz transformation of a vector $\boldsymbol{u}$ be expressed without referring to the basis vectors $\boldsymbol{x}, \boldsymbol{y}$ and $\boldsymbol{z}$?

2. **(a)** What is the spacetime split of $\mathfrak{v} = \gamma(\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}})$ in the $\boldsymbol{t}$-frame?

(b) What is the spacetime split of $\boldsymbol{t}$ in the $\mathfrak{v}$-frame?

(c) What are the relative vectors for $\boldsymbol{v} = \boldsymbol{t} + \underset{\sim}{\boldsymbol{v}}$ and $\mathfrak{v}$ in the $\boldsymbol{t}$-frame? Comment on any upper limit on their magnitudes.

3. Show that the projection of the vector $\boldsymbol{u}$ onto the orthogonal space of the time vector $\boldsymbol{\theta}$ is given by $\underset{\sim}{\boldsymbol{u}} = \boldsymbol{\theta}(\boldsymbol{u} \wedge \boldsymbol{\theta})$.

4. In Equation (10.44), the "two-sided" rotor equation $\boldsymbol{u}' = \boldsymbol{L}\boldsymbol{u}\boldsymbol{L}^{\dagger}$ is transformed into a single-sided one of the form $\boldsymbol{u}' = \boldsymbol{V}\boldsymbol{u}_{/\!/} + \boldsymbol{u}_{\perp}$. Comment on how this is possible.

5. Find the counterpart of Equation (10.35) when $\boldsymbol{U}$ is purely spacelike in the original $\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{t}$ basis.

6. Show how Equation (10.44) may be derived by employing Equation (10.23).

7. Use Equation (10.44) to find expressions for the transformed basis vectors $\boldsymbol{x}', \boldsymbol{y}', \boldsymbol{z}'$ when the Lorentz transformation parameter $\mathbf{v}$ lies anywhere in the $\mathbf{xy}$ plane.

Chapter 11

Application of the Spacetime Geometric Algebra to Basic Electromagnetics

The material presented in this chapter seeks to address two separate themes. The basic theme is concerned with showing how the equations of (3+1)D electromagnetic theory turn out in spacetime form and assessing the potential benefits. While this is a fairly obvious and straightforward thing to do, it offers only limited scope for getting a better understanding of electromagnetic theory. However, since this may be done without involving special relativity, the material covered in this theme is readily accessible provided the reader has gained a reasonable grasp of Chapters 1–6 and most of 7–8. Given that only one frame of reference is involved, that is to say the frame where the observer is, there will be only one time vector, $\boldsymbol{t}$, the role of which may be regarded as being largely symbolic. The physical motivation for spacetime will not be an issue; it is just to be regarded as a convenient mathematical framework that works well with geometric algebra. Even so, it will be apparent that even just rendering the equations into 4D form provides some interesting observations that are harder to grasp in (3+1)D.

The second theme works on two levels. On the first level, it extends the basic theme set out above to new situations in which changes of reference frame are involved, for example, finding the form of the electromagnetic field in a different frame. For this purpose, the material introduced in Sections 7.6–7.7 should be a sufficient prerequisite. On the second level, however, this theme also addresses the fundamental physical questions such as: "where does the magnetic field come from?" For example, it takes us from the properties of a point charge in its own rest frame, that is to say, its Coulomb field or scalar potential, to the field of a moving charge thereby allowing the true origin of the magnetic field to be exposed. Some familiarity with Chapters 9 and 10 is therefore recommended prior to attempting this more advanced level.

While Chapter 5 dealt with the magnetic field in the classical way by assuming that it originated from an entirely separate phenomenon attributed to electric

Understanding Geometric Algebra for Electromagnetic Theory, First Edition. John W. Arthur.

currents, the application of some geometric algebra nevertheless demonstrated that there was a connection between magnetic and electric effects. By applying it to the solution for the electric field of a quasistatic charge distribution, ρ, it was found that the very same solution could be made to apply to finding the magnetic field simply by replacing ρ with $-\mathbf{J}$, where $\mathbf{J}$ is the current density vector. In fact, in the form $\boldsymbol{F} = \mathbf{E} + \boldsymbol{B}$, the entire electromagnetic field immediately follows just by using the paravector $\boldsymbol{J} = \rho - \mathbf{J}$ to represent the total electromagnetic source density in Equation (5.13). Although this was only shown for steady-state conditions, it was all the same a strong indication that magnetic and electric effects have a common origin, the only difference being due to the state of motion of the charges involved. Geometric algebra simply made it easier to see that connection.

Spacetime takes this to its logical conclusion. Starting from the Coulomb field in the charges' own rest frame, observation in a different frame produces a different sort of field. The Coulomb field has the form of a timelike bivector, but in any other frame, it appears in a mixed bivector form, the spacelike part of which can be identified with the magnetic field. This is all implied in the Lorentz transformation, which was studied in some detail in Chapter 9. Those readers who are interested in the underlying physics, or even just in a philosophical outlook on the subject, may benefit from following through this approach.

From Chapter 8, we know in principle how to convert, or translate, between (3+1)D objects and spacetime objects based on the grade of object concerned. However, it was emphasized that the choice of whether to use either an even or odd mapping must be made predominantly on physical rather than mathematical grounds. We will therefore have to answer this question for the electromagnetic source density, potential, field, and force in the light of the equations that govern them. But this is not the only issue; we have encountered at least one example of an equation, Maxwell's equation, that has alternative forms, namely Equations (5.7) and (5.9). This traces back to the fact that there is no unique way to form a multivector equation. A multivector equation such as $a + \mathbf{b} + \boldsymbol{C} = p + \mathbf{q} + \boldsymbol{R}$ simply requires that the terms of each grade on both sides of the equation be equal. Therefore, $a + \mathbf{b} - \boldsymbol{C} = p + \mathbf{q} - \boldsymbol{R}$ and $-a + 2\mathbf{b} + \boldsymbol{C} = -p + 2\mathbf{q} + \boldsymbol{R}$ are both equally valid alternative forms and so, in the end, it is a matter of convention which one is to be preferred. It is yet another nuance that the preferred form in spacetime may be different, or at least appear to be different, from the conventional one in (3+1)D. The final point to be considered is the effect that chosen metric signature has on how the equations appear. All of these things put together means that we must be careful not only about the correct assignment of the grades, but also of the signs attributed to various quantities. We have said that the underlying physics plays a key role in determining the appropriate grades, but when it comes to the question of signs, the overall consistency of an entire system of related equations becomes crucial.

11.1 THE VECTOR POTENTIAL AND SOME SPACETIME SPLITS

In the case of (3+1)D, all the equations that are traditionally required to express the relationship between the scalar and vector potentials Φ and $\mathbf{A}$ to the fields $\mathbf{E}$ and

B, together with a gauge condition, are summarized in Equation (5.32). We may restate these here as

$$\begin{aligned} &\partial_t \Phi + \boldsymbol{\nabla} \cdot \mathbf{A} = 0 \\ &\mathbf{E} = -\boldsymbol{\nabla}\Phi - \partial_t \mathbf{A} \\ &\boldsymbol{B} = \boldsymbol{\nabla} \wedge \mathbf{A} \end{aligned} \tag{11.1}$$

It was then shown in Equation (5.30) that these may be encapsulated in a single equation of the form

$$(\boldsymbol{\nabla} - \partial_t)\boldsymbol{A} = \boldsymbol{F} \tag{11.2}$$

where $\boldsymbol{F}$ is equal to $\mathbf{E} + \boldsymbol{B}$, representing the entire electromagnetic field due to the multivector potential $\boldsymbol{A} = -\Phi + \mathbf{A}$ alone; that is, no external fields are involved. We find the spacetime form of Equation (11.2) as follows. Recall first the notation introduced in Section 7.3.1 whereby in a chosen frame, here the $\boldsymbol{t}$-frame, any spacetime vector $\boldsymbol{u}$ may be written as $u_t \boldsymbol{t} + \underset{\sim}{\boldsymbol{u}}$ where the spatial vector $\underset{\sim}{\boldsymbol{u}}$ is orthogonal to the time vector $\boldsymbol{t}$. The spacetime split of $\boldsymbol{u}$ in the $\boldsymbol{t}$-frame is then just $u_t + \mathbf{u}$ where $\mathbf{u} = \underset{\sim}{\boldsymbol{u}}\boldsymbol{t}$. Although this is just the same thing as saying $\mathbf{u} = \boldsymbol{u} \wedge \boldsymbol{t}$, it is often more convenient. As before, we can use this as a means to convert any (3+1)D vector or paravector to a spacetime form. Using Equation (11.2), we apply this in turn to $\boldsymbol{\nabla} - \partial_t$, $-\Phi + \mathbf{A}$, and $\mathbf{E}$ so as to give

$$\begin{aligned} &(\boldsymbol{\nabla} - \partial_t)(-\Phi + \mathbf{A}) = \mathbf{E} + \boldsymbol{B} \\ \Leftrightarrow\ &(\underset{\sim}{\boldsymbol{\nabla}}\boldsymbol{t} - \partial_t)(-\Phi + \underset{\sim}{\boldsymbol{A}}\boldsymbol{t}) = \underset{\sim}{\boldsymbol{E}}\boldsymbol{t} + \boldsymbol{B} \end{aligned} \tag{11.3}$$

According to the rules for the translation of bivectors, we now treat $\boldsymbol{B}$ as a spacelike bivector, that is, the basis elements $\mathbf{yz}, \mathbf{zx}, \mathbf{xy}$ are replaced by $\boldsymbol{yz}, \boldsymbol{zx}, \boldsymbol{xy}$. The next step is to pre- and postmultiply each side with $\boldsymbol{t}$ to achieve

$$\boldsymbol{t}(\underset{\sim}{\boldsymbol{\nabla}}\boldsymbol{t} - \partial_t)(-\Phi + \underset{\sim}{\boldsymbol{A}}\boldsymbol{t})\boldsymbol{t} = \boldsymbol{t}(\underset{\sim}{\boldsymbol{E}}\boldsymbol{t} + \boldsymbol{B})\boldsymbol{t} \tag{11.4}$$

It is now possible to rearrange this equation by using the fact that $\boldsymbol{t}$ commutes with bivectors but anticommutes with any spatial vector, whereupon

$$\begin{aligned} &(\underset{\sim}{\boldsymbol{\nabla}} - \partial_t \boldsymbol{t})(-\Phi\boldsymbol{t} - \underset{\sim}{\boldsymbol{A}}) = \underset{\sim}{\boldsymbol{E}}\boldsymbol{t} - \boldsymbol{B} \\ &\Leftrightarrow \boldsymbol{\nabla}\boldsymbol{A} = \boldsymbol{E} - \boldsymbol{B} = \boldsymbol{F} \end{aligned} \tag{11.5}$$

The spacetime form of $\boldsymbol{A}$ is therefore given as

$$\boldsymbol{A} = -\Phi\boldsymbol{t} - \underset{\sim}{\boldsymbol{A}} \tag{11.6}$$

As we have already noted, $\boldsymbol{B}$ is a spacelike bivector indistinguishable from its (3+1)D bivector counterpart, and $\boldsymbol{E} = \underset{\sim}{\boldsymbol{E}}\boldsymbol{t} = \mathbf{E}$ is a timelike bivector rather than a vector. While a vector result might have been expected, it must be remembered from Section 8.2 that a (3+1)D vector may originate from either a vector or a timelike

bivector. But again, it is the underlying physics that dictates the outcome, and for reasons that will become clear from the discussion in Section 11.5.1, it must indeed be a bivector. In spacetime, the electromagnetic field $\boldsymbol{F}$ therefore comprises a timelike bivector $\boldsymbol{E}$ for the electric field and spacelike bivector $-\boldsymbol{B}$ that gives the magnetic field.

From the foregoing, we see that $\boldsymbol{F}$ and $\boldsymbol{A}$ have special spacetime splits. The split for $\boldsymbol{F}$ is evident from comparing Equations (11.2) and (11.5)

$$\underbrace{\boldsymbol{F} = \boldsymbol{E} - \boldsymbol{B}}_{\text{spacetime}} \leftrightarrow \underbrace{\boldsymbol{F} = \mathbf{E} + \boldsymbol{B}}_{\text{(3+1)D}} \tag{11.7}$$

The change of sign of $\boldsymbol{B}$ is an artifact of choosing $(-+++)$ as the metric signature and does not occur in the case of $(+---)$. We will review this particular point in Section 11.2.1. The different conventions associated with metric signature result in several such nuances, as indeed we shall see from time to time.

In the case of $\boldsymbol{A}$, we can find the spacetime split from Equation (11.6)

$$\begin{aligned} & \boldsymbol{A} = -\Phi\boldsymbol{t} - \underset{\sim}{\boldsymbol{A}} \\ \Leftrightarrow\ & \boldsymbol{A}(-\boldsymbol{t}) = (-\Phi\boldsymbol{t} - \underset{\sim}{\boldsymbol{A}})(-\boldsymbol{t}) \\ & = \Phi\boldsymbol{t}^2 + \underset{\sim}{\boldsymbol{A}}\boldsymbol{t} \\ & = -\Phi + \mathbf{A} \end{aligned} \tag{11.8}$$

so that by the process of *postmultiplication* by $-\boldsymbol{t}$,

$$\underbrace{\boldsymbol{A} = -\Phi\boldsymbol{t} - \underset{\sim}{\boldsymbol{A}}}_{\text{spacetime}} \leftrightarrow \underbrace{-\Phi + \mathbf{A}}_{\text{(3+1)D}} \tag{11.9}$$

Since $\boldsymbol{A}^2 = (\Phi + \underset{\sim}{\boldsymbol{A}})^2 (-\boldsymbol{t})^2 = -\left(\Phi^2 + \mathbf{A}^2\right)$ is negative, the spacetime potential $\boldsymbol{A}$ is a timelike vector. Note that in (3+1)D, we used the same label $\boldsymbol{A}$ for the multivector potential, that is to say $-\Phi + \mathbf{A}$, but there is no problem having the same label $\boldsymbol{A}$ for its spacetime counterpart. This observation also applies to the labels I, $\boldsymbol{B}$, $\boldsymbol{F}$, and $\boldsymbol{J}$, but the intention should always be clear from the context.

Although $\boldsymbol{\nabla} + \partial_t$, rather than $\boldsymbol{\nabla} - \partial_t$, is the (3+1)D counterpart of the spacetime vector derivative, Equation (11.2) has been manipulated so that $\boldsymbol{\nabla}\boldsymbol{A}$ features in Equation (11.5). As a consequence, we have this irregular spacetime split in which $-\boldsymbol{t}$ occurs as a postmultiplier rather than a premultiplier, and, similar to the fact that the spacetime form of $\boldsymbol{F}$ is $\boldsymbol{E} - \boldsymbol{B}$ rather than $\boldsymbol{E} + \boldsymbol{B}$, we also have $\boldsymbol{A} = -\Phi\boldsymbol{t} - \underset{\sim}{\boldsymbol{A}}$ rather than $\boldsymbol{A} = -\Phi\boldsymbol{t} + \underset{\sim}{\boldsymbol{A}}$.

In the $(+---)$ signature, however, we get

$$\begin{aligned} \mathbf{E} + \boldsymbol{B} &= (\boldsymbol{\nabla} - \partial_t)(-\Phi + \mathbf{A}) \\ &= (\underset{\sim}{\boldsymbol{\nabla}}\boldsymbol{t} - \partial_t)(-\Phi + \underset{\sim}{\boldsymbol{A}}\boldsymbol{t}) \\ &= (\underset{\sim}{\boldsymbol{\nabla}}\boldsymbol{t} - \partial_t)\boldsymbol{t}\boldsymbol{t}(-\Phi + \underset{\sim}{\boldsymbol{A}}\boldsymbol{t}) \\ &= \left(\underset{\sim}{\boldsymbol{\nabla}}\boldsymbol{t}^2 - \partial_t\boldsymbol{t}\right)(-\Phi\boldsymbol{t} + \boldsymbol{t}\underset{\sim}{\boldsymbol{A}}\boldsymbol{t}) \end{aligned}$$

$$\Leftrightarrow (-\underset{\sim}{\nabla} + \partial_t \boldsymbol{t})(\Phi \boldsymbol{t} + \underset{\sim}{\boldsymbol{A}}) = \boldsymbol{E} + \boldsymbol{B}$$

$$\Leftrightarrow \qquad \boldsymbol{\nabla} \boldsymbol{A} = \boldsymbol{E} + \boldsymbol{B} = \boldsymbol{F} \tag{11.10}$$

so that in this case, $\boldsymbol{A} = \Phi \boldsymbol{t} + \underset{\sim}{\boldsymbol{A}}$ and $\boldsymbol{F} = \boldsymbol{E} + \boldsymbol{B}$. The fact that in the $(+---)$ metric signature $\boldsymbol{A}$ turns out to be the negative of its definition in the $(-+++)$ case may seem somewhat odd, but again this is due to the fact that $\boldsymbol{\nabla} - \partial_t$ does not translate directly to the spacetime vector derivative. In addition, $\boldsymbol{\nabla}^2$ is given in the $(-+++)$ metric signature by $\nabla^2 - \partial_t^2$ (recall $\nabla^2 = \boldsymbol{\nabla}^2 = \partial_x^2 + \partial_y^2 + \partial_z^2$), while in the case of $(+---)$, it has the opposite sign. Given that $\boldsymbol{\nabla}^2$ is a scalar operator and $\boldsymbol{\nabla}^2 = \nabla^2 - \partial_t^2$, we can anticipate that in spacetime, the (3+1)D wave equation $\left(\boldsymbol{\nabla}^2 - \partial_t^2\right)\boldsymbol{A} = \boldsymbol{J}$ is simply replaced by $\boldsymbol{\nabla}^2 \boldsymbol{A} = \boldsymbol{J}$. Since $\boldsymbol{A}$ and $\boldsymbol{\nabla}^2$ both change signs between the two signatures, $\boldsymbol{\nabla}^2 \boldsymbol{A}$ turns out to be exactly the same. This in turn implies that the form of $\boldsymbol{J}$ must also be exactly the same in both metric signatures, and this indeed turns out to be the case as we shall see when we come to Maxwell's equation.

For reasons that were explained earlier on in Section 7.8, the vector derivative is sensitive to the choice of metric in a way that does not affect ordinary vectors. While it does have its own special spacetime split, it is often best to address equations as a whole to see how they translate between spacetime and (3+1)D. Even so, be wary of the fact that the final form of an equation may not necessarily be unique; for example, Equations (5.7) and (5.9) above are equivalent to

$$(\boldsymbol{\nabla} + \partial_t)(\mathbf{E} + \boldsymbol{B}) = \rho - \mathbf{J} \quad \text{(i)}$$
$$(\boldsymbol{\nabla} - \partial_t)(\mathbf{E} - \boldsymbol{B}) = \rho + \mathbf{J} \quad \text{(ii)}$$

While (i) is the conventional form, (ii) is rarely used but equally valid because it simply corresponds to inverting all the vectors in (i). Taking the sum and difference of (i) and (ii) yields a pair of equations for ρ and $\mathbf{J}$ separately, from which Maxwell's equations may be returned in their usual form simply by expressing $\boldsymbol{\nabla}$ as $\nabla\cdot + I\boldsymbol{\nabla}\times$. It is of little surprise then that in spacetime, we can have an equation involving $\boldsymbol{\nabla}$, with subtly different forms that depend on the choice of metric signature. It is also clear that here, the possibility of the spacetime split of $\boldsymbol{\nabla}$ resulting in either $\boldsymbol{\nabla} + \partial_t$ or $\boldsymbol{\nabla} - \partial_t$ goes beyond the rule for even and odd multivectors discussed in Section 8.4.4. Not only can we do a spacetime split on each side of an equation, as discussed in the introduction to this chapter, we can also modify the form of the equation so as to have a different arrangement of signs. The crucial point is that an equation must remain an equation. Even restricting ourselves to the $\boldsymbol{t}$-frame, we cannot regard spacetime splits as always having the rigid form associated with the mapping of basis elements that is depicted in Figure 8.1 when we deal with equations as a whole. While such rules may be directly applicable to simple vectors, introducing an irregular vector like $\boldsymbol{\nabla}$ causes problems when it appears on only one side of an equation where we have only a regular spacetime split on the other. In these situations, as already mentioned it is best to translate the equation as a whole rather than to try to tackle each side separately with different spacetime

splits. If one side of the equation is rendered into the desired form, then the form that results on the other side must be accepted as the consequence. Although variations from an expected form may give rise to concerns that some error has been made, it will soon be found that these variations are limited in practice to a few well-known cases. We have already encountered two here in the electromagnetic field and the electromagnetic potential, and, unsurprisingly, the electromagnetic source density will prove to be the third case.

11.2 MAXWELL'S EQUATIONS IN SPACETIME FORM

11.2.1 Maxwell's Free Space or Microscopic Equation

The normal (3+1)D form taken by Maxwell's equation in free space is, from Equation (5.10):

$$(\partial_t + \boldsymbol{\nabla})(\mathbf{E} + \boldsymbol{B}) = \rho - \mathbf{J} \tag{11.11}$$

where we normally use the symbols $\boldsymbol{F}$ and $\boldsymbol{J}$ for the multivectors $\mathbf{E} + \boldsymbol{B}$ and $\rho - \mathbf{J}$, respectively. The problem is to decide how this should be translated into a spacetime form. Considering that $\boldsymbol{B}$ is a bivector, its contribution $(\partial_t + \boldsymbol{\nabla})\boldsymbol{B}$ on the left-hand side of the equation does not tally with the spacetime split of the vector derivative of an even multivector (Equation 8.26), which would require the form $(-\partial_t + \boldsymbol{\nabla})\boldsymbol{B}$. If, on the other hand, the spacetime form of the electric field is a vector then we do not have this problem, but let us put that aside for the time being. While Equation (11.11) is the standard form, as we have just been discussing in the previous section, there is an alternative form:

$$(-\partial_t + \boldsymbol{\nabla})(\mathbf{E} - \boldsymbol{B}) = \rho + \mathbf{J} \tag{11.12}$$

But this is in an appropriate form to relate to the spacetime split of $\boldsymbol{\nabla}\boldsymbol{B}$, and so, provided that in agreement with Equation (11.5) above we also associate the spacetime form of $\mathbf{E}$ with a bivector, it may be concluded that a valid spacetime form of Maxwell's equation is to be found as follows:

$$\begin{aligned}
&(-\partial_t + \boldsymbol{\nabla})(\mathbf{E} - \boldsymbol{B}) = \rho + \mathbf{J} \\
\Leftrightarrow\ &\boldsymbol{t}(-\partial_t + \boldsymbol{\nabla})(\mathbf{E} - \boldsymbol{B}) = \rho\boldsymbol{t} + \boldsymbol{t}\mathbf{J} \\
\Leftrightarrow\ &(-\partial_t\boldsymbol{t} + \boldsymbol{t}\boldsymbol{\nabla})(\mathbf{E} - \boldsymbol{B}) = \rho\boldsymbol{t} + \boldsymbol{t}\mathbf{J} \\
\Leftrightarrow\ &(-\partial_t\boldsymbol{t} + \boldsymbol{t}\underset{\sim}{\boldsymbol{\nabla}}\boldsymbol{t})(\boldsymbol{E} - \boldsymbol{B}) = \rho\boldsymbol{t} + \boldsymbol{t}\underset{\sim}{\boldsymbol{J}}\boldsymbol{t} \\
\Leftrightarrow\ &(-\partial_t\boldsymbol{t} + \underset{\sim}{\boldsymbol{\nabla}})(\boldsymbol{E} - \boldsymbol{B}) = \rho\boldsymbol{t} + \underset{\sim}{\boldsymbol{J}} \\
\Leftrightarrow\ &\boldsymbol{\nabla}(\boldsymbol{E} - \boldsymbol{B}) = \rho\boldsymbol{t} + \underset{\sim}{\boldsymbol{J}} = \boldsymbol{J}
\end{aligned} \tag{11.13}$$

Once again, we have used the method of representing a (3+1)D vector as a timelike bivector, for example $\mathbf{J} = \underset{\sim}{\boldsymbol{J}}\boldsymbol{t}$ and $\boldsymbol{\nabla} = \underset{\sim}{\boldsymbol{\nabla}}\boldsymbol{t}$, so that the time vector can be eliminated, reducing them to the spatial vectors $\underset{\sim}{\boldsymbol{J}}$ and $\underset{\sim}{\boldsymbol{\nabla}}$. On the other hand, $\boldsymbol{E}$ is simply the timelike bivector corresponding to $\mathbf{E}$, so that we may simply write $\boldsymbol{E} = \mathbf{E}$ in the same sense that $E_x\boldsymbol{xt} + E_y\boldsymbol{yt} + E_z\boldsymbol{zt} = E_x\mathbf{x} + E_y\mathbf{y} + E_z\mathbf{z}$. Furthermore, the individual spacetime splits of $\boldsymbol{F} = \boldsymbol{E} - \boldsymbol{B}$ and $\boldsymbol{J} = \rho\boldsymbol{t} + \underset{\sim}{\boldsymbol{J}}$ associate correctly with the terms in Equation (11.12) rather than with the usual (3+1)D definitions $\boldsymbol{F} = \mathbf{E} + \boldsymbol{B}$ and $\boldsymbol{J} = \rho - \mathbf{J}$. Equation (11.13), however, may be recast in the form $(\boldsymbol{\nabla t})(\boldsymbol{t}(\boldsymbol{E} - \boldsymbol{B})\boldsymbol{t}) = -(\rho\boldsymbol{t} + \underset{\sim}{\boldsymbol{J}})\boldsymbol{t}$, which does reveal the proper association with the standard (3+1)D counterparts through

$$\begin{aligned} \boldsymbol{\nabla t} &= \partial_t + \boldsymbol{\nabla} \\ \boldsymbol{t}(\boldsymbol{E} - \boldsymbol{B})\boldsymbol{t} &= \mathbf{E} + \boldsymbol{B} \\ \text{and } -(\rho\boldsymbol{t} + \underset{\sim}{\boldsymbol{J}})\boldsymbol{t} &= \rho - \mathbf{J} \end{aligned} \tag{11.14}$$

As discussed in the preceding section, the special metric of spacetime creates this nuance, and, as may be expected, the result is different in the $(+---)$ signature where $\boldsymbol{F} = \boldsymbol{E} + \boldsymbol{B}$ is the same as the (3+1)D form, while on the other hand, $\boldsymbol{J} = \rho\boldsymbol{t} + \underset{\sim}{\boldsymbol{J}}$ still holds. The reason that $\boldsymbol{F}$ reverts to the form $\boldsymbol{E} + \boldsymbol{B}$ here is due to the fact that, as will be discussed toward the end of Section 11.8.3, the basis elements of the spatial bivectors in the two metric signatures differ through a change of sign. Discussion of the spacetime forms of $\boldsymbol{F}$, $\boldsymbol{E}$, and $\boldsymbol{B}$ arose in Section 11.1 and will come up again from different viewpoints in Sections 11.8.3 and 11.5.1.

Once again, we see that it is always the underlying physics that dictates the relationship between spacetime and (3+1)D, in this case because we had to find an arrangement that was compatible both with the spacetime split of the vector derivative and the detailed structure of Maxwell's equations in (3+1)D. The resultant conclusion from Equation (11.13) is that the spacetime form of Maxwell's equation for free space takes the astonishingly simple form

$$\boldsymbol{\nabla F} = \boldsymbol{J} \tag{11.15}$$

where the bivector field $\boldsymbol{F}$ and the vector source density $\boldsymbol{J}$ are given by

$$\begin{aligned} \boldsymbol{F} &= \boldsymbol{E} - \boldsymbol{B} \qquad (-+++) \\ \boldsymbol{F} &= \boldsymbol{E} + \boldsymbol{B} \qquad (+---) \\ \boldsymbol{J} &= \rho\boldsymbol{t} + \underset{\sim}{\boldsymbol{J}} \qquad \text{(both)} \end{aligned} \tag{11.16}$$

Although Maxwell's equation in spacetime is similar to the standard (3+1)D version as given in Equation (11.11), it is obvious that some care is required with comparisons, particularly in the case of our metric signature in which not only the form of $\boldsymbol{J}$ is contrary to expectation but also the sign of $\boldsymbol{B}$. We have now encountered all of the important irregular spacetime splits, and so it is worth summarizing them:

$$\begin{aligned} \boldsymbol{J} &= \rho\boldsymbol{t} + \underset{\sim}{\boldsymbol{J}} \quad \leftrightarrow \quad \rho - \mathbf{J} \\ \boldsymbol{F} &= \boldsymbol{E} - \boldsymbol{B} \quad \leftrightarrow \quad \mathbf{E} + \boldsymbol{B} \\ \boldsymbol{A} &= -\Phi\boldsymbol{t} - \underset{\sim}{\boldsymbol{A}} \quad \leftrightarrow \quad -\Phi\boldsymbol{t} + \mathbf{A} \end{aligned} \tag{11.17}$$

where $\mathbf{J} = \underset{\sim}{\boldsymbol{J}}\boldsymbol{t}$, $\mathbf{A} = \underset{\sim}{\boldsymbol{A}}\boldsymbol{t}$, and $\mathbf{E} = \boldsymbol{E} = \underset{\sim}{\boldsymbol{E}}\boldsymbol{t}$. The simple result is that in each case, the higher-grade (3+1)D term of each multivector changes sign in the spacetime form.

In this instance, we tackled the problem of translating Maxwell's equation into spacetime by seeking a form of the (3+1)D equation that was amenable to applying a normal spacetime split, that is to say by using $\boldsymbol{\nabla} - \partial_t$ rather than the more familiar $\boldsymbol{\nabla} + \partial_t$. In practice, however, it is often more straightforward to tackle the (3+1)D equation directly as suggested in the closing remarks of Section 11.1. For example,

$$\begin{aligned} & (\boldsymbol{\nabla} + \partial_t)(\mathbf{E} + \boldsymbol{B}) = \rho - \mathbf{J} \\ \Leftrightarrow \quad & (\underset{\sim}{\boldsymbol{\nabla}}\boldsymbol{t} + \partial_t)(\boldsymbol{E} + \boldsymbol{B}) = \rho - \underset{\sim}{\boldsymbol{J}}\boldsymbol{t} \\ \Leftrightarrow \quad & (\underset{\sim}{\boldsymbol{\nabla}}\boldsymbol{t} + \partial_t)(\boldsymbol{E} + \boldsymbol{B})\boldsymbol{t} = (\rho - \underset{\sim}{\boldsymbol{J}}\boldsymbol{t})\boldsymbol{t} \\ \Leftrightarrow \quad & (\underset{\sim}{\boldsymbol{\nabla}}\boldsymbol{t} + \partial_t)(-\boldsymbol{t})(\boldsymbol{E} - \boldsymbol{B}) = \rho\boldsymbol{t} + \underset{\sim}{\boldsymbol{J}} \\ \Leftrightarrow \quad & (\underset{\sim}{\boldsymbol{\nabla}} - \partial_t\boldsymbol{t})(\boldsymbol{E} - \boldsymbol{B}) = \rho\boldsymbol{t} + \underset{\sim}{\boldsymbol{J}} \\ \Leftrightarrow \quad & \boldsymbol{\nabla}(\boldsymbol{E} - \boldsymbol{B}) = \boldsymbol{J} \end{aligned} \tag{11.18}$$

offers an alternative route to Equation (11.13). The main problem with this sort of approach is to know where to start, but this skill will develop with experience.

11.2.2 Maxwell's Equations in Polarizable Media

While Equation (11.15) (or its (3+1)D version in Equation 11.11) may be referred to as the fundamental or free space Maxwell's equation, it can be recast for macroscopic media by means of the standard technique of partitioning charge and current, including the intrinsic current of magnets, into free and bound quantities. Now, it will be recalled from Section 5.9 that one of the shortcomings of the (3+1)D approach is that while this can be done, it cannot be encoded in quite the same neat sort of way. Although the effect of the bound sources is described by a term of the form $-\langle(\boldsymbol{\nabla} + \partial_t)\boldsymbol{Q}\rangle_{0,1}$ where $\boldsymbol{Q} = \mathbf{P} - \boldsymbol{M}$ is the electromagnetic polarization multivector, the equations apparently cannot be expressed in terms of $\boldsymbol{F}$, the auxiliary field $\boldsymbol{G}$ (recall $\boldsymbol{G} = \mathbf{D} + \boldsymbol{H}$), and the free sources, $\boldsymbol{J}_{\text{free}}$, alone. Without using a grade selection filter, we find that, for example, the time derivatives of both $\boldsymbol{B}$ and $\mathbf{D}$ are left over, as in Equation (5.69). We may hazard a guess that because the spacetime vector derivative includes the time derivative, it may help to sweep up these troublesome terms, and therefore we now revisit this question with the benefit of the spacetime toolset.

Let us first consider how to represent the (3+1)D bound source density $\boldsymbol{J}_{\text{bound}} = -\langle(\boldsymbol{\nabla} + \partial_t)\boldsymbol{Q}\rangle_{0,1}$ in spacetime form. Considering $-(\boldsymbol{\nabla} + \partial_t)\boldsymbol{Q}$ as a whole, since $\boldsymbol{Q}$ and $\boldsymbol{F}$ are of identical forms, that is to say vector plus bivector, its conversion to spacetime must follow the same lines as $-(\boldsymbol{\nabla} + \partial_t)\boldsymbol{F}$. We therefore have

$$\begin{aligned}(\nabla + \partial_t)(\mathbf{E} + \boldsymbol{B}) &\leftrightarrow \nabla(\boldsymbol{E} - \boldsymbol{B}) \\ \Leftrightarrow -(\nabla + \partial_t)(\mathbf{P} - \boldsymbol{M}) &\leftrightarrow -\nabla(\boldsymbol{P} + \boldsymbol{M})\end{aligned} \tag{11.19}$$

so that the spacetime form of $\boldsymbol{Q}$ is $\boldsymbol{P} + \boldsymbol{M}$, where $\boldsymbol{P}$ is the timelike bivector corresponding to $\mathbf{P}$ and the bivector $\boldsymbol{M}$ transfers straight across as a spacelike bivector apart from the change of sign that also happens with $\boldsymbol{B}$. However, although $\nabla(\boldsymbol{E} - \boldsymbol{B})$ results in a vector, since $\boldsymbol{P}$ and $\boldsymbol{M}$ are quite arbitrary fields we cannot necessarily say the same for $-\nabla(\boldsymbol{P} + \boldsymbol{M})$. But since $\boldsymbol{P}$ and $\boldsymbol{M}$ are bivectors, it must be the case that $-\nabla \cdot (\boldsymbol{P} + \boldsymbol{M})$ will provide the necessary vector part that ought to be associated with the bound current density vector, $\boldsymbol{J}_{\text{bound}}$. While in (3+1)D it is necessary to use $\langle\ \rangle_{0,1}$ as a grade selection filter to pick out the paravector part of $-(\nabla + \partial_t)\boldsymbol{Q}$ that corresponds to the bound source density, we find that in spacetime $-\nabla \cdot \boldsymbol{Q}$ does the same thing by selecting the vector part of $-\nabla(\boldsymbol{P} + \boldsymbol{M})$.

Going back to Equation (11.15), Maxwell's equation in free space, we may split the source density into free and bound contributions and write it as $\nabla \boldsymbol{F} = \boldsymbol{J}_{\text{free}} + \boldsymbol{J}_{\text{bound}}$. At this point, we note that we can also split the equation in two by using $\nabla \boldsymbol{F} = \nabla \cdot \boldsymbol{F} + \nabla \wedge \boldsymbol{F}$ and noting that the only vector contribution comes from $\nabla \cdot \boldsymbol{F}$ so that $\nabla \wedge \boldsymbol{F}$ must vanish. That is to say,

$$\begin{aligned}\nabla \cdot \boldsymbol{F} &= \boldsymbol{J}_{\text{free}} + \boldsymbol{J}_{\text{bound}} \\ \nabla \wedge \boldsymbol{F} &= 0\end{aligned} \tag{11.20}$$

Substituting $-\nabla \cdot \boldsymbol{Q}$ for $\boldsymbol{J}_{\text{bound}}$ then gives us

$$\begin{aligned}&\nabla \cdot \boldsymbol{F} = \boldsymbol{J}_{\text{free}} - \nabla \cdot \boldsymbol{Q} \\ \Leftrightarrow\ &\nabla \cdot (\boldsymbol{F} + \boldsymbol{Q}) = \boldsymbol{J}_{\text{free}} \\ \Leftrightarrow\ &\nabla \cdot \boldsymbol{G} = \boldsymbol{J}_{\text{free}}\end{aligned} \tag{11.21}$$

which defines the spacetime form of the auxiliary electromagnetic field bivector, $\boldsymbol{G}$, in exactly the same way as the (3+1)D form. Reuniting this with the homogeneous part of the equation, $\nabla \wedge \boldsymbol{F} = 0$, gives the final form for Maxwell's equations in polarizable media,

$$\begin{aligned}\nabla \cdot \boldsymbol{G} &= \boldsymbol{J}_{\text{free}} \\ \nabla \wedge \boldsymbol{F} &= 0 \\ \boldsymbol{G} &= \boldsymbol{F} + \boldsymbol{Q}\end{aligned} \tag{11.22}$$

where

$$\begin{aligned}\boldsymbol{F} + \boldsymbol{Q} &= \boldsymbol{E} - \boldsymbol{B} + \boldsymbol{P} + \boldsymbol{M} \\ &= (\boldsymbol{E} + \boldsymbol{P}) - (\boldsymbol{B} - \boldsymbol{M}) \\ &= \boldsymbol{D} - \boldsymbol{H}\end{aligned} \tag{11.23}$$

The closest comparable (3+1)D forms to these are to be found in Equations (5.64) and (5.63), in which specific grade selection filters were required in order to

invoke the roles of the simple inner and outer products that we have here. The spacetime bivector fields $\boldsymbol{D} - \boldsymbol{H}$ and $\boldsymbol{P} + \boldsymbol{M}$ relate to their (3+1)D counterparts $\mathbf{D} + \boldsymbol{H}$ and $\mathbf{P} - \boldsymbol{M}$ in the same way that $\boldsymbol{E} - \boldsymbol{B}$ relates to $\mathbf{E} + \boldsymbol{B}$. In each case, the timelike bivector becomes a vector, whereas the spacelike bivector stays as a bivector but changes sign.

It is unsurprising that introducing an auxiliary electromagnetic field $\boldsymbol{G}$ causes Maxwell's equation to split into two separate equations. The bound source density in the form of $-\boldsymbol{\nabla} \cdot \boldsymbol{Q}$ can only be taken together with $\boldsymbol{\nabla} \cdot \boldsymbol{F}$, leaving $\boldsymbol{\nabla} \wedge \boldsymbol{F}$ unaltered. There is no necessity to have two separate equations, however, for

$$\boldsymbol{\nabla F} = \boldsymbol{J}_{\text{free}} - \boldsymbol{\nabla} \cdot \boldsymbol{Q} \tag{11.24}$$

is still completely valid and, in principle, no less easy to work with than the established form (Equation 11.22).

The discussion leading up to Equation (11.22), however, illustrates an interesting point about the spacetime vector derivative in that for any multivector $\boldsymbol{U}$ in general, $\boldsymbol{\nabla} \cdot \boldsymbol{U}$ and $\boldsymbol{\nabla} \wedge \boldsymbol{U}$ both involve the time derivative, whereas in (3+1)D the time derivative is separate. It is therefore impossible to do anything with Equations (5.64) and (5.61) that would give similar results to Equations (11.22) and (11.24) without resorting to a grade selection filter. In the spacetime geometric algebra, this is completely unnecessary so that Maxwell's macroscopic equations are expressed in a very much more succinct way than their (3+1)D form.

11.3 CHARGE CONSERVATION AND THE WAVE EQUATION

In Section 5.6, we speculated whether it would be possible to express the law of charge conservation in terms of the vector derivative $\boldsymbol{\nabla}$ and the source density vector $\boldsymbol{J}$ alone. In the case of (3+1)D, we recovered the continuity equation by turning Maxwell's equation into a wave equation. But ρ and $\boldsymbol{J}$ appear separately in each line of Equation (5.28). They do not appear together in the form of the paravector $\rho - \mathbf{J}$ that represents the total source density $\boldsymbol{J}$. Let us therefore investigate what happens in spacetime by expanding $\boldsymbol{\nabla J}$:

$$\begin{aligned} \boldsymbol{\nabla J} &= (-\boldsymbol{t}\partial_t + \underset{\sim}{\boldsymbol{\nabla}})(\rho \boldsymbol{t} + \underset{\sim}{\boldsymbol{J}}) \\ &= \underbrace{-\partial_t \rho \boldsymbol{t}^2 + \underset{\sim}{\boldsymbol{\nabla}} \cdot \underset{\sim}{\boldsymbol{J}}}_{\substack{\text{scalar} \\ \boldsymbol{\nabla} \cdot \boldsymbol{J}}} + \underbrace{\underset{\sim}{\boldsymbol{\nabla}} \wedge \underset{\sim}{\boldsymbol{J}} - \boldsymbol{t}\partial_t \underset{\sim}{\boldsymbol{J}} + \underset{\sim}{\boldsymbol{\nabla}} \boldsymbol{t} \rho}_{\substack{\text{bivector} \\ \boldsymbol{\nabla} \wedge \boldsymbol{J}}} \\ \leftrightarrow \quad & \underbrace{\partial_t \rho + \boldsymbol{\nabla} \cdot \mathbf{J}}_{\text{scalar}} + \underbrace{\partial_t \mathbf{J} + \boldsymbol{\nabla} \rho + \boldsymbol{\nabla} \wedge \mathbf{J}}_{\text{vector+bivector}} \end{aligned} \tag{11.25}$$

The (3+1)D expression here is nearly, but not quite, identical to that found in Equation (5.27). The scalar terms corresponding to $\boldsymbol{\nabla} \cdot \boldsymbol{J}$ and its spacetime split are what we are looking for, whereas the remaining terms correspond to $\boldsymbol{\nabla} \wedge \boldsymbol{J}$. Using

much the same argument as we employed before, we can use $\boldsymbol{\nabla}^2 \boldsymbol{F} = \boldsymbol{\nabla} \boldsymbol{J}$ to show that $\boldsymbol{\nabla} \cdot \boldsymbol{J} = 0$ simply because $\boldsymbol{\nabla}^2$ is a scalar operator, implying that $\boldsymbol{\nabla} \boldsymbol{J}$ must be the same grade as $\boldsymbol{F}$, a bivector. There can therefore be no scalar term in $\boldsymbol{\nabla} \boldsymbol{J}$ and consequently $\boldsymbol{\nabla} \cdot \boldsymbol{J} = 0$ and $\boldsymbol{\nabla} \boldsymbol{J} = \boldsymbol{\nabla} \wedge \boldsymbol{J}$. As long as $\boldsymbol{J}$ includes all the sources, this leaves us with

$$\boldsymbol{\nabla}^2 \boldsymbol{F} = \boldsymbol{\nabla} \wedge \boldsymbol{J} \tag{11.26}$$

as the spacetime form of the electromagnetic wave equation. Now, again because $\boldsymbol{\nabla}^2$ is scalar, $\boldsymbol{\nabla}^2 \boldsymbol{F} = \left(\underset{\sim}{\nabla}^2 - \partial_t^2\right)(\boldsymbol{E} - \boldsymbol{B})$ must have the spacetime split $\left(\boldsymbol{\nabla}^2 - \partial_t^2\right)(\mathbf{E} - \boldsymbol{B})$. From Maxwell's equation, we might expect this to be in terms of $\mathbf{E} + \boldsymbol{B}$ rather than $\mathbf{E} - \boldsymbol{B}$, but the scalar nature of $\boldsymbol{\nabla}^2$ rules out any special form of split and so there is no flipping of sign with the split of $\boldsymbol{B}$. For this reason, we find the sign of $\boldsymbol{\nabla} \wedge \mathbf{J}$ in Equation (11.25) differs from that of Equation (5.27). Nevertheless, both approaches lead to Equation (5.28) for the conservation of charge together with the two separate wave equations for $\boldsymbol{E}$ and $\mathbf{B}$. The only essential differences are that in spacetime, the bivector $\boldsymbol{\nabla} \wedge \boldsymbol{J}$ neatly replaces $\partial_t \mathbf{J} + \boldsymbol{\nabla} \rho - \boldsymbol{\nabla} \wedge \mathbf{J}$ as the source in a single multivector wave equation for $\boldsymbol{F}$ and conservation of charge may be expressed simply as

$$\boldsymbol{\nabla} \cdot \boldsymbol{J} = 0 \tag{11.27}$$

In the case of polarizable media where, from Equation (11.22), we have $\boldsymbol{\nabla} \cdot \boldsymbol{G} = \boldsymbol{J}_f$, it follows from the identity $\boldsymbol{u} \cdot (\boldsymbol{u} \cdot \boldsymbol{V}) = 0$ that $\boldsymbol{\nabla} \cdot (\boldsymbol{\nabla} \cdot \boldsymbol{G}) = \boldsymbol{\nabla} \cdot \boldsymbol{J}_f = 0$ so that, as might be expected, Equation (11.27) now becomes $\boldsymbol{\nabla} \cdot \boldsymbol{J}_f = 0$. This is a statement of the separate conservation of free charge and, since total charge is conserved, we must also have conservation of bound charge with $\boldsymbol{\nabla} \cdot \boldsymbol{J}_b = 0$.

11.4 PLANE ELECTROMAGNETIC WAVES

It was pointed out in Section 5.5 that one of the minor shortcomings of a (3+1)D treatment of plane waves is the inability to express the phase factor $\omega t - \mathbf{k} \cdot \mathbf{r}$ as an inner product directly between the paravectors $\boldsymbol{R}$ and $\boldsymbol{K}$ where $\boldsymbol{R} = t + \mathbf{r}$ is the independent variable representing time and position and $\boldsymbol{K} = \omega + \mathbf{k}$ represents the combined frequency and wave vector of the wave. It was therefore necessary to treat the time-dependent part of the phase factor separately from the inner product between $\mathbf{r}$ and $\mathbf{k}$. In light of experience so far, it would be reasonable to investigate the spacetime approach in the hope of resolving this issue.

We start from the spacetime history of the stationary observer located at $\mathbf{r}$, which we can take to be in the form $\boldsymbol{r} = t\boldsymbol{t} + \underset{\sim}{\boldsymbol{r}}$ where $\underset{\sim}{\boldsymbol{r}}\boldsymbol{t} = \mathbf{r}$. We then define the spacetime wave vector for our waves as $\boldsymbol{k} = \omega \boldsymbol{t} + \underset{\sim}{\boldsymbol{k}}$ where $\underset{\sim}{\boldsymbol{k}}\boldsymbol{t} = \mathbf{k}$ is the relative wave vector in the $\boldsymbol{t}$-frame. The pair of vectors $\boldsymbol{r}$ and $\boldsymbol{k}$ therefore corresponds to the original pair of paravectors $\boldsymbol{R}$ and $\boldsymbol{K}$, which are in fact their $\boldsymbol{t}$-frame spacetime splits. Because $\boldsymbol{k}$ includes the time dependence, waves having the wave vectors $\boldsymbol{k}$ and $-\boldsymbol{k}$ are both

valid solutions that propagate *in the same direction*, and so we must allow the possibility of ω being negative, the significance of which will emerge later.

The inner product between $\boldsymbol{k}$ and $\boldsymbol{r}$ gives $\boldsymbol{k}\cdot\boldsymbol{r} = (\omega\boldsymbol{t} + \underset{\sim}{\boldsymbol{k}})\cdot(t\boldsymbol{t} + \underset{\sim}{\boldsymbol{r}}) = \underset{\sim}{\boldsymbol{k}}\cdot\underset{\sim}{\boldsymbol{r}} - \omega t = \mathbf{k}\cdot\mathbf{r} - \omega t$, which, allowing for the sign, is exactly the form of phase factor that we require. The spacetime vectors $\boldsymbol{r}$ and $\boldsymbol{k}$ therefore simply replace their (3+1)D paravector counterparts $\boldsymbol{R}$ and $\boldsymbol{K}$, and likewise, the spacetime bivector $\boldsymbol{F}_0 = \boldsymbol{E}_0 - \boldsymbol{B}_0$ replaces the original (3+1)D vector + bivector field, giving us

$$\boldsymbol{F}(\boldsymbol{r},t) = \boldsymbol{F}_0 e^{-I\boldsymbol{k}\cdot\boldsymbol{r}} \tag{11.28}$$

Since $\boldsymbol{k}\cdot\boldsymbol{r}$ is a scalar, the phase factor $e^{-I\boldsymbol{k}\cdot\boldsymbol{r}}$ will evaluate to $\alpha + I\beta$ where $\alpha = \cos(\boldsymbol{k}\cdot\boldsymbol{r}) = \cos(\omega t - \mathbf{k}\cdot\mathbf{r})$ and $\beta = \sin(-\boldsymbol{k}\cdot\boldsymbol{r}) = \sin(\omega t - \mathbf{k}\cdot\mathbf{r})$. Also, since I commutes with bivectors, $\alpha + I\beta$ and $\boldsymbol{F}_0$ will commute so that it does not matter whether we write $e^{-I\boldsymbol{k}\cdot\boldsymbol{r}}\boldsymbol{F}_0$ or $\boldsymbol{F}_0 e^{-I\boldsymbol{k}\cdot\boldsymbol{r}}$. Apart from the time dependence now being implicit in $\boldsymbol{k}\cdot\boldsymbol{r}$, there is clearly very little change from (3+1)D, and it will be immediately obvious that $e^{-I\boldsymbol{k}\cdot\boldsymbol{r}}$ defines a traveling plane that is wave directly comparable to $e^{I(\omega t - \mathbf{k}\cdot\mathbf{r})}$.

We now have to find $\boldsymbol{\nabla}\boldsymbol{F}$. Cast in spacetime form, the discussion concerning Equation (5.19) is still valid so that we may simply restate it as

$$\boldsymbol{\nabla} e^{-I\boldsymbol{k}\cdot\boldsymbol{r}} = -I\boldsymbol{k}\boldsymbol{\nabla} e^{-I\boldsymbol{k}\cdot\boldsymbol{r}} \tag{11.29}$$

The major change that arises in the spacetime version goes almost unnoticed, for time dependence is now implicit as a result of time being embodied within both $\boldsymbol{k}\cdot\boldsymbol{r}$ and the spacetime derivative. Noting that for the purposes of differentiation it will be easier to work with $\boldsymbol{F}$ written as $e^{-I\boldsymbol{k}\cdot\boldsymbol{r}}\boldsymbol{F}_0$, from this result we find

$$\begin{aligned}\boldsymbol{\nabla}\boldsymbol{F} &= \boldsymbol{\nabla}\left(e^{-I\boldsymbol{k}\cdot\boldsymbol{r}}\boldsymbol{F}_0\right)\\ &= \boldsymbol{\nabla}\left(e^{-I\boldsymbol{k}\cdot\boldsymbol{r}}\right)\boldsymbol{F}_0\\ &= -I\boldsymbol{k}e^{-I\boldsymbol{k}\cdot\boldsymbol{r}}\boldsymbol{F}_0\\ &= -I\boldsymbol{k}\boldsymbol{F}\end{aligned} \tag{11.30}$$

which we may then apply directly Maxwell's equation in an empty space (Equation 11.15), much as we did in the case of (3+1)D:

$$\begin{aligned}&\boldsymbol{\nabla}\boldsymbol{F} = -I\boldsymbol{k}\boldsymbol{F} = 0\\ &\Rightarrow (\omega\boldsymbol{t} + \underset{\sim}{\boldsymbol{k}})(\boldsymbol{E}_0 - \boldsymbol{B}_0)e^{-I\boldsymbol{k}\cdot\boldsymbol{r}} = 0\\ &\Leftrightarrow (\omega\boldsymbol{t} + \underset{\sim}{\boldsymbol{k}})(\boldsymbol{E}_0 - \boldsymbol{B}_0) = 0\\ &\Leftrightarrow \underbrace{\underset{\sim}{\boldsymbol{k}}\cdot\boldsymbol{E}_0}_{\substack{\text{timelike}\\ \text{vector}}} + \underbrace{\omega\boldsymbol{t}\cdot\boldsymbol{E}_0 - \underset{\sim}{\boldsymbol{k}}\cdot\boldsymbol{B}_0}_{\substack{\text{spacelike}\\ \text{vector}}} + \underbrace{\underset{\sim}{\boldsymbol{k}}\wedge\boldsymbol{E}_0 - \omega\boldsymbol{t}\wedge\boldsymbol{B}_0}_{\substack{\text{timelike}\\ \text{trivector}}} - \underbrace{\underset{\sim}{\boldsymbol{k}}\wedge\boldsymbol{B}_0}_{\substack{\text{spacelike}\\ \text{trivector}}} = 0\end{aligned} \tag{11.31}$$

The second step requires us to note that $e^{-I\boldsymbol{k}\cdot\boldsymbol{r}}$ has an inverse, which we know is in fact $e^{+I\boldsymbol{k}\cdot\boldsymbol{r}}$. The grade of each resulting expression here may be determined by the step-up and step-down rules for inner and outer products, while the timelike or spacelike character is most easily verified by substituting typical basis elements for $\underset{\sim}{\boldsymbol{k}}$, $\boldsymbol{E}_0$, and $\boldsymbol{B}_0$, for example, $\boldsymbol{x}$ for $\boldsymbol{k}$, $\boldsymbol{xt}$ or $\boldsymbol{yt}$ for $\boldsymbol{E}_0$, and $\boldsymbol{xy}$ or $\boldsymbol{yz}$ for $\boldsymbol{B}_0$. Now since the entire expression for $\boldsymbol{\nabla F}$ must vanish, the terms of each grade must separately vanish, as must their separate timelike and spacelike parts:

$$\begin{aligned}
\underset{\sim}{\boldsymbol{k}}\cdot\boldsymbol{E}_0 = 0 &\Rightarrow \mathbf{k}\cdot\mathbf{E}_0 = 0 \\
\underset{\sim}{\boldsymbol{k}}\wedge\boldsymbol{E}_0 - \omega\boldsymbol{t}\boldsymbol{B}_0 = 0 &\Rightarrow \omega\boldsymbol{B}_0 = -\boldsymbol{t}\left(\underset{\sim}{\boldsymbol{k}}\wedge\boldsymbol{E}_0\right) \\
&\Rightarrow \omega\boldsymbol{B}_0 = \mathbf{k}\wedge\mathbf{E}_0 \\
&\Rightarrow \omega\mathbf{B}_0 = \mathbf{k}\times\mathbf{E_0} \\
\omega\boldsymbol{t}\boldsymbol{E}_0 + \underset{\sim}{\boldsymbol{k}}\cdot\boldsymbol{B}_0 = 0 &\Rightarrow \omega\boldsymbol{E}_0 = \boldsymbol{t}\left(\underset{\sim}{\boldsymbol{k}}\cdot\boldsymbol{B}_0\right) \\
&\Rightarrow \omega\mathbf{E}_\mathrm{o} = -\mathbf{k}\times\mathbf{B}_0 \\
\underset{\sim}{\boldsymbol{k}}\wedge\boldsymbol{B}_0 = 0 &\Rightarrow \mathbf{k}\cdot\mathbf{B}_0 = 0
\end{aligned} \tag{11.32}$$

In reaching these results, it is necessary to expand inner and outer products using the rules of Equation (4.6) and then to manipulate the time vector in order to split all the variables into their (3+1)D counterparts. For example,

$$\begin{aligned}
\boldsymbol{t}\left(\underset{\sim}{\boldsymbol{k}}\wedge\boldsymbol{E}_0\right) &= \tfrac{1}{2}\boldsymbol{t}\left(\underset{\sim}{\boldsymbol{k}}\boldsymbol{E}_0 + \boldsymbol{E}_0\underset{\sim}{\boldsymbol{k}}\right) \\
&= \tfrac{1}{2}\left(-(\underset{\sim}{\boldsymbol{k}}\boldsymbol{t})\boldsymbol{E}_0 + \boldsymbol{E}_0(\underset{\sim}{\boldsymbol{k}}\boldsymbol{t})\right) \\
&= -\tfrac{1}{2}\left(\mathbf{k}\mathbf{E}_0 - \mathbf{E}_0\mathbf{k}\right) \\
&= -\mathbf{k}\wedge\mathbf{E}_0
\end{aligned}$$

Having reached this point, the rest of the solution is as in the case of (3+1)D, the main difference being that in $-\boldsymbol{k}\cdot\boldsymbol{r}$, we have found the simplest way possible of expressing the phase factor. The direction of propagation and the frequency are given by $\boldsymbol{k}$, but note that $I\boldsymbol{k}$ is a trivector so that, unlike $I\mathbf{k}$, it does not represent a surface parallel to the wavefront.

As may be expected, by eliminating either one of $\mathbf{E}_0$ or $\mathbf{B}_0$ from Equation (11.32), we find $\omega^2 = \underset{\sim}{\boldsymbol{k}}^2 = \mathbf{k}^2$. This is highly significant in spacetime terms since the implication is that $\boldsymbol{k}$ must be a null vector, in effect demonstrating that any electromagnetic wave in free space must travel along the light cone originating from its source. The inner product of $\boldsymbol{k}$ with the observer's history $\boldsymbol{r}$ will determine how the observer sees the traveling wavefronts that are determined by any fixed value of $\boldsymbol{k}\cdot\boldsymbol{r}$. Since $\boldsymbol{r}$ can represent any observer's history, we can deduce from this the Doppler shift for an observer moving relative to the source, as discussed below.

Let us return to the description of a plane wave in the form $\boldsymbol{F} = (\alpha + I\beta)\boldsymbol{F}_0$ where $\alpha + I\beta = e^{-I\boldsymbol{k}\cdot\boldsymbol{r}}$, which is much the same as we had in Equation (5.24) except that it is now convenient to have the factor $\alpha + I\beta$ on the left-hand side of the bivector, with which it conveniently commutes. Since the scalars α and β vary as the cosine and sine of $\omega t - \mathbf{k}\cdot\mathbf{r}$ respectively, what is at one instant a timelike bivector

becomes a spacelike bivector one quarter of a cycle later, and vice versa. But since we have $\boldsymbol{\nabla F} = (\boldsymbol{\nabla}(\alpha + I\beta))\boldsymbol{F}_0 = -I\boldsymbol{k}(\alpha + I\beta)\boldsymbol{F}_0 = 0$, we must also have $\boldsymbol{kF}_0 = 0$. This puts a considerable constraint on the options for choosing the bivector $\boldsymbol{F}_0$ and, since $\boldsymbol{k}$ is null, we may guess that it must be of the form $-\boldsymbol{k}\underset{\sim}{\boldsymbol{f}}$ where $\underset{\sim}{\boldsymbol{f}}$ is any vector that is orthogonal to the $\boldsymbol{k} \wedge \boldsymbol{t}$ plane (since a null vector appears to be orthogonal to itself, it is awkward to specify something as being orthogonal to a null vector). This means that $-\boldsymbol{k}\underset{\sim}{\boldsymbol{f}}$ will be a bivector with the required property since $\boldsymbol{kF}_0 = -\boldsymbol{k}^2\underset{\sim}{\boldsymbol{f}} = 0$. For example, taking $\boldsymbol{x}$ as an arbitrary propagation direction, we have $\boldsymbol{k} = \omega(\underset{\sim}{\boldsymbol{t}} + \boldsymbol{x})$ so that $\underset{\sim}{\boldsymbol{f}}$ must be orthogonal to $\boldsymbol{k} \wedge \boldsymbol{t} = \omega\boldsymbol{xt}$, that is to say it must take the form $f_y\boldsymbol{y} + f_z\boldsymbol{z}$. Now, observe that $\boldsymbol{kxt} = \omega(\boldsymbol{t} + \boldsymbol{x})\boldsymbol{xt} = \omega(\boldsymbol{t} + \boldsymbol{x}) = \boldsymbol{k}$. Putting all this together, we obtain

$$\begin{aligned}
\boldsymbol{F}_0 &= -\boldsymbol{k}\underset{\sim}{\boldsymbol{f}} \\
&= -\boldsymbol{k}\left(f_y\boldsymbol{y} + f_z\boldsymbol{z}\right) \\
&= -\boldsymbol{k}\left(f_y + f_z I\boldsymbol{xt}\right)\boldsymbol{y} \\
&= -\left(f_y\boldsymbol{k} + If_z\boldsymbol{k}\right)\boldsymbol{y} \\
&= -\left(f_y + If_z\right)\boldsymbol{ky} \\
&= \left(f_y + If_z\right)(\boldsymbol{yt} - \boldsymbol{xy}) \qquad (11.33) \\
\Leftrightarrow (\alpha + I\beta)\boldsymbol{F}_0 &= (\alpha + I\beta)\left(f_y + If_z\right)(\boldsymbol{yt} - \boldsymbol{xy}) \\
\Leftrightarrow \qquad \boldsymbol{F} &= \underbrace{(\alpha + I\beta)\left(f_y + If_z\right)\boldsymbol{yt}}_{\boldsymbol{E}} - \underbrace{(\alpha + I\beta)\left(f_y + If_z\right)\boldsymbol{xy}}_{\boldsymbol{B}}
\end{aligned}$$

If we replace I with the imaginary unit j here, it will be appreciated that $(\alpha + \mathrm{j}\beta)\left(f_y + \mathrm{j}f_z\right)$ is simply complex multiplication in which the first complex term rotates the angle of the second, for example, $(\alpha + \mathrm{j}\beta)\left(f_y + \mathrm{j}f_z\right) = f_y' + \mathrm{j}f_z'$. In terms of geometric algebra, we have exactly the same behavior (in fact an isomorphism) but with the result $f_y' + If_z'$, that is to say, the polarization of the wave is rotated from $\left(f_y + If_z\right)(\boldsymbol{yt} - \boldsymbol{xy})$ to $\left(f_y' + If_z'\right)(\boldsymbol{yt} - \boldsymbol{xy})$. For example, if $f_z = 0$ but then after the phase rotation $f_y' = 0$, the polarization of $\boldsymbol{F}$ changes from $\boldsymbol{yt} - \boldsymbol{xy}$ to $I(\boldsymbol{yt} - \boldsymbol{xy}) = \boldsymbol{zt} + \boldsymbol{zx}$, which is the same as rotating $\boldsymbol{y}$ into $\boldsymbol{z}$. The significance of allowing ω to be negative is that by replacing $\boldsymbol{k}$ with $-\boldsymbol{k}$, we create a wave polarized in the opposite sense because it has the effect of changing the sign of β, but not α, so that the factor $\alpha + I\beta$ will now rotate the polarization of the wave in the opposite direction.

So far, there is nothing new here compared with the (3+1)D treatment in Section 5.5; nevertheless, the comparison of methods is interesting in its own right. Plane waves are just one possible solution to Maxwell's equation in free space. It is therefore worth considering the broader implications of the simple differential equation, $\boldsymbol{\nabla F} = 0$, that lies at the heart of it. The solutions to this equation generate the class of meromorphic functions, so that in general, $\boldsymbol{F}(\boldsymbol{r})$ must be identifiable with some meromorphic function. From Equations (11.29) and (11.31), meromorphic functions

clearly include $\boldsymbol{U}e^{-I\boldsymbol{k}\cdot\boldsymbol{r}}$ and its derivatives provided that $\boldsymbol{U}$ and $\boldsymbol{k}$ are constant and $\boldsymbol{Uk}$ is null. In 2D, the meromorphic functions belong to the analytic functions, while in 3D, they obey the familiar source-free vector field condition, that is to say, they have zero divergence and curl except where sources are present. This we already know is equivalent to writing $\nabla\mathbf{f} = 0$ in the form $\nabla\cdot\mathbf{f} + I\nabla\times\mathbf{f} = 0$. In spite of this achievement, there has been no need of the separate concept of complex numbers, which is otherwise practically inescapable in electromagnetic theory.

Finally, in the discussion of Equations (5.47)–(5.51), we interpreted $\frac{1}{2}\boldsymbol{F}\boldsymbol{F}^{\dagger}$ as a multivector $\mathcal{E} + \boldsymbol{g}$ that represents the energy and momentum densities of an electromagnetic wave. Surprisingly, this does not transpose directly to spacetime in the form $\frac{1}{2}\boldsymbol{F}\boldsymbol{F}^{\dagger}$ since $\boldsymbol{F}^{\dagger}$ is equal to $-\boldsymbol{F}$ (see Exercise 11.9.1). In addition, since, as discussed above, $\boldsymbol{F}$ is of the form $\boldsymbol{k}\underset{\sim}{\boldsymbol{f}}$, we find $\boldsymbol{F}\boldsymbol{F}^{\dagger} = \boldsymbol{k}\underset{\sim}{\boldsymbol{f}}\underset{\sim}{\boldsymbol{f}}\boldsymbol{k} = \underset{\sim}{\boldsymbol{f}}^{2}\boldsymbol{k}^{2} = 0$ so that, for plane waves at least, $\boldsymbol{F}$ itself is null. This is not particularly strange because $\boldsymbol{B}$ is determined by $\boldsymbol{E}$, and being bivectors of different sorts, one of which has a negative square and the other positive (for example, $\boldsymbol{Fyt}$ and $-\boldsymbol{Fxy}$), it is easily verified that $\boldsymbol{Fyt} - \boldsymbol{Fxy}$ is null. However, we find the required result in the form $\frac{1}{2}\boldsymbol{F}\boldsymbol{t}\boldsymbol{F}^{\dagger} = \mathcal{E}\boldsymbol{t} + \underset{\sim}{\boldsymbol{g}}$ with $\mathcal{E} = \frac{1}{2}\left(\boldsymbol{E}^{2} - \boldsymbol{B}^{2}\right)$ and $\underset{\sim}{\boldsymbol{g}}\boldsymbol{t} = \boldsymbol{E}\wedge\boldsymbol{B} = \mathbf{g} = \mathbf{E}\times\mathbf{B}$. Although this looks like a sort of spacetime split, it has deeper significance [27, section 7.2.3, pp. 237–238].

11.5 TRANSFORMATION OF THE ELECTROMAGNETIC FIELD

As we have seen in the earlier sections of this chapter, the electromagnetic field is represented in spacetime by the bivector field $\boldsymbol{F} = \boldsymbol{E} - \boldsymbol{B}$, where $\boldsymbol{E}$ represents the timelike bivector part of $\boldsymbol{F}$ and $-\boldsymbol{B}$ represents the spacelike bivector part. The split of a general bivector into timelike and spacelike parts, however, is a frame-dependent process since, as we have noted earlier on in Section 9.5, the two mix under a change of frame. The electric and magnetic fields are therefore always frame-dependent quantities. This conclusion tallies with the observed physical properties, for given either a pure electric or magnetic field in one frame, we can always find some mix of magnetic and electric fields in another, as Faraday's law of induction and Maxwell's fourth equation clearly demonstrate. We therefore now go on to consider the general spacetime split of the electromagnetic field, the effects of a Lorentz transformation on Maxwell's equation as a whole, and the detail of how $\boldsymbol{F}$ itself behaves under a Lorentz transformation.

11.5.1 A General Spacetime Split for F

We have found that, in our metric signature, the spacetime electromagnetic field $\boldsymbol{F}$ comprises timelike and spacelike parts $\boldsymbol{E}$ and $\boldsymbol{B}$ in the form $\boldsymbol{F} = \boldsymbol{E} - \boldsymbol{B}$, which has a generic spacetime split back to the standard (3+1)D vector plus bivector form $\boldsymbol{F} = \mathbf{E} + \boldsymbol{B}$ (Equation 11.7). It will clearly be useful to have a formal way of separating $\boldsymbol{F}$ into its timelike and spacelike bivector parts in any given frame, that is to say, the observed electric and magnetic fields.

To find a way of achieving this, note that in any frame, say the $\boldsymbol{\theta}$-frame, the time vector $\boldsymbol{\theta}$ commutes with all spacelike bivectors, such as $\boldsymbol{B}$, whereas it anticommutes with all the timelike sort, such as $\boldsymbol{E}$. For example, in the $\boldsymbol{t}$-frame, $\boldsymbol{t}(\boldsymbol{xy}) = (\boldsymbol{xy})\boldsymbol{t}$ whereas $\boldsymbol{t}(\boldsymbol{xt}) = -(\boldsymbol{xt})\boldsymbol{t}$. For any bivector $\boldsymbol{U}$ that is either purely spacelike or timelike in the $\boldsymbol{\theta}$-frame, we therefore have $\boldsymbol{\theta U \theta} = \pm \boldsymbol{U}$, with the sign being negative if $\boldsymbol{U}$ is purely spacelike and, conversely, positive if it is purely timelike. We may therefore recover $\boldsymbol{E}$ and $\boldsymbol{B}$ from $\boldsymbol{F}$ in the $\boldsymbol{\theta}$-frame as follows:

$$\begin{aligned} \mathbf{E} = \boldsymbol{E} &= \tfrac{1}{2}((\boldsymbol{E}-\boldsymbol{B})+(\boldsymbol{E}+\boldsymbol{B})) = \tfrac{1}{2}(\boldsymbol{F}+\boldsymbol{\theta F \theta}) = \boldsymbol{\theta}(\boldsymbol{F}\cdot\boldsymbol{\theta}) \\ \boldsymbol{B} &= \tfrac{1}{2}((\boldsymbol{E}-\boldsymbol{B})-(\boldsymbol{E}+\boldsymbol{B})) = -\tfrac{1}{2}(\boldsymbol{F}-\boldsymbol{\theta F \theta}) = \boldsymbol{\theta}(\boldsymbol{F}\wedge\boldsymbol{\theta}) \end{aligned} \tag{11.34}$$

This is therefore just a form of the split of any bivector into timelike and spacelike parts quoted in Equation (7.31). From it, we may construct a spacetime split for $\boldsymbol{F}$ in any frame:

$$\underbrace{\boldsymbol{F}}_{\text{spacetime}} \leftrightarrow \underbrace{\boldsymbol{\theta F \theta} = \mathbf{E} + \boldsymbol{B} = \mathbf{E} + I\mathbf{B}}_{\substack{\text{(3+1)D, relative} \\ \text{to the } \boldsymbol{\theta}\text{-frame}}} \tag{11.35}$$

If $\boldsymbol{F}$ is already expressed in the $\boldsymbol{\theta}$-frame, all that this split will do is to invert the sign of the spacelike bivector. How useful it will prove to be depends on the form in which $\boldsymbol{F}$ is given. For example, if the $\boldsymbol{t}'$-frame travels with velocity $v\boldsymbol{x}$ with respect to the $\boldsymbol{t}$-frame then, as per Equation (9.17), $\boldsymbol{t} = \gamma(\boldsymbol{t}' - v\boldsymbol{x}')$ where $\boldsymbol{x}' = \gamma(\boldsymbol{x} + v\boldsymbol{t})$ is the $\boldsymbol{t}'$-frame axis corresponding to $\boldsymbol{x}$. Suppose we have an electromagnetic field $\boldsymbol{F} = E\boldsymbol{yt}$ as seen in the $\boldsymbol{t}$-frame. In (3+1)D, this corresponds to a pure electric field of magnitude E along the y-axis. According to Equation (11.35), the spacetime split of $\boldsymbol{F}$ in the $\boldsymbol{t}'$-frame will then be

$$\begin{aligned} \mathbf{E}' + \boldsymbol{B}' &= \boldsymbol{t}'(E\boldsymbol{yt})\boldsymbol{t}' \\ &= \boldsymbol{t}'(E\boldsymbol{y}\gamma(\boldsymbol{t}' - v\boldsymbol{x}'))\boldsymbol{t}' \\ &= \gamma E\boldsymbol{yt}' + \gamma v E\boldsymbol{x}'\boldsymbol{t}'\boldsymbol{yt}' \\ &= \underbrace{\gamma E\mathbf{y}}_{\substack{\text{electric} \\ \text{field}}} + \underbrace{\gamma v E\mathbf{xy}}_{\substack{\text{magnetic} \\ \text{field}}} \end{aligned} \tag{11.36}$$

Here the split effectively only serves to change the sign of $\boldsymbol{B}$, for it is the change of basis vectors that actually splits the electric field into a vector plus a bivector. But on the other hand, if $\boldsymbol{F}$ is given in some basis-free form, then this does not apply. Instead, what we find is the route that actually led to the spacetime form of $(\boldsymbol{\nabla} - \partial_t)(-\Phi + \mathbf{A}) = \mathbf{E} + \boldsymbol{B}$. If $\mathbf{E} + \boldsymbol{B} = \boldsymbol{tFt}$, it then follows that $\boldsymbol{F} = \boldsymbol{t}(\mathbf{E} + \boldsymbol{B})\boldsymbol{t} = \boldsymbol{t}(\boldsymbol{\nabla} - \partial_t)(-\Phi + \mathbf{A})\boldsymbol{t}$, which takes us directly to Equation (11.4). As another example, we will find that the electromagnetic field at $\boldsymbol{r}$ due to a point charge moving with constant proper velocity $\boldsymbol{\upsilon}$ depends on $\boldsymbol{r} \wedge \boldsymbol{\upsilon}$. Our spacetime split for this term in the $\boldsymbol{t}$-frame will then be $\boldsymbol{t}(\boldsymbol{r} \wedge \boldsymbol{\upsilon})\boldsymbol{t}$. Using the usual $\boldsymbol{t}'$-frame

spacetime split $-\boldsymbol{tr} = t + \mathbf{r}$ for $\boldsymbol{r}$, *but a modified one*, $-\mathfrak{v}\boldsymbol{t} = \gamma(1 - \mathbf{v})$ for $\mathfrak{v}$, we find a very useful result:

$$\begin{aligned}
\boldsymbol{t}(\boldsymbol{r} \wedge \mathfrak{v})\boldsymbol{t} &= \boldsymbol{t}(\boldsymbol{r}\mathfrak{v} - \boldsymbol{r} \cdot \mathfrak{v})\boldsymbol{t} \\
&= \boldsymbol{tr}\mathfrak{v}\boldsymbol{t} + \boldsymbol{r} \cdot \mathfrak{v} \\
&= (-\boldsymbol{tr})(-\mathfrak{v}\boldsymbol{t}) + \boldsymbol{r} \cdot \mathfrak{v} \\
&= (t + \mathbf{r})\gamma(1 - \mathbf{v}) + \boldsymbol{r} \cdot \mathfrak{v} \\
&= \underbrace{\gamma(\mathbf{r} - \mathbf{v}t)}_{\text{vector}} + \underbrace{\gamma(\mathbf{r} \wedge \mathbf{v})}_{\text{bivector}} + \underbrace{\gamma(t - \mathbf{r} \cdot \mathbf{v}) + \boldsymbol{r} \cdot \mathfrak{v}}_{\text{scalar terms}}
\end{aligned} \tag{11.37}$$

The vector and bivector terms must relate to $\mathbf{E}$ and $\boldsymbol{B}$ respectively, but the scalar terms are easily disposed of. Since they cannot be part of the field, they must vanish. While this is only part of the calculation of the actual electromagnetic field, it provides a convenient way of introducing separate spacetime splits for each factor in a product such as $\boldsymbol{r}\mathfrak{v}$.

11.5.2 Maxwell's Equation in a Different Frame

Maxwell's equation is of such general significance that it cannot apply in only one frame, it must apply in any frame. A Lorentz transformation, that is to say a change of basis vectors, from one frame to another should therefore preserve the form of the equation, that is to say, if the equation is $\boldsymbol{\nabla F} = \boldsymbol{J}$ in the $\boldsymbol{t}$-frame, then in the $\boldsymbol{t}'$-frame it must be of the same form $\boldsymbol{\nabla' F'} = \boldsymbol{J'}$ where $\boldsymbol{\nabla'}$, $\boldsymbol{F'}$, and $\boldsymbol{J'}$ are the representations of $\boldsymbol{\nabla}$, $\boldsymbol{F}$, and $\boldsymbol{J}$ in this new frame. An equation of this type that maintains the same form in any inertial (nonaccelerating) frame is said to be covariant. We would therefore expect all the fundamental equations of the universe to be covariant, but we may not always be accustomed to them in their covariant form. For example, Coulomb's law cannot be covariant because it involves only the electric field, whereas in a different frame it would be forced to take a modified form that includes a magnetic interaction.

We will deal with the detail of how the electromagnetic field actually changes in the following sections, but in order to clearly establish the covariance of Maxwell's free space equation, we need to know what happens to $\boldsymbol{\nabla}$. Although the change of basis applies to $-\partial_t \boldsymbol{t} + \partial_x \boldsymbol{x} + \partial_x \boldsymbol{y} + \partial_x \boldsymbol{z}$ in the same way as it does for any vector, this still leaves the scalar derivatives $\partial_t, \partial_x, \partial_y \cdots$ in terms of the original frame variables t, x, y, z rather than the new frame variables t', x', y', z'. To find $\boldsymbol{\nabla'}$, therefore, we must also transform t, x, y, z.

First, we use Equation (9.17) to replace the basis vectors in $-\partial_t \boldsymbol{t} + \partial_x \boldsymbol{x} + \partial_x \boldsymbol{y} + \partial_x \boldsymbol{z}$ with the forms that express them in terms of the $\boldsymbol{t}'$-frame basis vectors, but in addition, we have to find the appropriate prescription for $\partial_t, \partial_x, \partial_y, \partial_z$ from Equation (9.29). We find in the usual way

$$\begin{aligned}
\partial_t &= (\partial_t t')\partial_{t'} + (\partial_t x')\partial_{x'} = \gamma\partial_{t'} - \gamma v \partial_{x'} \\
\partial_x &= (\partial_x t')\partial_{t'} + (\partial_x x')\partial_{x'} = \gamma\partial_{x'} - \gamma v \partial_{t'}
\end{aligned} \tag{11.38}$$

so that on performing all the required substitutions, we find

$$\begin{aligned}
\boldsymbol{\nabla} &= -\partial_t \boldsymbol{t} + \partial_x \boldsymbol{x} + \partial_x \boldsymbol{y} + \partial_x \boldsymbol{z} \\
\mapsto \boldsymbol{\nabla}' &= -\left(\gamma \partial_{t'} - \gamma v \partial_{x'}\right)\gamma\left(\boldsymbol{t}' - v\boldsymbol{x}'\right) + \left(\gamma \partial_{x'} - \gamma v \partial_{t'}\right)\gamma\left(\boldsymbol{x}' - v\boldsymbol{t}'\right) + \partial_y \boldsymbol{y} + \partial_z \boldsymbol{z} \\
&= -\gamma^2\left(1 - v^2\right)\partial_{t'}\boldsymbol{t}' + \gamma^2\left(1 - v^2\right)\partial_{x'}\boldsymbol{x}' + \left(\gamma^2 v \partial_{x'}\boldsymbol{t}' + \gamma^2 v \partial_{t'}\boldsymbol{x}' - \gamma^2 v \partial_{x'}\boldsymbol{t}'\right. \\
&\quad \left. - \gamma^2 v \partial_{t'}\boldsymbol{x}'\right) + \partial_y \boldsymbol{y} + \partial_z \boldsymbol{z} \\
&= -\partial_{t'}\boldsymbol{t}' + \partial_{x'}\boldsymbol{x}' + \partial_y \boldsymbol{y} + \partial_z \boldsymbol{z}
\end{aligned} \tag{11.39}$$

which is exactly the same form as it takes in the $\boldsymbol{t}$-frame; it only requires to be written with the basis vectors and scalar derivatives of the new frame substituted for those of the old one. But the principle of relativity would in any case insist on this—no choice of frame can be special, and so $\boldsymbol{\nabla}$ and $\boldsymbol{\nabla}'$ must have identical forms.

While we find that the electromagnetic field appears different in the $\boldsymbol{t}'$-frame, it is not because $\boldsymbol{F}$ itself is any different, it is because of the way it separates out into $\boldsymbol{E}$ and $\boldsymbol{B}$ is different. Similarly, the source density $\boldsymbol{J}$ is represented by a different mix of $\rho\boldsymbol{t}$ and $\underset{\sim}{\boldsymbol{J}}$. This may be thought of as a type of spacetime split in which the quantities involved are simply kept in terms of the spacetime basis elements rather than being replaced by their (3+1)D counterparts.

The crux of the matter is that $\boldsymbol{\nabla}$, $\boldsymbol{F}$, and $\boldsymbol{J}$ are all invariant under a change of orthonormal basis inasmuch as the change of basis does not affect them at all; it only affects the way in which they are represented in terms of basis elements. Consequently, they must still be related by the same equation whatever basis we choose, and so it does not matter whether we write $\boldsymbol{\nabla F} = \boldsymbol{J}$ or $\boldsymbol{\nabla}'\boldsymbol{F}' = \boldsymbol{J}'$. In the latter case, the only significance of the primes is that the $\boldsymbol{t}'$-frame *representations* of $\boldsymbol{\nabla}'$, $\boldsymbol{F}'$, and $\boldsymbol{J}'$ are different from their original $\boldsymbol{t}$-frame form.

There is, of course, another way to look at this question, that is to say, if we were actually to apply a Lorentz transformation, $\boldsymbol{L}$, to each of $\boldsymbol{\nabla}$, $\boldsymbol{F}$, and $\boldsymbol{J}$ so that in particular $\boldsymbol{F}'$ and $\boldsymbol{J}'$ would now be different from $\boldsymbol{F}$ and $\boldsymbol{J}$ rather than just being the same thing under a different representation. This, therefore, is a case that involves an active transformation. Whatever basis we may assume, the result of each transformation will still be expressed in terms of that basis. However, such transformations on functions and operators cannot be implemented simply by using formulas for transforming basic multivectors, for example, $\boldsymbol{U} \mapsto \boldsymbol{LUL}^\dagger$. Recalling the analogy between Lorentz transformation and spatial rotation, the rotor $\boldsymbol{R} = \frac{1}{\sqrt{2}}(1 - \boldsymbol{xy})$ rotates any vector in the $\boldsymbol{xy}$ plane by 90°, for example, $\boldsymbol{x} \mapsto \boldsymbol{RxR}^\dagger = \frac{1}{\sqrt{2}}(1 - \boldsymbol{xy})\boldsymbol{x}\frac{1}{\sqrt{2}}(1 + \boldsymbol{xy})$ $= \boldsymbol{y}$, but applied to the function $\boldsymbol{f}(x, y) = x\boldsymbol{x}$, the result is $\boldsymbol{f}'(x, y) = x\boldsymbol{y}$ rather than the intended rotated function $\boldsymbol{f}'(x, y) = y\boldsymbol{y}$. To achieve the purpose of rotating the entire function, the coordinates x and y must also be transformed, but in total, this would be the same thing as starting from the beginning and simply rotating the basis vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ by −90°.

Applying these same considerations to the Lorentz transformation of a multivector function or operator, the transformation may therefore be carried out as though we were transforming the basis vectors with a velocity parameter $-\mathbf{v}$ rather

than +**v**. For example, with a simple Lorentz transformation of the sort we encountered in Equation (9.11), the transformation for vectors in the $\boldsymbol{xt}$ plane is $\boldsymbol{t} \mapsto \gamma(\boldsymbol{t} + v\boldsymbol{x})$ and $\boldsymbol{x} \mapsto \gamma(\boldsymbol{x} + v\boldsymbol{t})$, while for the coordinates, we have $t \mapsto \gamma(t - vx)$ and $x \mapsto \gamma(x - vt)$ (Equation 9.29). To transform $\boldsymbol{f}(x, y) = x\boldsymbol{x}$, we therefore change the sign of v and apply $\boldsymbol{x} \mapsto \gamma(\boldsymbol{x} - v\boldsymbol{t})$ and $x \mapsto \gamma(x + vt)$, resulting in

$$\begin{aligned} \boldsymbol{f}'(x, y) &= \gamma(x + vt)\gamma(\boldsymbol{x} - v\boldsymbol{t}) \\ &= \gamma^2(x + vt)\boldsymbol{x} - v\gamma^2(x + vt)\boldsymbol{t} \end{aligned} \tag{11.40}$$

which is quite distinct from the result that would have been obtained from $\boldsymbol{f}(x) \mapsto \boldsymbol{L}\boldsymbol{f}(x)\boldsymbol{L}^{\dagger} = \gamma x(\boldsymbol{x} + v\boldsymbol{t})$. Since an active Lorentz transformation applied to a multivector function or operator may be thought as still being equivalent to a change of basis, Maxwell's equation remains an equation even when the intention is an active transformation. It will therefore be sufficient to restrict our comments to passive transformations alone. The key points are the following:

- While a passive transformation on $\boldsymbol{f}$ results in the same multivector function expressed in a different basis, an active transform on $\boldsymbol{f}$ implies a different, but entirely equivalent, function expressed in the same basis. This is a subtle and often confusing point.
- Covariance means retaining the same form under a change (orthogonal transformation) of basis vectors.
- More specifically, $\boldsymbol{\nabla}'\boldsymbol{F}' = \boldsymbol{J}'$ means the same thing as $\boldsymbol{\nabla}\boldsymbol{F} = \boldsymbol{J}$.
- This is because $\boldsymbol{\nabla}'$, $\boldsymbol{F}'$, and $\boldsymbol{J}'$ are actually the same thing as $\boldsymbol{\nabla}$, $\boldsymbol{F}$, and $\boldsymbol{J}$.
- The only difference between primed and unprimed objects here is their representation in terms of the alternative bases.
- When these quantities are written in terms of second-rank tensor and four-vectors, for example, $\boldsymbol{\nabla}$ is treated as ∂_α, $\boldsymbol{F}$ as $F_{\alpha\beta}$, and $\boldsymbol{J}$ as J_β for $\alpha, \beta = t, x, y, z$, we see only the components, not the basis vectors.
- Such objects *do change* under a Lorentz transformation.
- In the case of an active transformation, the transformation must also be applied to any variables that are linked to coordinates.

11.5.3 Transformation of F by Replacement of Basis Elements

Let us assume that the field $\boldsymbol{F}$ is originally observed in the $\boldsymbol{t}$-frame and it is required to find the field seen by an observer in, say, the $\boldsymbol{t}'$-frame. The most straightforward and informative way of proceeding is to take the field in the $\boldsymbol{t}$-frame as $\boldsymbol{F} = E_x\boldsymbol{xt} + E_y\boldsymbol{yt} + E_z\boldsymbol{zt} - B_x\boldsymbol{yz} - B_y\boldsymbol{zx} - B_z\boldsymbol{xy}$. Here we may readily identify the timelike and spacelike parts and associate them with the electric magnetic fields, respectively.

Following the same general procedure described in Section 9.5, we can construct the bivector basis elements for the $\boldsymbol{t}$-frame in terms of the $\boldsymbol{t}'$-frame basis elements, but this time using the *reverse* Lorentz transformation of the vector basis elements (Equation 9.17). This then gives us

$$\begin{aligned}
\boldsymbol{xt} &= \gamma(\boldsymbol{t}' - v\boldsymbol{x}')\gamma(\boldsymbol{x}' - v\boldsymbol{t}') = \boldsymbol{x}'\boldsymbol{t}' \\
\boldsymbol{yt} &= \boldsymbol{y}'\gamma(\boldsymbol{t}' - v\boldsymbol{x}') = \gamma(\boldsymbol{y}'\boldsymbol{t}' + v\boldsymbol{x}'\boldsymbol{y}') \\
\boldsymbol{zt} &= \boldsymbol{z}'\gamma(\boldsymbol{t}' - v\boldsymbol{x}') = \gamma(\boldsymbol{z}'\boldsymbol{t}' - v\boldsymbol{z}'\boldsymbol{x}') \\
\boldsymbol{yz} &= \boldsymbol{y}'\boldsymbol{z}' \\
\boldsymbol{zx} &= \boldsymbol{z}'\gamma(\boldsymbol{x}' - v\boldsymbol{t}') = \gamma(\boldsymbol{z}'\boldsymbol{x}' - v\boldsymbol{z}'\boldsymbol{t}') \\
\boldsymbol{xy} &= \gamma(\boldsymbol{x}' - v\boldsymbol{t}')\boldsymbol{y}' = \gamma(\boldsymbol{x}'\boldsymbol{y}' + v\boldsymbol{y}'\boldsymbol{t}')
\end{aligned} \tag{11.41}$$

It has already been mentioned in Section 10.8 that, as a general principle, any spacetime multivector in component form, for example, $u_t\boldsymbol{t} + u_x\boldsymbol{x}\ldots + U_{xt}\boldsymbol{xt} + U_{yt}\boldsymbol{yt}\ldots + U_{yz}\boldsymbol{yz} + U_{zx}\boldsymbol{zx}\ldots$ may be transformed to a new frame by substitution of the basis elements of the old frame by their representations in the new frame. In the case of a bivector, each $\boldsymbol{t}$-frame bivector basis element on the left of Equation (11.41) is to be replaced by the corresponding expression on the right, now stated in terms of the $\boldsymbol{t}'$-frame basis elements. Applying this to the spacetime forms for $\boldsymbol{E}$ and $\boldsymbol{B}$ given in the $\boldsymbol{t}$-frame, we obtain

$$\begin{aligned}
E_x\boldsymbol{xt} + E_y\boldsymbol{yt} + E_z\boldsymbol{zt} \mapsto\ & E_x\boldsymbol{x}'\boldsymbol{t}' + E_y\gamma(\boldsymbol{y}'\boldsymbol{t}' + v\boldsymbol{x}'\boldsymbol{y}') + E_z\gamma(\boldsymbol{z}'\boldsymbol{t}' - v\boldsymbol{z}'\boldsymbol{x}') \\
&= E_x\boldsymbol{x}'\boldsymbol{t}' + \gamma\left(E_y\boldsymbol{y}'\boldsymbol{t}' + E_z\boldsymbol{z}'\boldsymbol{t}'\right) \\
&\quad - \gamma v\left(E_z\boldsymbol{z}'\boldsymbol{x}' - E_y\boldsymbol{x}'\boldsymbol{y}'\right)
\end{aligned} \tag{11.42}$$

$$\begin{aligned}
B_x\boldsymbol{yz} + B_y\boldsymbol{zx} + B_z\boldsymbol{xy} \mapsto\ & B_x\boldsymbol{y}'\boldsymbol{z}' + B_y\gamma(\boldsymbol{z}'\boldsymbol{x}' - v\boldsymbol{z}'\boldsymbol{t}') + B_z\gamma(\boldsymbol{x}'\boldsymbol{y}' + v\boldsymbol{y}'\boldsymbol{t}') \\
&= B_x\boldsymbol{y}'\boldsymbol{z}' + \gamma\left(B_y\boldsymbol{z}'\boldsymbol{x}' + B_z\boldsymbol{x}'\boldsymbol{y}'\right) \\
&\quad + \gamma v\left(B_z\boldsymbol{y}'\boldsymbol{t}' - B_y\boldsymbol{z}'\boldsymbol{t}'\right)
\end{aligned} \tag{11.43}$$

Combining these, we find a result directly comparable to Equation (10.34),

$$\begin{aligned}
\boldsymbol{F} &= \boldsymbol{E}' - \boldsymbol{B}' \\
&= E_x\boldsymbol{x}'\boldsymbol{t}' + \gamma\left(E_y\boldsymbol{y}'\boldsymbol{t}' + E_z\boldsymbol{z}'\boldsymbol{t}'\right) - \gamma v\left(E_z\boldsymbol{z}'\boldsymbol{x}' - E_y\boldsymbol{x}'\boldsymbol{y}'\right) - B_x\boldsymbol{y}'\boldsymbol{z}' \\
&\quad - \gamma\left(B_y\boldsymbol{z}'\boldsymbol{x}' + B_z\boldsymbol{x}'\boldsymbol{y}'\right) - \gamma v\left(B_z\boldsymbol{y}'\boldsymbol{t}' - B_y\boldsymbol{z}'\boldsymbol{t}'\right) \\
&= E_x\boldsymbol{x}'\boldsymbol{t}' + \gamma\left(E_y\boldsymbol{y}'\boldsymbol{t}' + E_z\boldsymbol{z}'\boldsymbol{t}'\right) - \gamma v\left(B_z\boldsymbol{y}'\boldsymbol{t}' - B_y\boldsymbol{z}'\boldsymbol{t}'\right) - B_x\boldsymbol{y}'\boldsymbol{z}' \\
&\quad - \gamma\left(B_y\boldsymbol{z}'\boldsymbol{x}' + B_z\boldsymbol{x}'\boldsymbol{y}'\right) - \gamma v\left(E_z\boldsymbol{z}'\boldsymbol{x}' - E_y\boldsymbol{x}'\boldsymbol{y}'\right)
\end{aligned} \tag{11.44}$$

so that $\boldsymbol{E}'$ and $\boldsymbol{B}'$ can be found by grouping together the terms that are associated with the timelike and spacelike basis elements respectively. We may then apply the spacetime split in the $\boldsymbol{t}'$-frame simply by replacing $\boldsymbol{x}'\boldsymbol{t}', \boldsymbol{y}'\boldsymbol{t}', \boldsymbol{z}'\boldsymbol{t}', \boldsymbol{y}'\boldsymbol{z}', \boldsymbol{z}'\boldsymbol{x}', \boldsymbol{x}'\boldsymbol{y}'$ with $\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{yz}, \mathbf{zx}, \mathbf{xy}$ where the primes have been dropped since the (3+1)D basis vectors are always $\mathbf{x}, \mathbf{y}, \mathbf{z}$ (see Section 10.6.2). We therefore find

$$\begin{aligned}
\boldsymbol{E}' &= E_x \boldsymbol{x}'\boldsymbol{t}' + \gamma\left(E_y \boldsymbol{y}'\boldsymbol{t}' + E_z \boldsymbol{z}'\boldsymbol{t}'\right) - \gamma v\left(B_z \boldsymbol{y}'\boldsymbol{t}' - B_y \boldsymbol{z}'\boldsymbol{t}'\right) \\
&\leftrightarrow \quad \mathbf{E}_{/\!/} + \gamma \mathbf{E}_{\perp} + \gamma \mathbf{v} \times \mathbf{B} \\
\boldsymbol{B}' &= B_x \boldsymbol{y}'\boldsymbol{z}' + \gamma\left(B_y \boldsymbol{z}'\boldsymbol{x}' + B_z \boldsymbol{x}'\boldsymbol{y}'\right) + \gamma v\left(E_z \boldsymbol{z}'\boldsymbol{x}' - E_y \boldsymbol{x}'\boldsymbol{y}'\right) \qquad (11.45) \\
&\leftrightarrow \quad \boldsymbol{B}_{/\!/} + \gamma \boldsymbol{B}_{\perp} - \gamma \mathbf{v} \wedge \mathbf{E} \\
&\qquad = I\left(\mathbf{B}_{/\!/} + \gamma \mathbf{B}_{\perp} - \gamma \mathbf{v} \times \mathbf{E}\right)
\end{aligned}$$

All that is required to reinstate the suppressed constants in the result is to replace $\mathbf{v}$ with $\mathbf{v}/c^2$, whereafter these are the same as the textbook formulas to be found, for example, in Jackson [37, section 11.10, p. 380] and Stratton [35, section 1.23, p. 79]. Since the choice of $\mathbf{x}$ for the direction of motion was arbitrary, we have been able to generalize the result by taking $\mathbf{E}_{/\!/}$ and $\mathbf{B}_{/\!/}$ and $\mathbf{v}$ to be the parts parallel to $\mathbf{x}$, while $\mathbf{E}_{\perp}$ and $\mathbf{B}_{\perp}$ are made up from the parts along $\mathbf{x}$ and $\mathbf{y}$.

11.5.4 The Electromagnetic Field of a Plane Wave Under a Change of Frame

In Section 11.4, we discussed the spacetime form of a plane electromagnetic wave as represented in a given frame, the $\boldsymbol{t}$-frame, as being given by $\boldsymbol{F} = (\alpha + I\beta)\boldsymbol{F}_0$ where $\alpha + I\beta = e^{-I\boldsymbol{k}\cdot\boldsymbol{r}}$, $\boldsymbol{k}$ is null, and the scalars α and β vary as the cosine and sine of $-\boldsymbol{k}\cdot\boldsymbol{r} = \omega t - \mathbf{k}\cdot\mathbf{r}$, respectively. We now ask, what form will this representation take in a different frame?

We may start by noting that any scalar plus pseudoscalar factor such as $\alpha + I\beta$ is unaffected by changing to a different frame (see Section 9.6), and so we need to concern ourselves only with how $\boldsymbol{F}_0$ is affected, that is to say, how it separates into timelike and spacelike parts in that frame. This means that we can leave the phase factor as it stands and simply apply the procedures discussed in the preceding subsections directly to $\boldsymbol{F}_0$. Referring to Equation (11.33), which reveals the form of $\boldsymbol{F}_0$ when $\boldsymbol{x}$ is taken as the propagation direction, $f_y + If_z$ is also unaffected by a change of frame, and so the end result simply depends on how $\boldsymbol{y}\boldsymbol{t} - \boldsymbol{x}\boldsymbol{y}$ is represented in the new basis. It is left as an exercise to show that for motion along the propagation direction, that is to say $\boldsymbol{x}$, it turns out that $\boldsymbol{y}\boldsymbol{t} - \boldsymbol{x}\boldsymbol{y} = \gamma(1-v)(\boldsymbol{y}'\boldsymbol{t}' - \boldsymbol{x}'\boldsymbol{y}')$, so that the polarization of the wave *is unaffected* and only its magnitude is modified. This is a surprising result, because we would have expected timelike and spacelike parts to remix as usual, as per Equation (11.45). This does happen, but the separate changes to $\boldsymbol{E}$ and $\boldsymbol{B}$ cancel out.

In the case of an individual wave, $\boldsymbol{B}$ is always determined by $\boldsymbol{E}$, the only degree of freedom being whether the wave is right or left circularly polarized, as determined by the sign of $\boldsymbol{k}$. But as per the preceding discussion, if the phase factor $\alpha + I\beta$ is unaffected by a change of frame, this must imply $e^{-I\boldsymbol{k}'\cdot\boldsymbol{r}'} = e^{-I\boldsymbol{k}\cdot\boldsymbol{r}}$ or, more simply,

$$\boldsymbol{k}'\cdot\boldsymbol{r}' = \boldsymbol{k}\cdot\boldsymbol{r} \qquad (11.46)$$

where $\boldsymbol{k}'$ and $\boldsymbol{r}'$ refer to the representations of $\boldsymbol{k}$ and $\boldsymbol{r}$ in the basis of the new frame. It is left as an exercise to show this by working out $\boldsymbol{k}'$ and $\boldsymbol{r}'$ then evaluating both

inner products. From inspection of the temporal and spatial parts of $\boldsymbol{k}'$, it can be seen that the frequency and the magnitude of the wave vector are different in the new frame, which is the effect we call Doppler shift.

We may now ask, what is the speed of the wave as seen in the new frame? This may be worked out from the components of $\boldsymbol{k}'$, but there is also a more direct route. The fact that $\boldsymbol{k}$ is null gave us $\omega^2 = k^2$, that is to say $|\omega/k| = 1[c]$ where $|\omega/k|$ defines the speed of light (strictly speaking the phase velocity which is relevant here), and so a null wave vector implies that the associated wave is traveling at the speed of light in free space. This is the association between null vectors and lightlike vectors. Our question can therefore be asked another way: Is $\boldsymbol{k}'$ null? If so, then we must also have $\omega'/k' = 1$, that is to say, the phase velocity of the wave is unchanged. And there we have the answer in the simple fact that $\boldsymbol{k}'$ and $\boldsymbol{k}$ are just two representations of *the same vector* using a different basis, the prime signifies only that and no more. As discussed in Section 9.8.1, it would be wrong to try to think of $\boldsymbol{k}'$ as the Lorentz transformation of $\boldsymbol{k}$, because that does change the vector rather than the basis. So the basis changes and the components change, but they always represent the same vector. Therefore, if $\boldsymbol{k}$ is null then so must be its alternative representation $\boldsymbol{k}'$. It is not possible to have $\boldsymbol{k}^2 = 0$ on the one hand and $\boldsymbol{k}'^2 \neq 0$ on the other. Since $\boldsymbol{k}'$ must be null, we may immediately deduce $\omega'/k' = \omega/k = 1$, that is to say, the speed of light of a plane electromagnetic wave in free space is the same in any chosen frame.

This is a significant result that, rather than being difficult to work out, is relatively easily discovered from the fundamental properties of spacetime, or as we should really say, it is *built into* the fundamental properties of spacetime.

11.6 LORENTZ FORCE

We may recall that in (3+1)D, any attempt to express the Lorentz force in terms of the electromagnetic field $\boldsymbol{F}$ and the charge's velocity at best resulted in an equation of the form $\mathbf{f} = q\boldsymbol{F}\langle 1+\mathbf{v}\rangle_1$, meaning the vector part of $q\boldsymbol{F}(1+\mathbf{v})$. Furthermore, the algebraic form $\mathbf{f} = q(\mathbf{E}+\boldsymbol{B}\cdot\mathbf{v})$ does little more than transcribe the traditional form $\mathbf{f} = q(\mathbf{E}+\mathbf{v}\times\mathbf{B})$ into the language of geometric algebra. What we should have hoped to find, on a point of principle, is that the force should be algebraically expressible in terms of three variables, not four, namely the charge, its velocity, and the *complete* electromagnetic field. Let us therefore see if the spacetime approach will improve matters.

We may try to tackle the problem of finding the spacetime form of the Lorentz force in either of two ways. The first is to start from the (3+1)D Newtonian form and try to find an equivalent spacetime form that will provide $\mathbf{f} = q(\mathbf{E}+\boldsymbol{B}\cdot\mathbf{v})$ as its spacetime split. The second is to start in spacetime and apply first principles. Because force is involved explicitly, either way requires us to employ some basic spacetime dynamics, and so to begin with, we simply state the basis of relativistic force as being

$$\boldsymbol{f} = \dot{\boldsymbol{p}} = m\dot{\boldsymbol{\upsilon}} \tag{11.47}$$

where $\boldsymbol{f}$ is referred to as the Minkowski force, $\boldsymbol{p} = m\boldsymbol{\upsilon}$ is the proper momentum of the particle in question, $\dot{\boldsymbol{\upsilon}}$ is its acceleration in terms of the rate of change of the proper velocity $\boldsymbol{\upsilon}$ with respect to its proper time τ, and m is its rest mass (see the discussion in Section 10.7). This is natural enough, because it simply states that in the particle's rest frame, the $\boldsymbol{\upsilon}$-frame, the force is given by Newton's second law. The key thing is that, from this starting point, Newton's second law is generalized in a relativistically correct manner simply by changing to the observer's rest frame. For example, expressing Equation (11.47) in the $\boldsymbol{t}$-frame with $\boldsymbol{p} = m\boldsymbol{\upsilon} = m\gamma(\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}})$ results in the following expression for the rate of change of momentum:

$$\begin{aligned}
\dot{\boldsymbol{p}} = \frac{d\boldsymbol{p}}{dt}\cdot\frac{dt}{d\tau} &= (m\partial_t\boldsymbol{\upsilon})\gamma \\
&= \gamma m\partial_t(\gamma(\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}})) \\
&= \gamma m((\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}})\partial_t\gamma + \gamma\partial_t\boldsymbol{v}) \\
&= \gamma m\left((\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}})v\gamma^3\partial_t v + \gamma\partial_t\boldsymbol{v}\right) \\
&= \gamma^4 m((v\partial_t v)\boldsymbol{t} + (v\partial_t v)\underset{\sim}{\boldsymbol{v}}) + \gamma^2 m\partial_t\boldsymbol{v} \\
&= \gamma^4\partial_t\left(\tfrac{1}{2}mv^2\right)\boldsymbol{t} + \gamma^2 m\left(\boldsymbol{a}_{/\!/} + \gamma^2\boldsymbol{a}_\perp\right)
\end{aligned} \tag{11.48}$$

where $\partial_t\underset{\sim}{\boldsymbol{v}} = \partial_t\boldsymbol{v}$, which defines the acceleration as seen on the $\boldsymbol{t}$-frame clock, while its parts parallel and perpendicular to the direction of motion are $\boldsymbol{a}_{/\!/} = (\underset{\sim}{\boldsymbol{v}}/v)\partial_t v$ and $\boldsymbol{a}_\perp = \partial_t\boldsymbol{v} - \boldsymbol{a}_{/\!/}$, respectively. In the Newtonian limit, $\gamma = 1$, the coefficient of $\boldsymbol{t}$ is $\partial_t\left(\frac{1}{2}mv^2\right)$, which is clearly the rate of change of kinetic energy, while the spatial term reduces to the usual force term $m\boldsymbol{a}$ with the expected spacetime split $m\boldsymbol{a} \leftrightarrow m\boldsymbol{a}\boldsymbol{t} = m\mathbf{a} = m\partial_t\mathbf{v}$. The form of Equation (11.48) is consequently in agreement with a four-vector whose components are rate of change of momentum and rate of change of kinetic energy as expressed in the $\boldsymbol{t}$-frame. Note that in the particle's own rest frame, the kinetic energy term vanishes, which is consistent with the fact that here $\dot{\boldsymbol{p}}$ can be equated to $m\boldsymbol{a}$.

Now that the relativistic definition of force and its implications are clear, let us apply it to a charged particle under the influence of an electromagnetic field. We know that, as seen in its own rest frame, the particle will be influenced by the electric field alone. Nevertheless, we cannot simply write the resulting force as $q\boldsymbol{E}$ since this is a bivector expression. Maxwell's equation requires the electromagnetic field to be in bivector form and so there is no point in trying to find some way to reinvent it as a vector. If the particle were at rest in the $\boldsymbol{t}$-frame, we could write the force $\boldsymbol{f}$ as $q\underset{\sim}{\boldsymbol{E}}$ where $\underset{\sim}{\boldsymbol{E}} = \boldsymbol{t}\boldsymbol{E}$, given that this agrees with the (3+1)D view resulting from the simple spacetime split $\mathbf{f} = -\boldsymbol{t}\boldsymbol{f} = -q\boldsymbol{t}\underset{\sim}{\boldsymbol{E}} = q\mathbf{E}$. But in general, we may also do the same thing in the $\boldsymbol{\upsilon}$-frame, the charge's rest frame, simply by using $\boldsymbol{\upsilon}$ here instead of $\boldsymbol{t}$. We then have $\underset{\sim}{\boldsymbol{E}} = \boldsymbol{\upsilon}\boldsymbol{E}$ as a spatial vector representing the electric field in the $\boldsymbol{\upsilon}$-frame rather than the $\boldsymbol{t}$-frame. We may then state the force on the charge as being

$$\begin{aligned} \boldsymbol{f} &= m\dot{\boldsymbol{\upsilon}} \\ &= q\underset{\sim}{\boldsymbol{E}} \\ &= q\boldsymbol{\upsilon}\boldsymbol{E} \end{aligned} \tag{11.49}$$

where $\boldsymbol{E}$ is to be taken as the timelike part $\boldsymbol{F}$ in the $\boldsymbol{\upsilon}$-frame. While this is progress, the equation lacks generality because $\boldsymbol{E}$ is a frame-dependent quantity. It does not apply to any other frame, that is to say, it is not covariant. What is really needed is to have $\boldsymbol{F}$ featuring in this equation instead of $\boldsymbol{E}$. If we turn to Equation (11.34), however, we find that $\boldsymbol{E} = \boldsymbol{\upsilon}(\boldsymbol{F}\cdot\boldsymbol{\upsilon})$ provides just the mechanism, so that $\boldsymbol{\upsilon}\boldsymbol{E}$ is given in the $\boldsymbol{\upsilon}$-frame by

$$\begin{aligned} \boldsymbol{\upsilon}\boldsymbol{E} &= \boldsymbol{\upsilon}^2(\boldsymbol{F}\cdot\boldsymbol{\upsilon}) \\ &= -(\boldsymbol{F}\cdot\boldsymbol{\upsilon}) \\ &= \boldsymbol{\upsilon}\cdot\boldsymbol{F} \end{aligned} \tag{11.50}$$

Replacing $\boldsymbol{\upsilon}\boldsymbol{E}$ with $\boldsymbol{\upsilon}\cdot\boldsymbol{F}$ in Equation (11.49) therefore gives us the following succinct equation for the spacetime form of the Lorentz force:

$$\boldsymbol{f} = q\boldsymbol{\upsilon}\cdot\boldsymbol{F} \tag{11.51}$$

This is clearly a covariant equation because $\boldsymbol{f}$, $\boldsymbol{\upsilon}$, and $\boldsymbol{F}$ are all independent of any choice of basis, and so it is valid in any basis we choose, or for that matter, with none at all. As would be expected, it yields the rate of change of momentum of the charge, that is, the force acting on it. In addition, it reveals a timelike part of the force $\boldsymbol{f}$ that equates to rate of change of energy. Equation (11.51) completely gets around the difficulty that we have in (3+1)D where we are limited to an expression that works on $\boldsymbol{E}$ and $\boldsymbol{B}$ separately. When $v \to 0$, however, $\boldsymbol{\upsilon} \to \boldsymbol{t} + \underset{\sim}{\boldsymbol{v}}$ so that in the Newtonian limit

$$\begin{aligned} \lim_{v\to 0}\boldsymbol{f} &= q(\boldsymbol{t}+\underset{\sim}{\boldsymbol{v}})\cdot\boldsymbol{F} + \boldsymbol{O}\left(v^2\right) \\ &= q\boldsymbol{t}\cdot(\boldsymbol{E}-\boldsymbol{B}) + q\underset{\sim}{\boldsymbol{v}}\cdot(\boldsymbol{E}-\boldsymbol{B}) \\ &= q(\underset{\sim}{\boldsymbol{E}}-\underset{\sim}{\boldsymbol{v}}\cdot\boldsymbol{B}) + q\underset{\sim}{\boldsymbol{v}}\cdot(\underset{\sim}{\boldsymbol{E}}\boldsymbol{t}) \end{aligned}$$

$$\begin{aligned} f_t + \mathbf{f} = -\boldsymbol{t}\left(\lim_{v\to 0}\boldsymbol{f}\right) &= -\boldsymbol{t}q(\underset{\sim}{\boldsymbol{E}}-\underset{\sim}{\boldsymbol{v}}\cdot\boldsymbol{B}) - \boldsymbol{t}q\underset{\sim}{\boldsymbol{v}}\cdot(\underset{\sim}{\boldsymbol{E}}\boldsymbol{t}) \\ &= q(\underset{\sim}{\boldsymbol{E}}\boldsymbol{t} + \boldsymbol{t}(\underset{\sim}{\boldsymbol{v}}\cdot\boldsymbol{B})) + q(\underset{\sim}{\boldsymbol{v}}\boldsymbol{t})\cdot(\underset{\sim}{\boldsymbol{E}}\boldsymbol{t}) \\ &= q(\mathbf{E} + \boldsymbol{B}\times(\underset{\sim}{\boldsymbol{v}}\boldsymbol{t})) + q\mathbf{v}\cdot\mathbf{E} \\ &= \underbrace{q(\mathbf{E}+\boldsymbol{B}\cdot\mathbf{v})}_{\text{vector}} + \underbrace{q\mathbf{v}\cdot\mathbf{E}}_{\text{scalar}} \end{aligned} \tag{11.52}$$

Note that $\boldsymbol{t}\cdot\boldsymbol{B}$ vanishes whereas $\underset{\sim}{\boldsymbol{v}}\cdot\boldsymbol{E}$ does not. The conversion of the spacetime inner products to the (3+1)D forms takes a little care and while the commutator product of two bivectors crops up, it is immediately replaced by the inner product of (3+1)D bivector and vector. In the end, we recover the original (3+1)D form of

the Lorentz force together with the scalar term $f_t = q\mathbf{v}\cdot\mathbf{E}$, which represents the rate of work done by the force in moving the charge through the field.

Before leaving the subject of relativistic force, it is interesting to note that the problem of relating a vector, like force or acceleration, to a bivector field like $\boldsymbol{F}$ may be dealt with by defining $\boldsymbol{\Omega} \equiv \dot{\boldsymbol{\upsilon}}\boldsymbol{\upsilon}$ as an acceleration bivector [8; 27, section 5.2.7, p. 138; 33]. The acceleration vector $\dot{\boldsymbol{\upsilon}}$ is orthogonal to $\boldsymbol{\upsilon}$ (as we shall later prove), and so $\dot{\boldsymbol{\upsilon}}$ and $\boldsymbol{\upsilon}$ anticommute. From Equation (11.49), we then find $m\dot{\boldsymbol{\upsilon}}\boldsymbol{\upsilon} = -\boldsymbol{\upsilon}m\dot{\boldsymbol{\upsilon}} = -\boldsymbol{\upsilon}(q\boldsymbol{\upsilon}\boldsymbol{E}) = q\boldsymbol{E}$, with $\boldsymbol{E}$ being in the charge's rest frame, and so we now have a bivector on each side of the equation. The interpretation of Equation (11.51), on the other hand, is that the inner product $\boldsymbol{\upsilon}\cdot\boldsymbol{F}$ projects the bivector $\boldsymbol{F}$ down $\boldsymbol{\upsilon}$ onto $\underset{\sim}{\boldsymbol{E}}$, a spatial vector that matches the form of the force vector in the charge's own rest frame.

A truly significant conclusion may be drawn from Equation (11.51)—there is no separate mechanism for electricity and magnetism. The fundamental equations $\boldsymbol{f} = q\boldsymbol{\upsilon}\cdot\boldsymbol{F}$ and $\boldsymbol{\nabla F} = \boldsymbol{J}$ are both covariant. As such, they depend on the electromagnetic field only through the *frame-independent* bivector $\boldsymbol{F}$ as a whole, while the *frame-dependent* fields $\boldsymbol{E}$ and $\boldsymbol{B}$ simply result from how $\boldsymbol{F}$ happens to split into spacelike and timelike bivector parts in any particular frame.

11.7 THE SPACETIME APPROACH TO ELECTRODYNAMICS

There are two main approaches to solving electrodynamical problems. In the first approach we solve Maxwell's equations. It does not matter whether we do this explicitly or we do it implicitly, for example, by means of a Green's function that allows the solution to be obtained from the known solution for a point charge. We employed this method in Section 5.4 to find the electromagnetic field of a quasistatic source distribution. In a dynamical situation, however, such as finding the multivector potential of point charge undergoing arbitrary motion, it led to the introduction of retardation, that is to say, making due allowance for the time it takes for information to propagate from the source to the observer, as in Section 5.7. Nevertheless, the root of this concept stemmed directly from solving a wave equation rather than from any conscious desire to take into account special relativity.

In contrast, the second approach starts out from the principles of relativity. In the rest frame of a nonaccelerating charge, the charge's electromagnetic field is simply its own Coulomb field. This is a static situation, and the charge consequently gives rise to no magnetic field. On the other hand, from some other observer's point of view, the charge may be in motion. The observed field will certainly be different from the charge's own field, for we know that it will now include a magnetic component. But, rather than having to find the observed field either by solving Maxwell's equation or by using retarded Green's functions, we may find it directly from the spacetime representation of the original Coulomb field in the charge's rest frame. This may be done either by changing the basis elements employed from those of the one frame to those of the other, or by means of a spacetime split in the desired

frame. But what about accelerating charges? It turns out that while we may still apply the same principles, the implications of acceleration lead to a radiation field that is quite unlike any static field. Nevertheless, the spacetime geometric algebra provides the essential tools for analyzing this more complex scenario where radiation is involved, and this indeed is the subject of Chapter 12.

The link between the two approaches is clearly due to the fact that the concept of retardation is incorporated into the structure of spacetime. Connecting cause to effect by a null vector imposes a constraint that ensures retardation is automatically taken into account in a relativistically correct manner.

Let us discuss this further in relation to a source q with trajectory $\boldsymbol{r}_q(\tau)$ whose electromagnetic field is observed at some event on the observer's history $\boldsymbol{r}(t)$. Here the observation event on $\boldsymbol{r}(t)$ is usually the independent variable and we need to know what the corresponding value of $\boldsymbol{r}_q(\tau)$ will be, in other words, finding the source event that causes the observation event in question. As illustrated in Figure 11.1, there will be some unique value of τ, say τ_S, that corresponds to the required point on the source's trajectory. We may think of the electromagnetic information that is "caused" by the source at precisely τ_S as spreading out isotropically at the

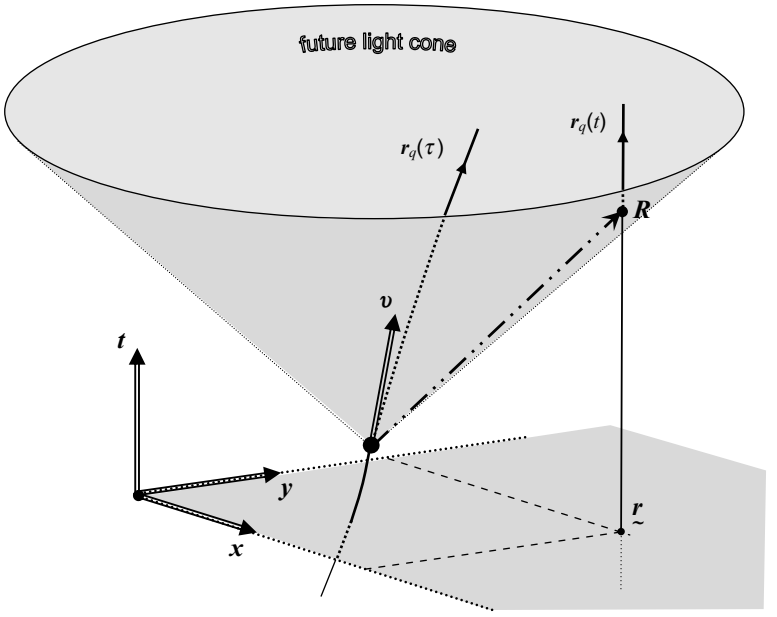

Figure 11.1 Information from a moving point charge with trajectory $\boldsymbol{r}_q(\tau)$ as seen by an observer at a fixed position $\mathbf{r}$. The time t at which a source event is observed is determined by the point at which the forward light cone originating from some source event on $\boldsymbol{r}_q(\tau)$ intersects the observer's history. The observer's history shown here is $\boldsymbol{r}(t) = \boldsymbol{t}(t + \mathbf{r})$, which corresponds to a fixed position $\mathbf{r} = \underset{\sim}{\boldsymbol{r}}\boldsymbol{t}$ in the $\boldsymbol{t}$-frame. A null vector $\boldsymbol{R}$, lying in the light cone and emanating from its origin, must connect the two events. It is therefore a lightlike vector.

speed of light. In 3D, this corresponds to the surface of a sphere whose radius expands at just that rate, while in spacetime, visualized as usual with one spatial dimension suppressed, it corresponds to a circle that expands as it travels up the surface of the forward light cone from $\boldsymbol{r}_q(\tau_S)$, again at the speed of light. The point at which this light cone meets $\boldsymbol{r}(t)$, the history of the observer, yields the observation event (time t_0 and location $\mathbf{r}$) that corresponds to τ_S. But the vector by which it has reached $\boldsymbol{r}(t)$ lies in the light cone and is therefore lightlike, that is to say, null, so that $\boldsymbol{r}(t_0) = \boldsymbol{r}_q(\tau_S) + \boldsymbol{R}$ for some vector $\boldsymbol{R}$ such that $\boldsymbol{R}^2 = 0$. An important fact about the light cone is that it is the same for any source and observer. Wherever in spacetime we choose the time axis to be, then that is the axis of the cone. However implausible this may at first seem, it follows from the fact that all the vectors in the light cone that emanate from some given event are lightlike and consequently null. In fact, we can describe the forward and reverse light cones as being the locus of all null vectors passing through that event.

Finding the required null vector $\boldsymbol{R}$ is remarkably easy since $\boldsymbol{R} = \boldsymbol{r}(t_0) - \boldsymbol{r}_q(\tau_0)$ implies $\big(\boldsymbol{r}(t_0) - \boldsymbol{r}_q(\tau_S)\big)^2 = 0$. The details are shown in Figure 11.2 where in (a), $\boldsymbol{R}$ is resolved in terms of the $\mathfrak{v}$-frame basis vectors, whereas in (b), it is resolved in terms of the $\boldsymbol{t}$-frame basis. The spacetime split of this result in the $\boldsymbol{t}$-frame leads to $-(t_0 - t_S)^2 + \big(\mathbf{r} - \mathbf{r}_q(t_S)\big)^2 = 0$, where t_S is the time in the $\boldsymbol{t}$-frame that corresponds to τ_0 at the source event. Put another way, $(t - t_S)^2 = \big(\mathbf{r} - \mathbf{r}_S(t_S)\big)^2$. Since this is the same result as obtained by the retardation approach, the claim that retardation is built into the framework of spacetime is clearly confirmed. Note that in version (b) of the figure, $d = (t_0 - t_S)$ and $\mathbf{d} = \mathbf{r} - \mathbf{r}_q(t_S) = \underset{\sim}{\boldsymbol{R}}\boldsymbol{t} = \boldsymbol{R} \wedge \boldsymbol{t}$, whereas in (a), we get the equivalent result for the primed quantities.

The above procedure may also be carried out in reverse, starting with some event on $\boldsymbol{r}$ and working back down the backward light cone to $\boldsymbol{r}_q$. In practical terms, the difference amounts only to which way round it is easier to solve the equations, which in turn depends on the nature of two trajectories involved. In cases where $\boldsymbol{r}$ is the independent variable, that is to say, we can treat the observer's history as being $t\boldsymbol{t} + \underset{\sim}{\boldsymbol{r}}$ with $\underset{\sim}{\boldsymbol{r}}$ fixed, it will usually be easier to work back down the light cone and calculate $\boldsymbol{r}_q$ for any given t. For example, given the relative $\mathbf{t}$-frame vector $\mathbf{r} = \underset{\sim}{\boldsymbol{r}}\boldsymbol{t}$ as the point of observation, and the charge's history $\boldsymbol{r}_q$ as a function of t_q, we can readily find the delay d that needs to be deducted from the observation time in order to get the corresponding source time, that is, $t_S = t - d$. We can evaluate d as before and express the observation event as $\boldsymbol{r} = (t_q + d)\boldsymbol{t} + \underset{\sim}{\boldsymbol{r}}$. It can easily be checked that the spacetime split of $\boldsymbol{r}$ gives us $-\boldsymbol{t}\boldsymbol{r} = (t_q + d) + \mathbf{r} = t + \mathbf{r}$ as required.

Either way, the procedure of finding the required null vector depends on the fact that given a history $\boldsymbol{r}_q(\tau)$ that is physically allowable but otherwise arbitrary, from any point $\boldsymbol{r}$ in spacetime it is possible to find exact solutions to $\big|\boldsymbol{r} - \boldsymbol{r}_q(\tau)\big| = 0$, one along a forward light cone and the other along a backward one. There can be no other solutions for it is not possible to intersect a forward or backward cone twice without exceeding the speed of light. While it may not always be possible to find algebraic solutions for $\boldsymbol{r}_q(\tau)$, we can see from the example given in Figure 11.3 that they can be found by construction, which, in computational terms, simply requires a root finder.

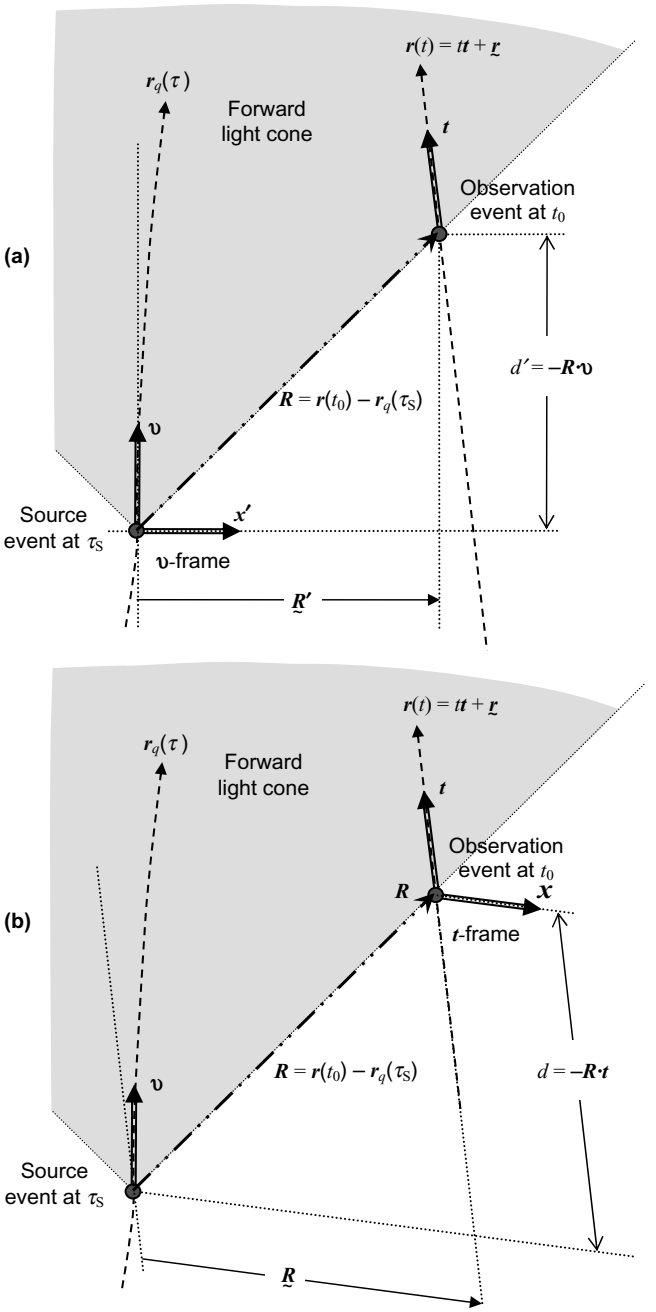
(a)
$r(t) = tt + \underset{\sim}{r}$
$r_q(\tau)$
Forward
light cone
t
Observation
event at t_0
$d' = -R \cdot \upsilon$
$R = r(t_0) - r_q(\tau_S)$
υ
Source
event at τ_S
x'
υ-frame
$\underset{\sim}{R}'$
(b)
$r(t) = tt + \underset{\sim}{r}$
$r_q(\tau)$
Forward
light cone
t
Observation
event at t_0
R
x
t-frame
$R = r(t_0) - r_q(\tau_S)$
$d = -R \cdot t$
υ
Source
event at τ_S
$\underset{\sim}{R}$

Figure 11.2 Separation between charge and observer as seen in the charge's rest frame and the observer's rest frame. In version (a), the figure is drawn from the perspective of the $\boldsymbol{\upsilon}$-frame, the rest frame of the charge, whereas in version (b) it is drawn from the perspective of the $\boldsymbol{t}$-frame, the rest frame of the observer. Otherwise, (a) and (b) both show the same situation, a section through Figure 11.1 such that the observer's history and the axis of the light cone from the source event both lie in the plane of the figure. The charge's history in its own rest frame is given in general by $\boldsymbol{r}_q(\tau)$, and the source event is taken to occur at τ_S. Only in the case of uniform motion can we express this history as $\tau\boldsymbol{\upsilon}$, and so $\boldsymbol{\upsilon}$ and $\boldsymbol{x}'$ are shown at some particular value of τ—here we have chosen τ_S. The source event at τ_S forms the origin of the light cone along which information propagates, thereby reaching the observer, whose history is as before $\boldsymbol{r}(t) = t\boldsymbol{t} + \underset{\sim}{\boldsymbol{r}}$, at $t = t_0$. This, the corresponding observation event, is therefore reached from the source event via the lightlike path $\boldsymbol{R}$. Since $\boldsymbol{R}$ is given by $\boldsymbol{R} = \boldsymbol{r}(t_0) - \boldsymbol{r}_q(\tau_S)$, the constraint $\boldsymbol{R}^2 = 0$ allows us to solve for the observation event. It is seen from (a) that this takes place in the $\boldsymbol{\upsilon}$-frame place with a delay $d' = t_0' - \tau_S = -\boldsymbol{R}\cdot\boldsymbol{\upsilon}$ after the source event. At that instant, the relative vector to the observer is $\mathbf{d}' = \underset{\sim}{\boldsymbol{R}}\boldsymbol{\upsilon} = \boldsymbol{R}\wedge\boldsymbol{\upsilon}$. In version (b), however, we obtain $d = t_0 - t_S = -\boldsymbol{R}\cdot\boldsymbol{t}$ and $\mathbf{d} = \underset{\sim}{\boldsymbol{R}}\boldsymbol{t} = \boldsymbol{R}\wedge\boldsymbol{t}$. The constraint $\boldsymbol{R}^2 = 0$ ensures $d^2 = \mathbf{d}^2$ and $d'^2 = \mathbf{d}'^2$; in other words, in either frame, the distance between source and observer equals the delay time multiplied by the speed of light. As before, we should not get any fixed ideas about what is orthogonal from how things appear on the page. Here we have made the basis vectors of the $\boldsymbol{\upsilon}$-frame appear to be orthogonal, whereas, for consistency, those of the $\boldsymbol{t}$-frame appear to be skewed. Nevertheless, both sets are in fact orthogonal.

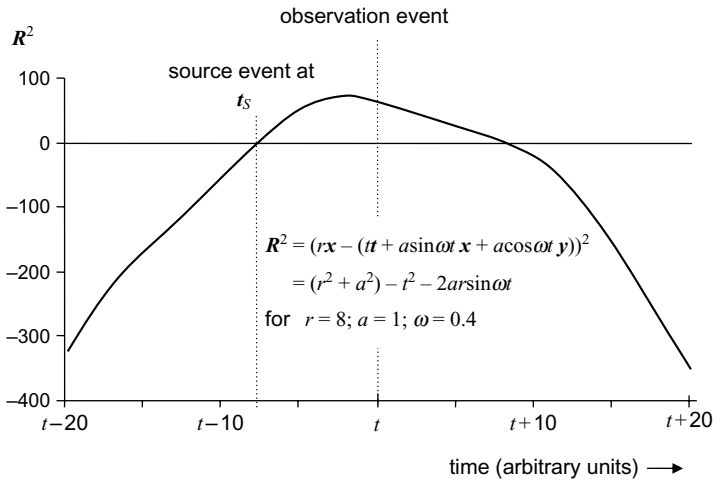

Figure 11.3 Finding where the separation vector is null. Finding the source event corresponding to a given observation time and position is discussed in Section 11.1. With a moving charge as the source, it is necessary to find the time of the source event t_S that corresponds to the time of the observation event, t, by finding where the separation vector $\boldsymbol{R}$ between source and event is null, that is to say, $\boldsymbol{R}^2 = 0$. With a history such as shown in Figure 7.1, where the charge is orbiting in some spatial plane, it is not possible to find the zeros of $\boldsymbol{R}$ analytically, but computing them is straightforward. Here we illustrate with a graph showing $\boldsymbol{R}^2$ as a function of time for the situation depicted in Figure 7.1. Since we are going back down the light cone from the observation event, $\boldsymbol{r}(t) = t\boldsymbol{t} + \underset{\sim}{\boldsymbol{r}}$, to the source event somewhere on $\boldsymbol{r}_q(t)$, we choose the solution for which $t_q < t$. Note that if we are given $\boldsymbol{r}_q(\tau)$, the particle still follows the same curve, but we do need to find the relationship linking t and τ. For the chosen parameters, the velocity of the charge is $0.4c$, and it completes an orbit in 16 time units. The resulting value of t_S will vary depending where the charge is on its orbit at $t = 0$. Here we have it at $\boldsymbol{r}_q(0) = \boldsymbol{y}$, which gives $t_S \approx t - 8.0$, meaning that the particle actually travels about half an orbit before information from it will reach the observer.

11.8 THE ELECTROMAGNETIC FIELD OF A MOVING POINT CHARGE

In Chapter 5, we discussed how the inherently relativistic problem of finding the electromagnetic field of a moving point charge could be solved exactly based on the scalar wave equation for its potential. However, this equation has its origin in Maxwell's equations, which predate the era of special relativity by some 40 years. We now address the same problem from the spacetime perspective. There are two main approaches to this, however. The first starts from the electric field $\mathbf{E}$ of a stationary point charge and then applies a Lorentz transformation to produce a different view as would be seen in a frame of reference moving with respect to the charge. The field we see in this view is no longer a pure electric field, a magnetic field $\mathbf{B}'$ is observed along with the electric field $\mathbf{E}'$. This is consistent with the force on a test charge q_0 being $q_0\mathbf{E}$ in its rest frame, whereas it is observed to be $q_0(\mathbf{E}'+\mathbf{v}\times\mathbf{B}')$ in a frame in which it has velocity $\mathbf{v}$. In the traditional approach to this problem based on tensors, it is customary to construct a 4D tensor $F_{\alpha\beta}$ to represent the electromagnetic field and then apply the Lorentz transformation in the form $F'_{\alpha\beta} = a_{\alpha\gamma}F_{\gamma\delta}a_{\beta\delta}$ where summation takes place over each pair of identical indices and $a_{\alpha\beta}$ is the transformation coefficient [37, section 11.7, pp. 371–374 and section 11.10, pp. 380–381; 44, section 8.3, p. 139; 50, chapter 27, section 26.3, pp. 26.5–26.10]. The analogy between matrix algebra and geometric algebra was discussed in Section 9.2, and it is therefore natural to suggest that the tensor approach will have a parallel in geometric algebra, as in fact discussed by Doran and Lasenby [27, section 7.1.2, pp. 232–233]. We would therefore expect to be able to replicate this approach using the methods in Section 11.5 in order to carry out the transformation of the electromagnetic field to the observer's frame.

The other method starts with the same initial concept, that the electromagnetic field and potential in the charge's rest frame is known, but instead of using a Lorentz transformation, a spacetime split is chosen as the method of expressing this in the observer's frame. This geometric approach, as described in the preceding section, seems more natural because it automatically deals with retardation. We will consider the application of this approach through spacetime diagrams and use it to solve the particular case of uniform motion, while in Chapter 12, we will go on to apply it to accelerating charges.

11.8.1 General Spacetime Form of a Charge's Electromagnetic Potential

According to one of the key principles of relativity, the potential we seek is to be found in the rest frame of the charge itself. To achieve our objective, we need to address only how this entirely scalar potential will be observed in our own rest frame. We start from the usual (3+1)D form of the scalar potential of a charge at rest at the origin

$$\Phi(t+\mathbf{r}) = \frac{1}{4\pi[\varepsilon_0]} \cdot \frac{q}{|\mathbf{r}|} \tag{11.53}$$

where $\mathbf{r}$ is the relative vector for the point of observation in the charge's own rest frame. We then use this to construct a compatible spacetime form. Since the relative quantity Φ is a frame-dependent scalar, the result we are looking for must be a vector along $\boldsymbol{\upsilon}$, the charge's local time vector, which we know is just the same as its proper velocity. Recall here that any (3+1)D scalar α may translate to spacetime as the vector $\alpha\boldsymbol{\theta}$ where $\boldsymbol{\theta}$ is the time vector of the *local* frame in which the scalar is observed. While we have usually used the $\boldsymbol{t}$-frame for this, it has often been stressed that the role of $\boldsymbol{t}$ is symbolic and we may use any frame. Given that the (3+1)D form of the multivector potential is $-\Phi + \mathbf{A}$, it therefore follows that in the charge's rest frame, where $\mathbf{A}$ must vanish, the equivalent spacetime form will be given by $\boldsymbol{A} = -\Phi\boldsymbol{\upsilon}$. Looking at this from the other direction, it may readily be verified that the spacetime split of $\boldsymbol{A} = -\Phi\boldsymbol{\upsilon}$ in the $\boldsymbol{\upsilon}$-frame is simply $-\Phi$, in agreement with our assumption. Accordingly, we may write

$$\boldsymbol{A} = \frac{-q}{4\pi} \cdot \frac{\boldsymbol{\upsilon}}{|\mathbf{d}'|} \tag{11.54}$$

Here $\mathbf{d}' = \mathbf{r}' - \mathbf{r}'_q$ must be the vector from the charge to the observer as seen in the charge's own rest frame, that is, the $\boldsymbol{\upsilon}$-frame, in which we conventionally identify relative vectors (other than basis vectors) with a prime. The next step is therefore to find an expression for $|\mathbf{r}'|$ in spacetime terms. This distance is measured along the lightlike path $\boldsymbol{R}$ that will be taken by any electromagnetic effect originating from the charge (Figure 11.1). Since, by definition, $\dot{\boldsymbol{r}}_q = \boldsymbol{\upsilon}$, the charge's history $\boldsymbol{r}_q$ must be expressible as a function of its proper time τ taken as the independent variable. Furthermore, we may express the null vector from the charge to the observation point, here denoted by $\boldsymbol{R}$, in terms of its temporal and spatial parts with respect to the $\boldsymbol{\upsilon}$-frame as $d'\boldsymbol{\upsilon} + \underset{\sim}{\boldsymbol{R}}'$, as shown in Figure 11.2(a). This in turn leads to the spacetime split of $\boldsymbol{R}$ with respect to the $\boldsymbol{\upsilon}$-frame as being $-\boldsymbol{\upsilon}\boldsymbol{R} = d' + \mathbf{d}'$ where $d' = -\boldsymbol{\upsilon}\cdot\boldsymbol{R} = -\boldsymbol{R}\cdot\boldsymbol{\upsilon}$ is effectively the light travel (delay) time along $\boldsymbol{R}$ and $\mathbf{d}' = \boldsymbol{R}\wedge\boldsymbol{\upsilon} = \underset{\sim}{\boldsymbol{R}}'\boldsymbol{\upsilon}$.

Now since $\boldsymbol{R}$ is a null vector, that is to say $\boldsymbol{R}^2 = 0$, we have $\boldsymbol{R}^2 = -d'^2 + \underset{\sim}{\boldsymbol{R}}'^2 = 0$ so that $d'^2 = \underset{\sim}{\boldsymbol{R}}'^2$. This means that the path can be measured either by the spatial distance $|\underset{\sim}{\boldsymbol{R}}'|$ or by the delay time, d'. But since $\mathbf{d}'^2 = \underset{\sim}{\boldsymbol{R}}'\boldsymbol{\upsilon}\underset{\sim}{\boldsymbol{R}}'\boldsymbol{\upsilon} = \underset{\sim}{\boldsymbol{R}}'^2$, it follows that $|\mathbf{d}'| = |\underset{\sim}{\boldsymbol{R}}'| = d' = -\boldsymbol{R}\cdot\boldsymbol{\upsilon} = |\boldsymbol{R}\cdot\boldsymbol{\upsilon}|$. This is very convenient since we already know what the charge's local time vector $\boldsymbol{\upsilon}$ is, and it also saves us having to know what any of the other $\boldsymbol{\upsilon}$-frame basis vectors might be. We are now in a position to state the concise spacetime form of the vector potential of a point charge in terms of its proper velocity $\boldsymbol{\upsilon}$ and the forward null vector $\boldsymbol{R}$ from the charge to the observer:

$$\boldsymbol{A} = \frac{-q}{4\pi} \cdot \frac{\boldsymbol{\upsilon}}{|\boldsymbol{R} \cdot \boldsymbol{\upsilon}|} \tag{11.55}$$

Note that we could equally well have arranged to tackle the problem by placing the charge in the $\boldsymbol{t}$-frame with the observer in the $\boldsymbol{\upsilon}$-frame. As far as relativity is concerned, however, that is only the same as swapping over the labels $\boldsymbol{t}$ and $\boldsymbol{\upsilon}$ and taking into account the velocity parameter will be $-v$ rather than $+v$.

To uncover the familiar (3+1)D scalar and vector potentials, we need only take the spacetime split of Equation (11.55) back into the observer's rest frame, as depicted in Figure 11.2. While Equation (11.55) is completely general, the sting in the tail is the evaluation of the term $|\boldsymbol{R} \cdot \boldsymbol{\upsilon}|$, and this is a straightforward task only for a limited number of scenarios. The problem is similar to working out the straight-line trajectory required to hit a moving target. For the case of uniform motion, at least, it is tractable, and so we now go on to explore the solution.

11.8.2 Electromagnetic Potential of a Point Charge in Uniform Motion

Equation (11.55) provides us with the general form of the electromagnetic potential due to a point charge, but in order to use it, we need to find $\boldsymbol{\upsilon}$ and $|\boldsymbol{R} \cdot \boldsymbol{\upsilon}|$ from the charge's history $\boldsymbol{r}_q(\tau)$. We can then find the observation event that corresponds to any given value of τ simply by determining where the forward light cone intercepts the observer's history as shown in, say, Figure 11.2(b). If we, the observer, are placed at a relative position $\mathbf{r}$ in the $\boldsymbol{t}$-frame, the (3+1)D paravector giving our position and the observation time is $t + \mathbf{r}$. This corresponds to the fact that we want to find $\boldsymbol{A}(\mathbf{r}, t)$ in terms of t, the observer's time, rather than in terms of the retarded time $t^* = t - d$ at the other end of the null vector $\boldsymbol{R}$. Since $t + \mathbf{r}$ translates to spacetime as $\boldsymbol{r}(t) = t\boldsymbol{t} + \underset{\sim}{\boldsymbol{r}}$ where $\underset{\sim}{\boldsymbol{r}}\boldsymbol{t} = \mathbf{r}$, this defines our, the observer's, history provided, of course, that we remain at the observation point $\mathbf{r}$. While this part is fairly straightforward, as mentioned in the previous section, the rest of the problem is less so. For a start, finding $\boldsymbol{r}_q(\tau)$ for a moving charge is generally not so simple since it needs to be found by integration of $\boldsymbol{\upsilon}(\tau)$ with respect to τ and that depends on what we know about its trajectory. For the case of a charge in uniform motion, $\boldsymbol{\upsilon}$ is constant, and so this also turns out to be quite easy. If we place the spatial origin of the $\boldsymbol{\upsilon}$-frame at the charge itself, integration yields $\boldsymbol{r}_q(\tau) = \tau\boldsymbol{\upsilon}$ in which, from Equation (10.5), $\boldsymbol{\upsilon} = \gamma(\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}})$ where $\underset{\sim}{\boldsymbol{v}}\boldsymbol{t} = \mathbf{v}$ and $\tau = \gamma^{-1}t$.

Having determined the histories of both charge and observer, the required null vector joining them at any time t is given by $\boldsymbol{R} = \boldsymbol{r} - \boldsymbol{r}_q = \boldsymbol{r}(t) - \tau\boldsymbol{\upsilon}$. Since $\boldsymbol{R}^2 = 0$, this provides the constraint that properly connects the observation and source events, which then allows τ to be determined from t for the particular configuration shown in Figure 11.4 by means of the general method discussed in Section 11.7, except that in this case, we can get an analytic solution for $\boldsymbol{R}^2 = 0$. Remembering that the charge's proper velocity is its normalized time vector, that is, $\boldsymbol{\upsilon}^2 = -1$, we find

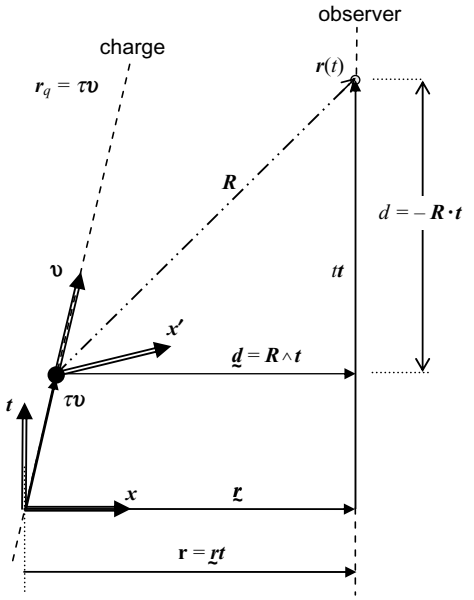

Figure 11.4 Separation between the observer and a charge in uniform motion. The charge discussed in Figures 11.1 and 11.2 now has a constant velocity **v** along **x**, giving it the straight-line trajectory $\tau\mathfrak{v}$ in the $\mathfrak{v}$-frame. The observer, on the other hand, is at $\boldsymbol{r}(t) = t\boldsymbol{t} + \underline{\boldsymbol{r}}$, which corresponds to the fixed location **r** in the $\boldsymbol{t}$-frame for all time t. It is then necessary to work back down the null vector $\boldsymbol{R}$ that connects charge and observer in order to eliminate τ and get a solution for $|\boldsymbol{R}\cdot\mathfrak{v}|$ that depends only on $\boldsymbol{r}$ and $\mathfrak{v}$. This gives the path length along $\boldsymbol{R}$, that is to say, the effective distance to the observer as seen from the charge in its own rest frame.

$$\begin{aligned}
&(\boldsymbol{r}-\tau\mathfrak{v})^2 = 0\\
\Leftrightarrow\ &\boldsymbol{r}^2 - \tau(\boldsymbol{r}\mathfrak{v}+\mathfrak{v}\boldsymbol{r}) + \tau^2\mathfrak{v}^2 = 0\\
\Leftrightarrow\ &\boldsymbol{r}^2 - 2\tau(\boldsymbol{r}\cdot\mathfrak{v}) - \tau^2 = 0\\
\Leftrightarrow\ &\tau = -\boldsymbol{r}\cdot\mathfrak{v} \pm \left((\boldsymbol{r}\cdot\mathfrak{v})^2 + \boldsymbol{r}^2\right)^{\frac{1}{2}}
\end{aligned} \tag{11.56}$$

and so we now know what τ must be for any given observation at $\boldsymbol{r}$. This is therefore just the information we need to evaluate the length of the path $|\boldsymbol{R}\cdot\mathfrak{v}|$. As a first step,

$$\begin{aligned}
|\boldsymbol{R}\cdot\mathfrak{v}| &= |(\boldsymbol{r}-\tau\mathfrak{v})\cdot\mathfrak{v}|\\
&= |\boldsymbol{r}\cdot\mathfrak{v} + \tau|
\end{aligned} \tag{11.57}$$

so that we then may substitute for τ by using Equation (11.56) so as to get

$$\begin{aligned}
|\boldsymbol{R}\cdot\mathfrak{v}| &= |\boldsymbol{r}\cdot\mathfrak{v} + \tau|\\
&= \left|\boldsymbol{r}\cdot\mathfrak{v} - \boldsymbol{r}\cdot\mathfrak{v} \pm \left((\boldsymbol{r}\cdot\mathfrak{v})^2 + \boldsymbol{r}^2\right)^{\frac{1}{2}}\right|
\end{aligned} \tag{11.58}$$

$$\Leftrightarrow |\boldsymbol{R}\cdot\mathfrak{v}| = \left((\boldsymbol{r}\cdot\mathfrak{v})^2 + \boldsymbol{r}^2\right)^{\frac{1}{2}}$$

To complete the process, we need to express this in our desired frame, the $\boldsymbol{t}$-frame, that of the observer. Since the motion is uniform, we need to introduce only

the usual relationship $\boldsymbol{\upsilon} = \gamma(\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}})$ and make use of the $\boldsymbol{t}$-frame spacetime splits $\underset{\sim}{\boldsymbol{r}}\boldsymbol{t} = \mathbf{r}$ and $\underset{\sim}{\boldsymbol{v}}\boldsymbol{t} = \mathbf{v}$ to obtain

$$\begin{aligned} \boldsymbol{r} \cdot \boldsymbol{\upsilon} &= \gamma(t\boldsymbol{t} + \underset{\sim}{\boldsymbol{r}}) \cdot (\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}}) \\ &= \gamma(-t + \underset{\sim}{\boldsymbol{r}} \cdot \underset{\sim}{\boldsymbol{v}}) \\ \Leftrightarrow \boldsymbol{r} \cdot \boldsymbol{\upsilon} &= \gamma(\mathbf{r} \cdot \mathbf{v} - t) \end{aligned} \tag{11.59}$$

Now since $\boldsymbol{r}^2 = \underset{\sim}{\boldsymbol{r}}^2 - t^2$, we also have the spacetime split $\boldsymbol{r}^2 = \mathbf{r}^2 - t^2$. Note that all the preceding splits can be taken as equalities. Substituting these results back into Equation (11.58) yields

$$\begin{aligned} |\boldsymbol{R} \cdot \boldsymbol{\upsilon}|^2 &= \gamma^2\left(t^2 - 2\mathbf{r} \cdot \mathbf{v} + (\mathbf{r} \cdot \mathbf{v})^2\right) + \left(\mathbf{r}^2 - t^2\right) \\ &= \gamma^2\left(v^2 t^2 - 2t(\mathbf{r} \cdot \mathbf{v}) + (\mathbf{r} \cdot \mathbf{v})^2\right) + \mathbf{r}^2 \end{aligned} \tag{11.60}$$

Putting this back in turn into Equation (11.55) together with the $\boldsymbol{t}$-frame representation for $\boldsymbol{\upsilon}$ in the numerator provides the result

$$\begin{aligned} \boldsymbol{A}(t) &= \frac{-q}{4\pi\varepsilon_0} \cdot \frac{\gamma(\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}})}{\left(\gamma^2\left(v^2 t^2 - 2t(\mathbf{r} \cdot \mathbf{v}) + (\mathbf{r} \cdot \mathbf{v})^2\right) + \mathbf{r}^2\right)^{\frac{1}{2}}} \\ &= \frac{-q}{4\pi\varepsilon_0} \cdot \frac{(\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}})}{\left(r^2 - 2t(\mathbf{r} \cdot \mathbf{v}) + (\mathbf{r} \cdot \mathbf{v})^2 + v^2\left(t^2 - r^2\right)\right)^{\frac{1}{2}}} \\ \Leftrightarrow \boldsymbol{A}(t) &= \frac{-q}{4\pi\varepsilon_0} \cdot \frac{(\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}})}{d_{\text{eff}}} \end{aligned} \tag{11.61}$$

where

$$d_{\text{eff}}(t) = \left(r^2\left(1 - v^2\right) - 2t(\mathbf{r} \cdot \mathbf{v}) + (\mathbf{r} \cdot \mathbf{v})^2 + v^2 t^2\right)^{\frac{1}{2}} \tag{11.62}$$

represents an effective distance that embodies the entire time dependency of the result. On completing the special spacetime split for the vector potential (Equation 11.9), we have it in the equivalent (3+1)D form

$$\boldsymbol{A}(\mathbf{r} + t) = -\Phi + \mathbf{A} = \frac{q}{4\pi} \cdot \frac{(\mathbf{v} - 1)}{d_{\text{eff}}} \quad \equiv \quad \begin{cases} \Phi = \dfrac{q}{4\pi[\varepsilon_0] d_{\text{eff}}} \\ \mathbf{A} = \Phi\mathbf{v} \end{cases} \tag{11.63}$$

The effective distance d_{eff} used here is equivalent to the term $\gamma^{-1}d$ appearing in Doran and Lasenby's result for the electromagnetic field [27, section 7.3.2, p. 244,

equation 7.98] and to κR_{ret} in Jackson's analysis [37, section 14.1, p. 465]. Taking $t = 0$ (at which time the charge will be at the origin of the $\boldsymbol{t}$-frame and $\mathbf{r}$ will then represent the vector from the charge to the point of observation) will allow us to make comparison with Equation (5.43) above. From Equation (11.62), then

$$d_0 = \left(\left(1-v^2\right)r^2 + \left(\mathbf{r}\cdot\mathbf{v}\right)^2\right)^{\frac{1}{2}} = r\left(1-v^2\sin^2\theta\right)^{\frac{1}{2}} \tag{11.64}$$

where θ is the angle between $\mathbf{r}$ and $\mathbf{v}$. This is exactly the same result as was found using Figure 5.2 to interpret Equation (5.43), except that here we have used the symbol r instead of R and d_0 is the effective distance at $t = 0$. For reasons that were discussed in Section 5.7.1, the effective distance d_{eff} corresponds to $\left|\mathbf{r}-\mathbf{r}_q^*\right| - \mathbf{v}\cdot\left(\mathbf{r}-\mathbf{r}_q^*\right)$ rather than retarded distance between charge and observer, which is simply $\left|\mathbf{r}-\mathbf{r}_q^*\right|$.

11.8.3 Electromagnetic Field of a Point Charge in Uniform Motion

It would at first appear that the procedure we have just used for finding the potential of a moving point charge could be applied to finding its electric and magnetic field but, unfortunately, this is limited to the case of constant charge velocity. The correct approach [8; 27, section 7.3.1, pp. 242 et seq.] is to differentiate the potential given in Equation (11.55) by means of the spacetime vector derivative $\boldsymbol{\nabla}$. The constant velocity solution is equivalent to the effect of a Lorentz transformation (change of basis) on the field of a charge at rest and is therefore of interest in its own right, whereas differentiation of the potential brings out an additional term due to the acceleration of the charge. It is clear from the principle of relativity that a charge moving with constant velocity cannot radiate electromagnetic energy; otherwise, a charge at rest could also do the same. On the other hand, the acceleration term is of considerable importance because it does result in electromagnetic radiation. We therefore address this in its own right in Chapter 12.

The question as to how the electric and magnetic fields are represented in spacetime has already been discussed in relation to the electromagnetic potential and Maxwell's equation. In each case, the result has been reasoned from how the (3+1)D equations translate into spacetime with the outcome in both cases that they may be represented by the timelike and spacelike bivectors $\boldsymbol{E}$ and $\boldsymbol{B}$, respectively. Let us now take a fresh look at the possibilities without the benefit of any particular (3+1)D equations to guide us, that is to say, from a fundamentally spacetime standpoint.

From Section 8.2, we know that there are only two spacetime elements that can give rise to a (3+1)D vector field. The electric field $\boldsymbol{E}$ must therefore be represented by either

- an odd element, that is to say a spacetime vector, or
- an even element, that is to say a timelike bivector.

Similarly, there are only two possibilities for a (3+1)D bivector field such as $\boldsymbol{B}$:

- an odd element, that is to say a spacetime pseudovector (trivector), and
- an even element, that is to say a spacelike bivector.

There are no separate considerations for the entire electromagnetic field $\boldsymbol{F}$ as it is determined purely by the sum of the appropriate representations of $\boldsymbol{E}$ and $\boldsymbol{B}$. However, we know from the Lorentz transformation of the basis vectors (Equation 9.11) that vectors are transformed into vectors; there is no mixing with other elements. Therefore, if the electric field were represented by a spacetime vector, it would always be represented by a vector in every frame. Now, the electromagnetic field of a nonaccelerating charged particle in its own rest frame is just its Coulomb field, a pure electric field. A magnetic component is only observed if, and only if, the charge is seen to be in motion, that is to say, in some other Lorentz frame. If $\boldsymbol{E}$ were a vector then $\boldsymbol{B}$ would be zero in every inertial frame, since otherwise this would require the Lorentz transformation of a basis vector to have either a bivector or pseudovector part, contrary to what we have just noted.

On the other hand, under a Lorentz transformation of basis bivectors (Equations 9.15 and 9.16), it is readily seen that the transformation of a timelike bivector gives rise to a spacelike part and vice versa. Therefore, let us retry the test on the alternative assumption that the electric field is represented by a timelike bivector. Starting as before in the charge's rest frame, in which the magnetic field is zero, let us transform to a different frame. We now see that under the transformation, the timelike bivector $\boldsymbol{E}$ gives rise to a spacelike bivector, which we can identify only with the magnetic field, $\boldsymbol{B}$. This agrees with the observed facts. In spacetime, therefore,

- the electromagnetic field $\boldsymbol{F}$ is a bivector;
- the bivector can be separated into *frame-dependent* timelike and spacelike parts; and
- in any given frame, the timelike part is the electric field whereas the spacelike part is the magnetic field.

Once again, there is clear evidence of the principle that the relationship between spacetime and (3+1)D must be determined from the underlying physics. It nevertheless leaves the question of how the separation of the field bivector into electric and magnetic fields is to be accomplished. However, a method of splitting any bivector such as $\boldsymbol{F}$ into its timelike and spacelike parts in a given frame has been already addressed in Sections 7.11 and 11.5.1. It is only necessary to replace $\boldsymbol{\theta}$ in Equation (11.34) with the time vector of the frame in question.

Let us now to determine the field. We should be able to proceed in much the same way as for the electromagnetic potential (Sections 11.8.1 and 11.8.2), the main difference being that we now need to translate the familiar (3+1)D vector expression for the electric field of a charge at the origin of its own rest frame, $\mathbf{E} = q\mathbf{r}'/\left(4\pi r'^3\right)$, into spacetime form. We call the charge's rest frame the $\boldsymbol{\upsilon}$-frame as before so that $\mathbf{r}'$ is a relative vector in that frame. Since the charge's proper velocity must be $\boldsymbol{\upsilon}$ and the observer is at rest in the $\boldsymbol{t}$-frame, we may construct their histories in the forms $\boldsymbol{r}_q = \boldsymbol{\upsilon}\tau$ and $\boldsymbol{r} = t\boldsymbol{t} + \underset{\sim}{\boldsymbol{r}}$, respectively. Referring again to Figure 11.4, a null vector $\boldsymbol{R}$ must link any source event on the charge's history $\boldsymbol{r}_q$ to the observation event on the observer's history $\boldsymbol{r}$ so that for any such observation event $\boldsymbol{R} = \boldsymbol{r} - \tau\boldsymbol{\upsilon}$. The spacetime split of $\boldsymbol{R}$ in the $\boldsymbol{\upsilon}$-frame then gives us both $\mathbf{r}' = -\boldsymbol{\upsilon} \wedge \boldsymbol{R} = \boldsymbol{R} \wedge \boldsymbol{\upsilon}$ and $|r'| = -\boldsymbol{R} \cdot \boldsymbol{\upsilon}$. In fact, all this is simply summarizing what we already know from our discussion of the potential. Quite conveniently, $\boldsymbol{R} \wedge \boldsymbol{\upsilon}$ is already in the bivector form we require for the electromagnetic field. We may now express the field of the point charge in the simple spacetime form

$$\boldsymbol{F}_{qs}(\boldsymbol{r}) = \frac{q}{4\pi[\varepsilon_0]} \frac{\boldsymbol{R} \wedge \boldsymbol{\upsilon}}{|\boldsymbol{R} \cdot \boldsymbol{\upsilon}|^3} \tag{11.65}$$

The subscript *qs* here implies that this is purely the *quasistatic* field that does not include the effects of the charge's acceleration. Now, since $\boldsymbol{R} = \boldsymbol{r} - \tau\boldsymbol{\upsilon}$ and the $\boldsymbol{t}$-frame representation of $\boldsymbol{\upsilon}$ can be taken to be $\gamma(\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}})$, we may readily express $\boldsymbol{F}(\boldsymbol{r})$ in the $\boldsymbol{t}$-frame. Starting with $\boldsymbol{R} \wedge \boldsymbol{\upsilon}$,

$$\begin{aligned}
\boldsymbol{R} \wedge \boldsymbol{\upsilon} &= (\boldsymbol{r} - \tau\boldsymbol{\upsilon}) \wedge \boldsymbol{\upsilon} \\
&= \boldsymbol{r} \wedge \boldsymbol{\upsilon} \\
&= (t\boldsymbol{t} + \underset{\sim}{\boldsymbol{r}}) \wedge \gamma(\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}}) \\
&= \underbrace{\gamma(\underset{\sim}{\boldsymbol{r}}\boldsymbol{t} - t\underset{\sim}{\boldsymbol{v}}\boldsymbol{t})}_{\substack{\text{timelike} \\ \text{bivector}}} + \underbrace{\gamma\underset{\sim}{\boldsymbol{r}} \wedge \underset{\sim}{\boldsymbol{v}}}_{\substack{\text{spacelike} \\ \text{bivector}}} \\
&= \gamma(\mathbf{r} - \mathbf{v}t) + \gamma\mathbf{r} \wedge \mathbf{v}
\end{aligned} \tag{11.66}$$

This spacetime split could also have been obtained using Equation (11.37) after the first step, but here we give it in full. Now, recalling the discussion of the potential, Equation (11.60) gave us a complete expression for $|\boldsymbol{R} \cdot \boldsymbol{\upsilon}|$, but in order to simplify matters, we defined an effective distance d_{eff} in Equation (11.62) to such that $\gamma/|\boldsymbol{R} \cdot \boldsymbol{\upsilon}| = 1/d_{\text{eff}}$. Since the configuration is the same, it will clearly be just as convenient to use this effective distance here so that by putting Equations (11.65) and (11.66) together, we find

$$\begin{aligned} \boldsymbol{F}_{qs} = \boldsymbol{E}_{qs} - \boldsymbol{B}_{qs} &= \frac{q}{4\pi[\varepsilon_0]} \frac{\boldsymbol{R} \wedge \boldsymbol{\upsilon}}{|\boldsymbol{R} \cdot \boldsymbol{\upsilon}|^3} \\ &= \frac{q}{4\pi} \cdot \frac{\gamma(\underset{\sim}{\boldsymbol{r}}\boldsymbol{t} - t\underset{\sim}{\boldsymbol{\upsilon}}\boldsymbol{t})}{\gamma^3 d_{\text{eff}}^3} + \frac{\gamma q}{4\pi} \cdot \frac{\gamma(\underset{\sim}{\boldsymbol{r}} \wedge \underset{\sim}{\boldsymbol{\upsilon}})}{\gamma^3 d_{\text{eff}}^3} \\ &= \underbrace{\frac{q}{4\pi} \cdot \frac{\underset{\sim}{\boldsymbol{r}}\boldsymbol{t} - t\underset{\sim}{\boldsymbol{\upsilon}}\boldsymbol{t}}{\gamma^2 d_{\text{eff}}^3}}_{\boldsymbol{E}_{qs}} - \underbrace{\frac{-q}{4\pi} \cdot \frac{(\underset{\sim}{\boldsymbol{r}}\boldsymbol{t}) \wedge (\underset{\sim}{\boldsymbol{\upsilon}}\boldsymbol{t})}{\gamma^2 d_{\text{eff}}^3}}_{\boldsymbol{B}_{qs}} \end{aligned} \tag{11.67}$$

To find the (3+1)D form of Equation (11.67), we need only to replace all the timelike bivectors with their corresponding relative vectors in the $\boldsymbol{t}$-frame as in the final line of Equation (11.66) and, as usual, change from $\boldsymbol{E} - \boldsymbol{B}$ to $\mathbf{E} + \boldsymbol{B}$:

$$\begin{aligned} \boldsymbol{F}_{qs} = \mathbf{E}_{qs} + \boldsymbol{B}_{qs} &= \frac{q}{4\pi} \cdot \frac{\mathbf{r} - \mathbf{v}t}{\gamma^2 d_{\text{eff}}^3} + \frac{q}{4\pi} \cdot \frac{\mathbf{v} \wedge \mathbf{r}}{\gamma^2 d_{\text{eff}}^3} \\ = \mathbf{E}_{qs} + I\mathbf{B}_{qs} &= \frac{q}{4\pi} \cdot \frac{\mathbf{r} - \mathbf{v}t}{\gamma^2 d_{\text{eff}}^3} + I\frac{q}{4\pi} \cdot \frac{\mathbf{v} \times \mathbf{r}}{\gamma^2 d_{\text{eff}}^3} \end{aligned} \tag{11.68}$$

If we let $\gamma \to 1$, this is directly identifiable with the standard nonrelativistic result.

Finally, a word of caution about bivector terms such as $\underset{\sim}{\boldsymbol{r}}\boldsymbol{t}$ or $\underset{\sim}{\boldsymbol{r}}'\boldsymbol{\upsilon}$ that may be generally identified with their (3+1)D bivector or relative vector counterparts. These examples are associated with the relative vectors $\mathbf{r}$ and $\mathbf{r}'$ in the $\boldsymbol{t}$- and $\boldsymbol{\upsilon}$-frames, respectively. While it can be useful to substitute one form for the other, for example, $\mathbf{r}$ for $\underset{\sim}{\boldsymbol{r}}\boldsymbol{t}$ or even $\mathbf{r} \wedge \mathbf{v}$ for $(\underset{\sim}{\boldsymbol{r}}\boldsymbol{t}) \wedge (\underset{\sim}{\boldsymbol{\upsilon}}\boldsymbol{t})$, some care is required to avoid confusing spacetime entities with those of (3+1)D. That is to say, if the $\boldsymbol{t}$-frame relative vector $\mathbf{r}$ is used in a spacetime expression, then it must be taken to mean $\underset{\sim}{\boldsymbol{r}}\boldsymbol{t}$ or $\boldsymbol{r} \wedge \boldsymbol{t}$ as opposed to $\underset{\sim}{\boldsymbol{r}}$ or $\boldsymbol{r}$.

11.9 EXERCISES

1. While for a plane wave $\boldsymbol{F}^2 = 0$, show that in general, $\boldsymbol{F}^2$ must be invariant under a Lorentz transformation and express the result in terms of $\mathbf{E}$ and $\mathbf{B}$. What are the conditions on $\mathbf{E}$ and $\mathbf{B}$ such that $\boldsymbol{F}$ will be null?

2. Find a wave equation from Maxwell's equations for the case of a linear, homogeneous, isotropic, polarizable medium where there are no free sources. Use $\boldsymbol{G} = \varepsilon\boldsymbol{E} + \boldsymbol{B}/\mu$ where ε and μ are constants.

3. Evaluate $\frac{1}{2}\boldsymbol{F}\boldsymbol{t}\boldsymbol{F}^\dagger$ in terms of $\mathbf{E}$ and $\boldsymbol{B}$. Confirm the results for $\mathcal{E}$ and $\mathbf{g}$ that were discussed at the end of Section 11.4.

4. **(a)** Show that if the $\boldsymbol{t}'$-frame is moving with velocity $v\mathbf{x}$ relative to the $\boldsymbol{t}$-frame, then $\boldsymbol{F}_0 = \boldsymbol{y}\boldsymbol{t} - \boldsymbol{x}\boldsymbol{y}$ is given in terms of the $\boldsymbol{t}'$-frame basis as $\boldsymbol{F}_0 = \gamma(1 - v)(\boldsymbol{y}'\boldsymbol{t}' - \boldsymbol{x}'\boldsymbol{y}')$.

(b) What is the result if the motion is instead along $\mathbf{y}$ or $\mathbf{z}$?

(c) Comment on the results in relation to plane waves.

5. **(a)** Show that $\boldsymbol{k}' \cdot \boldsymbol{r}' = \boldsymbol{k} \cdot \boldsymbol{r}$ as in Equation (11.46) by working out $\boldsymbol{k}'$ and $\boldsymbol{r}'$ and then evaluating the inner product.

(b) Work out the Doppler shift seen in the $\boldsymbol{t}'$-frame by comparing the components of $\boldsymbol{k}'$ with those of $\boldsymbol{k}$.

(c) Verify that $|\omega'/k'| = |\omega/k| = 1$. Note that since the wave vector $\boldsymbol{k}$ is null, k in this context means $|\mathbf{k}|$ or $|\underset{\sim}{k}|$ rather than $|\boldsymbol{k}|$.

6. **(a)** Prove for any spacetime vector $\boldsymbol{r}$ and any time vector $\boldsymbol{\theta}$ that $(\boldsymbol{r} \wedge \boldsymbol{\theta})^2 = (\boldsymbol{r} \cdot \boldsymbol{\theta})^2 + \boldsymbol{r}^2$.

(b) If $\boldsymbol{R}$ is a null vector connecting $\boldsymbol{r}$ and $\mathfrak{v}\tau$, prove $(\boldsymbol{R} \cdot \mathfrak{v})^2 = (\boldsymbol{r} \wedge \mathfrak{v})^2$.

(c) Use this to show $\boldsymbol{F}_{qs}(\boldsymbol{r})$ in Equation (11.65) may be written as

$$\frac{q}{4\pi} \frac{\boldsymbol{r} \wedge \mathfrak{v}}{|\boldsymbol{r} \wedge \mathfrak{v}|^3}$$

(d) What does this result mean?

Chapter 12

The Electromagnetic Field of a Point Charge Undergoing Acceleration

Calculating the electromagnetic field $\boldsymbol{F}$ of a point charge as we did in Section 11.8.3 does not extend to the possibility of electromagnetic radiation, and so the results obtained in the form of Equations (11.65) et seq. are valid only in circumstances where we can neglect radiation, for example, constant velocity. To calculate the field including the radiated part, it is necessary to differentiate the electromagnetic potential $\boldsymbol{A}$, as for example shown by Jackson [37, section 14.1, pp. 465–467], using a four-vector approach, or by Gull et al. [8] or Doran and Lasenby [27, section 7.3.1, pp. 242–243] using geometric algebra. While our approach may seem long-winded by comparison, it is because our aim here, naturally, is not just to focus on the use of geometric algebra, which actually offers a concise derivation [8], but also to promote understanding of how it works in such a context by expanding on the intermediate steps and explaining the frequent subtleties that arise, particularly those due to the choice of metric signature. In addition, once we have found the spacetime form of $\boldsymbol{F}$ for an accelerating charge, we go on to find the observed field in both the charge's and the observer's rest frames. The field of an accelerating charge as seen in its own instantaneous rest frame is by no means a trivial notion and, in both cases, we show that the results obtained are equivalent to those found by Jackson.

12.1 WORKING WITH NULL VECTORS

We have already encountered null vectors and have used them in Section 11.8 for the calculation of the electromagnetic potential and field due to a moving point charge. Irrespective of the history of source and observer, their construction in spacetime is the key to finding the causal relationship between any source and

Understanding Geometric Algebra for Electromagnetic Theory, First Edition. John W. Arthur.

observation event. Once constructed, we may then obtain the relative vectors that correspond to these events in any frame; in other words, we may obtain the relationship as seen in the more familiar (3+1)D world. It will therefore be useful to understand more about how null vectors may be manipulated in expressions. For example, the spacetime split of some given null vector $\boldsymbol{R}$ generates expressions such as $-\boldsymbol{R}\cdot\boldsymbol{t}$, $\boldsymbol{R}\wedge\boldsymbol{t}$, $-\boldsymbol{R}\cdot\boldsymbol{\upsilon}$, and $\boldsymbol{R}\wedge\boldsymbol{\upsilon}$, which give us the relative time delays and separation vectors between these events as seen in the $\boldsymbol{t}$-frame and $\boldsymbol{\upsilon}$-frame respectively.

We can begin by proving a simple relationship between the inner and outer products of any vector $\boldsymbol{u}$ with a null vector $\boldsymbol{R}$. First, we recall the key properties that $\boldsymbol{u}^2$ is a scalar, which will generally be nonzero, whereas by definition $\boldsymbol{R}^2=0$. From this, it then follows that

$$\begin{aligned} 4|\boldsymbol{R}\cdot\boldsymbol{u}|^2 &= \tfrac{1}{4}(\boldsymbol{R}\boldsymbol{u}+\boldsymbol{u}\boldsymbol{R})(\boldsymbol{R}\boldsymbol{u}+\boldsymbol{u}\boldsymbol{R}) \\ &= \boldsymbol{R}\boldsymbol{u}\boldsymbol{R}\boldsymbol{u}+\boldsymbol{R}\boldsymbol{u}^2\boldsymbol{R}+\boldsymbol{u}\boldsymbol{R}^2\boldsymbol{u}+\boldsymbol{u}\boldsymbol{R}\boldsymbol{u}\boldsymbol{R} \\ &= \boldsymbol{R}\boldsymbol{u}\boldsymbol{R}\boldsymbol{u}+\boldsymbol{u}\boldsymbol{R}\boldsymbol{u}\boldsymbol{R} \\ 4|\boldsymbol{R}\wedge\boldsymbol{u}|^2 &= \tfrac{1}{4}(\boldsymbol{R}\boldsymbol{u}-\boldsymbol{u}\boldsymbol{R})(\boldsymbol{R}\boldsymbol{u}-\boldsymbol{u}\boldsymbol{R}) \\ &= \boldsymbol{R}\boldsymbol{u}\boldsymbol{R}\boldsymbol{u}-\boldsymbol{R}\boldsymbol{u}^2\boldsymbol{R}-\boldsymbol{u}\boldsymbol{R}^2\boldsymbol{u}+\boldsymbol{u}\boldsymbol{R}\boldsymbol{u}\boldsymbol{R} \\ &= \boldsymbol{R}\boldsymbol{u}\boldsymbol{R}\boldsymbol{u}+\boldsymbol{u}\boldsymbol{R}\boldsymbol{u}\boldsymbol{R} \end{aligned} \tag{12.1}$$

Therefore, in general,

$$|\boldsymbol{R}\cdot\boldsymbol{u}|=|\boldsymbol{R}\wedge\boldsymbol{u}| \tag{12.2}$$

Note that results such as these do not require any choice of basis. By letting $\boldsymbol{u}$ be the local time vector in whatever frame we choose, it then follows that the magnitudes of the temporal and spatial parts of any null vector in any frame are always equal, as we have already inferred for particular examples. This is the very property that guarantees the invariance of the speed of light. If the null vector $\boldsymbol{R}$ connecting two events is taken to be of the form $\boldsymbol{R}=d\boldsymbol{t}+\underset{\sim}{\boldsymbol{d}}$ where $\underset{\sim}{\boldsymbol{d}}$ is purely spatial vector in the $\boldsymbol{t}$-frame, then $\underset{\sim}{\boldsymbol{d}}\cdot\boldsymbol{t}=0$ and the corresponding relative paravector for $\boldsymbol{R}$ in that frame is $\boldsymbol{D}=d+\mathbf{d}$ where $\mathbf{d}=\underset{\sim}{\boldsymbol{d}}\boldsymbol{t}$. The fact that $\boldsymbol{R}$ is a null vector imposes only the constraint that $d^2=\underset{\sim}{\boldsymbol{d}}^2=\mathbf{d}^2$. It is readily checked that this guarantees $\boldsymbol{R}^2=0$, but note that for a forward-pointing null vector, d is positive (that is to say, $\boldsymbol{R}\cdot\boldsymbol{t}$ is negative), whereas for a backward-pointing one, it is negative. We could also state the foregoing in a more general way for any frame, say the $\boldsymbol{\theta}$-frame, as

- $d=-\boldsymbol{R}\cdot\boldsymbol{\theta}$ (in the $(-+++)$ metric signature),
- $\mathbf{d}=\boldsymbol{R}\wedge\boldsymbol{\theta}$, and
- $|\boldsymbol{R}\cdot\boldsymbol{\theta}|=|\boldsymbol{R}\wedge\boldsymbol{\theta}|$.

For example, in the $\boldsymbol{\upsilon}$-frame, we would simply replace $\boldsymbol{t}$ or $\boldsymbol{\theta}$ with $\boldsymbol{\upsilon}$ and express the result as $\boldsymbol{R}=d'\boldsymbol{\upsilon}+\underset{\sim}{\boldsymbol{d}}'$, but we will always have $d'^2=\underset{\sim}{\boldsymbol{d}}'^2=\mathbf{d}'^2$ in any frame. The

primes here are not important as they serve only to distinguish relative quantities in different frames.

From the above, it is clear that under a given choice of frame, the spacetime split of a null vector $\boldsymbol{R}$ is a paravector of the form $d+\mathbf{d}$ where d and $\mathbf{d}$ are the light travel time and the relative vector (the directed distance) respectively, corresponding to $\boldsymbol{R}$. But does a null vector translate into a null paravector? No, since in the case of $\boldsymbol{R} \leftrightarrow d+\mathbf{d}$, we have $(d+\mathbf{d})^2 = 2d(d+\mathbf{d})$, and, given $|d| = |\mathbf{d}|$, this cannot vanish unless $\mathbf{d}$ itself vanishes. However, the product $(d+\mathbf{d})(d-\mathbf{d})$ does vanish, and so paravectors of this form do have a property that is somewhat akin to nullness.

Let us now try to find the relationship between the spacetime splits of a null vector in two different frames. In the $\boldsymbol{t}$-frame and $\boldsymbol{\upsilon}$-frame already referred to, the delays seen in each frame are given by $d = -\boldsymbol{R}\cdot\boldsymbol{t}$ and $d' = -\boldsymbol{R}\cdot\boldsymbol{\upsilon}$, respectively, while the corresponding relative vectors are $\mathbf{d} = \boldsymbol{R}\wedge\boldsymbol{t}$ and $\mathbf{d}' = \boldsymbol{R}\wedge\boldsymbol{\upsilon}$. We can now make use of a general result that applies to any pair of vectors $\boldsymbol{u}$ and $\boldsymbol{w}$. We see that the inner product $\boldsymbol{u}\cdot\boldsymbol{w}$ may be rearranged with the aid of any non-null unit vector $\boldsymbol{\theta}$ (i.e., $\boldsymbol{\theta}^2 = \pm 1$) as follows:

$$\begin{aligned}
\boldsymbol{u}\cdot\boldsymbol{w} &= \tfrac{1}{2}(\boldsymbol{u}\boldsymbol{w}+\boldsymbol{w}\boldsymbol{u}) \\
&= \tfrac{1}{2}\boldsymbol{\theta}^2(\boldsymbol{u}\boldsymbol{\theta}\boldsymbol{\theta}\boldsymbol{w}+\boldsymbol{w}\boldsymbol{\theta}\boldsymbol{\theta}\boldsymbol{u}) \\
&= \tfrac{1}{2}\boldsymbol{\theta}^2((\boldsymbol{u}\cdot\boldsymbol{\theta}+\boldsymbol{u}\wedge\boldsymbol{\theta})(\boldsymbol{w}\cdot\boldsymbol{\theta}-\boldsymbol{w}\wedge\boldsymbol{\theta})+(\boldsymbol{w}\cdot\boldsymbol{\theta}+\boldsymbol{w}\wedge\boldsymbol{\theta})(\boldsymbol{u}\cdot\boldsymbol{\theta}-\boldsymbol{u}\wedge\boldsymbol{\theta})) \\
&= \boldsymbol{\theta}^2(\boldsymbol{u}\cdot\boldsymbol{\theta})(\boldsymbol{w}\cdot\boldsymbol{\theta})-\boldsymbol{\theta}^2(\boldsymbol{u}\wedge\boldsymbol{\theta})\cdot(\boldsymbol{w}\wedge\boldsymbol{\theta})
\end{aligned} \tag{12.3}$$

Taking $\boldsymbol{\theta}$ as the $\boldsymbol{t}$-frame time vector, the right-hand side of here can be stated relative to the $\boldsymbol{\theta}$-frame as

$$\boldsymbol{u}\cdot\boldsymbol{w} = \boldsymbol{\theta}^2\left(uw-\mathbf{u}\cdot\mathbf{w}\right) \tag{12.4}$$

where $u = \boldsymbol{\theta}^2(\boldsymbol{u}\cdot\boldsymbol{\theta})$ and $\mathbf{u} = \boldsymbol{u}\wedge\boldsymbol{\theta}$, and similarly for w and $\mathbf{w}$. But we may also use this as a spacetime split to express $\boldsymbol{u}\cdot\boldsymbol{w}$ in any desired frame simply by replacing $\boldsymbol{\theta}$ with the appropriate time vector. Some interesting results follow, including

- the sign of $\boldsymbol{u}\cdot\boldsymbol{w}$ depends on the chosen signature ($\boldsymbol{\theta}^2 = +1$ or -1);
- if two spacetime vectors $\boldsymbol{u}$ and $\boldsymbol{w}$ are orthogonal, then for some frame $\boldsymbol{\upsilon}$, their relative vectors $\mathbf{u}$ and $\mathbf{w}$ are orthogonal if and only if either $\boldsymbol{u}\cdot\boldsymbol{\upsilon} = 0$ or $\boldsymbol{w}\cdot\boldsymbol{\upsilon} = 0$;
- if the spacetime split of $\boldsymbol{u}$ in the $\boldsymbol{\upsilon}$-frame is $u+\mathbf{u}$, then $\boldsymbol{u}^2 = \boldsymbol{\upsilon}^2\left(u^2-\mathbf{u}^2\right)$; and
- in particular, $\boldsymbol{u}^2 = 0 \Leftrightarrow \mathbf{u}^2 = u^2$, in agreement with our previous result.

The last three results are independent of the chosen metric signature, and the ploy here of inserting a term like $\boldsymbol{\theta}^2$ between products of spacetime vectors often proves fruitful for manipulating such expressions, as we shall again see below.

Let us try to put this to use in finding out how $-\boldsymbol{R}\cdot\boldsymbol{\upsilon}$ and $\boldsymbol{R}\wedge\boldsymbol{\upsilon}$ are related to $-\boldsymbol{R}\cdot\boldsymbol{t}$ and $\boldsymbol{R}\wedge\boldsymbol{t}$. Since Equation (10.5) gives us $\boldsymbol{\upsilon}$ in terms of $\boldsymbol{t}$-frame basis vectors, the spacetime split of $\boldsymbol{\upsilon}$ in the $\boldsymbol{t}$-frame is found to be

$$\begin{aligned} -\boldsymbol{t}\boldsymbol{\upsilon} &= -\boldsymbol{t}\gamma(\boldsymbol{t}+\underset{\sim}{\boldsymbol{v}}) \\ &= \gamma(1+\underset{\sim}{\boldsymbol{v}}\boldsymbol{t}) \\ &= \gamma(1+\mathbf{v}) \end{aligned} \tag{12.5}$$

so that $-\boldsymbol{\upsilon}\cdot\boldsymbol{t}=\gamma$ and $\boldsymbol{\upsilon}\wedge\boldsymbol{t}=\gamma\mathbf{v}$.

It will be helpful at this stage to recall here the two separate definitions of velocity discussed in Sections 7.5 and 10.5. If $\boldsymbol{r}_q$ is some point at rest in the $\boldsymbol{\upsilon}$-frame, then from Equation (7.15), the velocity of $\boldsymbol{r}_q$ *with respect to the* $\boldsymbol{t}$*-frame* is defined by $\boldsymbol{v}=\partial_t\boldsymbol{r}_q$, where t is the $\boldsymbol{t}$-frame's time parameter. In spacetime terms, the vector $\boldsymbol{v}$ therefore corresponds to the usual meaning of velocity. It relates to (3+1)D through the spacetime split $\boldsymbol{v}\leftrightarrow 1+\mathbf{v}$, where $\boldsymbol{v}=\boldsymbol{t}+\underset{\sim}{\boldsymbol{v}}$ and $\mathbf{v}=\underset{\sim}{\boldsymbol{v}}\boldsymbol{t}$ is the relative velocity of $\boldsymbol{r}_q$ with respect to the $\boldsymbol{t}$-frame. On the other hand, the proper velocity $\boldsymbol{\upsilon}$ is defined by Equation (10.4) as $\dot{\boldsymbol{r}}_q=\partial_\tau\boldsymbol{r}_q$, that is, by differentiating not with respect to t but with respect to the $\boldsymbol{\upsilon}$-frame's *own local time* τ. While $\boldsymbol{\upsilon}$ is similarly associated with the motion of the $\boldsymbol{\upsilon}$-frame, it is consequently different from $\boldsymbol{v}$ by the factor of γ, as given in Equation (10.5). Using Equation (12.3), we then have

$$\begin{aligned} d' &= -\boldsymbol{R}\cdot\boldsymbol{\upsilon} \\ &= (\boldsymbol{R}\cdot\boldsymbol{t})(\boldsymbol{\upsilon}\cdot\boldsymbol{t})-(\boldsymbol{R}\wedge\boldsymbol{t})\cdot(\boldsymbol{\upsilon}\wedge\boldsymbol{t}) \\ &= ((d\boldsymbol{t}+\underset{\sim}{\boldsymbol{d}})\cdot\boldsymbol{t})(\gamma(\boldsymbol{t}+\underset{\sim}{\boldsymbol{v}})\cdot\boldsymbol{t})-((d\boldsymbol{t}+\underset{\sim}{\boldsymbol{d}})\wedge\boldsymbol{t})\cdot(\gamma(\boldsymbol{t}+\underset{\sim}{\boldsymbol{v}})\wedge\boldsymbol{t}) \\ &= (-d)(-\gamma)-(\underset{\sim}{\boldsymbol{d}}\boldsymbol{t})\cdot(\gamma\underset{\sim}{\boldsymbol{v}}\boldsymbol{t}) \\ &= \gamma(d-\mathbf{d}\cdot\mathbf{v}) \end{aligned} \tag{12.6}$$

As previously noted, the signs of $\boldsymbol{R}\cdot\boldsymbol{\upsilon}$ and $\boldsymbol{R}\cdot\boldsymbol{t}$ come out opposite to their values in the $(+---)$ signature.

As to $\boldsymbol{R}\wedge\boldsymbol{\upsilon}$, we can follow the same approach. Let us again start with the general case

$$\begin{aligned} \boldsymbol{u}\wedge\boldsymbol{w} &= \tfrac{1}{2}(\boldsymbol{u}\boldsymbol{w}-\boldsymbol{w}\boldsymbol{u}) \\ &= \tfrac{1}{2}\boldsymbol{\theta}^2(\boldsymbol{u}\boldsymbol{\theta}\boldsymbol{\theta}\boldsymbol{w}-\boldsymbol{w}\boldsymbol{\theta}\boldsymbol{\theta}\boldsymbol{u}) \\ &= \tfrac{1}{2}\boldsymbol{\theta}^2((\boldsymbol{u}\cdot\boldsymbol{\theta}+\boldsymbol{u}\wedge\boldsymbol{\theta})(\boldsymbol{w}\cdot\boldsymbol{\theta}-\boldsymbol{w}\wedge\boldsymbol{\theta})-(\boldsymbol{w}\cdot\boldsymbol{\theta}+\boldsymbol{w}\wedge\boldsymbol{\theta})(\boldsymbol{u}\cdot\boldsymbol{\theta}-\boldsymbol{u}\wedge\boldsymbol{\theta})) \\ &= -\boldsymbol{\theta}^2(\boldsymbol{u}\cdot\boldsymbol{\theta})(\boldsymbol{w}\wedge\boldsymbol{\theta})+\boldsymbol{\theta}^2(\boldsymbol{w}\cdot\boldsymbol{\theta})(\boldsymbol{u}\wedge\boldsymbol{\theta})-\tfrac{1}{2}\boldsymbol{\theta}^2[(\boldsymbol{u}\wedge\boldsymbol{\theta})(\boldsymbol{w}\wedge\boldsymbol{\theta})-(\boldsymbol{w}\wedge\boldsymbol{\theta})(\boldsymbol{u}\wedge\boldsymbol{\theta})] \\ &= \boldsymbol{\theta}^2(\boldsymbol{w}\cdot\boldsymbol{\theta})(\boldsymbol{u}\wedge\boldsymbol{\theta})-\boldsymbol{\theta}^2(\boldsymbol{u}\cdot\boldsymbol{\theta})(\boldsymbol{w}\wedge\boldsymbol{\theta})-\boldsymbol{\theta}^2(\boldsymbol{u}\wedge\boldsymbol{\theta})\wedge(\boldsymbol{w}\wedge\boldsymbol{\theta}) \end{aligned} \tag{12.7}$$

As before, taking $\boldsymbol{\theta}$ to be a time vector and using relative vectors and scalars in the $\boldsymbol{\theta}$-frame, this can be written as

$$\boldsymbol{u}\wedge\boldsymbol{w}=(w\mathbf{u}-u\mathbf{w})-\boldsymbol{\theta}^2(\mathbf{u}\wedge\mathbf{w}) \tag{12.8}$$

Note that $\boldsymbol{\theta}^2$ appears in only the bivector part of the result since it is taken into account within the definition of the scalars u and w. This is a reflection of the fact that only in the $(-+++)$ metric signature do the bivectors pass into (3+1)D unchanged. In the $(+---)$ signature, the spacelike bivectors change sign, for example, $\boldsymbol{xy}\leftrightarrow-\mathbf{xy}$ in order to agree with $\mathbf{xy}=\boldsymbol{xtyt}=-\boldsymbol{xy}$.

Applying this in order to find $\boldsymbol{R}\wedge\boldsymbol{\upsilon}$, however, requires some care. In Section 10.6.5, a similar problem was discussed in relation to transforming relative vectors to a different frame. The issue concerned is the potential pitfall of creating unwanted spatial bivector terms, which is avoided by noting that any relative vector perpendicular to the transformation parameter $\mathbf{v}$ is unaffected. As a result, if the vector $\mathbf{r}$ is split into parts $\mathbf{r}_{/\!/}$ and $\mathbf{r}_\perp$ that are parallel and perpendicular to $\mathbf{v}$ respectively, then $\mathbf{r}_\perp=\boldsymbol{r}_\perp\boldsymbol{t}$, where $\boldsymbol{r}_\perp$ is a spacetime vector that is perpendicular to both $\boldsymbol{t}$ and $\boldsymbol{\upsilon}$ (we do not need to write this as $\underset{\sim}{\boldsymbol{r}}_\perp$, which is clumsy). The transformation to the $\boldsymbol{\upsilon}$-frame for the perpendicular part then simply amounts to substituting the new time vector for the old such that $\mathbf{r}_\perp=\boldsymbol{r}_\perp\boldsymbol{t}\mapsto\mathbf{r}'_\perp=\boldsymbol{r}_\perp\boldsymbol{\upsilon}$, but the critical point is that in (3+1)D, the relative vectors $\mathbf{r}'_\perp$ and $\mathbf{r}_\perp$ are to be regarded as exactly the same thing because, from a local perspective at least, time vectors are indistinguishable. Time vector $\boldsymbol{\upsilon}$ is exactly the same thing to the $\boldsymbol{\upsilon}$-frame as time vector $\boldsymbol{t}$ is to the $\boldsymbol{t}$-frame. Within our own rest frame, it does not matter whether our time vector is $\boldsymbol{t}$, $\boldsymbol{\upsilon}$, or anything else—from a (3+1)D perspective, we are no longer connected with it. From the standpoint of each rest frame, $\boldsymbol{t}$, $\boldsymbol{\upsilon}$, and so on, are all the same thing, only going under different labels. The discussion on relative basis vectors in Section 10.6.2 may also prove helpful in getting to grips with this rather tricky point.

In order that that we may properly apply Equation (12.8), therefore, let us partition $\boldsymbol{R}$ into $\boldsymbol{R}_{/\!/}+\boldsymbol{R}_\perp$ where $\boldsymbol{R}_{/\!/}$ lies in the $\boldsymbol{\upsilon}\wedge\boldsymbol{t}$ plane and $\boldsymbol{R}_\perp$ is perpendicular to it. Note that $\boldsymbol{\upsilon}\wedge\boldsymbol{t}\,/\!/\,\boldsymbol{v}\wedge\boldsymbol{t}=\mathbf{v}$, and so this plane is readily identified. The relative vectors for $\boldsymbol{R}_{/\!/}$ and $\boldsymbol{R}_\perp$ are then parallel and perpendicular to $\mathbf{v}$, respectively, both in the $\boldsymbol{t}$-frame and the $\boldsymbol{\upsilon}$-frame. Note also that in both frames $\boldsymbol{R}_\perp$ is purely spatial, while $\boldsymbol{R}_{/\!/}$ includes the time vector, that is $\boldsymbol{R}_\perp\cdot\boldsymbol{\upsilon}=0=\boldsymbol{R}_\perp\cdot\boldsymbol{t}$, whereas $\boldsymbol{R}_{/\!/}\cdot\boldsymbol{\upsilon}\neq0\neq\boldsymbol{R}_{/\!/}\cdot\boldsymbol{t}$.

We are now in a position to apply this to finding $\mathbf{d}'$ in terms of $\boldsymbol{t}$-frame parameters. Following the partitioning of $\boldsymbol{R}$ into $\boldsymbol{R}_{/\!/}$ and $\boldsymbol{R}_\perp$, we have

$$\begin{aligned}
\mathbf{d}'_{/\!/}&=\boldsymbol{R}_{/\!/}\wedge\boldsymbol{\upsilon}\\
&=\boldsymbol{t}^2(\gamma(\boldsymbol{t}+\underset{\sim}{\boldsymbol{v}})\cdot\boldsymbol{t})((d\boldsymbol{t}+\underset{\sim}{\boldsymbol{d}}_{/\!/})\wedge\boldsymbol{t})-\boldsymbol{t}^2((d\boldsymbol{t}+\underset{\sim}{\boldsymbol{d}}_{/\!/})\cdot\boldsymbol{t})(\gamma(\boldsymbol{t}+\underset{\sim}{\boldsymbol{v}})\wedge\boldsymbol{t})\\
&\quad-\boldsymbol{t}^2((d\boldsymbol{t}+\underset{\sim}{\boldsymbol{d}}_{/\!/})\wedge\boldsymbol{t})\wedge(\gamma(\boldsymbol{t}+\underset{\sim}{\boldsymbol{v}})\wedge\boldsymbol{t})\\
&=-(-\gamma)(\underset{\sim}{\boldsymbol{d}}_{/\!/}\wedge\boldsymbol{t})+(-\boldsymbol{d})(\gamma\underset{\sim}{\boldsymbol{v}}\wedge\boldsymbol{t})+(\underset{\sim}{\boldsymbol{d}}_{/\!/}\wedge\boldsymbol{t})\wedge(\gamma\underset{\sim}{\boldsymbol{v}}\wedge\boldsymbol{t})\\
&=(\gamma\mathbf{d}_{/\!/}-\gamma d\mathbf{v}+\gamma\mathbf{d}_{/\!/}\wedge\mathbf{v})\\
&=\gamma(\mathbf{d}_{/\!/}-\gamma d\mathbf{v})
\end{aligned} \tag{12.9}$$

$$\mathbf{d}'_\perp=\mathbf{d}_\perp$$

Here we have $d = -\boldsymbol{R}\cdot\boldsymbol{t}$ as before, but because it is the entire vector $\boldsymbol{R}$ that is null, we have $d^2 = (\boldsymbol{R}\wedge\boldsymbol{t})^2 = \mathbf{d}^2$ rather than $d^2 = \mathbf{d}_{/\!/}^2$. Putting these results together, if some event is seen in the $\boldsymbol{t}$-frame as occurring at $d+\mathbf{d}$, it will be seen in the $\boldsymbol{\upsilon}$-frame as occurring at $d'+\mathbf{d}'$ where, from Equations (12.6) and (12.9),

$$d'+\mathbf{d}' = \gamma\big((d-\mathbf{d}\cdot\mathbf{v})+(\mathbf{d}_{/\!/}-d\mathbf{v})\big)+\mathbf{d}_{\perp} \tag{12.10}$$

Finally, recall the definitions of timelike and future pointing in Section 7.11, but now with $\boldsymbol{\upsilon}$ being the time vector. In fact, in our signature, any unit vector with a negative square will qualify. We can say that in general, a vector $\boldsymbol{u}$ is timelike *and future pointing* if $\boldsymbol{u}\cdot\boldsymbol{\upsilon}$ is nonzero and has the same sign as $\boldsymbol{\upsilon}\cdot\boldsymbol{\upsilon}$.

12.2 FINDING F FOR A MOVING POINT CHARGE

In Section 11.8.1, we found the spacetime form of the electromagnetic potential of a point charge moving with proper velocity $\boldsymbol{\upsilon}$ and, consequently, at rest in the $\boldsymbol{\upsilon}$-frame. Since $\boldsymbol{R}\cdot\boldsymbol{\upsilon}$ is always negative in our signature for any forward-pointing null vector $\boldsymbol{R}$, it will make things more straightforward if we replace $|\boldsymbol{R}\cdot\boldsymbol{\upsilon}|$ with $-\boldsymbol{R}\cdot\boldsymbol{\upsilon}$ thereby eliminating the need to take the modulus in the denominator of equations such as Equation (11.55). We then have

$$\begin{aligned}\boldsymbol{A} &= -\Phi\boldsymbol{\upsilon}\\ &= \frac{q}{4\pi[\varepsilon_0]}\cdot\frac{\boldsymbol{\upsilon}}{\boldsymbol{R}\cdot\boldsymbol{\upsilon}}\end{aligned} \tag{12.11}$$

Since the charge's electromagnetic field is given in general by Equation (11.5) as $\boldsymbol{F} = \boldsymbol{\nabla}\boldsymbol{A}$, we have to evaluate the derivative in the form of

$$\boldsymbol{F} = \boldsymbol{\nabla}\boldsymbol{A} = \frac{q}{4\pi}\boldsymbol{\nabla}\left(\frac{\boldsymbol{\upsilon}}{\boldsymbol{R}\cdot\boldsymbol{\upsilon}}\right) \tag{12.12}$$

Here, as before, $\boldsymbol{R} = \boldsymbol{r}-\boldsymbol{r}_q$ is the future-pointing null vector from the source, q, to $\boldsymbol{r}$, the "observation event" at which the right-hand side of Equation (12.12) is evaluated. It is important to note that $\boldsymbol{r}$ is also the variable with respect to which we are differentiating. We give the charge the trajectory $\boldsymbol{r}_q(\tau)$ in terms of its own proper time τ so that $\boldsymbol{\upsilon} = \partial_\tau\boldsymbol{r}_q = \dot{\boldsymbol{r}}_q$.

Applying the product rule, Equation (7.23), to Equation (12.12) we find

$$\boldsymbol{\nabla}\boldsymbol{A} = \boldsymbol{\nabla}_{\boldsymbol{r}}\boldsymbol{A}(\boldsymbol{r},\boldsymbol{r}_q,\boldsymbol{\upsilon}) = \frac{q}{4\pi}\left[-\frac{\boldsymbol{\nabla}(\boldsymbol{R}\cdot\boldsymbol{\upsilon})\boldsymbol{\upsilon}}{(\boldsymbol{R}\cdot\boldsymbol{\upsilon})^2}+\frac{\boldsymbol{\nabla}\boldsymbol{\upsilon}}{\boldsymbol{R}\cdot\boldsymbol{\upsilon}}\right] \tag{12.13}$$

and so we now only have to differentiate the two simpler expressions $\boldsymbol{\upsilon}$ and $\boldsymbol{R}\cdot\boldsymbol{\upsilon}$. The former is a function of the scalar variable τ, and so the chain rule, Equation

(7.24), may be used, that is, $\boldsymbol{\nabla}\boldsymbol{\upsilon} = (\boldsymbol{\nabla}\tau)\dot{\boldsymbol{\upsilon}}$. The expression $\boldsymbol{\nabla}(\boldsymbol{R}\cdot\boldsymbol{\upsilon})$, however, requires some simplification:

$$\begin{aligned}\boldsymbol{\nabla}(\boldsymbol{R}\cdot\boldsymbol{\upsilon}) &= \boldsymbol{\nabla}\left[\tfrac{1}{2}(\boldsymbol{R}\boldsymbol{\upsilon}+\boldsymbol{\upsilon}\boldsymbol{R})\right] \\ &= \tfrac{1}{2}\left[(\boldsymbol{\nabla}\boldsymbol{R})\boldsymbol{\upsilon}+\mathring{\boldsymbol{\nabla}}\boldsymbol{R}\mathring{\boldsymbol{\upsilon}}+(\boldsymbol{\nabla}\boldsymbol{\upsilon})\boldsymbol{R}+\mathring{\boldsymbol{\nabla}}\boldsymbol{\upsilon}\mathring{\boldsymbol{R}}\right]\end{aligned} \tag{12.14}$$

We now need expressions for the individual terms, for which the other identities in Section 7.9 come in useful, so that we have

$$\begin{aligned}(\boldsymbol{\nabla}\boldsymbol{R})\boldsymbol{\upsilon} &= \boldsymbol{\nabla}(\boldsymbol{r}-\boldsymbol{r}_q)\boldsymbol{\upsilon} \\ &= (4-\boldsymbol{\nabla}\tau\,\boldsymbol{\upsilon})\boldsymbol{\upsilon} \\ &= 4\boldsymbol{\upsilon}+\boldsymbol{\nabla}\tau \\ \\ \mathring{\boldsymbol{\nabla}}\boldsymbol{R}\mathring{\boldsymbol{\upsilon}} &= \boldsymbol{\nabla}\tau\,\boldsymbol{R}\dot{\boldsymbol{\upsilon}} \\ \\ (\boldsymbol{\nabla}\boldsymbol{\upsilon})\boldsymbol{R} &= \boldsymbol{\nabla}\tau\,\dot{\boldsymbol{\upsilon}}\boldsymbol{R} \\ \\ \mathring{\boldsymbol{\nabla}}\boldsymbol{R}\mathring{\boldsymbol{r}} &= -2\boldsymbol{R} \\ \\ \mathring{\boldsymbol{\nabla}}\boldsymbol{\upsilon}\mathring{\boldsymbol{R}} &= \mathring{\boldsymbol{\nabla}}\boldsymbol{\upsilon}\mathring{\boldsymbol{r}}-\mathring{\boldsymbol{\nabla}}\boldsymbol{\upsilon}\mathring{\boldsymbol{r}}_q \\ &= -2\boldsymbol{\upsilon}-\boldsymbol{\nabla}\tau\,\boldsymbol{\upsilon}^2 \\ &= -2\boldsymbol{\upsilon}+\boldsymbol{\nabla}\tau\end{aligned} \tag{12.15}$$

We have used here the standard result $\boldsymbol{\nabla}\boldsymbol{r} = 4$ and the chain rule comes in useful again, for example, for $\boldsymbol{\nabla}\boldsymbol{r}_q = (\boldsymbol{\nabla}\tau)\partial_\tau\boldsymbol{r}_q = (\boldsymbol{\nabla}\tau)\boldsymbol{\upsilon}$, and $\mathring{\boldsymbol{\nabla}}\boldsymbol{R}\mathring{\boldsymbol{\upsilon}} = \boldsymbol{\nabla}\tau\,\boldsymbol{R}\dot{\boldsymbol{\upsilon}}$, while both $\mathring{\boldsymbol{\nabla}}\boldsymbol{R}\mathring{\boldsymbol{r}} = -2\boldsymbol{R}$ and $\mathring{\boldsymbol{\nabla}}\boldsymbol{\upsilon}\mathring{\boldsymbol{r}} = -2\boldsymbol{\upsilon}$ come from Equation (7.27). An expression for $\boldsymbol{\nabla}\tau$, however, has to be obtained by the somewhat devious route of differentiating $\boldsymbol{\nabla}\left(\boldsymbol{R}^2\right)$ [8]. Since $\boldsymbol{R}$ is null, this of course must evaluate to zero. With the aid of this crucial starting point, however, the actual working is once again quite straightforward:

$$\begin{aligned}0 &= \boldsymbol{\nabla}\left(\boldsymbol{R}^2\right) \\ &= (\boldsymbol{\nabla}\boldsymbol{R})\boldsymbol{R}+\mathring{\boldsymbol{\nabla}}\boldsymbol{R}\mathring{\boldsymbol{R}} \\ &= \boldsymbol{\nabla}(\boldsymbol{r}-\boldsymbol{r}_q)\boldsymbol{R}+\mathring{\boldsymbol{\nabla}}\boldsymbol{R}\left(\mathring{\boldsymbol{r}}-\mathring{\boldsymbol{r}}_q\right) \\ &= (4-\boldsymbol{\nabla}\tau\,\boldsymbol{\upsilon})\boldsymbol{R}-2\boldsymbol{R}-\boldsymbol{\nabla}\tau\boldsymbol{R}\boldsymbol{\upsilon} \\ &= 2\boldsymbol{R}-2\boldsymbol{\nabla}\tau(\boldsymbol{R}\cdot\boldsymbol{\upsilon}) \\ \\ \Rightarrow \boldsymbol{\nabla}\tau &= \frac{\boldsymbol{R}}{\boldsymbol{R}\cdot\boldsymbol{\upsilon}}\end{aligned} \tag{12.16}$$

We may now use these results in Equation (12.14) to find

$$\begin{aligned}\nabla(\boldsymbol{R}\cdot\boldsymbol{\upsilon}) &= \tfrac{1}{2}[(4\boldsymbol{\upsilon}+\nabla\tau)+\nabla\tau\boldsymbol{R}\dot{\boldsymbol{\upsilon}}+\nabla\tau\dot{\boldsymbol{\upsilon}}\boldsymbol{R}+(-2\boldsymbol{\upsilon}+\nabla\tau)] \\ &= \boldsymbol{\upsilon}+\tfrac{1}{2}\nabla\tau(2+\boldsymbol{R}\dot{\boldsymbol{\upsilon}}+\dot{\boldsymbol{\upsilon}}\boldsymbol{R}) \\ &= \frac{\boldsymbol{\upsilon}(\boldsymbol{R}\cdot\boldsymbol{\upsilon})}{\boldsymbol{R}\cdot\boldsymbol{\upsilon}}+\frac{1}{2}\frac{2\boldsymbol{R}+\boldsymbol{R}(\boldsymbol{R}\dot{\boldsymbol{\upsilon}}+\dot{\boldsymbol{\upsilon}}\boldsymbol{R})}{\boldsymbol{R}\cdot\boldsymbol{\upsilon}} \\ &= \frac{\boldsymbol{\upsilon}\boldsymbol{R}\boldsymbol{\upsilon}+\boldsymbol{\upsilon}^2\boldsymbol{R}+2\boldsymbol{R}+\boldsymbol{R}\dot{\boldsymbol{\upsilon}}\boldsymbol{R}}{2\boldsymbol{R}\cdot\boldsymbol{\upsilon}} \\ &= \frac{\boldsymbol{\upsilon}\boldsymbol{R}\boldsymbol{\upsilon}+\boldsymbol{R}+\boldsymbol{R}\dot{\boldsymbol{\upsilon}}\boldsymbol{R}}{2\boldsymbol{R}\cdot\boldsymbol{\upsilon}}\end{aligned} \tag{12.17}$$

This in turn may be substituted into Equation (12.13), and on carrying this through, we obtain

$$\begin{aligned}\nabla\boldsymbol{A} &= \frac{q}{4\pi}\left[\frac{\boldsymbol{R}\dot{\boldsymbol{\upsilon}}}{(\boldsymbol{R}\cdot\boldsymbol{\upsilon})^2}-\frac{(\boldsymbol{\upsilon}\boldsymbol{R}\boldsymbol{\upsilon}+\boldsymbol{R}+\boldsymbol{R}\dot{\boldsymbol{\upsilon}}\boldsymbol{R})\boldsymbol{\upsilon}}{2(\boldsymbol{R}\cdot\boldsymbol{\upsilon})^3}\right] \\ &= \frac{q}{4\pi}\left[\frac{\boldsymbol{R}\dot{\boldsymbol{\upsilon}}(\boldsymbol{R}\boldsymbol{\upsilon}+\boldsymbol{\upsilon}\boldsymbol{R})}{2(\boldsymbol{R}\cdot\boldsymbol{\upsilon})^3}-\frac{-\boldsymbol{\upsilon}\boldsymbol{R}+\boldsymbol{R}\boldsymbol{\upsilon}+\boldsymbol{R}\dot{\boldsymbol{\upsilon}}\boldsymbol{R}\boldsymbol{\upsilon}}{2(\boldsymbol{R}\cdot\boldsymbol{\upsilon})^3}\right] \\ &= \frac{q}{4\pi}\left[\frac{\boldsymbol{R}\dot{\boldsymbol{\upsilon}}\boldsymbol{\upsilon}\boldsymbol{R}}{2(\boldsymbol{R}\cdot\boldsymbol{\upsilon})^3}-\frac{(\boldsymbol{R}\boldsymbol{\upsilon}-\boldsymbol{\upsilon}\boldsymbol{R})}{2(\boldsymbol{R}\cdot\boldsymbol{\upsilon})^3}\right] \\ &= \frac{q}{4\pi}\left[\frac{\boldsymbol{R}\dot{\boldsymbol{\upsilon}}\boldsymbol{\upsilon}\boldsymbol{R}}{2(\boldsymbol{R}\cdot\boldsymbol{\upsilon})^3}-\frac{\boldsymbol{R}\wedge\boldsymbol{\upsilon}}{(\boldsymbol{R}\cdot\boldsymbol{\upsilon})^3}\right]\end{aligned} \tag{12.18}$$

Whereas $\dot{\boldsymbol{\upsilon}}$ denotes the proper acceleration, it is helpful here to employ the acceleration bivector $\boldsymbol{\Omega}\equiv\dot{\boldsymbol{\upsilon}}\boldsymbol{\upsilon}$ (see Section 11.6) since it equates to the relative vector representing the acceleration $\mathbf{a}'$ experienced by the charge in its own rest frame. This quantity is of obvious physical significance. As to Equation (12.18), it is clear that only the contribution involving $\dot{\boldsymbol{\upsilon}}\boldsymbol{\upsilon}$ can be relevant to the radiated part of the field, which we may denote by $\boldsymbol{F}_{rad}$, because we know that the other contribution involving $\boldsymbol{R}\wedge\boldsymbol{\upsilon}$ is simply the field for the case of constant velocity (Equation 11.65), which we may refer to as the quasistatic field, $\boldsymbol{F}_{qs}$. Since $\partial_t\boldsymbol{F}_{qs}$ is not essentially zero, it would be misleading to refer to it as truly static, but in any case, as is well known, no radiation is emitted by a point charge in uniform motion. Radiation is associated with a $1/\mathbf{r}^2$ dependency of the radiated power, and consequently, the magnitude of the radiated field should decrease as $1/|\mathbf{r}|$. To check the situation here, let us recall that $|\boldsymbol{R}\wedge\boldsymbol{\upsilon}|=|\boldsymbol{R}\cdot\boldsymbol{\upsilon}|=d'$, and $\boldsymbol{R}\wedge\boldsymbol{\upsilon}$ is simply the relative vector $\mathbf{r}'-\mathbf{r}'_q$ from the charge to the observer as seen in the $\boldsymbol{\upsilon}$-frame at the retarded proper time

$\tau = -\boldsymbol{\upsilon}\cdot\boldsymbol{r}_q$. The numerator $\boldsymbol{R}\dot{\boldsymbol{\upsilon}}\boldsymbol{\upsilon}\boldsymbol{R}$ in $\boldsymbol{F}_{rad}$, however, is a null bivector, and so we cannot readily use $(\boldsymbol{R}\dot{\boldsymbol{\upsilon}}\boldsymbol{\upsilon}\boldsymbol{R})^2$ as a means of estimating its magnitude. Nevertheless, by using $\boldsymbol{R}\boldsymbol{\Omega}\boldsymbol{R} = -\boldsymbol{R}\boldsymbol{\upsilon}^2\,\dot{\boldsymbol{\upsilon}}\boldsymbol{\upsilon}\boldsymbol{R} = (\boldsymbol{R}\boldsymbol{\upsilon})\boldsymbol{\Omega}(\boldsymbol{\upsilon}\boldsymbol{R})$ together with $\boldsymbol{R}\boldsymbol{\upsilon} = (\boldsymbol{R}\cdot\boldsymbol{\upsilon} + \boldsymbol{R}\wedge\boldsymbol{\upsilon})$ and $\boldsymbol{\upsilon}\boldsymbol{R} = (\boldsymbol{R}\cdot\boldsymbol{\upsilon} - \boldsymbol{R}\wedge\boldsymbol{\upsilon})$, we may write it as the sum of four terms:

$$\begin{aligned}\boldsymbol{R}\boldsymbol{\Omega}\boldsymbol{R} = {} & (\boldsymbol{R}\cdot\boldsymbol{\upsilon})\boldsymbol{\Omega}(\boldsymbol{R}\cdot\boldsymbol{\upsilon}) - (\boldsymbol{R}\cdot\boldsymbol{\upsilon})\boldsymbol{\Omega}(\boldsymbol{R}\wedge\boldsymbol{\upsilon}) \\ & + (\boldsymbol{R}\wedge\boldsymbol{\upsilon})\boldsymbol{\Omega}(\boldsymbol{R}\cdot\boldsymbol{\upsilon}) - (\boldsymbol{R}\wedge\boldsymbol{\upsilon})\boldsymbol{\Omega}(\boldsymbol{R}\wedge\boldsymbol{\upsilon})\end{aligned} \tag{12.19}$$

Since $|\boldsymbol{R}\wedge\boldsymbol{\upsilon}| = |\boldsymbol{R}\cdot\boldsymbol{\upsilon}| = d'$, the magnitude of each of the four terms in this expression depends on d'^2. Taking the denominator into account, this result is sufficient to establish that $\boldsymbol{F}_{rad}$ must depend on $1/d'$, as required.

Summarizing, our result is

$$\begin{aligned}\boldsymbol{F}(\boldsymbol{r},\boldsymbol{r}_q,\boldsymbol{\upsilon},\boldsymbol{\Omega}) &= \boldsymbol{F}_{rad} + \boldsymbol{F}_{qs} \\ \boldsymbol{F}_{rad} &= \frac{1}{8\pi}\frac{\boldsymbol{R}\boldsymbol{\Omega}\boldsymbol{R}}{(\boldsymbol{R}\cdot\boldsymbol{\upsilon})^3} \\ \boldsymbol{F}_{qs} &= \frac{-1}{4\pi}\frac{\boldsymbol{R}\wedge\boldsymbol{\upsilon}}{(\boldsymbol{R}\cdot\boldsymbol{\upsilon})^3}\end{aligned} \tag{12.20}$$

The negative sign arises in the expression for $\boldsymbol{F}_{qs}$ because $\boldsymbol{R}\cdot\boldsymbol{\upsilon}$, as we have shown, is negative in the $(-+++)$ metric signature while it is positive for $(+---)$. On the other hand, from Equation (12.8), the sign of the timelike bivector part of $\boldsymbol{R}\wedge\boldsymbol{\upsilon}$ is the same in both signatures, while in contrast, the sign of spacelike bivector part changes. In $\boldsymbol{F}_{qs}$, the timelike bivector part of $\boldsymbol{R}\wedge\boldsymbol{\upsilon}$ is associated with $\boldsymbol{E}$, while the spacelike bivector part is associated with $\boldsymbol{B}$, so that the behavior of the signs here is entirely consistent with the fact that $\boldsymbol{F}$ is given by $\boldsymbol{E}-\boldsymbol{B}$ in one signature and $\boldsymbol{E}+\boldsymbol{B}$ in the other. As to $\boldsymbol{F}_{rad}$, while the expression given above has the same sign in both signatures, we still have $\boldsymbol{R}\cdot\boldsymbol{\upsilon}$ resulting in a negative value in our case while it is positive in the other. This can only be the case if the sign of $\boldsymbol{R}\boldsymbol{\Omega}\boldsymbol{R}$ also turns out differently. In the $(+---)$ metric signature, $\boldsymbol{R}\boldsymbol{\Omega}\boldsymbol{R} = \boldsymbol{R}\boldsymbol{\upsilon}^2\,\dot{\boldsymbol{\upsilon}}\boldsymbol{\upsilon}\boldsymbol{R} = -(\boldsymbol{R}\boldsymbol{\upsilon})\boldsymbol{\Omega}(\boldsymbol{\upsilon}\boldsymbol{R})$ so that this is indeed the case. Here the issue of signs is therefore quite tricky, but it is just another example of the type of nuance that occurs as a result of having different forms of mixed metric signatures. In the end, referring to Equation (12.2) and allowing for the fact $\boldsymbol{R}\cdot\boldsymbol{\upsilon}$ is negative, we may replace it with $-|\boldsymbol{R}\wedge\boldsymbol{\upsilon}|$ so that we have in both signatures

$$\boldsymbol{F}_{qs} = \frac{q}{4\pi}\frac{\boldsymbol{R}\wedge\boldsymbol{\upsilon}}{|\boldsymbol{R}\wedge\boldsymbol{\upsilon}|^3} \tag{12.21}$$

Now since $\boldsymbol{R}\wedge\boldsymbol{\upsilon}$ is none other than the relative vector $\mathbf{r}' - \mathbf{r}'_q$ between observer and charge as seen in the $\boldsymbol{\upsilon}$-frame, Equation (12.21) is easily recognized as the

charge's Coulomb field expressed in its own rest frame, our starting point when we set out in the other direction to find the field of a charge moving with constant velocity.

Recall that in the rest frame of a charge moving with uniform velocity, only the Coulomb field is observed, nothing else. The quantitative form of the spacetime split of $\boldsymbol{F}_{qs}$ is therefore only of interest in some other frame, say the $\boldsymbol{t}$-frame, and this has already been discussed. Similarly, the usual approach to the spacetime split of $\boldsymbol{F}_{rad}$ is to take it from the observer's viewpoint, and the standard relativistically correct result for this obtained by means of traditional vector analysis is given by Jackson [37, section 14.1, p. 467]. But given the charge is *accelerating*, what will *it* observe? If it radiates, then it can no longer see its original Coulomb field. We therefore explore both situations.

12.3 $\boldsymbol{F}_{rad}$ IN THE CHARGE'S REST FRAME

Recall that since $\boldsymbol{F}$ itself is frame independent, $\boldsymbol{F}_{rad}$ and $\boldsymbol{F}'_{rad}$ are the same thing, and the only thing that changes is how $\boldsymbol{F}_{rad}$ splits into $\boldsymbol{E}'_{rad} - \boldsymbol{B}'_{rad}$ in the charge's rest frame as compared with $\boldsymbol{E}_{rad} - \boldsymbol{B}_{rad}$ in the observer's rest frame. To find how $\boldsymbol{F}_{rad}$ splits into $\boldsymbol{E}'_{rad} - \boldsymbol{B}'_{rad}$, we must therefore express Equation (12.19) in terms of the relative vectors and scalars appropriate to the $\boldsymbol{\upsilon}$-frame. Since relative vectors are by definition frame dependent, we distinguish those of the $\boldsymbol{\upsilon}$-frame from the observer's rest frame, the $\boldsymbol{t}$-frame, by using the customary prime. Because $\boldsymbol{\Omega} = \dot{\boldsymbol{\upsilon}}\boldsymbol{\upsilon}$ is a timelike bivector, it translates directly into a relative vector so that as previously mentioned, $\boldsymbol{\Omega}$ equates to $\mathbf{a}'$, the acceleration seen from the charge's viewpoint, while for the vector $\boldsymbol{R}$, we have $\boldsymbol{R} \leftrightarrow -\boldsymbol{R}\cdot\boldsymbol{\upsilon} + \boldsymbol{R}\wedge\boldsymbol{\upsilon} = d' + \mathbf{d}'$ as discussed above. All these relative quantities require to be measured in the $\boldsymbol{\upsilon}$-frame at a time τ such that the result would be observed in the $\boldsymbol{t}$-frame at $t + \mathbf{r}$. As previously discussed, starting from $t + \mathbf{r}$, there will generally be no closed form for $\tau + \mathbf{r}_q$ as this depends on the given trajectory of the charge, and so we take the simpler approach starting from $\tau + \mathbf{r}'_q$, treating $\mathbf{r}$ as fixed and then finding the time t corresponding to a given value of τ as in Figures 11.1 and 11.2.

On substituting these relative quantities into Equation (12.19) in place of $\boldsymbol{R}\cdot\boldsymbol{\upsilon}$, $\boldsymbol{R}\wedge\boldsymbol{\upsilon}$, and $\boldsymbol{\Omega}$, we find

$$\begin{aligned}\boldsymbol{R}\boldsymbol{\Omega}\boldsymbol{R} &= (\boldsymbol{R}\cdot\boldsymbol{\upsilon})\boldsymbol{\Omega}(\boldsymbol{R}\cdot\boldsymbol{\upsilon}) - (\boldsymbol{R}\cdot\boldsymbol{\upsilon})\boldsymbol{\Omega}(\boldsymbol{R}\wedge\boldsymbol{\upsilon}) + (\boldsymbol{R}\wedge\boldsymbol{\upsilon})\boldsymbol{\Omega}(\boldsymbol{R}\cdot\boldsymbol{\upsilon}) - (\boldsymbol{R}\wedge\boldsymbol{\upsilon})\boldsymbol{\Omega}(\boldsymbol{R}\wedge\boldsymbol{\upsilon}) \\ &= d'^2\mathbf{a}' + 2d'\mathbf{a}'\wedge\mathbf{d}' - \mathbf{d}'\mathbf{a}'\mathbf{d}' \\ &= d'^2\mathbf{a}' + 2d'\mathbf{a}'\wedge\mathbf{d}' - \mathbf{d}'(2\mathbf{a}'\cdot\mathbf{d}' - \mathbf{d}'\mathbf{a}') \\ &= 2\left(d'^2\mathbf{a}' - (\mathbf{a}'\cdot\mathbf{d}')\mathbf{d}' + d'(\mathbf{a}'\wedge\mathbf{d}')\right)\end{aligned} \tag{12.22}$$

As this turns out to be in the form of relative vector plus bivector, the vector part may readily be associated with $\boldsymbol{E}'_{rad}$ and the bivector part with $-\boldsymbol{B}'_{rad}$ so that after taking care to note that $R\cdot\boldsymbol{\upsilon} = -d'$, Equation (12.20) gives us

$$\begin{aligned}\boldsymbol{F}_{rad} = \boldsymbol{E}'_{rad} - \boldsymbol{B}'_{rad} &= \frac{q}{4\pi}\cdot\frac{d'^2\mathbf{a}' - (\mathbf{a}'\cdot\mathbf{d}')\mathbf{d}' + d'(\mathbf{a}'\wedge\mathbf{d}')}{-d'^3} \\ &= \underbrace{\frac{q}{4\pi}\cdot\frac{(\mathbf{a}'\cdot\mathbf{d}')\mathbf{d}' - d'^2\mathbf{a}'}{d'^3}}_{\mathbf{E}'_{rad}} - \underbrace{\frac{q}{4\pi}\cdot\frac{(\mathbf{a}'\wedge\mathbf{d}')}{d'^2}}_{\boldsymbol{B}'_{rad}}\end{aligned} \tag{12.23}$$

On the right-hand side here, we have simply replaced $\boldsymbol{E}'_{rad}$ with its (3+1)D counterpart $\mathbf{E}'_{rad}$. Now in the case of $\mathbf{E}'_{rad}$, it turns out that $(\mathbf{a}'\cdot\mathbf{d}')\mathbf{d}' - d'^2\mathbf{a}'$ simplifies to $-d'^2\mathbf{a}'_\perp$ where $\mathbf{a}'_\perp$ is the part of $\mathbf{a}'$ that is perpendicular to $\mathbf{d}'$, the apparent vector from the charge to the point of observation, while $\boldsymbol{B}'_{rad}$ is readily expressed as $\hat{\mathbf{d}}'\wedge\mathbf{E}'_{rad}$. The final result may be stated quite simply (with the supressed factors of ε_0 and c restored):

$$\begin{aligned}\mathbf{E}'_{rad} &= \frac{-q\mathbf{a}'_\perp}{4\pi[\varepsilon_0 c^2]d'} \\ \boldsymbol{B}'_{rad} &= \left[\tfrac{1}{c}\right]\hat{\mathbf{d}}'\wedge\mathbf{E}'_{rad}\end{aligned} \tag{12.24}$$

While in this result we could have replaced $\left(\varepsilon_0 c^2\right)^{-1}$ with μ_0, it is easier to check it dimensionally in the form given. The dimensions of $\varepsilon_0 E'_{rad}$ are $[Q][L]^{-2}$, and it is really verified that the dimensions of $\mathbf{a}'_\perp/c^2 d'$ are $\left[Q\right]\left[LT^{-2}\right]\left[L\right]^{-1}\left[LT^{-1}\right]^{-2} = \left[Q\right]\left[L\right]^{-2}$ as required.

Equation (12.24) agrees with Jackson's result with $\boldsymbol{\beta} = 0$ and $\mathbf{n}\times(\mathbf{n}\times\dot{\boldsymbol{\beta}}) \equiv -\mathbf{a}'_\perp$. As discussed, the devil in the detail is that $\mathbf{d}'$ must be evaluated relativistically from the particle's trajectory in relation to the observer's. It is clear from this result that the distinction between the radiation and quasistatic fields from a point source lies not only in the fact that the electric part of the field depends on $1/d$ rather than $1/d^2$, but that the magnetic part of the field is given by $\hat{\mathbf{d}}\wedge\mathbf{E}$ rather than $\mathbf{v}\wedge\mathbf{E}$. The $\hat{\mathbf{d}}\wedge\mathbf{E}$ dependence is, however, consistent with the behavior of plane waves, Equation (5.21b) with $\hat{\mathbf{k}} = \hat{\mathbf{d}}$, so that plane waves are consistent with radiation rather than some sort of quasistatic field. The fact that that their electric field is independent of d is only due to the fact that they correspond to an extended plane source rather than a point source. All sources have finite dimensions, however, and ultimately, when d is very much greater than the dimensions of the source, the $1/d$ dependence is restored.

From a physical viewpoint, it is of great significance that an accelerating charge radiates irrespective of its observed velocity. But this must be so, because a little thought on the subject assures us that radiation cannot be made to disappear simply by a change of inertial reference frame—once the radiation has left the source, it cannot be put back again. Nor is it necessary for charge to be oscillating in order to radiate, as is generally assumed to be the case, this only happens to be one of the most common circumstances. The charge in a spark gap undergoes acceleration for the duration of its transit across the gap; lighting is the same phenomenon on a grand

scale; a pulse of current in an electrical circuit must also have accelerating electrons within it; the electron beam in a cathode ray tube accelerates toward the anode before abruptly decelerating on impact with it, not to mention the deflection it undergoes to make it scan. All of these, to a greater or lesser extent, are sources of electromagnetic radiation that we normally see from their contributions to radio interference.

A charge undergoing circular motion in some orbital plane can be thought of as undergoing simple harmonic motion in two dimensions, but from the charge's point of view, it is simply accelerating under the influence of whatever centripetal force compels it to remain in orbit. Consider what an observer would see from the center of the charge's orbit. The radius of the orbit therefore corresponds to d', and to keep matters very simple, we assume that the time that light takes to travel from the charge to the observer is much less than the orbital period. If they look roughly in the direction of the charge, the observer will see the magnitude of the charge's Coulomb field reduced by $|\boldsymbol{E}'_{rad}|$, that is to say, by a factor of $[1/c^2]|a'_\perp d'|$, which simply reduces to v^2/c^2. But this observer sees this "radiated field" as a static field! And, if it can actually be classed as radiation, it also appears to be longitudinal. On the other hand, if the observer does not look in the direction of the charge but looks along some fixed direction in our $\boldsymbol{t}$-frame, $\boldsymbol{E}'_{rad}$ is harmonic since its polarization rotates at the same frequency as the charge's orbit.

12.4 F_{rad} IN THE OBSERVER'S REST FRAME

To evaluate $\boldsymbol{R\Omega R}$ in the observer's frame, we may once more employ the technique used in Equations (12.3) and (12.7):

$$\begin{aligned}\boldsymbol{R\Omega R} &= \boldsymbol{R}(\boldsymbol{tt})\dot{\boldsymbol{\upsilon}}\boldsymbol{\upsilon}(\boldsymbol{tt})\boldsymbol{R}\\ &= (\boldsymbol{Rt})(-\boldsymbol{t}\dot{\boldsymbol{\upsilon}})(\boldsymbol{\upsilon t})(-\boldsymbol{tR})\end{aligned} \tag{12.25}$$

Taking the factors one at a time,

$$\begin{aligned}-\boldsymbol{tR} &= \mathbf{d} + d\\ \boldsymbol{Rt} &= \mathbf{d} - d\end{aligned} \tag{12.26}$$

$$\begin{aligned}\boldsymbol{\upsilon t} &= \gamma(\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}})\boldsymbol{t}\\ &= \gamma(\mathbf{v} - 1)\end{aligned} \tag{12.27}$$

The acceleration term may be dealt with following Hestenes [51]:

$$\begin{aligned}-\boldsymbol{t}\dot{\boldsymbol{\upsilon}} &= -\partial_\tau(\boldsymbol{t\upsilon})\\ &= -\partial_\tau(\boldsymbol{t}\gamma(\boldsymbol{t} + \underset{\sim}{\boldsymbol{v}}))\\ &= \partial_\tau(\gamma(1 + \mathbf{v}))\\ &= \dot{\gamma}(\mathbf{v} + 1) + \gamma\dot{\mathbf{v}}\end{aligned} \tag{12.28}$$

Putting the last two results together,

$$\begin{aligned}(-\boldsymbol{t}\,\dot{\boldsymbol{\upsilon}})(\boldsymbol{\upsilon t}) &= (\dot\gamma(\mathbf{v}+1)+\gamma\dot{\mathbf{v}})\gamma(\mathbf{v}-1)\\ &= \gamma\dot\gamma\left(v^2-1\right)+\gamma^2\,\dot{\mathbf{v}}(\mathbf{v}-1)\\ &= \underbrace{\gamma\dot\gamma\left(v^2-1\right)+\gamma^2\,\dot{\mathbf{v}}\cdot\mathbf{v}}_{\text{scalar}} - \underbrace{\gamma^2(\dot{\mathbf{v}}-\dot{\mathbf{v}}\wedge\mathbf{v})}_{\text{vector + bivector}}\end{aligned} \tag{12.29}$$

By the same route, however, we find

$$\begin{aligned}\dot{\boldsymbol{\upsilon}}\boldsymbol{t} &= \partial_\tau(\boldsymbol{\upsilon t})\\ &= \partial_\tau(\gamma(\boldsymbol{t}+\underline{\boldsymbol{v}})\boldsymbol{t})\\ &= \partial_\tau(\gamma(-1+\mathbf{v}))\\ &= \dot\gamma(\mathbf{v}-1)+\gamma\dot{\mathbf{v}}\end{aligned} \tag{12.30}$$

so that

$$\begin{aligned}-\boldsymbol{t}\boldsymbol{\upsilon}\,\dot{\boldsymbol{\upsilon}}\boldsymbol{t} &= -\gamma(\mathbf{v}+1)(\dot\gamma(\mathbf{v}-1)+\gamma\dot{\mathbf{v}})\\ &= \gamma\dot\gamma\left(v^2-1\right)+\gamma^2(\mathbf{v}+1)\dot{\mathbf{v}}\\ &= \underbrace{\gamma\dot\gamma\left(v^2-1\right)+\gamma^2\mathbf{v}\cdot\dot{\mathbf{v}}}_{\text{scalar}} + \underbrace{\gamma^2(\dot{\mathbf{v}}-\dot{\mathbf{v}}\wedge\mathbf{v})}_{\text{vector}+\text{bivector}}\end{aligned} \tag{12.31}$$

Since $\dot{\boldsymbol{\upsilon}}\perp\boldsymbol{\upsilon}$, we must have $-\boldsymbol{t}\,\dot{\boldsymbol{\upsilon}}\boldsymbol{\upsilon t}=\boldsymbol{t}\boldsymbol{\upsilon}\,\dot{\boldsymbol{\upsilon}}\boldsymbol{t}$, a fact that leads to the bottom lines of Equations (12.29) and (12.31) of necessity being equal and opposite, and this can only be the case if their scalar parts vanish since these are in fact equal but of the same sign. This requires

$$\begin{aligned}&\gamma\dot\gamma\left(v^2-1\right)+\gamma^2\mathbf{v}\cdot\dot{\mathbf{v}}=0\\ \Leftrightarrow\ &\dot\gamma=\frac{\gamma\mathbf{v}\cdot\dot{\mathbf{v}}}{\left(1-v^2\right)}=\gamma^3\mathbf{v}\cdot\dot{\mathbf{v}}\end{aligned} \tag{12.32}$$

which is an interesting but purely incidental result. The result we require therefore reduces to the vector + bivector part of Equation (12.29),

$$(-\boldsymbol{t}\,\dot{\boldsymbol{\upsilon}})(\boldsymbol{\upsilon t})=-\gamma^3\left(\mathbf{a}-\mathbf{a}\wedge\mathbf{v}\right) \tag{12.33}$$

where we have taken the opportunity to substitute $\gamma\mathbf{a}$ for $\dot{\mathbf{v}}$ where $\mathbf{a}$ is the acceleration as seen in the $\boldsymbol{t}$-frame, as opposed to the $\boldsymbol{\upsilon}$-frame, and is therefore different from $\dot{\boldsymbol{\upsilon}}$ by a factor of $\partial_\tau t=\gamma$.

We can now introduce the other terms from Equation (12.26). This leads to

$$\begin{aligned}\boldsymbol{R\Omega R} &= (\boldsymbol{Rt})(-\boldsymbol{t}\,\dot{\mathbf{v}})(\mathbf{v}\boldsymbol{t})(-\boldsymbol{tR}) \\ &= -(\mathbf{d}-d)\left(\gamma^3(\mathbf{a}-\mathbf{a}\wedge\mathbf{v})\right)(\mathbf{d}+d) \\ &= \gamma^3\left(-\mathbf{dad}+d\mathbf{ad}-d\mathbf{da}+d^2\mathbf{a}+\mathbf{d}(\mathbf{a}\wedge\mathbf{v})\mathbf{d}-d(\mathbf{a}\wedge\mathbf{v})\mathbf{d}+d\mathbf{d}(\mathbf{a}\wedge\mathbf{v})-d^2\mathbf{a}\wedge\mathbf{v}\right)\end{aligned} \tag{12.34}$$

The first term within the main brackets may be simplified to result in a pure vector,

$$\begin{aligned}\mathbf{dad} &= (2\mathbf{a}\cdot\mathbf{d}-\mathbf{ad})\mathbf{d} \\ &= 2(\mathbf{a}\cdot\mathbf{d})\mathbf{d}-d^2\mathbf{a}\end{aligned} \tag{12.35}$$

while, for the fifth and eighth terms, rearranging $\mathbf{d}(\mathbf{a}\wedge\mathbf{v})\mathbf{d}-d^2\mathbf{a}\wedge\mathbf{v}$ leads to a form that will subsequently help to simplify the end result. We can achieve the necessary rearrangement by noting that for any relative vector $\mathbf{d}$ and bivector $\boldsymbol{U}$,

$$\begin{aligned}(\mathbf{d}\cdot\boldsymbol{U})\wedge\mathbf{d} &= \tfrac{1}{2}(\mathbf{d}\boldsymbol{U}-\boldsymbol{U}\mathbf{d})\wedge\mathbf{d} \\ &= \tfrac{1}{4}((\mathbf{d}\boldsymbol{U}-\boldsymbol{U}\mathbf{d})\mathbf{d}-\mathbf{d}(\mathbf{d}\boldsymbol{U}-\boldsymbol{U}\mathbf{d})) \\ &= \tfrac{1}{2}\mathbf{d}\boldsymbol{U}\mathbf{d}-\mathbf{d}^2\boldsymbol{U}\end{aligned} \tag{12.36}$$

so that

$$\mathbf{d}(\mathbf{a}\wedge\mathbf{v})\mathbf{d}-d^2\mathbf{a}\wedge\mathbf{v} = 2(\mathbf{d}\cdot(\mathbf{a}\wedge\mathbf{v}))\wedge\mathbf{d} \tag{12.37}$$

Bringing these together, Equation (12.34) finally simplifies to

$$\begin{aligned}\boldsymbol{R\Omega R} &= \gamma^3\left(-2(\mathbf{a}\cdot\mathbf{d})\mathbf{d}+d\mathbf{ad}-d\mathbf{da}+2d^2\mathbf{a}+2\mathbf{d}\cdot(\mathbf{a}\wedge\mathbf{v})\wedge\mathbf{d}-d(\mathbf{a}\wedge\mathbf{v})\mathbf{d}+d\mathbf{d}(\mathbf{a}\wedge\mathbf{v})\right) \\ &= 2\gamma^3\left(\underbrace{d^2\mathbf{a}-(\mathbf{a}\cdot\mathbf{d})\mathbf{d}+d\mathbf{d}\cdot(\mathbf{a}\wedge\mathbf{v})}_{\text{vector}}+\underbrace{d\mathbf{a}\wedge\mathbf{d}+\mathbf{d}\cdot(\mathbf{a}\wedge\mathbf{v})\wedge\mathbf{d}}_{\text{bivector}}\right)\end{aligned} \tag{12.38}$$

On comparison of the relative vector and bivector contributions above, it is clear that they are related since

$$\left(d^2\mathbf{a}-(\mathbf{a}\cdot\mathbf{d})\mathbf{d}+d\mathbf{d}\cdot(\mathbf{a}\wedge\mathbf{v})\right)\wedge\frac{\mathbf{d}}{d} = d\mathbf{a}\wedge\mathbf{d}+\mathbf{d}\cdot(\mathbf{a}\wedge\mathbf{v})\wedge\mathbf{d} \tag{12.39}$$

The bivector contribution is simply the outer product of the relative vector contribution with $\hat{\mathbf{d}}$. We can state this more formally as

$$\langle\boldsymbol{R\Omega R}\rangle_2 = \langle\boldsymbol{R\Omega R}\rangle_1\wedge\hat{\mathbf{d}} \tag{12.40}$$

For the time being, therefore, we need to pursue only the relative vector part $\langle\boldsymbol{R\Omega R}\rangle_1$ since we can always generate the associated bivector later on.

Using the identity that for any three vectors $\boldsymbol{a}, \boldsymbol{b}$, and $\boldsymbol{c}$, we have $\boldsymbol{a}\cdot(\boldsymbol{b}\wedge\boldsymbol{c}) = (\boldsymbol{a}\cdot\boldsymbol{b})\boldsymbol{c} - (\boldsymbol{a}\cdot\boldsymbol{c})\boldsymbol{b}$, together with the traditional form $\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = (\mathbf{a}\cdot\mathbf{c})\mathbf{b} - (\mathbf{a}\cdot\mathbf{b})\mathbf{c}$, we find that $\langle \boldsymbol{R}\boldsymbol{\Omega}\boldsymbol{R}\rangle_1$ may be written as

$$\begin{aligned}\langle \boldsymbol{R}\boldsymbol{\Omega}\boldsymbol{R}\rangle_1 &= 2\gamma^3\left((\mathbf{d}\cdot\mathbf{d})\mathbf{a} - (\mathbf{a}\cdot\mathbf{d})\mathbf{d} + d\mathbf{d}\cdot(\mathbf{a}\wedge\mathbf{v})\right)\\ &= 2\gamma^3\left(-\mathbf{d}\times(\mathbf{d}\times\mathbf{a}) + d\mathbf{d}\times(\mathbf{v}\times\mathbf{a})\right)\\ &= -2\gamma^3\mathbf{d}\times((\mathbf{d} - d\mathbf{v})\times\mathbf{a})\end{aligned} \tag{12.41}$$

Returning to Equation (12.20), we may now get the radiated electric field from the relative vector part of $\boldsymbol{F}_{rad}$ by recalling the result $-\boldsymbol{R}\cdot\boldsymbol{\upsilon} = d' = \gamma(d - \mathbf{d}\cdot\mathbf{v})$ from Equation (12.6), whereupon

$$\begin{aligned}\mathbf{E}_{rad} &= \frac{q}{8\pi}\frac{\langle \boldsymbol{R}\boldsymbol{\Omega}\boldsymbol{R}\rangle_1}{(\boldsymbol{R}\cdot\boldsymbol{\upsilon})^3}\\ &= \frac{-2q\gamma^3}{8\pi}\frac{\mathbf{d}\times((\mathbf{d} - d\mathbf{v})\times\mathbf{a})}{-\gamma^3(d - \mathbf{d}\cdot\mathbf{v})^3}\\ &= \frac{q}{4\pi}\frac{\hat{\mathbf{d}}\times\left(\left(\hat{\mathbf{d}} - \mathbf{v}\right)\times\mathbf{a}\right)}{\kappa^3 d}\end{aligned} \tag{12.42}$$

where $\kappa = \left(1 - \hat{\mathbf{d}}\cdot\mathbf{v}\right)$. This compares directly with Jackson's standard result with $\boldsymbol{\beta} = \mathbf{v}$ and $\mathbf{n} = \hat{\mathbf{d}}$. We may now use the relationship between the vector and bivector parts (Equation 12.40) to state the magnetic part of the radiation field as simply being $\boldsymbol{B}_{rad} = \hat{\mathbf{d}}\wedge\mathbf{E}_{rad}$, given that with our metric signature the bivector part of $\boldsymbol{F}$ is $-\boldsymbol{B}$ rather than $+\boldsymbol{B}$ while additionally $\boldsymbol{R}\cdot\boldsymbol{\upsilon} = -d'$ rather than d'.

It will be seen from Equation (12.42) that $\mathbf{E}_{rad}$ may be split into two terms, one depending on $\hat{\mathbf{d}}\times\left(\hat{\mathbf{d}}\times\mathbf{a}\right)$ and the other on $\hat{\mathbf{d}}\times(\mathbf{v}\times\mathbf{a})$. Only the former term is significant at nonrelativistic charge velocities so that the maximum radiated field is observed in a direction perpendicular to the acceleration. This is not the case with the other term for which the maximum occurs when the velocity and acceleration are mutually perpendicular and the direction of observation lies in the plane of the charge's trajectory.

We will say no more on the result itself; what is more important is the manner by which it was derived. The important result is actually Equation (12.20), the spacetime form of the electromagnetic field of a point charge on a smooth but otherwise arbitrary trajectory. The key step that enabled us to reach that point was simply working through the application of the differentiation process to the charge's spacetime electromagnetic potential, $\boldsymbol{A}$. Carrying this out is made a little tricky by the way that the product rule works in a noncommuting algebra, that is to say, the reason why we need to use expressions such as $\mathring{\boldsymbol{\nabla}}\boldsymbol{u}\mathring{\boldsymbol{v}}$. Noncommutation also creates extra work when it comes to simplifying expressions. Working without basis vectors can also be challenging because it is not possible just to "turn the handle"; it is necessary to find some more clever means of simplifying and rearranging expressions.

Calculating the radiation in the charge's frame proved not to be too difficult as most of the effort involved working out $\boldsymbol{R\Omega R}$ in terms of relative vectors and that was accomplished in only four lines. The working from Equations (12.25)–(12.42), however, simply amounts to the chore involved in finding the spacetime split of Equation (12.20) in the observer's frame. Even here, an essentially frame-free approach was taken. Now, we must admit that at this stage, it is impossible to avoid working with the time vectors $\boldsymbol{t}$ and $\boldsymbol{\upsilon}$ because the frames of the observer and charge need to be specified, but neither $\boldsymbol{A}$ nor $\boldsymbol{\nabla}$, nor $\boldsymbol{F}$, was represented in terms of a basis. The process was fairly tedious but nevertheless quite straightforward. But neither is it particularly easy to work this out any other way because solutions for the differentiation and retardation problems have to be found and worked out in detail. It requires considerable skill to tackle these problems without the guidance of a tried and trusted textbook. Spacetime geometric algebra, however, gives us a fairly straightforward and systematic approach to such problems—it is perhaps noteworthy that there were no integral signs in either this section or the previous one.

12.5 EXERCISES

1. Demonstrate that the following are true for any two spacetime vectors $\boldsymbol{u}$ and $\boldsymbol{w}$:

(a) The sign of $\boldsymbol{u}\cdot\boldsymbol{w}$ depends on the chosen metric signature; that is, it depends on whether $\boldsymbol{\theta}^2 = +1$ or -1 for any given time vector $\boldsymbol{\theta}$.

(b) If $\boldsymbol{u}\perp\boldsymbol{w}$, then the relative vectors $\mathbf{u}$ and $\mathbf{w}$ in the $\boldsymbol{\theta}$-frame are orthogonal if and only if $\boldsymbol{u}\cdot\boldsymbol{\theta} = 0$ or $\boldsymbol{w}\cdot\boldsymbol{\theta} = 0$.

(c) If the spacetime split of $\boldsymbol{u}$ in the $\boldsymbol{\theta}$-frame is $u_\theta + \mathbf{u}$, then $\boldsymbol{u}^2 = \boldsymbol{\theta}^2\left(u_\theta^2 - \mathbf{u}^2\right)$.

(d) In particular, $\boldsymbol{u}^2 = 0 \Leftrightarrow \mathbf{u}^2 = u_\theta^2$.

2. **(a)** Show that $\boldsymbol{\upsilon}(\tau) = \cosh(a\tau)\boldsymbol{t} + \sinh(a\tau)\boldsymbol{x}$ represents the proper velocity of a uniformly accelerating particle. In what sense is the acceleration uniform?

(b) What is the relationship between τ and t, the $\boldsymbol{t}$-frame time parameter?

(c) Find the velocity of the particle as seen in the $\boldsymbol{t}$-frame.

3. A charge is initially at rest at the origin but then accelerates uniformly for a time T after which it remains at a constant velocity $\mathbf{v}_1$ such that $|\mathbf{v}_1| \ll 1[c]$.

(a) Using convenient first order approximations, describe the electromagentic field as seen at some distant point $\mathbf{d}$.

(b) Find the resulting Poynting vector.

4. The electromagnetic field of a moving charge may be readily partitioned into a quasistatic field plus a radiated field, which have overtly different characteristics. How might the electromagnetic potential be partitioned so as to draw the same distinction?

Chapter 13

Conclusion

For the physicist and engineer alike, mathematics provides the formal framework for describing the way things work. Given some sort of system governed by physical processes, we use mathematics in two ways. First, we use it as a way of making a model of the system by writing down the objects involved and the rules they obey in some suitable mathematical form, then we test the rules and try them out with some hypothetical data so that we may analyze the behavior of the system. The mathematical tools we use can make a big difference to how easy or hard this process may be. When the mathematical tools at James Clerk Maxwell's disposal could not deal effectively with vector analysis, he turned to Hamilton's quaternions. That was not the end of the story because it largely fell to Gibbs to provide us with the traditional toolset we use today, and it was Heaviside who brought it to bear on the usual present day form of Maxwell's equations. It is perhaps ironic that the foundation of geometric algebra by Grassmann and Clifford dates back to the era of Maxwell and Hamilton, but its potential as a toolset went largely unnoticed and only started to emerge about a century later. Even before that came about, however, Cartan had devised another sophisticated toolset, differential forms, with features and analytic power that are somewhat comparable to geometric algebra. It is, however, a subset of tensor analysis which is, perhaps, the best known and most widely used toolset for dealing with objects of arbitrary rank or grade.

Most of us have dealt with tensors in some shape or form, for example, in matrix algebra where column and row vectors are of rank 1, and matrices themselves are of rank 2. The rank of a tensor is effectively the number of independent subscripts that it has, and so a tensor of rank 3 requires us only to add an additional subscript, and so on. By way of example, Maxwell's equation appears in tensor form as

$$\begin{aligned} \sum_{\nu} \partial_{\nu} F_{\mu\nu} &= J_{\mu} \\ \sum_{\lambda} \sum_{\mu} \sum_{\nu} \varepsilon_{\kappa\lambda\mu\nu} \partial_{\lambda} F_{\mu\nu} &= 0 \end{aligned} \tag{13.1}$$

Understanding Geometric Algebra for Electromagnetic Theory, First Edition. John W. Arthur.

where each of the subscripts runs from 1 to 4 and the fourth rank tensor $\varepsilon_{\kappa\lambda\mu\nu}$, called the Levi-Civita tensor, takes the value +1 when $\kappa\lambda\mu\nu$ forms an even permutation of 1234, −1 for an odd permutation, and 0 otherwise. For example, $\varepsilon_{2341} = 1; \varepsilon_{3241} = -1$ and $\varepsilon_{3341} = 0$. It is clear, therefore, that if this is in some way comparable to writing $\nabla F = J$, the relationships between ∇, F and J are expressed entirely in terms of their coordinates with respect to some implied basis. These relationships then amount to a generalization of ordinary matrix multiplication to cases involving more than two subscripts. In geometric algebra, however, the multiplication between ∇ and F is defined *algebraically* with no need for coordinates, that is to say, without having to specify basis vectors at all. This, then, is the fundamental difference between geometric algebra and tensors. Even so, there is a psychological factor that is in favor of geometric algebra, because $\nabla F = J$ is easy to recognize for what it is, namely the vector derivative of the electromagnetic field bivector equals the electromagnetic source density, but how about the same equation in tensor form? Not only is it in two parts that look different, but also the homogeneous part that includes the Levi-Civita tensor looks quite complicated. It is therefore hard to take in what the equation actually says, let alone what it is supposed to mean.

It would therefore seem that geometric algebra is the natural expression of what many previous tool developers and users were perhaps seeking but did not quite achieve. Their toolsets were useful in many respects but lacking in others. Geometric algebra combines the notions of a graded structure, in which we have scalars, vectors, bivectors, and so on, with an *algebraic* form of vector multiplication that gives rise to a natural metric. It therefore has no need of an imposed coordinate system. Coordinate systems are useful but optional in geometric algebra. The advantage of being coordinate free is the complete generality of results, and, should we wish to bring in a coordinate system at some stage, there is the freedom to do this how and when we like.

Chapters 1–6 presented the foundations of a geometric algebra and illustrated the main features of its application to electromagnetic theory in comparison with traditional methods. This was referred to as the (3+1)D approach in which, for example, scalar time t and vector position $\mathbf{r}$ may be combined as $R = t + \mathbf{r}$, a form known as a paravector, so that R defines an event rather than just a position. Likewise, the time derivative and vector derivative combine to form the paravector operator $\nabla + \left[\frac{1}{c}\right]\partial_t$, while the electromagnetic quantities such as charge ρ and current $\mathbf{J}$ combine into the paravector quantities $J = [1/\varepsilon_0]\rho - [Z_0]\mathbf{J}$ and the electric and magnetic fields $\mathbf{E}$ and B combine to form the multivector $F = \mathbf{E} + [c]B$. The key results are summarized in Table 6.1. The noteworthy points are that in free space, Maxwell's equations are reduced to just one that has a single field quantity, F, and a single source, J. Rather than being completely independent, the part of the solution for the magnetic field is directly related to the solution for the electric field, for while the latter depends purely on the charge distribution, the magnetic solution requires that its motion be taken into account. Looked at in this way, this is a clear trait of special relativity. With other mathematical representations of Maxwell's equation, this point may not come across at all so clearly, but here we can write $J = \rho(1 - \mathbf{v})$

where the factor $(1-\mathbf{v})$ simply modifies an originally static charge distribution. Perhaps surprisingly, it was only necessary to include this factor in the equation for the electric field, Equation (5.12), in order to obtain the solution for the complete electromagnetic field. Although we only showed this for the quasistatic limit where $\partial_t \to 0$, it nevertheless establishes the important principle that the magnetic field is not a separate phenomenon in its own right. Rather, it may be deduced from the same origins as the electric field simply by taking velocity into account.

Moving on to other results, it turns out that plane wave solutions of Maxwell's equation are readily represented in exponential form and naturally exhibit left or right circular polarization. We now have a single electromagnetic potential $\boldsymbol{A}$ in paravector form, which obeys the wave equation with $\boldsymbol{J}$ as source. Finally, the quantity $\frac{1}{2}\boldsymbol{F}\boldsymbol{F}^{\dagger}$ yields both energy density and momentum, which could be taken as yet another hint about underlying themes from relativity. Somewhat less satisfactory is the fact that both the Lorentz force and Maxwell's equations in polarizable media can only be expressed algebraically by dealing with $\mathbf{E}$ and $\boldsymbol{B}$ separately. Nor can we have a unified vector derivative when time is treated on a different footing from space, and so it seems that, useful though this (3+1)D regime may be, we come back to the central issue—Are we using the best toolset?

Evidently, (3+1)D geometric algebra is nearly, but not quite, an optimal toolset. As we have seen, the spacetime geometric algebra does better. Superficially, it may appear to provide only a neater form of the equations, but even so, this would seem to suggest some affinity between its mathematical structure and the underlying physical processes. In fact, the significance of the spacetime geometric algebra is quite clear in that it provides a simple yet relativistically sound mathematical model for fundamental physical phenomena. As to our particular cause, it provides us with an elegant yet complete model for classical electromagnetic theory. But it needs to be emphasized that the spacetime geometric algebra should not be considered as essentially being a tool for those who want to work with relativity. Putting relativity aside, there is a certain advantage in packaging up time and space as a 4D vector space with the time vector on an equal standing to spatial ones. Had we explored a 4D space with a standard Euclidean norm, we would have found that $\nabla^2\boldsymbol{F}=0$ is no longer a wave equation, since with such a norm ∇^2 equates to $\boldsymbol{\nabla}^2+\partial_t^2$ rather than the required $\boldsymbol{\nabla}^2-\partial_t^2$. We would also have found $I^2=+1$ so that it would be necessary to introduce complex scalars in order to be able to represent a plane wave. Things do not work out properly unless we use the non-Euclidean spacetime norm so that spacetime is as much about Maxwell's equation as it is about special relativity. With hindsight, it could be said that this is no surprise, for the two themes are intrinsically coupled, but when we use Maxwell's equation, we mostly do so without thinking in the least about special relativity. If we have no interest in special relativity we may nevertheless accept that the spacetime geometric algebra is the proper framework for electromagnetic theory and simply use it as a practical tool. If we always work in one frame, our $\boldsymbol{t}$-frame, it matters little and so, rather than using spacetime splits, we can stick with the translation process, as we have called it, to find results in (3+1)D. Since this is basically a replacement scheme for basis elements, we can use it without thinking of frames and relativity. But should the

question arise as to results in some other frame, the spacetime split is such a simple concept that this would generally be a straightforward matter.

Turning now from the conceptual benefits of using spacetime geometric algebra for electromagnetic theory, either with or without special relativity, let us consider some of the practical benefits. The reason that the vector derivative, plane waves, the Lorentz force and Maxwell's equations all appear in a neater, more compact form both in free space and in polarizable media, is that it provides not just a form of window dressing but actually a better fit to the theory. Better encoding sums up this idea very effectively. As an example, in the form $\nabla F = J$, Maxwell's free space equation reaches the simplicity of being no special equation at all. It is now just a generic equation that is analogous to specifying an analytic function in 4D rather than 2D, and clearly has the potential to describe a broad range of mathematical and physical phenomena. The only thing special about it is that, through the spacetime norm, it gives rise to wave solutions that propagate along null vectors. This is the reason why it exhibits features of special relativity, something that is effectively indiscernible in other wave systems such as sound waves. To give a somewhat grander example, the formal solution of $\nabla F = J$ can be stated in an equally simple way, namely $F = \nabla^{-1} J$. Since the electromagnetic source density associated with a charge density ρ in motion is given by $\rho \mathfrak{v}$, we may combine this solution with the Lorentz force expressed as a force density, $\mathcal{F} = \rho \mathfrak{v} \cdot F$. Here ρ and $\mathfrak{v}$ will generally both depend on space and time. Under suitable assumptions, we find a self-consistent formal equation for the evolution of the charge distribution in the form of $\dot{\mathfrak{v}} = (q/m)\mathfrak{v} \cdot \nabla^{-1}(\rho \mathfrak{v})$ where m/q is the ratio of mass to charge that obtains within the source density. While it is not our intention here to go any further with this idea, it is nevertheless intriguing to see how the basic laws of mechanics and electricity can be brought together so easily and in a relativistically proper manner. Without the power of geometric algebra to enlighten us, working out such a relationship would have seemed a major undertaking, but here it seems rather simple.

Despite the possibility of just working with the spacetime geometric algebra as a convenient way of dealing with electromagnetic theory, it is inescapable that we are implicitly taking account of the laws of special relativity. In the example we have just given, no special effort was needed to make the treatment "relativistic"; it is simply inherent in the structure. This may appeal to many readers who previously were not interested in, or were even put off, special relativity but perhaps now see it as less of a problem. The field of a moving charge and the transformation of the electromagnetic field from one frame to another are among the most important issues in electromagnetic theory, the understanding of which is key to understanding the nature of the electromagnetic field. But this is relativity; it is inescapable. Even in (3+1)D, there were clear hints of this. As discussed in Reference 2 and elsewhere, it is rarely mentioned that the magnetic field itself is prima facie evidence for special relativity. In the case of the Coulomb field, what is a purely electric field in a charge's own rest frame has an accompanying magnetic field as seen in some other frame, and what was just a charge now has an associated current. Hence, the observed magnetism is associated with the observed current, and all this stems from the Coulomb field of a static charge. There are, in effect, no separate mechanisms for

electricity and magnetism. A more sophisticated way of referring to this behavior is to say that Maxwell's free space equation is covariant, that is to say, it remains an equation in any Lorentz frame. The independent and dependent variables may change their form, but the equation always holds good. Even if this is of little relevance to the reader, the fact that the magnetic field is an essentially relativistic phenomenon should not be ignored. Only then can it be appreciated why the spacetime geometric algebra is such an effective toolset for electromagnetic theory.

The practical benefits of a 4D treatment may indeed seem superficial; nevertheless, it clearly offers greatly enhanced physical interpretation. In the (3+1)D approach, finding the electromagnetic field of a point charge in motion effectively requires the solution of time-dependent wave equation. The spacetime approach starts from the known electrostatic solution for the potential and merely projects it from the charge's rest frame into the observer's frame. What, we may ask, is the link? The answer is simply that, as we have seen, Maxwell's equation conforms to the principles of relativity and therefore produces wave equations that give results that are equivalent to the more direct spacetime approach. For example, both methods properly account for the time it takes for an effect at the source to reach the observer, and also they both result in a magnetic field when source and observer are in relative motion.

The relationship between (3+1)D and spacetime involves some intriguing subtleties, mostly attributable to the vagaries of metric signature, which we have taken some time to explain. We have also discussed the special role of the time vector in generating the relative vectors that we perceive in the Newtonian view of the world, $\mathbf{x} = x\boldsymbol{t}$, $\mathbf{y} = y\boldsymbol{t}$ and so on. The freedom to choose different time vectors here corresponds to the choice of different inertial frames. Replacing the local time vector $\boldsymbol{t}$ with $\boldsymbol{\upsilon}$, the proper velocity of an observer in the chosen alternative frame, generates a different set of relative vectors, $\mathbf{x}' = x\boldsymbol{\upsilon}$, $\mathbf{y}' = y\boldsymbol{\upsilon}$ and so on, appropriate to that frame. This result certainly causes a deal of thought as to the simplicity of the approach and only serves to underline the appropriateness of geometric algebra as a mathematical toolset. By comparison, in the (3+1)D approach we have to find forms of transformations between the relative vectors and scalars in the different frames. This process is actually less intuitive than the spacetime approach and is the cause of difficulty for many undergraduates and graduates alike.

The key electromagnetic equations in spacetime are summarized in Table 13.1. Those equations that do not explicitly involve a time vector appear to be only marginally different from their (3+1)D counterparts so that the absence of separate time derivatives is the only real clue. But for those equations that do involve a time vector, the underlying physical significance is clear; this is what allows them to work in *any* frame.

As we have seen, there are many toolsets, some better than others, some specialized and others more general purpose, some relatively easy to access and others quite daunting. On the face of it, there seems to be only a fine difference between geometric algebra and the traditional approach in that it regularizes vector multiplication, a property that allows it to generate new entities that are neither vectors nor scalars. Yet this is clearly a difference that has enormous consequences in that it is closer by far to being an ideal language for physics and engineering. Though it

wields the power of tensor analysis, it is not constrained by the need to reduce everything to component form. As a toolset, this gives it far greater freedom. That it languished as "just another algebra" for the best part of a century is almost the antithesis of a discovery. This inordinate delay in its uptake can only have had a negative effect on our current way of thinking and ability to deal with these subjects. Most of us are still talking the language of Gibbs, many of us probably teach it, and some of us may never know anything else. Hestenes and his followers made it clear that this needs to change.

Table 13.1 Summary of the Key Results of Spacetime Electromagnetic Theory

The Lorentz force	$\boldsymbol{f} = q\,\boldsymbol{\upsilon}\cdot\boldsymbol{F}$
Maxwell's equation in free space	$\boldsymbol{\nabla F} = \boldsymbol{J}$
Maxwell's equations in polarizable media	$\boldsymbol{\nabla}\cdot\boldsymbol{G} = \boldsymbol{J}$ $\boldsymbol{\nabla}\wedge\boldsymbol{F} = 0$
Wave equation and conservation of charge	$\boldsymbol{\nabla}^2\boldsymbol{F} = \boldsymbol{\nabla J}$
Vector potential	$\boldsymbol{\nabla}^2\boldsymbol{A} = \boldsymbol{J}$ $\boldsymbol{F} = \boldsymbol{\nabla A}$
The electromagnetic field of circularly polarized plane waves	$\boldsymbol{F}(\boldsymbol{r}) = \boldsymbol{F}_0 e^{\pm I\boldsymbol{k}\cdot\boldsymbol{r}}$
The vector potential of a moving point charge	$\boldsymbol{A} = \dfrac{q\boldsymbol{\upsilon}}{4\pi\boldsymbol{R}\cdot\boldsymbol{\upsilon}}$
The electromagnetic field of a moving point charge	$\boldsymbol{F}_{qs} = \dfrac{-1}{4\pi}\dfrac{\boldsymbol{R}\wedge\boldsymbol{\upsilon}}{(\boldsymbol{R}\cdot\boldsymbol{\upsilon})^3}$ $\boldsymbol{F}_{rad} = \dfrac{1}{8\pi}\dfrac{\boldsymbol{R\Omega R}}{(\boldsymbol{R}\cdot\boldsymbol{\upsilon})^3}$
Electromagnetic energy and momentum in the $\boldsymbol{\theta}$-frame	$\frac{1}{2}\boldsymbol{F\theta F}^{\dagger} = \mathcal{E}\boldsymbol{\theta} + \boldsymbol{g}$

All the equations in the table are in terms of the modified variables of Table 5.1. Those equations that do not explicitly involve a time vector appear to be only marginally different from their (3+1)D counterparts so that, apart from the minor notational difference, the absence of separate time derivatives is often the only real clue.

Chapter 14

Appendices

14.1 GLOSSARY

The glossary is a guide to terms that may be unfamiliar to the reader for various reasons. The list begins with symbols.

$\cdot$	A dot placed between two objects indicates the inner product. A dot placed over an object indicates differentiation with respect to proper time, ∂_τ.
$\circ$	An open dot placed over an operator and on an object is used here to imply that the operator is to act on that object even when the object is not to the operator's immediate right, for example, as in $\mathring{\nabla}\boldsymbol{u}\mathring{\boldsymbol{v}}$. Other authors use different symbols, including an ordinary dot or a prime, both of which could be confused with other notations used here.
$\hat{}$	The caret placed over an object is the normalization operator.
$\perp$	Symbol denoting orthogonality.
//	Symbol denoting parallelism.
$\wedge$	The wedge placed between two objects indicates the outer product.
$\times$	Vector cross product.
$\times$	Commutator product.
$\dagger$	The superscript dagger placed after an object indicates the reverse operator. See also ~.
*	(1) The superscript asterisk placed after an object is commonly used for the inversion operator (involution). (2) A superscript asterisk placed after a variable that is a function of time or distance indicates that **retardation** is to be applied. (3) Scalar product.
~	(1) A tilde placed over an object, or as superscript on the right, usually means the same as † (reverse), particularly in a spacetime context. (2) A tilde placed under a vector is used here to indicate the purely spatial vector in a given frame. See also **spatial**.

(Continued)

Understanding Geometric Algebra for Electromagnetic Theory, First Edition. John W. Arthur.

$\leftrightarrow$	$\boldsymbol{A} \leftrightarrow \boldsymbol{B}$ is used here to mean that the spacetime object $\boldsymbol{A}$ maps onto the (3+1)D object $\boldsymbol{B}$, and vice versa. See also **translate** and **spacetime split**.
$\lvert\ \rvert$	Magnitude of an object, for example, length of a vector or the area of a bivector; absolute value of a scalar. It is the measure of the object taken as a positive scalar. The symbol $\lVert\ \rVert$ is sometimes used, scalars excepted.
[ ]	Square brackets enclosing physical constants such as c, ε_0, and μ_0 are used here to highlight constants that would normally be hidden in simplified notation (*q.v.*).
$[\ ]_{ret}$	Square brackets with the subscript *ret* indicates that **retardation** is to be applied to the expression within.
$\langle\ \rangle\ \langle\ \rangle_k$	Grade selection filter. Without a subscript, it selects the scalar part of the enclosed object (grade 0); otherwise, the subscripts indicate the grade or grades to be selected.
∂	(1) Indicates the boundary of a region; for example, ∂A is the line that borders the area A and ∂V is the surface that encloses the volume V. (2) An alternative representation for ∇.
∂_k	Differential operator that acts on the object on its right. The subscript indicates the variable of differentiation, for example, $\partial_t u = \partial u/\partial t$, or in the case of an index, $\partial_k u = \partial u/\partial x_k$.
$\boldsymbol{\nabla}$, ∇, $\boldsymbol{\nabla}$, ∂, $\Box$	Vector derivative. While $\boldsymbol{\nabla}$ is the traditional 3D form, the general form in use is ∇. For the sake of readability, we use $\boldsymbol{\nabla}$ rather than ∇ for the spacetime derivative. ∂ without a subscript may also be seen and $\Box$ was previously common in a spacetime context.
3D	The familiar concept of 3D space, or more specifically, a 3D vector space in which the vectors represent position, velocity, and so on.
(3+1)D	A representation of space and time in which 3D spatial vectors and scalar time may be jointly represented as a paravector, for example, $\mathbf{r} + t$, as in a 3D geometric algebra.
(3+1)D frame	The (3+1)D reference frame that corresponds directly to a given spacetime frame.
4-vector	There are two meanings, for which see **four-vector** and ***n*-vector**.
Absolute spacetime	Referring to a central theme of spacetime in which events are represented by vectors that are independent of the frame of reference of any observer. The observation of any event, however, *is* frame dependent.
Absolute value	The absolute value of a scalar a is its magnitude, denoted by $\lvert a \rvert$.
Acceleration bivector	If $\boldsymbol{\upsilon}$ is the proper velocity of a particle, then its acceleration bivector is defined as is, $\dot{\boldsymbol{\upsilon}}\boldsymbol{\upsilon}$. See also **proper velocity**, **proper time**, and $\dot{}$ (overdot).
Active transformation	A transformation that acts directly on an object. See also **passive transformation**.

Algebra — In broad terms, it may be said that an algebra is a set of elements together with a set of operations defined by rules under which the result of any allowed operation on one or more of the elements is itself a member of the algebra.

Anticommutative — An operation or operands for which the result is antisymmetric; that is, reversing the order of the operands changes the sign of the result as in $\boldsymbol{v} \wedge \boldsymbol{u} = -\boldsymbol{u} \wedge \boldsymbol{v}$. See also **commutative** and **non commutative**.

Auxiliary field (electromagnetic, electric, magnetic) — A field such as **D**, $\boldsymbol{H}$, or $\boldsymbol{G}$ that is required only to deal with bound sources. See also **auxiliary electromagnetic field multivector**.

Auxiliary electromagnetic field multivector — A multivector, $\boldsymbol{G}$, that combines the auxiliary electric field (displacement) vector **D** and the auxiliary magnetic field (magnetic field intensity) bivector $\boldsymbol{H}$. In spacetime, $\boldsymbol{G}$ is a pure bivector.

Axial vector — A 3D vector formed as the cross product of true vectors. See also **true vector**.

Basis, basis vector — A set of vectors that span a vector space, meaning that any given vector in the space may be formed by a linear combination of the vectors in the set. A basis vector is any one of the given set.

Basis element — One of a set of elements that span a geometric algebra such that any multivector in the geometric algebra may be formed by a linear combination of these elements.

Blade — An object in a geometric algebra that may be formed by the outer product of k vectors, where k is the grade of the blade. For the sake of completeness, grade 0 refers to the scalars.

Bivector — Also known as a 2-vector. In general, a bivector is any linear combination of blades of grade 2. In 3D, all bivectors are blades, but not in 4D.

Bound source — Electromagnetic bound sources are associated with distributions of polarization and magnetization within matter or at its surfaces. Bound sources are important because their distribution not only depends on the electromagnetic field but contributes to it. See also **free source**.

Circuital — See **solenoidal**.

Commutative — An operation or operands for which the result is symmetric; that is, reversing the order of the operands leaves the result unchanged as in $\boldsymbol{v} \cdot \boldsymbol{u} = \boldsymbol{u} \cdot \boldsymbol{v}$.
See also **anticommutative** and **noncommutative**.

Commutator product — The commutator product of any two objects $\boldsymbol{U}$ and $\boldsymbol{V}$ is denoted by the symbol $\times$ and is defined as $\boldsymbol{U} \times \boldsymbol{V} = \frac{1}{2}(\boldsymbol{UV} - \boldsymbol{VU})$. Not to be confused with $\boldsymbol{U} \wedge \boldsymbol{V}$, which must be formed on a grade-by-grade basis.

Covariant equation — An equation that retains the same form under a change of basis vectors (orthogonal transformation).

(Continued)

Cross product	A product between 3D vectors as defined through Equation (2.1).
Derived	Obtained by a process such as differentiation. For example, a velocity vector is derived from an event vector by differentiation with respect to time. Relative vectors of derived vectors require special consideration.
Direct product	The geometric product, for example, $\boldsymbol{UV}$, as opposed to inner product or outer product.
Dot	The inner product operator or operation. See also **wedge**.
Dot product	Strictly speaking, the inner product between vectors that is always commutative.
Dual	The dual of any multivector $\boldsymbol{U}$ is effectively $\pm I\boldsymbol{U}$.
Electromagnetic source density	The spacetime vector or (3+1)D multivector that combines charge and current densities. Usual symbol: $\boldsymbol{J}$.
Electromagnetic polarization	A spacetime bivector or (3+1)D multivector that combines both electric and magnetic polarization (magnetization). Symbol used here: $\boldsymbol{Q}$.
Euclidean	A description of any N-dimensional space in which Pythagoras' theorem applies, that is, $\lvert\boldsymbol{u}\rvert^2 = \sum_{k=1}^{N} u_k^2$ for any vector $\boldsymbol{u}$. Otherwise, the space is referred to as non-Euclidean.
Event	A specific position *and* time, particularly with reference to spacetime. See also **history** and **trajectory**.
Event vector	The vector that specifies a given spacetime event.
Even subalgebra	A subalgebra in which all the elements have even grade.
Forward light cone	That part of the light cone that is in the future.
Four-vector	A spacetime vector expressed as a row or column vector, for example, $\boldsymbol{r} = (jct, x, y, z)$. See, for example, References 37, section 11.8, pp. 374–377; and 48, chapter XIII, pp. 127–130.
Frame	A given set of basis vectors. In the spacetime context, the frame may be fixed with respect to our rest frame, or travel along with some particle or observer. See also **(3+1)D frame**, **Lorentz frame**.
Free source	A source that may be independently varied, either physically or as the independent variable in a set of equations. See also **bound source**.
Future pointing	Pointing in the same direction as some time vector.
Gauge	For example, the Lorenz condition and the Coulomb gauge, which provide alternative ways of fixing the electromagnetic potential.
Geometric algebra	Given a vector space, then the larger vector space formed by introducing a geometric product between all elements forms a geometric algebra.
Geometric product	A binary operation associated with the multiplication of vectors. Also referred to here as the direct product. See **geometric algebra**.
Grade	The number of vectors required to form a given blade, or, in the case of a homogeneous multivector, the grade of each of the blades that form it.

Green's function or Green function This is essentially the solution of a linear differential equation for a unit point source located at an arbitrary point $\boldsymbol{u}'$. The variables can either be (3+1)D or spacetime vectors, as appropriate. For example, Green's function $\mathbf{G}(\mathbf{u},\mathbf{u}')$ for the static electromagnetic field results from

$$\nabla\mathbf{G}(\mathbf{u},\mathbf{u}')=\delta(\mathbf{u}-\mathbf{u}')\Leftrightarrow\mathbf{G}(\mathbf{u},\mathbf{u}')=\frac{1}{4\pi}\frac{\mathbf{u}-\mathbf{u}'}{|\mathbf{u}-\mathbf{u}'|^3}$$

History The spacetime path followed by any point object. It describes the evolution of the event vector giving the object's position at any time. Also known as trajectory or world line.

Homogeneous (1) Being all the same kind, for example, all of the same grade.
(2) Of a differential equation, where the source terms are all zero.

Homogeneous multivector A multivector composed of only one grade of object, for example, $\mathbf{a}+\mathbf{b}+\mathbf{c}$, which is composed only of vectors or $\mathbf{ab}+\mathbf{cd}$, which is composed only of bivectors. See also the related concept of ***n*-vector**.

Inertial Non-accelerating.

Inner product Symbol $\cdot$. The inner product of a 1-vector with any n-vector is the part of their geometric product having grade $n-1$. Consequently, it is often referred to as a step-down operator. In linear algebra, the inner product between 1-vectors is often called the dot product.

Involution A more precise term for what is commonly called inversion.

Inversion Under inversion, a vector $\boldsymbol{u}$ is replaced by $-\boldsymbol{u}$.

Isomorphism One-to-one correspondence. Two sets $\boldsymbol{A}$ and $\boldsymbol{A}'$ are isomorphic if each element of $\boldsymbol{A}$ has a unique matching element in $\boldsymbol{A}'$ and vice versa.

Length Magnitude of a 1-vector.

Light cone A surface that represents the locus of all null paths to (backward cone) or from (forward cone) some given event. It equates to the characteristic surface of the electromagnetic wave equation in free space.

Lightlike See **null**. See also **spacelike** and **timelike**.

Lorentz frame One of an infinite set of frames that are all related to each other by Lorentz transformation. Each frame in the set may be identified by its local time vector, for example, the $\boldsymbol{t}$-frame or $\mathfrak{v}$-frame.

Lorentz transformation A transformation between two frames of reference that conforms with the requirements of special relativity. In spacetime, it is an orthogonal transformation affecting the vectors in some timelike plane, cf. a rotation that affects vectors in some spacelike plane. See also **passive transformation** and **active transformation**.

Magnitude Symbol $|\ |$. The magnitude of a scalar is its absolute value. In a geometric algebra, the magnitude of a vector $\boldsymbol{u}$ is defined as $|\boldsymbol{u}|=|\boldsymbol{u}^2|^{1/2}$. The magnitude of some other object $\boldsymbol{U}$ may be defined as $|\boldsymbol{U}|=|\langle\boldsymbol{U}\boldsymbol{U}^\dagger\rangle|^{1/2}$.

(Continued)

Measure Generic term for length, area volume, and so on. The measure of an object in a geometric algebra is given by its **magnitude**.

Meromorphic The generalization of the term analytic to any dimension of space. A function $\boldsymbol{f}$ is meromorphic obeys $\nabla \boldsymbol{f} = 0$ in some N-dimensional region. In 3D, this is equivalent to the condition $\boldsymbol{\nabla} \cdot \mathbf{f} = \boldsymbol{\nabla} \times \mathbf{f} = 0$.

Metric A scheme or formula used as the basis for measurement.

Metric signature The metric signature identifies whether each basis vector has been given a positive or negative square.

Modified variable A variable that has been combined with a physical constant so as to eliminate such constants from common equations. See also **simplified equation** and Table 5.1.

Multivector The most general entity in a geometric algebra. It is a linear combination of entities of any grade. See also **homogeneous multivector**.

***n*-Vector** A homogeneous multivector that has the specific grade n. This is often implied in names like bivector and trivector, while 0-vectors and 1-vectors refer to scalars and vectors, respectively. This is not to be confused with the use of the term 4-vector in the tensor formulation of special relativity.

Newtonian A term here describing the intuitive pre-relativity view of ordinary space and time, which may be described as being (3+1)D.

Noncommutative Having no specific commutation property. For example, the geometric product for which in general $\boldsymbol{uv} \neq \boldsymbol{vu} \neq -\boldsymbol{uv}$. See also **commutative** and **anticommutative**.

Norm **Metric** or **measure**.

Normal (1) Of unit magnitude. See also **normalize** and **orthonormal**.
(2) A vector that is perpendicular to a given surface.

Normalize To scale to unit magnitude. The caret placed over an object indicates normalization, that is, unless $\boldsymbol{U}$ is null then $|\hat{\boldsymbol{U}}| = 1$.

Null An object is null if it has zero magnitude.

Observation event Any event at which some form of information is observed. See also **source event**.

Origin In spacetime, the event $\boldsymbol{r} = 0$. In (3+1)D, however, we simply mean the location $\mathbf{r} = 0$. See also **spatial origin**.

Orthogonal Symbol $\perp$. Two objects are orthogonal if their inner product vanishes.

Orthogonal space The subspace formed by all vectors that are orthogonal to a given object.

Orthonormal A set of objects are orthonormal if they are all mutually orthogonal and have unit magnitude.

Outer product Symbol: $\wedge$. The outer product of a 1-vector with any n-vector is the part of their geometric product with grade $n+1$. It is consequently referred to as a step-up operator.

Overdot See $\cdot$ and $\circ$.

Parallel	Symbol: //. Two objects of the same grade are parallel if their geometric product results a scalar. A vector $\boldsymbol{u}$ is parallel a bivector $\boldsymbol{V}$ if $\boldsymbol{u} \wedge \boldsymbol{V} = 0$.
Paravector	A multivector in the form of a vector plus a scalar. Usually, refers to (3+1)D in which paravectors may correspond to spacetime vectors.
Parity	The property of being either even or odd.
Passive transformation	A transformation that acts only on basis elements. It therefore affects only the representation of an object, not the object itself. See also **active transformation**.
Perpendicular	Orthogonal in a geometric sense.
Polar	Associated with a field that has zero curl everywhere but has non-zero divergence at least somewhere. This applies to the fields of both electric charges and magnetic poles.
Polar vector	Same as a **true vector**.
Polarization multivector	See **electromagnetic polarization multivector**.
Postmultiply	Multiply on the right, for example, $\boldsymbol{X}$ postmultiplied by $\boldsymbol{Y}$ is $\boldsymbol{XY}$.
Premultiply	Multiply on the left, for example, $\boldsymbol{X}$ premultiplied by $\boldsymbol{Y}$ is $\boldsymbol{YX}$.
Proper	Describes a property that an object has in its own rest frame. See also **proper time** and **proper velocity**.
Proper time	The local time measured on a clock that is always at rest within a given frame, for example, the rest frame of a moving particle.
Proper velocity	The proper velocity of a particle is the rate of change of its event vector with respect to its proper time. If a particle follows the history $\boldsymbol{r}(\lambda)$, then its proper velocity $\boldsymbol{v}$ is $\partial_\tau \boldsymbol{r}(\lambda)$ where τ is its proper time at the given value of λ. See also **velocity** and **spatial velocity**.
Pseudoscalar	The element having the highest possible grade in a given geometric algebra. While it has some of the characteristics of a scalar, it often has a negative square.
Pseudovector	An element of a geometric algebra that can be written as the product of a pseudoscalar and a vector.
Quaternion	A mathematical object with vector and scalar parts (similar to a paravector). The quaternions form an algebra that allows multiplication and addition (see Appendix 14.5).
Quasistatic	Describes a slowly changing situation that may nevertheless be treated as static, for example, when the time derivatives in Maxwell's equation are small enough to be ignored.
Relative vector	A 3D vector that equates to a bivector in the even subalgebra of spacetime, for example, $\mathbf{x}$ and $\boldsymbol{xt}$.
Retarded, retardation	Implying that the time at the source is to be evaluated by subtracting the propagation delay from the time of observation, for example, $t^* = t - R/c$.

(Continued)

Reverse The reverse of a product. The symbol is generally † or ~, for example, $\boldsymbol{ABC}^{\dagger} = \boldsymbol{CBA}$.

Rotor A multivector $\boldsymbol{R}$ that generates the rotation of an object $\boldsymbol{U}$ by means of an expression of the form $\boldsymbol{RUR}^{\dagger}$, or in some cases just $\boldsymbol{RU}$.

Scalar In a geometric algebra, the scalars are 0-vectors, that is, objects of grade 0, generally speaking, a real number.

Scalar product The scalar product of any two objects $\boldsymbol{U}$ and $\boldsymbol{V}$ is represented as $\boldsymbol{U} * \boldsymbol{V}$, defined by $\boldsymbol{U} * \boldsymbol{V} = \langle \boldsymbol{UV} \rangle$, that is, the scalar part of $\boldsymbol{UV}$.

Separation The length of the **separation vector** between two spacetime events. Corresponds to distance in 3D.

Separation vector The vector joining one spacetime event to another. See also **separation**.

Simple Free of any frame-dependent parameter, c.f. **Derived**.

Simplified equation An equation in which the appearance of physical constants has been supressed for reasons of clarity and simplicity by means of a system of **modified variables**.

Solenoidal Resulting from a field that has zero divergence everywhere but at least somewhere has a nonzero curl. Also, circuital.

Source event The event at which some form of observable information originates. See also **observation event**.

Spacelike Given $\boldsymbol{\theta}$ is a time vector, in any given metric signature a vector or bivector $\boldsymbol{U}$ is spacelike if $\boldsymbol{UU}^{\dagger}$ has the opposite sign to $\boldsymbol{\theta}^2$. In the $(-+++)$ metric signature, this means $0 < \boldsymbol{U}^2$ for vectors, whereas for bivectors, it is the opposite, $\boldsymbol{U}^2 < 0$. See also **lightlike** and **timelike**.

Spacetime A representation of the natural world as a 4D vector space in which both time and position are vectors and the non-Euclidean metric conforms with the requirements of special relativity.

Spacetime split The projection of a spacetime object either into some given (3+1)D frame or into an alternative spacetime frame.

Spatial Referring to position independent of time. The concept is frame dependent and a vector $\boldsymbol{r}$ is spatial in the $\boldsymbol{\theta}$-frame only if $\boldsymbol{r} \cdot \boldsymbol{\theta} = 0$. See also the ~ symbol and **temporal**.

Spatial origin Of the $\boldsymbol{\theta}$-frame, is the point $\boldsymbol{\Omega}$ with the history $\boldsymbol{\Omega} = \lambda \boldsymbol{\theta}$, where λ is the time parameter of that frame.

Spatial velocity The spatial part of the spacetime velocity in some given frame. See also **velocity** and **proper velocity**.

Subalgebra A subset of an algebra that is in itself an algebra.

Symmetric product The symmetric product of any two objects $\boldsymbol{U}$ and $\boldsymbol{V}$ is defined as $\frac{1}{2}(\boldsymbol{UV} + \boldsymbol{VU})$ [27, section 4.1.3].

Temporal Referring only to time. The concept is frame dependent and a vector $\boldsymbol{r}$ is temporal in the $\boldsymbol{\theta}$-frame only if $\boldsymbol{r} = \lambda \boldsymbol{\theta}$ where λ is scalar. See also **spatial**.

Timelike	Given $\boldsymbol{\theta}$ is a time vector, in any given metric signature a vector or bivector $\boldsymbol{U}$ is timelike if $\boldsymbol{U}\boldsymbol{U}^{\dagger}$ has the same sign as $\boldsymbol{\theta}^2$. In the $(-+++)$ metric signature, this means $\boldsymbol{U}^2 < 0$ for vectors, whereas for bivectors, it is the opposite, $0 < \boldsymbol{U}^2$. See also **lightlike** and **spacelike**.
Trajectory	In a spacetime context, the same as history or world line.
Translate	Herein, $\boldsymbol{A}$ translates to $\boldsymbol{B}$ means that spacetime object $\boldsymbol{A}$ is algebraically equivalent to the (3+1)D object $\boldsymbol{B}$. This relationship is symmetric, whereas the **spacetime split** is one way. Symbol $\leftrightarrow$.
Trivector	A 3-vector. See also ***n*-vector**.
True vector	A simple vector that changes sign on inversion of the basis vectors. Also known as a **polar vector**. See also **axial vector**.
Vector	Vectors are objects that may be combined by addition or multiplied by a scalar such that the result is also a vector. In the context of a geometric algebra, the term more specifically refers to the 1-vectors.
Vector space	A set spanned by vectors using the operations of addition and multiplication by scalars alone.
Vector derivative	Although it is not actually frame dependent, it is convenient to express the vector derivative in terms of standard basis vectors, for example, $\nabla = \partial_x \mathbf{x} + \partial_y \mathbf{y} + \partial_z \mathbf{z}$ for 3D and $\boldsymbol{\nabla} = -\partial_t \boldsymbol{t} + \partial_x \boldsymbol{x} + \partial_y \boldsymbol{y} + \partial_z \boldsymbol{z}$ for spacetime.
Velocity (spacetime)	Of a particle, the rate of change of its history with respect to the time parameter of the frame *from which it is observed*. See also **proper velocity** and **spatial velocity**.
Wedge	The outer product operator or operation. See also **dot**.
World line	Also referred to as history or trajectory.

14.2 AXIAL VERSUS TRUE VECTORS

Under the operation of inversion, $\mathbf{r} \mapsto -\mathbf{r}$, that is to say, the position vector $\mathbf{r}$ is replaced by $-\mathbf{r}$. This change of sign under inversion is the hallmark of a true vector. It is readily found that velocity, acceleration, and inertial force are also examples of true vectors. For example, under inversion, the position vector on the right-hand side of Coulomb's law changes sign, as does the force on the left-hand side. Consequently, both the electric force and the electric field are true vectors. But if we take the case of a current traveling around a circular loop, we find that after inverting all the vectors involved, that is, the position of each element of current and the direction of its flow, the sense of the current is unaltered. This is demonstrated in Figure 14.1 where we see a current element $d\mathcal{S}$ at some chosen position $\mathbf{r}$ on the loop on the right. On the left, we have exactly the same thing but with all position vectors and current elements on the loop inverted. We conclude that the original current loop and its inverse have the current circulating in the same sense, and so they both produce a magnetic field in the same direction—out of the plane of the page within

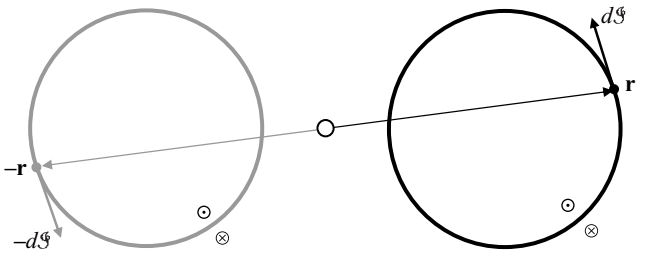

Figure 14.1 Demonstration that the inversion of the spatial vectors leaves the sense of a circulating current unchanged. The solid figure on the right shows a current loop lying in the plane of the paper. The vector **r** touches some point on the loop. For simplicity, the origin also lies in the plane of the paper, but this is not essential to the argument. When the spatial vectors in the same plane are inverted, they are effectively rotated through 180°. In this way, **r** and every other point on the loop, produce the inverted image on the left. The fact that the loop is also translated across the page does not matter. A current element $d\mathfrak{g}$ on the solid loop therefore appears as $-d\mathfrak{g}$ on the opposite side of the inverted loop resulting in no change in the sense of current circulation—anticlockwise in both cases.

the loop and into the page elsewhere. The magnetic field is therefore unaffected by spatial inversion and so it cannot be represented by a true vector, but rather by an axial one. This agrees with the Biot and Savart law due to the fact that the cross product of true vectors results in an axial vector. Torque is another example of an axial vector.

14.3 COMPLEX NUMBERS AND THE 2D GEOMETRIC ALGEBRA

Given the property that all geometric algebras have a unit pseudoscalar I that often obeys $I^2 = -1$, it seems obvious that a 2D geometric algebra should generate complex numbers based on treating 1 and I as the real and imaginary units. The 2D geometric algebra's basis elements are $\{1; \boldsymbol{x}, \boldsymbol{y}; I\}$ where we take $\boldsymbol{x}$ and $\boldsymbol{y}$ as being orthonormal so that $\boldsymbol{x}^2 = \boldsymbol{y}^2 = 1$. It then follows that $\boldsymbol{xy} = -\boldsymbol{yx} = I$ where $I^2 = \boldsymbol{xyxy} = -1$. As a result, complex arithmetic uses only half of the geometric algebra, in fact the fairly trivial 1D even subalgebra comprising the basis elements $\{1; I\}$. By analogy with the relationship that exists between the even subalgebra of spacetime and the remaining odd elements to which the vectors belong, $\{1; I\}$ can be mapped onto $\{\boldsymbol{x}; \boldsymbol{y}\}$. In spacetime, we premultiplied by $\boldsymbol{t}$ to perform the mapping, but here $\boldsymbol{x}1 = \boldsymbol{x}$ and $\boldsymbol{x}I = \boldsymbol{xxy} = \boldsymbol{y}$ so that it is clear that we should premultiply with $\boldsymbol{x}$. This means that any complex number $a + \mathrm{j}b$ can be mapped onto a vector $a\boldsymbol{x} + b\boldsymbol{y}$ simply by substituting I for j and then doing the required premultiplication by $\boldsymbol{x}$.

Mapping from complex numbers to vectors in this way may seem to be just another way of producing an Argand diagram, but note that $(a + \mathrm{j}b)^2 = (a^2 - b^2) + 2ab\mathrm{j}$, whereas $(a\boldsymbol{x} + b\boldsymbol{y})^2 = a^2 + b^2$. The vector form consequently does not replicate complex arithmetic. The problem is that multiplying two vectors, say $\boldsymbol{x}u$ and $\boldsymbol{x}v$, which are mapped from the complex numbers u and v, results in $(\boldsymbol{x}u)(\boldsymbol{x}v) = u^*\boldsymbol{xx}v = u^*v$. Here the complex numbers are in the form $a + Ib$, which uses I as the "imaginary unit" and so bearing in mind that $\boldsymbol{x}I = -I\boldsymbol{x}$, we find

$\boldsymbol{x}(a+Ib)=a\boldsymbol{x}+b\boldsymbol{x}I=a-bI\boldsymbol{x}=(a-Ib)\boldsymbol{x}$. It therefore follows that commuting $\boldsymbol{x}$ past a complex number transforms that number to its complex conjugate. If, however, we take $(\boldsymbol{x}u)^{\dagger}(\boldsymbol{x}v)$, we find that this works out as $(u\boldsymbol{x})(\boldsymbol{x}v)=u\boldsymbol{x}\boldsymbol{x}v=uv$, which is actually the result we require. At the root of the problem is the fact that $\boldsymbol{x}$ and $\boldsymbol{y}$ do not commute. Complex multiplication cannot be replicated directly by vector multiplication and vice versa; rather, it is necessary to introduce an extra step. This is a clear illustration that a pseudoscalar cannot be thought of in the same way as an imaginary scalar, which is why we must not confuse I and j.

14.4 THE STRUCTURE OF VECTOR SPACES AND GEOMETRIC ALGEBRAS

14.4.1 A Vector Space

A vector space over the real numbers (scalars) is a set of elements called vectors having the following properties [10, section 6; 11, chapter VII].

Given any vectors $\{\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \ldots\}$ and any scalars $\{a, b, c \ldots\}$, then

Property

1. $\boldsymbol{u} + \boldsymbol{v}$ is a vector in the space (vector addition),
2. $a\boldsymbol{u}$ is a vector in the space and $a\boldsymbol{u} = \boldsymbol{u}a$ (scalar multiplication),
3. $1\boldsymbol{u} = \boldsymbol{u}$ (the unit scalar is the identity for scalar multiplication),
4. there exists a vector 0 such that $\boldsymbol{u} + 0 = \boldsymbol{u}$ (identity element for vector addition),
5. there exists a vector $-\boldsymbol{u}$ such that $\boldsymbol{u}+(-\boldsymbol{u})=0$ (existence of inverses for vector addition),
6. $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$ (vector addition is commutative …),
7. $\boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) = (\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w}$ (… and associative),
8. $a(\boldsymbol{u} + \boldsymbol{v}) = a\boldsymbol{u} + a\boldsymbol{v}$ (scalar multiplication is distributive over vector addition),
9. $(a + b)\boldsymbol{u} = a\boldsymbol{u} + b\boldsymbol{u}$ (scalar multiplication is distributive over scalar addition), and
10. $a(b\boldsymbol{u}) = (ab)\boldsymbol{u}$ (scalar multiplication is associative).

No basis or metric is mentioned above. While none need be defined, any suitable basis and/or metric may be imposed provided it does not infringe any of the above properties. When a vector space also constitutes a geometric algebra, the metric is implied by properties (24) and (25) given in Section 14.4.2.

14.4.2 A Geometric Algebra

A geometric algebra of dimension N is a graded vector space in which the vectors $\{\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W} \ldots\}$ are more generally called multivectors. A geometric algebra has the

following additional properties that are more concise than, but nevertheless equivalent to, the introductory forms given in Chapter 2.

Property

11. The grades, labeled 0, 1, 2 … N each form separate subsets of the multivectors.

12. A multivector of grade n is called an n-vector.

13. Each grade n is closed with respect to scalar multiplication and n-vector addition.

14. Any multivector in the geometric algebra may be expressed as a linear combination of n-vectors of each grade.

15. The 0-vectors are synonymous with scalars.

16. 1-vectors are synonymous with vectors in the ordinary sense (unless stated to the contrary, *vector* generally implies 1-*vector*).

17. The geometric product of any two multivectors $\boldsymbol{U}$ and $\boldsymbol{V}$ is also multivector and is written as $\boldsymbol{UV}$.

18. Each grade shares the same 0 element with respect to multiplication and addition.

19. Because multivectors are still vectors in the general sense, the usual rules of vector spaces including scalar multiplication and vector addition apply.

20. In keeping with scalar multiplication, geometric multiplication is associative and is also distributive over scalar and vector addition.

21. However, in contrast with scalar multiplication, geometric multiplication is not in general commutative.

22. Any two non-null vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal $\Leftrightarrow \boldsymbol{uv} = -\boldsymbol{vu}$ (i.e., they anticommute).

23. Any two nonzero vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are parallel $\Leftrightarrow \boldsymbol{uv} = \boldsymbol{vu}$ (i.e., they commute).

24. For any vector $\boldsymbol{u}$, $\boldsymbol{u}^2$ is a scalar.

25. The length of any vector $\boldsymbol{u}$ is given by $|\boldsymbol{u}| = |\boldsymbol{u}^2|^{1/2}$.

26. A 1-vector is a blade of grade 1, while for $n \leq 2$, a multivector is a blade of grade n if and only if it can be expressed as the geometric product of n mutually orthogonal 1-vectors.

27. Any n-vector is a linear combination of blades of grade n.

While inner and outer products have not been defined, properties (22) and (23) clearly provide a basis for them. Given any two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, then $(\boldsymbol{u}+\boldsymbol{v})^2 = \boldsymbol{u}^2 + \boldsymbol{uv} + \boldsymbol{vu} + \boldsymbol{v}^2$. Since the square of a vector is associated with the square of its length, Pythagoras' theorem, stated in the form $(\boldsymbol{u}+\boldsymbol{v})^2 = \boldsymbol{u}^2 + \boldsymbol{v}^2 \Leftrightarrow \boldsymbol{u} \perp \boldsymbol{v}$ is equivalent to $\boldsymbol{u} \perp \boldsymbol{v} \Leftrightarrow \boldsymbol{uv} + \boldsymbol{vu} = 0$, in keeping with the definition of orthogonality

in property (22). Property (23), however, is in keeping with the fact that we would expect $\boldsymbol{u} /\!/ \boldsymbol{v}$ to imply $\boldsymbol{v} = \lambda \boldsymbol{u}$ for some nonzero scalar λ (see Theorem (2)).

We now give some simple theorems and their proofs as an illustration of these properties.

Theorem (1) Given any two 1-vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ then $\boldsymbol{uv} + \boldsymbol{vu}$ is a scalar.

Proof:

(i) For any two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, their sum $\boldsymbol{u} + \boldsymbol{v}$ is a vector.
$\Rightarrow (\boldsymbol{u} + \boldsymbol{v})^2 = c$ for some scalar c (by property (24)).

(ii) Vector multiplication is distributive over vector addition.
$\Rightarrow (\boldsymbol{u} + \boldsymbol{v})^2 = (\boldsymbol{u} + \boldsymbol{v})(\boldsymbol{u} + \boldsymbol{v}) = \boldsymbol{u}^2 + \boldsymbol{uv} + \boldsymbol{vu} + \boldsymbol{v}^2 = c$ for some scalar c.

But $\boldsymbol{u}^2 = a$ and $\boldsymbol{v}^2 = b$ for some scalars a and b, again by property (24).
$\Rightarrow \boldsymbol{uv} + \boldsymbol{vu} = c - a - b$, which is also scalar, **QED**.

Theorem (2) For any two nonzero vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, $\boldsymbol{u} /\!/ \boldsymbol{v} \Leftrightarrow \boldsymbol{v} = \lambda \boldsymbol{u}$ for some nonzero scalar λ.

Proof:
We restrict our proof to the case that neither $\boldsymbol{u}$ nor $\boldsymbol{v}$ is null.

By property (23), $\boldsymbol{u} /\!/ \boldsymbol{v} \Leftrightarrow \boldsymbol{uv} = \boldsymbol{vu}$. But by Theorem (1), $\boldsymbol{uv} + \boldsymbol{vu}$ must be a scalar, and so it follows that $\boldsymbol{u} /\!/ \boldsymbol{v} \Leftrightarrow 2\boldsymbol{uv} = 2\boldsymbol{vu} = a$ for some scalar a.

Let $\boldsymbol{u}^2 = b$ for some nonzero scalar b:

$$\begin{aligned} \boldsymbol{u} /\!/ \boldsymbol{v} &\Leftrightarrow 2\boldsymbol{uv} = a \\ &\Rightarrow 2\boldsymbol{u}^2\boldsymbol{v} = a\boldsymbol{u} \\ &\Rightarrow \boldsymbol{v} = \lambda \boldsymbol{u} \quad \text{where} \quad \lambda = \frac{a}{2\boldsymbol{u}^2} = \frac{a}{2b} \end{aligned}$$

This proves $\boldsymbol{u} /\!/ \boldsymbol{v} \Rightarrow \boldsymbol{v} = \lambda \boldsymbol{u}$. Since $\boldsymbol{u} = (1/\lambda)\boldsymbol{v}$ then we also have $\boldsymbol{u} /\!/ \boldsymbol{v} \Leftrightarrow \boldsymbol{v} /\!/ \boldsymbol{u}$.
Now, if $\boldsymbol{v} = \lambda \boldsymbol{u}$, we have

$$\begin{aligned} \boldsymbol{uv} &= \boldsymbol{u}(\lambda \boldsymbol{u}) = \lambda \boldsymbol{u}^2 \\ \boldsymbol{vu} &= (\lambda \boldsymbol{u})\boldsymbol{u} = \lambda \boldsymbol{u}^2 \\ \Rightarrow \boldsymbol{uv} &= \boldsymbol{vu} \\ \Rightarrow \quad \boldsymbol{u} /\!/ \boldsymbol{v}&, \text{ by Property (23)} \end{aligned}$$

This proves $\boldsymbol{u} /\!/ \boldsymbol{v} \Leftarrow \boldsymbol{v} = \lambda \boldsymbol{u}$, and we have already shown $\boldsymbol{u} /\!/ \boldsymbol{v} \Rightarrow \boldsymbol{v} = \lambda \boldsymbol{u}$, **QED**.

Theorem (3) Every non-null vector $\boldsymbol{u}$ has a multiplicative inverse $\boldsymbol{u}^{-1}$ such that $\boldsymbol{u}^{-1}\boldsymbol{u} = \boldsymbol{u}\boldsymbol{u}^{-1} = 1$

Proof:
From property (24), let $\boldsymbol{u}^2 = a$ where, by assumption, the scalar a is nonzero.

Then,

$$\begin{aligned} \boldsymbol{u}^2 = a &\Leftrightarrow \boldsymbol{uu} = a \\ &\Leftrightarrow a^{-1}\boldsymbol{uu} = 1 \\ &\Leftrightarrow \left(a^{-1}\boldsymbol{u}\right)\boldsymbol{u} = 1 \end{aligned}$$

Since the same procedure may equally well be used to show $\boldsymbol{u}\left(a^{-1}\boldsymbol{u}\right) = 1$, it follows that $\boldsymbol{u}^{-1} = a^{-1}\boldsymbol{u}$ is the right and left inverse of $\boldsymbol{u}$, **QED**.

Theorem (4) Given any two nonzero vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, then $\boldsymbol{uv}$ is a scalar $\Leftrightarrow \boldsymbol{uv} = \boldsymbol{vu}$.

Proof:
Take the case first that $\boldsymbol{uv} = a$ for some scalar a. Since $\boldsymbol{u}$ is nonzero, by Theorem (3) it possesses an inverse. Then,

$$\begin{aligned} \boldsymbol{uv} = a &\Leftrightarrow \boldsymbol{v} = \boldsymbol{u}^{-1}a \\ &\Leftrightarrow \boldsymbol{vu} = a\boldsymbol{u}^{-1}\boldsymbol{u} \\ &\Leftrightarrow \boldsymbol{vu} = a \\ &\qquad = \boldsymbol{uv} \end{aligned}$$

Now take the case that $\boldsymbol{uv} = \boldsymbol{vu}$ from which it follows that $\frac{1}{2}(\boldsymbol{uv} + \boldsymbol{vu}) = \boldsymbol{uv} = \boldsymbol{vu}$.

But from Theorem (1), $\boldsymbol{uv} + \boldsymbol{vu}$ is a scalar and therefore so are $\boldsymbol{uv}$ and $\boldsymbol{vu}$.

Consequently, $\boldsymbol{uv}$ is a scalar $\Leftrightarrow \boldsymbol{vu} = \boldsymbol{uv}$. It then must follow that $\boldsymbol{vu}$ is also a scalar, **QED**.

Theorem (5) In any geometric algebra, each grade of n-vectors forms a vector space in its own right (subspace).

Proof:
Property (13) states that each grade (subset of n-vectors) must be closed. Each grade therefore obeys all the required properties of a vector space with each n-vector therein being a vector in the abstract sense. Note that it is not necessary to exclude the scalars (0-vectors) from this.

Theorem (6) Given any pair of nonzero vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, $\boldsymbol{v}$ may be uniquely expressed in the form $\boldsymbol{v} = \boldsymbol{v}_{//} + \boldsymbol{v}_{\perp}$ where $\boldsymbol{v}_{//} = \dfrac{\boldsymbol{v}\cdot\boldsymbol{u}}{\boldsymbol{u}^2}\boldsymbol{u}$ is parallel to $\boldsymbol{u}$ and $\boldsymbol{v}_{\perp} = \dfrac{\boldsymbol{v}\wedge\boldsymbol{u}}{\boldsymbol{u}^2}\boldsymbol{u}$ is orthogonal to $\boldsymbol{u}$.

Proof:
In this proof we introduce $\boldsymbol{uv} = \boldsymbol{u}\cdot\boldsymbol{v} + \boldsymbol{u}\wedge\boldsymbol{v}$ where $\boldsymbol{u}\cdot\boldsymbol{v} \equiv \frac{1}{2}(\boldsymbol{uv} + \boldsymbol{vu})$ and $\boldsymbol{u}\wedge\boldsymbol{v} \equiv \frac{1}{2}(\boldsymbol{uv} - \boldsymbol{vu})$. From Theorem (1), it is clear that $\boldsymbol{u}\cdot\boldsymbol{v}$ is a scalar but as yet we make no claim about the nature of $\boldsymbol{u}\wedge\boldsymbol{v}$.

Assume the proposition to be true. Theorem (2), however, states $\boldsymbol{v}_{//} = a\boldsymbol{u}$ for some scalar a. We then have

$$\begin{aligned} \boldsymbol{uv} &= a\boldsymbol{u}^2 + \boldsymbol{uv}_\perp \\ \boldsymbol{vu} &= a\boldsymbol{u}^2 + \boldsymbol{v}_\perp\boldsymbol{u} \end{aligned} \tag{i}$$

from which $\boldsymbol{uv} + \boldsymbol{vu} = 2a\boldsymbol{u}^2 + (\boldsymbol{uv}_\perp + \boldsymbol{v}_\perp\boldsymbol{u})$. Now since $\boldsymbol{v}_\perp \perp \boldsymbol{u}$, by property (22), $\boldsymbol{uv}_\perp + \boldsymbol{v}_\perp\boldsymbol{u} = 0$ from which it follows that $\boldsymbol{uv} + \boldsymbol{vu} = 2a\boldsymbol{u}^2$ or, put another way, $a\boldsymbol{u}^2 = \boldsymbol{u}\cdot\boldsymbol{v} = \boldsymbol{v}\cdot\boldsymbol{u}$. But given that $\boldsymbol{u}$ is nonzero, by Theorem (3), it possesses an inverse, $\boldsymbol{u}^{-1} = \frac{1}{\boldsymbol{u}^2}\boldsymbol{u}$, so that

$$a\boldsymbol{u}^2 = \boldsymbol{v}\cdot\boldsymbol{u} \Leftrightarrow a\boldsymbol{u} = (\boldsymbol{v}\cdot\boldsymbol{u})\boldsymbol{u}^{-1} = \frac{(\boldsymbol{v}\cdot\boldsymbol{u})\boldsymbol{u}}{\boldsymbol{u}^2}$$

But $\boldsymbol{v}_{//} = a\boldsymbol{u}$ so that $\boldsymbol{v}_{//} = \dfrac{\boldsymbol{v}\cdot\boldsymbol{u}}{\boldsymbol{u}^2}\boldsymbol{u}$.

Now consider $\boldsymbol{uv} - \boldsymbol{vu}$. From (i) above, we have

$$\begin{aligned} \boldsymbol{uv} - \boldsymbol{vu} &= \left(a\boldsymbol{u}^2 + \boldsymbol{uv}_\perp\right) - \left(a\boldsymbol{u}^2 + \boldsymbol{v}_\perp\boldsymbol{u}\right) \\ &= \boldsymbol{uv}_\perp - \boldsymbol{v}_\perp\boldsymbol{u} \\ &= 2\boldsymbol{uv}_\perp \end{aligned}$$

$$\Leftrightarrow \boldsymbol{u}(\boldsymbol{uv} - \boldsymbol{vu}) = 2\boldsymbol{u}^2\boldsymbol{v}_\perp$$

$$\begin{aligned} \Leftrightarrow \boldsymbol{v}_\perp &= \frac{\boldsymbol{u}}{\boldsymbol{u}^2}\frac{(\boldsymbol{uv} - \boldsymbol{vu})}{2} \\ &= \left(\frac{\boldsymbol{vu} - \boldsymbol{uv}}{2}\right)\frac{\boldsymbol{u}}{\boldsymbol{u}^2} \\ &= \frac{\boldsymbol{v}\wedge\boldsymbol{u}}{\boldsymbol{u}^2}\boldsymbol{u} \end{aligned}$$

Therefore, if the conjecture is true, $\boldsymbol{v}_{//}$ and $\boldsymbol{v}_\perp$ are determined as above. It is only necessary to show they have the required properties:

$$\begin{aligned} \boldsymbol{v}_{//} + \boldsymbol{v}_\perp &= \frac{\boldsymbol{v}\cdot\boldsymbol{u}}{\boldsymbol{u}^2}\boldsymbol{u} + \frac{\boldsymbol{v}\wedge\boldsymbol{u}}{\boldsymbol{u}^2}\boldsymbol{u} \\ &= \frac{\boldsymbol{v}\cdot\boldsymbol{u} + \boldsymbol{v}\wedge\boldsymbol{u}}{\boldsymbol{u}^2}\boldsymbol{u} \\ &= \frac{\boldsymbol{vu}}{\boldsymbol{u}^2}\boldsymbol{u} \\ &= \boldsymbol{v} \end{aligned} \tag{ii}$$

Since $\boldsymbol{v}_{//} = \dfrac{\boldsymbol{v}\cdot\boldsymbol{u}}{\boldsymbol{u}^2}\boldsymbol{u}$ and $\dfrac{\boldsymbol{v}\cdot\boldsymbol{u}}{\boldsymbol{u}^2}$ is a scalar, by virtue of Theorem (2),

$$\boldsymbol{v}_{//} \,//\, \boldsymbol{u} \tag{iii}$$

Now from (i) above, $\boldsymbol{v}_{\perp} = \boldsymbol{v} - \boldsymbol{v}_{//}$ so that

$$\begin{aligned} \boldsymbol{u}\boldsymbol{v}_{\perp} + \boldsymbol{v}_{\perp}\boldsymbol{u} &= \boldsymbol{u}(\boldsymbol{v} - \boldsymbol{v}_{//}) + (\boldsymbol{v} - \boldsymbol{v}_{//})\boldsymbol{u} \\ &= \boldsymbol{u}\boldsymbol{v} + \boldsymbol{v}\boldsymbol{u} - \boldsymbol{u}\boldsymbol{v}_{//} - \boldsymbol{v}_{//}\boldsymbol{u} \\ &= \boldsymbol{u}\boldsymbol{v} + \boldsymbol{v}\boldsymbol{u} - \boldsymbol{u}\frac{\boldsymbol{v}\cdot\boldsymbol{u}}{\boldsymbol{u}^2}\boldsymbol{u} - \frac{\boldsymbol{v}\cdot\boldsymbol{u}}{\boldsymbol{u}^2}\boldsymbol{u}\boldsymbol{u} \\ &= 2\boldsymbol{u}\cdot\boldsymbol{v} - 2\boldsymbol{v}\cdot\boldsymbol{u} \\ &= 0 \\ \Leftrightarrow \quad \boldsymbol{v}_{\perp}\boldsymbol{u} &= -\boldsymbol{u}\boldsymbol{v}_{\perp} \\ \Leftrightarrow \quad \boldsymbol{v}_{\perp} &\perp \boldsymbol{u} \quad \text{(Property 22)} \end{aligned} \tag{iv}$$

From (ii), (iii), and (iv), the vectors $\boldsymbol{v}_{//}$ and $\boldsymbol{v}_{\perp}$ have all the required properties, **QED**.

Several other results follow as corollaries.

Theorem (7) Given any pair of vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ then, provided that it does not vanish, $\boldsymbol{u} \wedge \boldsymbol{v}$ is a blade of grade 2.

Proof:
By Theorem (6), we can write $\boldsymbol{v} = \boldsymbol{v}_{//} + \boldsymbol{v}_{\perp}$ where $\boldsymbol{v}_{//} // \boldsymbol{u}$ and $\boldsymbol{v}_{\perp} \perp \boldsymbol{u}$. Since $\boldsymbol{v}_{\perp}$ is given by $(\boldsymbol{v} \wedge \boldsymbol{u})/(\boldsymbol{u}^2)\boldsymbol{u}$, it follows that $\boldsymbol{v} \wedge \boldsymbol{u} = \boldsymbol{v}_{\perp}\boldsymbol{u}$. By property (26), however, $\boldsymbol{v}_{\perp}\boldsymbol{u}$ must be a blade of grade 2, so that $\boldsymbol{v} \wedge \boldsymbol{u}$ is likewise a blade of grade 2. By property (27), it is also a 2-vector, **QED**.

Theorem (8) Given any pair of vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, $\boldsymbol{u}\boldsymbol{v}$ may be uniquely expressed as $\boldsymbol{u}\cdot\boldsymbol{v} + \boldsymbol{u}\wedge\boldsymbol{v}$ where $\boldsymbol{u}\cdot\boldsymbol{v} \equiv \frac{1}{2}(\boldsymbol{u}\boldsymbol{v} + \boldsymbol{v}\boldsymbol{u})$ is a scalar and $\boldsymbol{u}\wedge\boldsymbol{v} \equiv \frac{1}{2}(\boldsymbol{u}\boldsymbol{v} - \boldsymbol{v}\boldsymbol{u})$ is a 2-blade.

Proof:
By Theorems (1) and (7), $\boldsymbol{u}\cdot\boldsymbol{v}$ is a scalar while $\boldsymbol{u}\wedge\boldsymbol{v}$ is a 2-blade, **QED**.

This provides a basis for the definition of the inner and outer product for vectors, which may then be generalized to inner and outer products between vectors and n-vectors, and from there to inner and outer products between different grades of n-vectors.

Some fundamental theorems from linear vector spaces may also be invoked, for example:

Theorem (9) In any geometric algebra, each grade n may be spanned by a set of linearly independent n-vectors.

In an N-dimensional geometric algebra, the number of independent n-vectors required is $\binom{N}{n}$, with grades 0 and N being trivial cases. In the case of grade 1, we already have definitions of orthogonality so that this theorem immediately extends to the following:

Theorem (10) In any geometric algebra, the vector subspace of grade 1 may be spanned by an orthonormal set of vectors.

In principle, these ideas may be extended to higher grades by generalizing properties (22), (23), and (25) in a suitable way, for example:

Property

22′. Any two non-null multivectors $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthogonal $\Leftrightarrow \langle \boldsymbol{UV} \rangle = 0$.

23′. Any two blades $\boldsymbol{U}$ and $\boldsymbol{V}$ are parallel $\Leftrightarrow \boldsymbol{UV}$ is a scalar.

25′. The measure of any n-vector $\boldsymbol{U}$ is given by $|\boldsymbol{U}| = \left|\langle \boldsymbol{UU}^{\dagger} \rangle\right|^{1/2}$ where $\boldsymbol{U}^{\dagger}$ is the reverse of $\boldsymbol{U}$.

Property (23′) is by no means essential, but it does, for example, provide a basis for using the term parallel with bivectors. Properties (22′) and (25′), however, allow Theorem (10) to apply to subspaces of every grade, so that it is always possible to find a complete set of orthonormal basis elements for a geometric algebra. This may be stated as follows:

Theorem (11) In any geometric algebra, each subspace of grade n may be spanned by an orthonormal set of n-blades.

It will be appreciated that grades 0 and N have to be treated as trivial cases. While the generalizations (22′), (23′), and (25′) are interesting to explore, they introduce practical and conceptual complications, and so it is less confusing to focus on the basic properties.

14.5 QUATERNIONS COMPARED

Although they appear to have had a completely separate origin, Hamilton's quaternions [15] may be regarded as an offshoot of geometric algebra. In the absence of property (24) of a geometric algebra, it would be permissible for the product of any two orthogonal vectors to result in a third vector rather than a bivector. This makes it possible to close the structure with just two grades, 0 and 1, resulting in the creation of the quaternions as a related, but separate, sort of 3D algebra in which property (26) is replaced by the following:

Any two orthogonal vectors together with their quaternion product make up an orthogonal triad of vectors.

A quaternion $\boldsymbol{Q}$ is therefore of the form

$$\boldsymbol{Q} = p + x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k} \tag{14.1}$$

where p, x, y, z are real scalars and, in the customary notation, $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$ are the three basis vectors and we take $\boldsymbol{k} = \boldsymbol{ij}$. It then follows that $\boldsymbol{i}^2 = \boldsymbol{j}^2 = \boldsymbol{k}^2 = -1$; in other words, the metric signature is $(---)$. Although the quaternions are similar in concept to multivectors, Hamilton principally saw them as an extension of complex numbers from 2D to (3+1)D. So defined, the product of any two vector quaternions $\boldsymbol{u}$ and $\boldsymbol{v}$ may be expressed as $\boldsymbol{uv} = \boldsymbol{u} \times \boldsymbol{v} - \boldsymbol{u} \cdot \boldsymbol{v}$, which is clearly analogous to a 3D geometric algebra where we have $\mathbf{uv} = I\,\mathbf{u} \times \mathbf{v} + \mathbf{u} \cdot \mathbf{v}$. Due to the $(---)$ metric signature, the vector derivative needs to take the form $\nabla = -\boldsymbol{i}\partial_x - \boldsymbol{j}\partial_y - \boldsymbol{k}\partial_z$, which gives $\nabla \boldsymbol{Q} = -\boldsymbol{\nabla} p + \boldsymbol{\nabla} \cdot \mathbf{q} - \boldsymbol{\nabla} \times \mathbf{q}$ where $\boldsymbol{Q} = p + \mathbf{q}$ is some differentiable quaternion function. Signs apart, the three terms on the right-hand side will be immediately recognized as gradient, divergence, and curl.

The quaternions are structurally different from a geometric algebra because they lack both bivectors and pseudoscalars. One way of looking at it is that these elements have been mapped back onto the vectors and scalars through $I(a + \mathbf{u}) \mapsto -a + \mathbf{u}$. But while the inherently negative metric signature of the quaternions seems little more than nuance, it gives a clue to an alternative and altogether different view [27, section 2.4.2, p. 34]. Recall that the basis bivectors in a 3D geometric algebra obey $(\mathbf{yz})^2 = (\mathbf{zx})^2 = (\mathbf{xy})^2 = -1$, which is clearly just like $\boldsymbol{i}^2 = \boldsymbol{j}^2 = \boldsymbol{k}^2 = -1$. Unfortunately, however, in comparison with $\boldsymbol{ijk}$, $(\mathbf{yz})(\mathbf{zx})(\mathbf{xy})$ yields +1 rather than −1. But if we alter the handedness of these bivectors to be a left-handed set, say by replacing $\mathbf{xy}$ with $\mathbf{yx}$, we still get $(\mathbf{yz})^2 = (\mathbf{zx})^2 = (\mathbf{yx})^2 = -1$ but now we also get the remaining essential property, $(\mathbf{yz})(\mathbf{zx})(\mathbf{yx}) = -1$. The quaternions are therefore a subalgebra of the 3D geometric algebra *in which only the even elements have been retained* (interestingly, the bivector with the swapped sign plays the part of the pseudoscalar). As far as the quaternion algebra is concerned, these bivectors are just vectors in the broader abstract sense, just as we find in the even subalgebra of spacetime (Section 8.1.1).

It is easy to see that quaternions are more than a mere curiosity. Their physical significance is even more apparent if we replace the arbitrary scalar p in Equation (14.1) with t for time, which leads straight to Hamilton's own concept of (3+1)D as discussed in Section 3.1.[1] They provide a useful means of describing rotations and certain extensions of analytic functions [28, sections 5.4 and 5.8] and in fact, in his treatise, Maxwell adopted them as his toolset for vector analysis. On the other hand, they fall short of the sort of graded structure that can differentiate between 1-vectors and bivectors, which is one of the things that make geometric algebra so useful when it comes to encoding the electromagnetic field. Although quaternions are essentially (3+1)D, they do not form part of any obvious overarching structure corresponding to the spacetime geometric algebra where the ability to treat time as a vector holds the key to the unambiguous representation of different Lorentz frames.

[1] It is interesting that this idea of Hamilton's developed some 60 years before relativity and spacetime, a prime example of an idea being before its time!

14.6 EVALUATION OF AN INTEGRAL IN EQUATION (5.14)

It was asserted in Section 5.4 that the integral

$$\frac{-1}{4\pi}\int_V \frac{(\mathbf{r}-\mathbf{r}')\cdot\mathbf{J}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3}d^3r'$$

which appears in the second line of Equation (5.14), vanishes under the assumptions that $\partial_t = 0$ and that $\boldsymbol{J}$ vanishes beyond a finite region of space V, albeit that this can be as large as we please. This result is actually implied in Jackson's evaluation of $\boldsymbol{\nabla}\times\mathbf{B}$ for the magnetostatic field arising from $\mathbf{J}$ [37, section 5.4, p. 138], but here we give the details. Starting with the identity $\mathbf{r}/r^3 = \boldsymbol{\nabla}\left(r^{-1}\right)$, we may put the integral in a slightly different form:

$$\int_V \frac{(\mathbf{r}-\mathbf{r}')\cdot\mathbf{J}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3}d^3r' = \int_V \left(\boldsymbol{\nabla}'|\mathbf{r}-\mathbf{r}'|^{-1}\right)\cdot\mathbf{J}(\mathbf{r}')d^3r' \tag{14.2}$$

This allows us to make use of the identity $(\boldsymbol{\nabla}a)\cdot\mathbf{J} + a(\boldsymbol{\nabla}\cdot\mathbf{J}) = \boldsymbol{\nabla}\cdot(a\mathbf{J})$, and since we have assumed $\partial_t = 0$, the continuity equation $\boldsymbol{\nabla}\cdot\mathbf{J} + \partial_t\rho = 0$ reduces to $\boldsymbol{\nabla}\cdot\mathbf{J} = 0$ so that here we simply have $(\boldsymbol{\nabla}a)\cdot\mathbf{J} = \boldsymbol{\nabla}\cdot(a\mathbf{J})$. We may then put Equation (14.2) into the form

$$\int_V \frac{(\mathbf{r}-\mathbf{r}')\cdot\mathbf{J}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3}d^3r' = \int_V \boldsymbol{\nabla}'\cdot\left(|\mathbf{r}-\mathbf{r}'|^{-1}\,\mathbf{J}(\mathbf{r}')\right)d^3r' \tag{14.3}$$

The integrand on the right-hand side here is a divergence so that, by Gauss' theorem, we may reduce the volume integral to an integral over the surface S that encloses V:

$$\int_V \left(\boldsymbol{\nabla}'|\mathbf{r}-\mathbf{r}'|^{-1}\right)\cdot\mathbf{J}(\mathbf{r}')d^3r' = \int_S |\mathbf{r}-\mathbf{r}'|^{-1}\,\mathbf{J}(\mathbf{r}')\cdot d\mathbf{s} \tag{14.4}$$

If we take V as being a sphere of a radius R that is very much larger than both r and r', it follows that if $\mathbf{J}$ vanishes on this surface, or at least as long as its magnitude decays faster than $1/R$ as $R\to\infty$, this surface integral also vanishes. This therefore proves the assertion that the first integral in the second line of Equation (5.14) vanishes.

It is interesting, however, to go back to the general case where $\partial_t \neq 0$. We then find that while the integrals in Equation (14.3) still vanish, we no longer have $\boldsymbol{\nabla}\cdot\mathbf{J} = 0$ so that, on bringing back in the factor of $-1/4\pi$, we find

$$\begin{aligned}\frac{-1}{4\pi}\int_V \frac{(\mathbf{r}-\mathbf{r}')\cdot\mathbf{J}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3}d^3r' &= \frac{1}{4\pi}\int_V |\mathbf{r}-\mathbf{r}'|^{-1}\nabla'\cdot\mathbf{J}(\mathbf{r}')d^3r' \\ &= \frac{-1}{4\pi}\int_V |\mathbf{r}-\mathbf{r}'|^{-1}\partial_t\rho\, d^3r' \\ &= -\partial_t\left(\frac{1}{4\pi}\int_V \rho|\mathbf{r}-\mathbf{r}'|^{-1}d^3r'\right) \\ &= -\partial_t\Phi(\mathbf{r})\end{aligned} \tag{14.5}$$

By including this contribution in Jackson's evaluation of $\nabla\times\mathbf{B}$, the result becomes $\nabla\times\mathbf{B}=\mathbf{J}-\nabla(\partial_t\Phi)$, and since $-\nabla\Phi=\mathbf{E}$, this is readily rearranged to give $\nabla\times\mathbf{B}=\mathbf{J}+\partial_t\mathbf{E}$, Maxwell's fourth equation in free space. From a geometric algebra perspective, if Equation (5.14) is to represent a magnetic field, the first integral on its second line *must* vanish because it results in a scalar. From Equation (14.5), the scalar in question is clearly $-\partial_t\Phi(\mathbf{r})$. Equation (5.13) may therefore be stated in completely general terms as

$$\boldsymbol{F}(\mathbf{r})=\frac{1}{4\pi}\int_V \frac{(\mathbf{r}-\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3}\mathbf{J}(\mathbf{r}')d^3r'-\partial_t\boldsymbol{A} \tag{14.6}$$

Since from Equation (5.30) $\boldsymbol{F}$ may be written as $(\nabla-\partial_t)\boldsymbol{A}$, without the above amendment Equation (5.13) simply corresponds to $\boldsymbol{F}=\nabla\boldsymbol{A}$ rather than $\boldsymbol{F}=(\nabla-\partial_t)\boldsymbol{A}$. This is, of course, consistent with the fact that it involves integration only over space, not time. It is also consistent with the fact that in magnetostatics, we may express $\mathbf{A}$ as

$$\mathbf{A}(\mathbf{r})=\frac{1}{4\pi}\int_V \frac{\mathbf{J}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d^3r' \tag{14.7}$$

for it can be seen that Equation (5.13) follows simply from taking the vector derivative of both sides here with respect to $\mathbf{r}$.

14.7 FORMAL DERIVATION OF THE SPACETIME VECTOR DERIVATIVE

Here we provide a more rigorous derivation of the form of the spacetime derivative given in Section 7.8 above. We follow a similar argument to the one outlined by Doran and Lasenby [27, pp. 100 and 168] without actually mentioning reciprocal vectors. It is useful to employ indices on the basis vectors so we use $\boldsymbol{e}_0,\boldsymbol{e}_1,\boldsymbol{e}_1,\boldsymbol{e}_3$ corresponding to our usual $\boldsymbol{t},\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}$ and metric signature so that $\boldsymbol{e}_1^2=\boldsymbol{e}_2^2=\boldsymbol{e}_3^2=-\boldsymbol{e}_0^2$.

We have to be careful with $\boldsymbol{\nabla}$ because the differentiation is with respect to the components of the vector $\boldsymbol{r} = \sum_k r_k \boldsymbol{e}_k$, and if we change the basis, the components change accordingly. We start out with the definition of $\boldsymbol{\nabla}$ based on a Euclidean norm $\boldsymbol{\nabla} = \sum_k \mathbf{e}_k \partial_k$ and then seek the result of using the $\boldsymbol{e}_k$ basis instead. Let us refer to the components of $\boldsymbol{r}$ expressed in that basis as r'_k and ∂ / r'_k as ∂'_k. The $\boldsymbol{e}_k$ basis is orthogonal, and so for each vector in the set, we can take $\boldsymbol{e}_k$ to be along $\mathbf{e}_k$, for only its measure is different. If we then let $\boldsymbol{e}_k = h_k \mathbf{e}_k$, where we put no constraints on the coefficient h_k other than it should not vanish, it must follow that $r_k \mathbf{e}_k = r'_k \boldsymbol{e}_k = r'_k h_k \mathbf{e}_k$ so that $r'_k h_k = r_k$. Turning now to the derivative, by the chain rule we find $\partial'_k = \partial / \partial r'_k = (\partial r_k / \partial r'_k) \partial / \partial r_k = h_k \partial_k$, and so putting together these two results, we find

$$\mathbf{e}_k \partial_k = \frac{1}{h_k^2} \boldsymbol{e}_k \partial'_k = \frac{1}{\boldsymbol{e}_k^2} \boldsymbol{e}_k \partial'_k \tag{14.8}$$

The last step requires $\boldsymbol{e}_k^2 = h_k^2 \mathbf{e}_k^2 = h_k^2$, which simply follows from $\boldsymbol{e}_k = h_k \mathbf{e}_k$. Now, so long as it does not vanish, we may choose h_k in any way we please. We may therefore take Equation (14.8) as being generally valid even if we only specify h_k^2, that is to say, $\boldsymbol{e}_k^2$. We may now write down $\boldsymbol{\nabla}$ for any given metric signature:

$$\nabla = \sum_k \mathbf{e}_k \partial_k = \sum_k \frac{1}{\boldsymbol{e}_k^2} \boldsymbol{e}_k \partial'_k \tag{14.9}$$

We will have no further need to refer to the $\mathbf{e}_k$ basis, and so the prime on ∂ has become redundant. For our particular case, we may therefore write

$$\begin{aligned} \boldsymbol{\nabla} &= -\boldsymbol{e}_0 \partial_0 + \sum_{k=1}^{3} \boldsymbol{e}_k \partial_k \\ &\equiv -\boldsymbol{t} \partial_t + \boldsymbol{x} \partial_x + \boldsymbol{y} \partial_y + \boldsymbol{z} \partial_z \end{aligned} \tag{14.10}$$

as required.

References

1. J. C. Maxwell, "On Faradays Lines of Force," Trans. Camb. Philos. Soc., **X**(1), pp. 25–83, 1856. Cambridge University Press, 1864 (see p. 208). Available at http://www.biodiversitylibrary.org/item/19852#7. Also reproduced in *The Scientific Papers of James Clerk Maxwell*, W. D. Niven, ed., pp. 155–229, Dover, New York, 1965 (two volumes bound as one). Available at http://www.archive.org/details/scientificpapers01maxw.
2. J. W. Arthur, "The Fundamentals of Electromagnetic Theory Revisited," IEEE Antennas Propag. Mag., **50**(1), pp. 19–65, February 2008 and "Correction," IEEE Antennas Propag. Mag., **50**(4), p. 65, August 2008.
3. K. F. Warnick, R. H. Selfridge, and D. V. Arnold, "Teaching Electromagnetic Field Theory Using Differential Forms," IEEE Trans. Ed., **40**(1), pp. 53–68, 1997.
4. I. V. Lindell, *Differential Forms in Electromagnetics*, IEEE Press Series on Electromagnetic Wave Theory, Wiley, New York, 2004.
5. H. Flanders, *Differential Forms*, Academic Press, New York, 1963.
6. D. Hestenes, "Oersted Medal Lecture 2002: Reforming the Mathematical Language of Physics," Am. J. Phys., **71**(2), 2003, pp. 104–121.
7. D. Hestenes, *New Foundations of Classical Mechanics*, 2nd ed., Kluwer Academic, Dordrecht, 1992.
8. S. Gull, A. Lasenby, and C. Doran, "Imaginary Numbers Are Not Real—The Geometric Algebra of Spacetime," Found. Phys., **23**(19), pp. 1175–1202, 1993.
9. R. Rynasiewicz, in *Stanford Encyclopedia of Philosophy*, "Supplement to Newton's Views on Space, Time, and Motion—Newton's Scholium on Time, Space, Place and Motion," E. N. Zalta, ed., 2007. Available at http://plato.stanford.edu/index.html.
10. J. T. Moore, *Elements of Abstract Algebra*, Chapter 6, Macmillan, New York, 1961.
11. G. Birkhoff and S. MacLane, *A Survey of Modern Algebra*, 3rd edn., Macmillan, New York, 1997.
12. W. Kaplan, *Advanced Calculus*, Chapters 2–5, pp. 32–292, Adisson-Wesley, Reading, MA, 1959.
13. J. C. Maxwell, "A Dynamical Theory of the Electromagnetic Field," Trans. R. Soc., **155**, pp. 459–512, 1865.
14. J. C. Maxwell, *A Treatise on Electricity and Magnetism*, (1873), two volumes bound as one, 3rd ed., Dover, New York, 1954.
15. W. R. Hamilton, "On a New Species of Imaginary Quantities Connected with a Theory of Quaternions," Proc. R. Ir. Acad., **2**, pp. 424–434, 1844. Available at http://www.emis.de/classics/Hamilton/Quatern1.pdf.
16. O. Heaviside, *Electrical Papers*, Vol. 1, McMillan, London, 1892. Reprinted by Chelsea, New York, 1970.
17. J. W. Gibbs, in *The Scientific Papers of J. Willard Gibbs*, Vol. II, "Dynamics,

Understanding Geometric Algebra for Electromagnetic Theory, First Edition. John W. Arthur.

Vector Analysis and Multiple Algebra Electromagnetic Theory of Light etc," H. A. Bumstead and R. G. Van Name, eds., Dover, New York, 1961.

18. J. C. Kolecki, "An Introduction to Tensors for Students of Physics and Engineering," NASA Report NASA TM 2002-211716, 2002. Available at http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20020083040_2002137571.pdf.
19. E. Butkov, *Mathematical Physics*, Addison-Wesley, Reading, MA, 1968.
20. D. H. Menzel, *Mathematical Physics*, Dover, New York, 1961.
21. L. Brand, *Vector and Tensor Analysis*, John Wiley & Sons, New York, 1953.
22. H. Cartan, *Differential Forms*, Dover, New York, 2006.
23. H. G. Grassmann, *Die Lineale Ausdehnungslehre, ein Neuer Zweig der Mathematik*, Wigand, Leipzig, 1844.
24. W. K. Clifford, *Applications of Grassmann's Extensive Algebra*, Am. J. Math., **1**, pp. 350–358, 1878.
25. D. Hestenes, *Spacetime Algebra*, Gordon and Breach, New York, 1966.
26. W. E. Baylis, ed., *Clifford (Geometric) Algebras with Applications to Physics Mathematics and Engineering*, Birkhauser, Boston, MA, 1996.
27. C. Doran and A. Lasenby, *Geometric Algebra for Physicists*, Cambridge University Press, Cambridge, 2003.
28. P. Lounesto, *Clifford Algebras and Spinors*, 2nd ed., Cambridge University Press, Cambridge, 2001.
29. R. P. Graves, *Life of Sir William Rowan Hamilton, Andrews Professor of Astronomy in the University of Dublin, and Royal Astronomer of Ireland, Including Selections from His Poems, Correspondence, and Miscellaneous Writings*, Vols. 2 and 3, Hodges & Figgis, Dublin, 1885 and 1889.
30. H. Weyl, *Space Time Matter*, translated by H. L. Brose, Methuen, London, 1922. Available online at Canadian Libraries Archive: http://www.archive.org/details/spacetimematter00weyluoft.
31. A. S. Eddington, *The Mathematical Theory of Relativity*, 2nd ed., Cambridge University Press, London, 1924.
32. J. Lasenby, L. Dorst, and C. Doran, eds., *Applications of Geometric Algebra in Computer Science and Engineering*, Birkhauser, Boston, MA, 2002.
33. D. Hestenes and G. Sobczyk, *Clifford Algebra to Geometric Calculus: A Unified Language for Mathematics and Physics*, Kluwer Academic, Dordrecht, 1992.
34. D. Hestenes, "Spacetime Physics with Geometric Algebra," Am. J. Phys., **71**(7), pp. 1–57, 2003.
35. J. A. Stratton, *Electromagnetic Theory*, McGraw-Hill, New York, 1941.
36. L. V. Lorenz, "On the Identity of the Vibrations of Light with Electrical Currents," Phil. Mag., Series 3, **34**, pp. 287–301, 1867.
37. J. D. Jackson, *Classical Electrodynamics*, 3rd ed., Wiley, New York, 1999.
38. J. W. Arthur, "An Elementary View of Maxwell's Displacement Current," IEEE Antennas Propag. Mag., **51**(6), pp. 58–68, December 2009.
39. A.-M. Liénard, "Champ electrique et magnetique produit par une charge electrique concentré en un point et animée d'un mouvement quelconque," Eclairage Electrique, **XVI**, pp. 5–14, 53–59 & 106–112, 1898.
40. E. Wiechert, "Electrodynamishe Elementargesetze," Archives Néerlandaises des Sciences Exactes et Naturelles Serié 2, **5**, Livre Jubilaire Dedié á H. A. Lorentz, pp. 549–573, 1900.
41. S. A. Matos, M. A. Rebeiro, and C. R. Paiva, "Anisotropy without Tensors: A

Novel Approach Using Geometric Algebra," Opt. Express, **15**(23), pp. 175–186, November 2007.

42. A. Einstein, H. A. Lorentz, H. Weyl, and H. Minkowski, including notes by A. Sommerfeld, *The Principle of Relativity: A Collection of Original Memoirs on the Special and General Theory of Relativity*, translated by W. Perrett and G. B. Jeffrey, Dover, New York, 1952.
43. L. Pyenson, "Hermann Minkowski and Einstein's Special Theory of Relativity," Arch. Hist. Exact Sci., Springer, **17**(1), pp. 71–95, 1977.
44. N. M. J. Woodhouse, *Special Relativity*, Springer Verlag, Undergrad. Maths Series, London, 2003.
45. B. Laurent, *Introduction to Spacetime: A First Course in Relativity*, World Scientific, Singapore, 1994.
46. D. Hestenes, "New Foundations for Mathematical Physics," (unpublished) 1998. Available at http://geocalc.clas.asu.edu/html/NFMP.html.
47. D. Hestenes, "Spacetime Calculus", 1998. Available at http://geocalc.clas.asu.edu/html/STC.html.
48. T. M. Helliwell, *Introduction to Relativity*, Allyn and Bacon, Boston, MA, 1966.
49. E. A. Milne, *Gravitation, Relativity and World-Structure*, Oxford University Press, Oxford, 1935.
50. R. P. Feynman, R. B. Leighton, and M. Sands, in *The Feynman Lectures on Physics—The Electromagnetic Field*, "Field Energy and Field Momentum," Addison-Wesley, Reading, MA, 1965.
51. D. Hestenes, "Proper Particle Mechanics," J. Math. Phys., **15**(10), pp. 1768–1777, October 1974.

Further Reading

In addition to the books and articles on the subject that are cited in the references, the following is a selection, listed by leading author, of the available material on geometric algebra and its applications to physics and engineering. Also listed are some related works on electromagnetic fundamentals and differential forms.

1. R. Abramowitz and G. Sobczyk, eds., *Lectures on Clifford (Geometric) Algebras and Applications*, Birkhauser, Boston, MA, 2004.
2. E. Artin, "Geometric Algebra," Interscience, 1957. Available at the Universal Library Internet Archive http://www.archive.org/details/geometricalgebra033556mbp.
3. W. E. Baylis, *Electrodynamics: A Modern Geometric Approach*, Birkhauser, Boston, MA, 2002.
4. E. B. Corrochano and G. Sobczyk, eds., *Geometric Algebra with Applications in Science and Engineering*, Birkhauser, Boston, MA, 2001.
5. J. Denker, "Introduction to Clifford Algebra," The Clifford algebra formulation of electromagnetism and other interesting items. Available at http://www.av8n.com/physics.
6. V. De Sabbata and B. K. Datta, *Geometric Algebra and Applications to Physics*, Taylor and Francis, CRC Press, Boca Raton, FL, 2002.
7. L. Dorst, C. Doran, and J. Lasenby, eds., *Applications of Geometric Algebra in Computer Science and Engineering*, Birkhauser, Boston, MA, 2002.
8. P. R. Girard, *Quaternions Clifford Algebras and Relativistic Physics*, Birkhauser, Boston, MA, 2007.
9. R. G. Harke, "An Introduction to the Mathematics of Space-Time Algebra," 1998. Available at www.harke.org/ps/intro.ps.gz.
10. F. W. Hehl and Y. N. Obukhov, *Foundations of Classical Electrodynamics—Charge, Flux and Metric*, Progress in Mathematical Physics, Vol. 33, Birkhauser, Boston, MA, 2003.
11. D. Hestenes, *A Unified Language for Mathematics and Physics* in *Clifford Algebras and Their Applications in Mathematical Physics*, J.S.R. Chisholm and A.K. Common eds., pp. 1–23, Kluwer, Dordrecht, 1986.
12. D. Hestenes, "Multivector Calculus," J. Math. Anal. Appl., **24**(2), pp. 313–325, November 1968.
13. D. Hestenes, "Multivector Functions," J. Math. Anal. Appl., **24**(3), pp. 467–473, December 1968.
14. B. Jancewicz, *Multivectors and Clifford Algebra in Electrodynamics*, World Scientific, Singapore, 1989.
15. J. Lasenby, A. Lasenby, and C. Doran, "A Unified Mathematical Language for

Understanding Geometric Algebra for Electromagnetic Theory, First Edition. John W. Arthur.

Physics and Engineering in the 21st Century," Philos. Trans. R. Soc. Lond. A, **358**, pp. 21–39, January 2000.

16. J. Snygg, *Clifford Algebra: A Computational Tool for Physicists*, Oxford University Press, Oxford, 1997.
17. T. G. Vold, *An introduction to geometric calculus and its application to electrodynamics*, Am. J. Phys., **61**(6), pp. 505–513, February 1993.
18. D. R. Rowland, *On the value of geometric algebra for spacetime analyses using an investigation of the form of the self-force on an accelerating charged particle as a case study*, Am. J. Phys., **78**(2), pp. 187–194, February 2010.

Index

Understanding Geometric Algebra for Electromagnetic Theory, First Edition. John W. Arthur.