Special Relativity
Special Relativity
Preface

These lectures were given to third-year mathematics undergraduates at Oxford in the late 1970s and early 1980s. The notes were produced originally in mimeographed form by the Mathematical Institute at Oxford in 1977, and in a revised edition in 1980. I have made further minor changes and corrections in this edition, and added some examples and exercises from problem sheets given out in lectures by Roger Penrose and Paul Tod.

Special relativity provides one of the more interesting pedagogical challenges. This particular course was given to students with a strong mathematical background who already had a good grounding in classical mathematical physics, but who had not yet met relativity. The emphasis is on the use of coordinate-free and tensorial methods: I tried to avoid the traditional arguments based on the standard Lorentz transformation, and to encourage students to look at problems from a four-dimensional point of view. I did not attempt to 'derive' relativity from a minimal set of axioms, but instead concentrated on stating clearly the basic principles and assumptions. Elsewhere in the world, relativity is usually introduced in a more elementary way earlier in undergraduate courses, and even at Oxford, it is now part of the second-year syllabus in mathematics. I doubt, therefore, that anyone would contemplate giving a lecture course exactly along these lines. Nevertheless, I hope that the notes may provide one or two ideas.

I have not attempted to produce a polished textbook. There are enough good ones in print already. The notes are more or less as I originally wrote them to supplement the lectures. They clearly owe much to other books, notably those by Bondi¹, Rindler², and Synge³. It is difficult after so many years to track down all the original sources, particularly when it comes to problems, which are often borrowed from textbooks and set anonymously as examination questions. I hope that I will be forgiven for not acknowledging in detail ideas that are part of the common currency of the subject. Many of the exercises were set in the Mathematics Final Honour School at Oxford.

I am grateful to Wolfgang Rindler for suggesting to Springer that the notes should be published more widely. He kindly acknowledged some arguments from these notes in Introduction to special relativity. The major influence, however, was in the other direction.

Oxford, July 1991

N. M. J. Woodhouse

² W. Rindler, Essential relativity, Springer-Verlag, Berlin, 1977; Special relativity, Oliver and Boyd, Edinburgh, 1960; Introduction to special relativity, Oxford University Press, Oxford, 1982. Exercises and examples taken directly from Special relativity are marked with a dagger (†).
³ J. L. Synge, Relativity, the special theory, North-Holland, Amsterdam, 1955.
# Contents

1 Space, Time and Maxwell's Equations ........................................... 1
  1.1 Introduction ........................................................................... 1
  1.2 Maxwell's Equations and Galilean Transformations ................... 1
  1.3 The Newtonian View of Space and Time ................................. 3

2 Inertial Coordinates ................................................................. 5
  2.1 Basic Principles ..................................................................... 5
  2.2 Lorentz Transformations ....................................................... 8
  2.3 Time Dilatation ..................................................................... 11
  2.4 The Standard Lorentz Transformation .................................. 12
  2.5 The Lorentz Contraction ....................................................... 14
  2.6 Transformation of Volumes ................................................... 15
  2.7 Addition of Velocities ......................................................... 16
  2.8 The Michelson-Morley Experiment ....................................... 17

3 Vectors in Space-Time ............................................................... 19
  3.1 4-Vectors ............................................................................ 19
  3.2 The Metric ........................................................................... 20
  3.3 The Causal Structure of Minkowski Space ............................... 21
  3.4 Orthonormal Frames ............................................................ 22
  3.5 Decomposition of 4-Vectors .................................................. 24
  3.6 Vector Fields ........................................................................ 25
  3.7 4-Velocity and Proper Time .................................................. 25
  3.8 4-Acceleration ...................................................................... 26
  3.9 Relative Speed ...................................................................... 29
  3.10 The 4-Gradient .................................................................... 29
  3.11 The Frequency 4-Vector ....................................................... 30
  3.12 The Appearance of Moving Objects ..................................... 31
  3.13 Examples and Exercises ...................................................... 33

4 Relativistic Particles ................................................................. 37
  4.1 Conservation of 4-Momentum ................................................. 37
  4.2 Equivalence of Mass and Energy .......................................... 39
  4.3 The 4-Momentum of a Photon ............................................. 39
  4.4 4-Force ................................................................................. 40
  4.5 Examples and Exercises ...................................................... 40

5 Tensors in Space-Time ............................................................. 45
  5.1 Covectors .............................................................................. 45
  5.2 Raising and Lowering of Indices ......................................... 46
## Contents

5.3 Tensors ................................................. 46
5.4 Operations on Tensors ................................. 47
5.5 The Alternating Tensor ................................. 49
5.6 Tensor Fields ............................................ 49
5.7 Example and Exercises ................................. 50

6 Electrodynamics ........................................... 53
6.1 Tensor Formulation of Maxwell's Equations ........ 53
6.2 Gauge Transformations ................................ 55
6.3 The Dual Electromagnetic Field Tensor ............. 55
6.4 Advanced and Retarded Solutions .................... 56
6.5 The Field of a Uniformly Moving Charge ............ 60
6.6 The Equation of Motion of a Test Charge .......... 61
6.7 Plane Waves ........................................... 62
6.8 Geometric Optics .................................... 63
6.9 Examples and Exercises ............................... 65

7 Energy-Momentum Tensors ............................... 69
7.1 Dust ................................................... 69
7.2 The Ideal Gas ........................................ 70
7.3 Conservation of Energy ............................... 71
7.4 Another Form of Energy Conservation ................ 72
7.5 Fluids ................................................ 72
7.6 The Electromagnetic Energy-Momentum Tensor ..... 74

8 Symmetries and Conservation Laws ..................... 77
8.1 Isometries ............................................ 77
8.2 Killing Vectors ...................................... 77
8.3 Translations and Lorentz Rotations .................. 79
8.4 Active and Passive Transformations ................. 81
8.5 Conservation of Momentum and Angular Momentum 81
8.6 Exercises ............................................ 82

Index ...................................................... 85
1 Space, Time and Maxwell’s Equations

1.1 Introduction

The mathematical content of special relativity is not hard to understand. The basic principles can be written down simply and concisely, and, with a little practice and some not very involved calculations, it is easy to derive from them the main predictions of the theory. What is less easy is to appreciate that these predictions are correct, even though they imply that space and time have properties which are contrary to intuition and to the established framework of Newtonian mechanics: within the limits of modern experimental physics, the theory of relativity is an accurate model of the real world.

Before launching into the formal development, therefore, it is helpful to reconsider the problems that originally forced theoretical physicists to abandon the classical view of space and time. These arose from attempts to fit Newton’s mechanics and Maxwell’s electrodynamics into a single consistent physical theory. In particular, there was the problem of describing electromagnetic processes in uniformly moving frames of reference.

1.2 Maxwell’s Equations and Galilean Transformations

In empty space, Maxwell’s equations take the form

\[
\text{div} \mathbf{B} = 0, \quad \text{curl} \mathbf{B} = \frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t}, \quad \text{div} \mathbf{E} = 0, \quad \text{curl} \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t},
\]

from which it follows that the components of the electric and magnetic fields, \( \mathbf{E} \) and \( \mathbf{B} \), satisfy the wave equation

\[
\frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2} - \frac{\partial^2 \phi}{\partial x^2} - \frac{\partial^2 \phi}{\partial y^2} - \frac{\partial^2 \phi}{\partial z^2} = 0. \tag{1.1}
\]

This has particular solutions which represent plane waves propagating with velocity \( c \)—the velocity of light in vacuo.

How does light travel relative to an observer moving along the \( x \)-axis with velocity \( v \)? To find out, introduce new coordinates

\[
\tilde{x} = x - vt, \quad \tilde{y} = y, \quad \tilde{z} = z, \quad \tilde{t} = t, \tag{1.2}
\]

relative to which the observer is at rest (this is an example of a Galilean transformation; the general definition is given in Exercise 1.1). A beam of light which has velocity \((u_1, u_2, u_3)\) in the original coordinates, where \(u_1^2 + u_2^2 + u_3^2 = c^2\), should have velocity

\[
(\tilde{u}_1, \tilde{u}_2, \tilde{u}_3) = (u_1 - v, u_2, u_3) \tag{1.3}
\]

relative to the new coordinates. When \(\tilde{u}_2 = \tilde{u}_3 = 0\), we have \(\tilde{u}_1 = \pm c - v\), and when \(\tilde{u}_1 = \tilde{u}_3 = 0\), we have \(\tilde{u}_1 = v\) and \(\tilde{u}_2 = u_2 = \pm \sqrt{c^2 - v^2}\). Thus light moving parallel to
the \( \hat{x} \)-axis should appear to the observer to have velocity \( c - v \), and light moving parallel to the \( \hat{y} \)-axis should appear to have velocity \( \sqrt{c^2 - v^2} \).

This does not seem a surprising or controversial result. In Maxwell's original theory, the electromagnetic equations did not take the form (1.1) in all frames of reference: they described the propagation of light through a medium (the luminiferous ether) and modified equations were needed in a frame moving through the ether. Unfortunately, there are serious objections to this point of view.

First, if (1.3) gave the correct transformation of the velocity of light, then it should be possible to detect the effect of motion through the ether on electromagnetic processes. For example, the earth cannot be at rest relative to the ether throughout the year. Thus the velocity of light in a fixed direction (that is, fixed relative to the earth) should appear to change during the year. Or, more simply, when the earth is moving through the ether, the velocity of light in different directions should be different. This was shown to be false in 1887 by the Michelson-Morley experiment, since repeated many times with very much greater accuracy.\(^4\)

Simplifying a little, the apparatus in the experiment consisted of a source \( S \) of light of wavelength \( \lambda \), two plane mirrors \( A \) and \( B \), a half-silvered mirror \( C \), and a screen \( T \). The light from \( S \) was split into two beams at \( C \). One arrived at \( T \) after reflection at \( A \) and \( C \), the other after reflection at \( C \) and \( B \), so that interference fringes could be seen at \( T \).

The whole apparatus was attached to a stone disc which could be rotated. If the earth had been moving through the ether with velocity \( v \) and if originally \( AC \) had been in the direction of the earth's motion, then the interference pattern should have shifted through approximately \( n = v^2(AC + CB)/\lambda c^2 \) fringes when the stone was rotated through \( 90^\circ \), for small \( v^2/c^2 \) (Exercise 1.3). No such shift was observed.

Second, Newton's laws of dynamics are invariant under Galilean transformations. They hold in any inertial frame—that is, in any nonaccelerating frame of reference. Newton's laws describe the dynamics of rigid bodies, and we know that the forces which act in collisions between rigid bodies are electromagnetic in origin. It is unreasonable that Newton's laws should contrive to take the same form in all inertial frames while the basic equations of electromagnetism do not.

Third, many other electromagnetic phenomena show an inexplicable indifference to motion through the ether. For example, the current induced in a conductor by a magnet is the same whether the conductor is at rest and the magnet is moving, or the magnet is at rest and the conductor is moving. This example was cited by Einstein in his 1905 paper On the electrodynamics of moving bodies.\(^5\)

Before 1905, the main point of attack was Maxwell's theory, which received some elaborate modifications. H. A. Lorentz suggested that the Michelson-Morley experiment gave a null result because rigid bodies are contracted in the direction of their motion through the ether by a factor \( \sqrt{1 - v^2/c^2} \). This would cancel the effect of the earth's motion on the light rays in the experiment, at least to the second order in \( v/c \). Lorentz's idea

\(^4\) See A. Brilet and J. L. Hall Improved laser test for the isotropy of space. Phys. Rev. Lett. 42 549–52 (1979). They find that the observed fringe shift is less than \( 2.5 \times 10^{-7} \) times the value predicted by the ether theory.

was not obviously absurd: since the internal forces in a rigid body are electromagnetic, it would not be surprising if they were affected by motion through the ether. However, one might call it a 'conspiracy theory' since it implies that electromagnetic phenomena are affected by motion through the ether, but, as in the Michelson-Morley experiment, the effects conspire to cancel, so the motion is undetectable.

Lorentz later extended his idea to explain the failure of other ether-detection experiments, and Poincaré stated it as a general principle that no experiment can detect absolute motion. That is, motion relative to the ether.

1.3 The Newtonian View of Space and Time

In Newtonian physics, one has the notions of absolute time and absolute distance. In a uniformly moving frame of reference, each event (an event is a particular point at a particular time) is labelled by four coordinates: the time $t$ and the Cartesian coordinates $x$, $y$ and $z$ of the point. Given two events, $A$ and $A'$, with coordinates $(t, x, y, z)$ and $(t', x', y', z')$, the time separation $t - t'$ is an invariant. It depends only on $A$ and $A'$, and not on the particular frame the coordinates refer to. It is called the absolute time between $A$ and $A'$. Similarly, if $A$ and $A'$ are simultaneous ($t = t'$), then the invariant

$$s = \sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2}$$

is called the absolute distance between $A$ and $A'$. Note that absolute distance is not defined in a frame-independent way for events that are not simultaneous. For example, the distance from Oxford at 1 pm to Berlin at 2 pm is about 600 miles in a frame fixed
relative to the earth, but some 65,000 thousand miles in a frame fixed relative to the sun, since the earth orbits the sun at about 18 miles per second.

Lorentz asserted that a rod moving through the ether contracts. Put another way, a rod is at its longest when it is at rest relative to the ether. How can this be proved? Not by measuring with a ruler since the ruler will also contract; nor by interference methods, since the contraction is just that required to give a null result in the Michelson-Morley experiment. Thus Lorentz’s idea undermines the physical meaning of ‘absolute distance’. If the Lorentz contraction is accepted uncritically, then only length measurements made in the rest frame of the ether are ‘correct’. But if no measurement will detect the ether, then how are absolute distances to be measured?

Einstein was the first to understand this difficulty and to see clearly that the solution lay not in modifications to Maxwell’s theory, but in a careful re-examination of the physical meaning of distance measurements, and, with it, of the physical meaning of simultaneity. In special relativity, we work from the premise that all but the most basic concepts in a physical theory must be given operational definitions. This means, for example, that if you cannot describe an experiment that will detect the ether, then the ether has no place in the theory. The same goes for time and distance measurements: in particular, you cannot talk about ‘distance’ until you have decided what it means to say in operational terms that two events are simultaneous.

Exercise 1.1 Equation (1.2) is one example of a Galilean transformation. In general, a Galilean transformation is any linear transformation of the coordinates \(x, y, z\) and \(t\) which preserves absolute time, and absolute distance for simultaneous events. Show that the general form of such a transformation is

\[
\begin{pmatrix}
  t \\
  x \\
  y \\
  z
\end{pmatrix}
= \begin{pmatrix}
  1 & 0 & 0 & 0 \\
  v_1 & H & & \\
  v_2 & & & \\
  v_3 & & & 
\end{pmatrix}
\begin{pmatrix}
  i \\
  \hat{x} \\
  \hat{y} \\
  \hat{z}
\end{pmatrix}
+ \begin{pmatrix}
  T^0 \\
  T^1 \\
  T^2 \\
  T^3
\end{pmatrix},
\]

where the \(v_1\)s and \(T^q\)s are constants, and \(H\) is a \(3 \times 3\) orthogonal matrix. What is the physical interpretation of the \(v_1\)s? What difference does it make if the absolute time between two events is defined to be \(|t - t'|\) rather than \(t - t'|\)?

Exercise 1.2 Show that the plane wave \(\phi = A \cos(\omega(t - c^{-1}e \cdot r) + \epsilon)\) is a solution of the wave equation, where \(A, \omega, \epsilon\) and \(e\) are constant, with \(e \cdot e = 1\) (\(A\) is the amplitude, \(\omega\) is the angular frequency, and \(e\) is the direction of propagation). Sketch the surfaces of constant \(\phi\) at a fixed value of \(t\), and explain how they move as \(t\) increases.

Exercise 1.3 By considering the distances light rays travel in a frame of reference fixed relative to the ether, show that the fringe shift expected in the Michelson-Morley experiment according to the ether theory is \(v^2(CA + CB)/\lambda c^2\) (keep terms of order \(v^2/c^2\), but ignore higher powers of \(v/c\)).

Exercise 1.4 Show that under the Galilean transformation (1.2) the wave equation becomes

\[
\frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2} - \frac{2v}{c^2} \frac{\partial^2 \phi}{\partial \hat{x} \partial \hat{t}} - \frac{c^2 - v^2}{c^2} \frac{\partial^2 \phi}{\partial \hat{x}^2} - \frac{\partial^2 \phi}{\partial \hat{y}^2} - \frac{\partial^2 \phi}{\partial \hat{z}^2} = 0.
\]
2 Inertial Coordinates

2.1 Basic Principles

Both special and general relativity begin by introducing *space-time*. This is the four-dimensional manifold of all events. It is denoted by $M$. By saying that $M$ is a four-dimensional manifold, we mean that events can be labelled by four parameters $x^a$, $a = 0, 1, 2, 3$. These might be three position coordinates and one time coordinate or they might be some complicated combinations of position and time coordinates. We do not yet think of the $x^a$s as having dimensions of time or distance: for the moment they are simply labels assigned to events.

The basic principles of special relativity are formulated in terms of certain idealized physical objects in space-time: free particles, clocks and photons (particles of light). As in Newton's first law, a free particle is a 'small body far removed from the influence of other bodies'. In general relativity, which deals with gravitational fields, this definition must be examined more critically. For the present purpose, its meaning should be clear.

Each free particle is equipped with a clock of some standard construction which measures the time $\tau$ of events in the particle's history: $\tau$ is called the particle's *proper time*. From a physical point of view, this picture is not unreasonable: even a single atom can act as a clock. The history of a free particle is represented by a curve in space-time, called its *worldline*, which is given by four equations of the form

$$x^a = x^a(\tau); \quad a = 0, 1, 2, 3.$$  

To start with, we shall consider only observations made by *inertial observers*, which are observers in free fall. That is to say, they move in the same way as free particles. An inertial observer can measure proper time along his own worldline (that is at events at his own location) by using a standard clock, and he can send out and receive back light signals, which are transmitted by photons. The photons are localized in space and their histories are again represented by worldlines in space-time.

The properties of the basic physical objects are summarized as follows.

P1 Free particles and photons appear to inertial observers to travel in straight lines at constant speeds.

P2 Photons appear to all inertial observers to travel at the same constant speed.

P3 The standard clock of one inertial observer appears to any other inertial observer to run at a constant rate (but the clocks of different observers do not necessarily appear to run at the same rate).

P4 Free particles never travel faster than photons.

To these is added the *principle of relativity*.

R Only the relative motion of inertial observers is detectable.
The first is Newton's first law and the second is an extrapolation from the failure of the ether-detection experiments. We cannot yet give precise meanings to (P1)–(P4) since they involve velocities and time measurements. Our strategy will be to introduce operational definitions of time and distance at the same time as building up a model of space-time in such a way that (P1)–(P4) and (R) hold automatically.

The fourth property (P4) is the easiest to interpret. The path of a photon in space-time must be independent of the motion of its source, or (P2) could not be true. Therefore the photon worldlines that pass through an event $A$ form a three-dimensional surface in space-time, which we call the light-cone of $A$. Roughly one can think of each section of the future half of the light-cone as the wave-front of a flash of light emitted at $A$. As time progresses (upwards in the diagram, in which one spatial dimension has been suppressed), the wavefront spreads out. Property (P4) states that the worldlines of free particles that pass through $A$ lie inside the light-cone of $A$.

![Diagram](image)

**Fig. 2.1.** The worldlines of free particles lie inside the light-cone at $A$

Before interpreting (P1) and (P2), we must say how distances are to be measured. As in the classical theory, distance involves the idea of simultaneity. How is an inertial observer $O$ to decide if two events are simultaneous? The answer is, by the radar method: $O$ sends out a photon at time $t_1$ (measured on his clock). This is reflected at an event $A'$ and arrives back at his worldline at time $t_2$. On the assumption that the velocity of light is constant, $O$ reckons that $A'$ is simultaneous with the event $A$ at his own location which happens at time $\frac{1}{2}(t_1 + t_2)$, and he assigns this time to $A'$. In this way, $O$ labels each event by his own time, and he reckons that two events are simultaneous if they

---

6 For observational support, see K. Brecher *Is the speed of light independent of the velocity of the source?* Phys. Rev. Lett. 39 1051–4 (1977), in which it is deduced from observations of regularly pulsating X-ray sources in binary star systems that $|\delta c/v| < 2 \times 10^{-9}$, where $\delta c$ is the change in the velocity of light for a source with velocity $v$. 
happen at the same time. He also assigns to each event a distance from his own location, on the assumption that the photon takes the same time to cover the outward journey as the return one. On this basis, he reckons that the distance from $A$ to $A'$ is $\frac{1}{2}c(t_2 - t_1)$. Here $c$, the velocity of light, is an arbitrary constant. If $t$ is measured in years and $c$ is taken to be 1, then the unit of distance is called the light year. If $t$ is measured in seconds and $c$ is taken to be $3 \times 10^8$, then the unit of distance is called the metre, and so on.\(^7\) These definitions are illustrated in Fig. 2.2: again, this is a space-time diagram, with time running up the page.

Fig. 2.2. The operational definition of simultaneity and distance

Fig. 2.3. The observer $\hat{O}$ reckons that $A_1$ happens before $A_2$

An immediate consequence of these operational definitions is that simultaneity is relative: there is no absolute time. Consider, for example, two observers $O$ and $\hat{O}$ who are initially coincident at an event $A_0$. Each measures time on his own standard clock, taking the time at $A_0$ to be zero. Let $A_1$ and $A_2$ be two events which are simultaneous, both according to $O$'s reckoning. Two photons are sent out from $A_0$ towards $A_1$ and $A_2$. After reflection at $A_1$ and $A_2$, respectively, they arrive back at $O$'s worldline at the same event. If the second observer $\hat{O}$ is moving relative to $O$ in the direction from $A_0$ to $A_1$, then the reflected photon from $A_1$ will arrive at $\hat{O}$'s worldline before that from $A_2$. Therefore $\hat{O}$ will reckon that $A_1$ happened before $A_2$ (see the second space-time diagram, Fig. 2.3).

If an inertial observer $O$ picks two nearby free particles $P$ and $P'$ which are at rest relative to himself (i.e. to which the distances are constant), then he can also measure distance from these. Knowing the distances of events from three points, he can set up a Cartesian coordinate system. He can then label each event in $M$ by its time $t$ and its Cartesian coordinates $x$, $y$ and $z$. If he takes his own location as the origin of the spatial coordinates, then his worldline is given by

$$t = \tau, \quad x = 0, \quad y = 0, \quad z = 0,$$

where $\tau$ is his proper time.

\(^7\) The metre is in fact defined so that the exact value of $c$ is $299792458 \frac{m}{s}$.
It is conventional to put \( x^0 = ct, x^1 = x, x^2 = y, x^3 = z \). Then all four coordinates \( x^a (a = 0, 1, 2, 3) \) have dimensions of distance. The \( x^a \)s are called \textit{inertial coordinates}: 'inertial' because they are coordinates in which Newton's first law holds. We also use the term \textit{inertial frame} for the coordinate system set up by an inertial observer.

### 2.2 Lorentz Transformations

The properties (P1)–(P4) of free particles and photons can now be restated in a more precise form.

1. In an inertial coordinate system, the worldline of a free particle is given by equations of the form
   \[
   x^a = V^a \tau + T^a, \quad a = 0, 1, 2, 3,
   \]  
   where \( \tau \) is the proper time along the worldline of the particle, and the \( V^a \)s and \( T^a \)s are constants, with
   \[
   (V^1)^2 + (V^2)^2 + (V^3)^2 < (V^0)^2.
   \]  

2. Two events \( A \) and \( B \) lie on the worldline of a photon if and only if their coordinates \( x^a \) and \( y^a \) are related by
   \[
   (x^0 - y^0)^2 - (x^1 - y^1)^2 - (x^2 - y^2)^2 - (x^3 - y^3)^2 = 0.
   \]

Equation (2.1) contains (P1) and (P3). The inequality (2.2) is a restatement of (P4) and (2.3) is a restatement of (P2): it is the condition that the distance from \( A \) to \( B \) in the coordinate system should be \( c \) times the time interval between \( A \) and \( B \). For fixed \( x^a \), (2.3) defines the \textit{light-cone} of \( A \), which is made up of all photon worldlines through \( A \).

Our first task is to find the relationship between the inertial coordinate systems of different observers. To do this, we need some notation.

1. \textit{The metric coefficients}. The quantities \( g_{ab} \) and \( g^{ab} \) are defined by
   \[
   \begin{pmatrix}
   g_{00} & g_{01} & g_{02} & g_{03} \\
   g_{10} & g_{11} & g_{12} & g_{13} \\
   g_{20} & g_{21} & g_{22} & g_{23} \\
   g_{30} & g_{31} & g_{32} & g_{33}
   \end{pmatrix}
   =
   \begin{pmatrix}
   1 & 0 & 0 & 0 \\
   0 & -1 & 0 & 0 \\
   0 & 0 & -1 & 0 \\
   0 & 0 & 0 & -1
   \end{pmatrix}
   =
   \begin{pmatrix}
   g^{00} & g^{01} & g^{02} & g^{03} \\
   g^{10} & g^{11} & g^{12} & g^{13} \\
   g^{20} & g^{21} & g^{22} & g^{23} \\
   g^{30} & g^{31} & g^{32} & g^{33}
   \end{pmatrix}.
   \]

   Note that
   \[
   \sum_{b=0}^{3} g_{ab} g^{bc} = \delta^c_a,
   \]  
   where \( \delta^c_a \) is the Kronecker delta (equal to one if \( a = c \) and to zero otherwise).

2. \textit{The range convention}. The indices \( a, b, c, \ldots \) always run over the range \( 0, 1, 2, 3 \).

3. \textit{The summation convention}. If an index appears in an expression as an upper index and as a lower index, then a sum over the values \( 0, 1, 2, 3 \) is implied.

For example, with these conventions, (2.4) becomes \( g_{ab} g^{bc} = \delta^c_a \), and (2.3) becomes
\[
\sum_{b=0}^{3} (x^a - y^a)(x^b - y^b) = 0.
\]
Let $O$ and $\tilde{O}$ be two inertial observers who have set up inertial coordinate systems $x^a$ and $\tilde{x}^a$, as above, and suppose that each takes his own position as origin of his spatial coordinates. Along the worldline of any free particle $P$, we have, by (2.1),

$$\frac{d^2 x^a}{d\tau^2} = 0 = \frac{d^2 \tilde{x}^a}{d\tau^2},$$

where $\tau$ is the particle's proper time. But

$$\frac{d^2 x^a}{d\tau^2} = \frac{\partial^2 x^a}{\partial x^b \partial x^c} \frac{d\tilde{x}^b}{d\tau} \frac{d\tilde{x}^c}{d\tau} + \frac{\partial x^a}{\partial \tilde{x}^b} \frac{d^2 \tilde{x}^b}{d\tau^2}.$$ 

Hence

$$\frac{\partial^2 x^a}{\partial \tilde{x}^b \partial \tilde{x}^c} \frac{d\tilde{x}^b}{d\tau} \frac{d\tilde{x}^c}{d\tau} = 0.$$ 

Since this holds for any free particle,

$$\frac{\partial^2 x^a}{\partial \tilde{x}^b \partial \tilde{x}^c} = 0.$$ 

Therefore the dependence of the $x^a$s on the $\tilde{x}^a$s is linear and can be expressed in the form

$$x^a = K^a_b \tilde{x}^b + T^a, \quad (2.6)$$

where $K^a_b$ and $T^a$ are constant.

Suppose that $A$ and $B$ are events on the worldline of a photon. If the coordinates of $A$ and $B$ are $x^a$ and $y^a$ in $O$'s coordinate system and $\tilde{x}^a$ and $\tilde{y}^a$ in $\tilde{O}$'s, then, from (2.5),

$$g_{ab}(x^a - y^a)(x^b - y^b) = 0 = g_{ab}(\tilde{x}^a - \tilde{y}^a)(\tilde{x}^b - \tilde{y}^b). \quad (2.7)$$

By substituting from (2.6) into the first equation in (2.7),

$$g_{ab} K^a_d K^b_e (\tilde{x}^d - \tilde{y}^d)(\tilde{x}^e - \tilde{y}^e) = 0.$$ 

But if this holds whenever the second equation in (2.7) holds, then $g_{ab} K^a_d K^b_e$ must be proportional to $g_{de}$. Therefore we can write $K^a_b = k L^a_b$ where

$$g_{ab} L^a_d L^b_e = g_{de}, \quad (2.8)$$

and $k \in \mathbb{R}$. In matrix form, (2.8) is $L^t g L = g$, where $g = (g_{ab})$, $L = (L^a_b)$, and $L^t$ is the transpose of $L$ (we call $L$ the transformation matrix).

Let $\tau$ and $\tilde{\tau}$ denote the proper time parameters along the worldlines of $O$ and $\tilde{O}$. Then the two worldlines are given by

$$(O) \quad x^0 = c\tau, \quad x^1 = x^2 = x^3 = 0$$

$$(\tilde{O}) \quad \tilde{x}^0 = c\tilde{\tau}, \quad \tilde{x}^1 = \tilde{x}^2 = \tilde{x}^3 = 0.$$ 

In $O$'s coordinates, the second set of equations becomes

$$x^a = ck L^a_0 \tilde{\tau} + T^a. \quad (2.9)$$

Thus, to $O$, $\tilde{O}$ appears to be moving with velocity
Inertial Coordinates

\[ \mathbf{v} = \frac{d\mathbf{\tau}}{dt} \left( \frac{dx}{d\tau}, \frac{dy}{d\tau}, \frac{dz}{d\tau} \right) = \frac{c}{L_0} (L_0^1, L_0^2, L_0^3). \]

But, by taking \( d = e = 0 \) in (2.8),

\[ (L_0^0)^2 - (L_0^1)^2 - (L_0^2)^2 - (L_0^3)^2 = 1. \]

Hence \( O \)'s speed relative to \( O \) is

\[ v = \sqrt{\mathbf{v} \cdot \mathbf{v}} = c\sqrt{1 - (L_0^0)^2} \]

and \( L_0^0 \) is given in terms of \( v \) by \( L_0^0 = \gamma(v) \), where

\[ \gamma(v) = \frac{1}{\sqrt{1 - v^2/c^2}}. \]

Conversely, it follows from (2.6) that the \( \tilde{x}^a \)'s are given in terms of the \( x^a \)'s by

\[ \tilde{x}^a = k^{-1}M_a^b(x^b - T^b), \]

where \( M_a^b = g^{ae}g_{bd}L_e^d \) (the matrix \( (M_a^b) \) is the inverse of \( L = (L_a^b) \); see Exercise 2.1). In particular, \( M_0^0 = L_0^0 \). Hence \( v \) is also \( O \)'s speed relative to \( \hat{O} \).

From (2.9), the proper time \( \tilde{\tau} \) along \( \hat{O} \)'s worldline is related to \( O \)'s time coordinate \( x^0/c \) by

\[ x^0 = \frac{ck\tilde{\tau}}{\sqrt{1 - v^2/c^2}} + \text{constant}, \quad (2.10) \]

and, similarly, along \( O \)'s worldline, \( \tau \) is related to \( \tilde{x}^0 \) by

\[ \tilde{x}^0 = \frac{c\tau}{k\sqrt{1 - v^2/c^2}} + \text{constant}. \quad (2.11) \]

But the principle of relativity states that only the relative motion of \( O \) and \( \hat{O} \) is detectable. Thus nothing should alter when \( O \) and \( \hat{O} \) are interchanged. Hence we must have \( k = \pm 1 \). Finally the sign of \( k \) is fixed by requiring that all inertial observers should agree on the direction of time. If the positive square roots are taken in (2.10) and (2.11), then we have \( k = 1 \) and \( L_0^0 > 0 \).

To summarize, the two inertial coordinate systems are related by a transformation of the form

\[ x^a = L_a^b \tilde{x}^b + T^a, \]

where \( g_{ab}L_a^dL_b^e = g_{de} \) and \( L_0^0 > 0 \). When \( T^a = 0 \), this is a Lorentz transformation. Otherwise, it is a Poincaré or inhomogeneous Lorentz transformation. (Strictly, a Lorentz transformation does not have to satisfy \( L_0^0 > 0 \). When this does hold, the transformation is said to be orthochronous. If, in addition, \( \det L = 1 \), then the transformation is said to be proper and orthochronous. In this case, it also preserves the orientation of the spatial axes. Note that, in any case, \( |\det L| = 1 \) as a consequence of (2.8). In these notes, all Lorentz transformations will be proper and orthochronous.) Two special cases are

1. When \( L_a^b = \delta_a^b \), the transformation translates the origin of the spatial and temporal coordinates.
2. When $T^a = 0$, $L^a_0 = \delta^a_0$, and $L^0_a = \delta^0_a$, the transformation is a rotation of the spatial coordinate axes (Exercise 2.3).

Exercise 2.1 Show that if $g_{ab} L^a_d L^b_e = g_{de}$, then $g^{ae} g_{bd} L^d_e L^b_f = \delta^a_f$ and $g^{de} L^a_d L^b_e = g^{ab}$. (Hint: you are asked to show that $g L^t g L = 1$ and $L g L^t = g$.)

Exercise 2.2 In an inertial coordinate system, the worldline of a free particle is given by $x^a = V^a \tau + T^a$ where $\tau$ is the particle's proper time, and $V^a$ and $T^a$ are constants. Show that $g_{ab} V^a V^b = c^2$. (Hint: consider the equation of the worldline in the particle's rest frame.)

Exercise 2.3 Show that a transformation such that $T^a = 0$ and $L^a_0 = \delta^a_0$, $L^0_a = \delta^0_a$ is of the form

\[
\begin{pmatrix}
ct \\
x \\
y \\
z
\end{pmatrix} = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 0 & H & 0 \\
0 & 0 & 1 & 0
\end{pmatrix} \begin{pmatrix}
ct' \\
\tilde{x} \\
\tilde{y} \\
\tilde{z}
\end{pmatrix},
\]

where $H$ is a $3 \times 3$ orthogonal matrix.

Exercise 2.4 Which of the following are the matrices of proper orthochronous transformations?

\[
\begin{pmatrix}
\sqrt{2} & 1 & 0 & 0 \\
1 & \sqrt{2} & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}, \quad \begin{pmatrix}
2 & 0 & 1 & -1 \\
1 & 1 & 1 & -1 \\
-1 & 1 & -1 & 1 \\
0 & 0 & 1 & 1
\end{pmatrix}, \quad \begin{pmatrix}
\frac{-2}{\sqrt{2}} & 1 & 0 & -1 \\
-1 & 1 & 1 & -1 \\
1 & -1 & 1 & 1 \\
0 & 1 & 0 & 1
\end{pmatrix}
\]

(Hint: you must decide whether or not $L^t g L = g$, as well as looking at the conditions on $\det L$ and $L^0_0$.)

Exercise 2.5 Show that the statement including (2.3) implies that photons travel in straight lines. (This is not easy.)

2.3 Time Dilatation

One consequence of the Lorentz transformation law follows immediately from (2.10): the proper time $\tilde{\tau}$ of the observer $\tilde{O}$ is related to $O$’s time coordinate $t = \tilde{x}^0/c$ by

\[
t = \frac{\tilde{\tau}}{\sqrt{1 - v^2/c^2}}.
\]

Measured against the time coordinate set up by $O$, $\tilde{O}$’s clock runs slow by a factor $\sqrt{1 - v^2/c^2}$. Similarly, compared with $\tilde{O}$’s time coordinate $\tilde{t} = \tilde{\tilde{x}}^0/c$, $O$’s clock runs slow
by the same factor. This is the time dilatation effect. Note that it is symmetric between the two observers.

For example, the time between two events in a spacecraft travelling away from the earth is longer, by a factor $\gamma(v)$, when measured by an observer on earth by the radar method than when measured by an astronaut in the spacecraft by using a standard clock. This result is paradoxical only if one insists on talking about ‘time’ independently of the process of measurement. The apparent asymmetry between the astronaut and the observer on earth arises because both are measuring the time between events on the astronaut’s worldline.

Fig. 2.4. The time dilatation effect, with $\gamma(v) = 2$

\[ \tau = 2 \quad t = 3 \]
\[ \tau = 1 \quad t = 2 \]
\[ \tau = 0 \quad t = 1 \]

Fig. 2.5. The standard Lorentz transformation

2.4 The Standard Lorentz Transformation

To understand the meaning of the Lorentz transformation, consider the special case in which $O$ and $\tilde{O}$ are initially coincident at the event $A$, and suppose that $O$ and $\tilde{O}$ both take $A$ as the origin of their coordinates. Suppose further that they align the axes of their coordinates so that, relative to $O$, $\tilde{O}$ has velocity $(v, 0, 0)$, and, relative to $\tilde{O}$, $O$ has velocity $(-v, 0, 0)$. This fixes the directions of the $x^1$ and $\tilde{x}^1$ axes, but leaves the freedom to make rotations $x^2$, $x^3$ and $\tilde{x}^2$, $\tilde{x}^3$ planes.

The two coordinate systems are related by

\[
x^a = L^a_b \tilde{x}^b \quad \text{and} \quad \tilde{x}^a = M^a_b x^b,
\]

where $g_{ab} L^a_d L^b_e = g_{de}$, $M^a_b = g^{ae} g_{bd} L^d_e$, $L^0_0 = \gamma(v)$, and

\[
(L^1_0, L^2_0, L^3_0) = \gamma(v)(v/c, 0, 0)
\]
\[
(M^1_0, M^2_0, M^3_0) = (-L^0_1, -L^0_2, -L^0_3) = \gamma(v)(-v/c, 0, 0).
\]
By taking \( d = 0, e = 1 \) in \( g_{ab}L^a_dL^b_e = g_{de} \), one obtains \( L^1_1 = \gamma(v) \). Then by taking \( d = e = 1 \), one obtains \( L^2_1 = L^3_1 = 0 \). Similarly, by Exercise 2.1, \( L^1_2 = L^3_3 = 0 \). We also have, \((L^2_2)^2 + (L^3_2)^2 = 1, (L^2_3)^2 + (L^3_3)^2 = 1, \) and \( L^2_2L^2_3 + L^3_2L^3_3 = 0 \). Therefore,

\[
L = \begin{pmatrix}
L^0_0 & L^0_1 & L^0_2 & L^0_3 \\
L^1_0 & L^1_1 & L^1_2 & L^1_3 \\
L^2_0 & L^2_1 & L^2_2 & L^2_3 \\
L^3_0 & L^3_1 & L^3_2 & L^3_3
\end{pmatrix}
= \begin{pmatrix}
\gamma(v) & \gamma(v)v/c & 0 & 0 \\
\gamma(v)v/c & \gamma(v) & 0 & 0 \\
0 & 0 & \cos \theta & \sin \theta \\
0 & 0 & -\sin \theta & \cos \theta
\end{pmatrix},
\]

where \( \theta \) is undetermined. By making a rotation in the \( x^2, x^3 \) plane, we can set \( \theta = 0 \). Then, with \((x^0, x^1, x^2, x^3) = (ct, x, y, z) \) and \((\bar{x}^0, \bar{x}^1, \bar{x}^2, \bar{x}^3) = (\bar{c}t, \bar{x}, \bar{y}, \bar{z}) \),

\[
t = \gamma(v)(\bar{t} + v\bar{x}/c^2), \quad x = \gamma(v)(\bar{x} + v\bar{t}), \quad y = \bar{y}, \quad z = \bar{z}.
\]

This is the standard Lorentz transformation.

Exercise 2.6 Show that if \( L \) is the \( 4 \times 4 \) matrix of a proper orthochronous Lorentz transformation (i.e. \( L^t gL = g, L^0_0 > 0 \), and \( \det L = 1 \)), then there exist \( 3 \times 3 \) proper orthogonal matrices \( H \) and \( \bar{H} \) such that

\[
\begin{pmatrix}
1 & 0 \\
0 & H
\end{pmatrix}
L
\begin{pmatrix}
1 & 0 \\
0 & \bar{H}
\end{pmatrix},
\]

is the matrix of a standard Lorentz transformation. Deduce that a general Lorentz transformation can be represented as a rotation, followed by a standard Lorentz transformation, followed by a rotation. (Hint: by making rotations of the \( x, y, z \) and \( \bar{x}, \bar{y}, \bar{z} \) axes, it is sufficient to consider the case \( L^2_0 = L^3_0 = L^0_2 = L^0_3 = 0 \). Show that in this case, \( L^2_1 = L^3_1 = L^1_2 = L^1_3 = 0 \) as well.)

Exercise 2.7 Show that for large \( c \), the standard Lorentz transformation reduces to a Galilean transformation. (Remember that \( x^0 \) is \( ct \): you must express \( t, x, y, z \) in terms of \( \bar{t}, \bar{x}, \bar{y}, \bar{z} \) before taking the limit \( c \to \infty \).)

Exercise 2.8 A spy passes his spymaster in the street at a relative speed of about 5 mph. The spymaster is walking directly towards the Andromeda galaxy and the spy is walking in the opposite direction. As they pass, the spy whispers 'at this very moment, the battle fleet from Andromeda is setting out to wipe out our galaxy'. How long ago did the fleet set out in the spymaster's frame? (Andromeda is about 2,200,000 light years away.)

Exercise 2.9 Show that the wave operator

\[
\frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial z^2}
\]

is invariant under the standard Lorentz transformation.
2.5 The Lorentz Contraction

Consider a rigid rod at rest relative to $\check{O}$ and lying along the $\check{x}^1$-axis. Suppose, for simplicity, that one end of the rod is at the origin of $\check{O}$'s spatial coordinate. Then the worldlines $W_1$ and $W_2$ of the two ends of the rod are given in $\check{O}$'s coordinates by

\[
(W_1) \quad \check{x}^0 = c\check{\tau}, \quad \check{x}^1 = \check{x}^2 = \check{x}^3 = 0
\]
\[
(W_2) \quad \check{x}^0 = c\check{\tau}, \quad \check{x}^1 = D, \quad \check{x}^2 = \check{x}^3 = 0,
\]

where $\check{\tau}$ is the proper time parameter along $W_1$ and $W_2$, and $D$ is the length of the rod, as measured by $\check{O}$. If we assume that $\check{O}$'s frame is related to $\check{O}$'s by a standard Lorentz transformation, then in $O$'s coordinates, the two worldlines are given by

\[
(W_1) \quad x^0 = c\gamma(v)\check{\tau}, \quad x^1 = \gamma(v)\check{v}\check{\tau}, \quad x^2 = x^3 = 0,
\]
\[
(W_2) \quad x^0 = c\gamma(v)(\check{\tau} + \check{v}D/c^2), \quad x^1 = \gamma(v)(D + \check{v}\check{\tau}), \quad x^2 = x^3 = 0.
\]

What is the length of the rod as measured by $O$? That is, what is the distance according to $O$ between two simultaneous events, one on $W_1$ and the other on $W_2$? When $x^0 = ct$, one end of the rod is at the event $A$ with coordinates $(x^0, x^1, x^2, x^3) = (ct, vt, 0, 0)$ and the other is at the event $B$ with coordinates

\[
(x^0, x^1, x^2, x^3) = (ct, vt + D\gamma(v)(1 - v^2/c^2), 0, 0)
\]
\[
= (ct, vt + D\sqrt{1 - v^2/c^2}, 0, 0)
\]

(by solving $ct = c\gamma(v)(\check{\tau} + \check{v}D/c^2)$ for $\check{\tau}$, and substituting into the equations for $W_2$ in $O$'s coordinate system). Therefore $O$ reckons that the rod is of length $D\sqrt{1 - v^2/c^2}$. It appears to be shortened by a factor $\gamma(v)^{-1}$. This is the Lorentz contraction.

Although it is given by the same formula as the contraction suggested by Lorentz, the interpretation is rather different. The relativistic contraction is a symmetric effect between any pair of observers (a rod at rest relative to $O$ similarly appears to $\check{O}$ to be contracted). The effect is not due to motion relative to the 'ether'.

Exercise 2.10 Show that there is no contraction when the rod is at right angles to the $\check{x}$ axis.

Exercise 2.11† An athlete carrying a 20 foot pole $PQ$ runs with speed $\sqrt{3} \cdot c/2$ into a 10 foot room. As the end $Q$ reaches the far wall, the end $P$ reaches the door, which is immediately closed. Explain with the aid of a space-time diagram, which should show the worldlines of the ends of the rod as well as the histories of the door and the far wall, how this is possible when in the athlete’s frame, the room is only 5 feet long. (Hint: you must consider which events are simultaneous with $Q$ striking the wall in the athlete's frame and in the frame of the room. If you think you have this sorted out, consider the following supplementary questions. If the athlete holds the end $P$ against his chest, he will be impaled on the pole. Is he inside or outside the room when this happens, according to his own definition of simultaneity? Is he still alive when he reaches the door?)

Exercise 2.12† The inertial coordinate systems $ct, x, y, z$ and $ct, \check{x}, \check{y}, \check{z}$ are related by a standard Lorentz transformation. In the second system, two photons travel along the $\check{x}$
Fig. 2.6. The Lorentz contraction: $A$ and $B$ are simultaneous in relative to $O$, $A$ and $\tilde{B}$ are simultaneous relative to $\tilde{O}$.

axis separated by a distance $D$. Show that the distance between them measured in the first system is $D\sqrt{(1 - v/c)/(1 + v/c)}$.

2.6 Transformation of Volumes

Consider a solid cube which is at rest relative to the coordinates $ct, \tilde{z}, \tilde{y}, \tilde{z}$. Suppose that its edges are of length $D$ and that four of its vertices, $V_1, V_2, V_3, V_4$, are at the points $(0, 0, 0), (D, 0, 0), (0, D, 0)$, and $(0, 0, D)$ in the $\tilde{x}, \tilde{y}, \tilde{z}$ coordinates. The worldlines $W_1$ and $W_2$ of $V_1$ and $V_2$ are as in §2.5. The worldlines $W_3$ and $W_4$ of $V_3$ and $V_4$ are given in the coordinates $ct, x, y, z$ by

\[
\begin{align*}
(W_3) \quad x^0 &= c\gamma(v)\tau, \quad x^0 = 0, \quad x^2 = D, \quad x^3 = 0 \\
(W_4) \quad x^0 &= c\gamma(v)\tau, \quad x^1 = 0, \quad x^2 = 0, \quad x^3 = D.
\end{align*}
\]

As measured by $O$, the edges parallel to the $x$-axis are contracted by a factor $\gamma(v)^{-1}$, while the edges parallel to the $y$ and $z$-axes are the same length $D$ as in the rest frame. Therefore $O$ measures the volume to be $\gamma(v)^{-1}D^3$ at any instant. Since any solid can be approximated by a number of small cubes, the volume of an arbitrary body in uniform motion is contracted by the same factor.

Exercise 2.13 Let $ct, x, y, z$ and $ct, \tilde{x}, \tilde{y}, \tilde{z}$ be the respective inertial coordinate systems of two inertial observers $O$ and $\tilde{O}$. Suppose that they are related by a general Lorentz transformation with $\gamma(v) = L^0_0 = v$. Let $A$ be the event with coordinates $(0, \tilde{x}, \tilde{y}, \tilde{z})$ in $\tilde{O}$'s system. Define a linear mapping $\mathbb{R}^3 \to \mathbb{R}^3$ by sending $(\tilde{x}, \tilde{y}, \tilde{z})$ to $(x, y, z)$, where $0, x, y, z$
are the coordinates in $O$'s system of the event $A$ which is (i) simultaneous in $O$'s system with the event $(0,0,0,0)$, and (ii) happens in the same place as $\hat{A}$ in $\hat{O}$'s system. Show that this linear mapping has determinant $\gamma(v)^{-1}$, and hence deduce that a body which is at rest relative to $\hat{O}$ has its volume reduced by a factor $\gamma(v)^{-1}$ when measured by $O$. (Hint: $\det L = 1$. A determinant is unchanged when a multiple of one row is added to another.)

### 2.7 Addition of Velocities

Let $O$, $\hat{O}$, and $\overline{O}$ be three inertial observers. Suppose that they set up inertial coordinate systems $x^a$, $\hat{x}^a$ and $\overline{x}^a$ which are related by

$$x^a = L_{\hat{b}}^a \hat{x}^\hat{b} \quad \text{and} \quad \overline{x}^a = L_d^{\prime a} \overline{x}^d,$$

with $L$ and $L'$ the standard Lorentz transformations

$$L = \begin{pmatrix} \gamma(v) & v\gamma(v)/c & 0 & 0 \\ v\gamma(v)/c & \gamma(v) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad L' = \begin{pmatrix} \gamma(v') & v'\gamma(v')/c & 0 & 0 \\ v'\gamma(v')/c & \gamma(v') & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

where $v$ is the speed of $\hat{O}$ relative to $O$ and $v'$ is the speed of $\overline{O}$ relative to $\hat{O}$. What is the speed of $\overline{O}$ relative to $O$? To find out, we combine the two transformations to get $x^a = L_{\hat{b}}^a L_d^{\prime b} \overline{x}^d$. But the matrix $(L_{\hat{b}}^a L_d^{\prime b})$ is

$$LL' = \begin{pmatrix} \gamma(v)\gamma(v')(1 + vv'/c^2) & \gamma(v')\gamma(v')(v + v')/c & 0 & 0 \\ \gamma(v)\gamma(v')(v + v')/c & \gamma(v')\gamma(v')(1 + vv'/c^2) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} \gamma(v'') & v''\gamma(v'')/c & 0 & 0 \\ v''\gamma(v'')/c & \gamma(v'') & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

where

$$v'' = \frac{c^2(v + v')}{(c^2 + vv')}.$$  (2.12)

This is again a standard Lorentz transformation. Hence $\overline{O}$'s speed relative to $O$ is $v'' = (v + v')/(1 + vv'/c^2)$. If $v$ and $v'$ are small compared with $c$, then $v'' = v + v'$ and the velocities add in the same way as in the classical picture. But when $v$ and $v'$ are large, this is no longer true. As $v, v' \to c$, also $v'' \to c$.

**Exercise 2.14** Show that a standard Lorentz transformation matrix can be written in the form

$$L = \begin{pmatrix} \cosh \theta & \sinh \theta & 0 & 0 \\ \sinh \theta & \cosh \theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$
where \( \theta = \tanh^{-1}(v/c) \) (\( \theta \) is called the \textit{rapidity} or \textit{pseudo-velocity} of the transformation). Show that the relativistic addition law for velocities is equivalent to addition of rapidity. Sketch a diagram in the \( t, x \) plane showing how the the \( t, \tilde{x} \) axes move as \( \theta \) varies. Compare this with the behaviour of the \( x, y \) axes in the plane under rotations.

2.8 The Michelson-Morley Experiment

How should we now understand the null result of the Michelson-Morley experiment? The key issue is: what does it mean to say that the lengths \( CA \) and \( CB \) are unchanged as the apparatus is rotated? According to our operational definition of distance, that \( CA \) is unchanged means that light takes the same time to travel from \( C \) to \( A \) and back before as it does after the rotation. Thus one would conclude that the distances \( CA \) and \( CB \) had not been changed by the rotation if the interference pattern did not shift. From a relativistic perspective, therefore, if the pattern did shift, the conclusion would not be that the velocity of light relative to the apparatus was different in different directions, but that the stone was not rigid. The null result indicates that under reasonable conditions, 'rigid' rulers give the same measure of distance as the radar method.
3 Vectors in Space-Time

3.1 4-Vectors

In Euclidean space, vectors can be defined by their lengths and directions. We do not yet have analogous geometric concepts in space-time, so we shall fall back on a less geometric definition of space-time vectors in terms of the transformation properties of their components. The analogous approach in Euclidean space would be to define a vector as an array of three numbers (the components in an orthonormal basis), together with a transformation rule under rotations of the axes.

The prototype of a vector in space-time is the displacement between two events. If $A$ and $B$ have coordinates $x^a$ and $y^a$ in some inertial coordinate system, then the quantities $X^a = y^a - x^a$, which measure the displacement in time and space of $B$ from $A$, transform under the inhomogeneous Lorentz transformation $x^a \rightarrow \tilde{x}^a$ by $X^a \rightarrow \tilde{X}^a$, where

$$ \text{if } x^a = L^a_b \tilde{x}^b + T^a, \text{ then } X^a = L^a_b \tilde{X}^b. \quad (3.1) $$

As a generalization, we define a 4-vector $X$ to be a map that assigns an ordered set of four numbers $(X^a) = (X^0, X^1, X^2, X^3)$ to each inertial coordinate system. The $X^a$s are called the components of $X$. The components in different coordinate systems must be related by the transformation rule (3.1).

The set of 4-vectors is a four-dimensional vector space under the obvious component-by-component definitions of addition and scalar multiplication. We shall often drop the prefix '4' when there is no danger of confusion. Sometimes we shall describe vectors in three-dimensional Euclidean space as '3-vectors' to distinguish them from space-time 4-vectors. We set 3-vectors in bold type.

Example 3.1 If $X^a = y^a - x^a$, where $x^a$ and $y^a$ are the coordinates of $A$ and $B$, then $X$ is called the displacement 4-vector from $A$ to $B$.

Example 3.2 Consider a curve in $M$ with parametric representation $x^a = x^a(s)$. Suppose that all observers use the same parameter $s$ to label the events on the curve. The tangent 4-vector to the curve at an event is the 4-vector $V$ with components $V^a = dx^a/ds$. These transform in the right way under change of inertial coordinates because if $x^a = L^a_b \tilde{x}^b + T^a$, with $L^a_b$ and $T^a$ constant, then

$$ \frac{dx^a}{ds} = L^a_b \frac{d\tilde{x}^b}{ds}. $$

Note that $V$ depends on the choice of parametrization. If the parameter is changed, then the new tangent vector is proportional to the old one.
3.2 The Metric

Let $X$ and $Y$ be vectors with components $X^a$ and $Y^a$. The scalar $g(X, Y)$ is defined by

$$g(X, Y) = g_{ab}X^aY^b = X^0Y^0 - X^1Y^1 - X^2Y^2 - X^3Y^3.$$ 

Under an inhomogeneous Lorentz transformation $x^a = L^a_b\tilde{x}^b + T^a$,

$$g_{ab}X^aY^b = g_{ab}L^a_cL^b_d\tilde{X}^c\tilde{Y}^d = g_{ab}\tilde{X}^a\tilde{Y}^b.$$ 

Therefore $g(X, Y)$ is an invariant: it is the same in every inertial coordinate system. We call $g$ the metric and we call $g(X, Y)$ the inner or scalar product of $X$ and $Y$. The scalar product is a symmetric bilinear form on the space of 4-vectors, but, unlike the Euclidean scalar product (the dot product), it is not positive definite.

Space-time, with the structure given to it by the metric $g$, is called Minkowski space. The vectors $X$ and $Y$ are said to be orthogonal if $g(X, Y) = 0$. If $g(X, X) > 0$, then $X$ is said to be timelike. If $g(X, X) = 0$, then $X$ is said to be null, and if $g(X, X) < 0$, then $X$ is said to be spacelike. Note that null vectors are orthogonal to themselves—a point that often causes confusion. If $X$ is timelike or null and $X^0 > 0$ ($< 0$), then $X$ is said to be future-pointing (past-pointing). Note that this last definition does not depend on the choice of coordinates since we allow only orthochronous transformations (Exercise 3.2).

![Fig. 3.1. The classification of 4-vectors; $X$ and $Y$ are orthogonal](image)

**Exercise 3.1** Show that if $X$ and $Y$ are timelike or null, and $X^0 > 0$, $Y^0 > 0$, then $g(X, Y) > 0$. (Hint: use the Cauchy-Schwarz inequality.)

**Exercise 3.2** Show that if $X$ is a timelike or null 4-vector and if $X^0 > 0$ in one inertial frame, then $X^0 > 0$ in every inertial frame. (Hint: let $Y$ be the 4-vector with components
(L^{0}_0, -L^{0}_1, -L^{0}_2, -L^{0}_3); apply the result of previous exercise.) Show by counterexample that this is not true for a spacelike vector.

**Exercise 3.3** Show that

(i) the sum of two future-pointing timelike vectors is future-pointing timelike;
(ii) the sum of two future-pointing null vectors is future-pointing and either timelike or null;
(iii) every nonzero vector orthogonal to a timelike vector is spacelike;
(iv) every nonzero vector orthogonal to a null vector X is either spacelike or else a multiple of X.

Under what conditions is the sum in (ii) null?

**Exercise 3.4** Let X and Y be future-pointing timelike vectors and let \( Z = X + Y \). Show that

\[
\sqrt{g(Z, Z)} \geq \sqrt{g(X, X)} + \sqrt{g(Y, Y)}.
\]

What is the analogous inequality for 3-vectors?

### 3.3 The Causal Structure of Minkowski Space

Let A be an event. In the classical picture, the events that are simultaneous with A lie on the 3-dimensional ‘surface’ of constant time through A. This 3-surface partitions space-time into three regions: (i) the future (the events above the surface in Fig. 3.2), (ii) the past (the events below the 3-surface), and (iii) the present (the events that lie on the surface). If B lies in the first region, then what happens at A can influence what happens at B, and if B lies in the second region, then what happens at B can influence what happens at A.

The relativistic picture is less simple. The 3-plane of constant time through A is different in different inertial coordinate systems, and so has no universal significance. There is, however, an analogous invariant partitioning of space-time by the light-cone of A. Let B an event distinct from A and let X be the displacement vector from A to B. There are five possibilities (Fig. 3.3).

1. X is future-pointing and timelike. Then A happens before B, and it is possible to travel from A to B at less than the velocity of light (both statements are true in any inertial coordinate system); that is, B lies inside the future light-cone of A.
2. X is future-pointing and null. Then B lies on the future light-cone of A and there is a photon worldline from A to B.
3. X is spacelike. No signal can propagate between A and B at or below the speed of light. The events A and B are simultaneous, or A happens before B, or A happens after B, according to the choice of inertial coordinate system.
4. X is past-pointing null.
5. X is past-pointing timelike.

Cases (4) and (5) are the same as (2) and (1), except that A and B are interchanged.
Fig. 3.2. The 3-surface of constant time through A in classical space-time

Fig. 3.3. In special relativity, there are five possibilities for the displacement vector from A to B

In cases (1) and (2), B is inside or on the future light-cone of A and all observers agree that A happens before B. Since it is possible to travel from A to B at or below the speed of light, what happens at A can influence what happens at B. Similarly, in cases (4) and (5), B is unambiguously to the past of A. In case (3), on the other hand, the issue is undecided, and observers disagree on the order of A and B. This is less paradoxical than it may seem: when X is spacelike, it is necessary to travel faster than light to get from A and B. Thus no physical influence can propagate between the two events.

3.4 Orthonormal Frames

Let $T^1$ denote the space of 4-vectors (the reason for this notation will become clear in §5.1). An orthonormal frame is a set of four 4-vectors $X_0$, $X_1$, $X_2$, $X_3$ such that

$$g(X_a, X_b) = g_{ab}.$$  \hspace{1cm} (3.2)

The brackets around the indices are to indicate that the $X_a$s are four different vectors, and not the four components of a single vector. The components of the vector $X_a$ are denoted by $X_a^b$.

**Example 3.3** The vectors with components
\[ X^b_{(a)} = \delta^b_a \]
in an inertial coordinate system \( x^a \) are orthonormal: they make up the \textit{coordinate basis} of \( T^1 \). In the \( x^a \) coordinates, \( X(0) \) has components \((1,0,0,0)\), \( X(1) \) has components \((0,1,0,0)\), and so on.

If \( \tilde{X}_{(a)} \) is a second orthonormal frame, then \( X_{(b)} = L^a_b \tilde{X}_{(a)} \) for some matrix \((L^a_b)\) such that

\[ g_{ab} L^a_c L^b_d = g_{cd} \]

(from 3.2). That is, \( L = (L^a_b) \) is a Lorentz transformation matrix: the Lorentz transformations bear the same relation to the metric \( g \) as the spatial rotations do to the dot product in Euclidean space.

If \( X_{(a)} \) is the coordinate basis of some inertial coordinate system, and if \( L \) is proper and orthochronous, then we say that \( \tilde{X}_{(a)} \) is an \textit{inertial frame}. Clearly any inertial frame is the coordinate basis of some inertial coordinate system, and conversely. Note that in any inertial frame, \( X(0) \) is future-pointing and timelike. The components of a vector in an inertial frame are the same as its components in the corresponding coordinate system. Inertial coordinate systems with the same coordinate basis differ only in the location of their origins. Usually we do not distinguish between the terms ‘inertial coordinate system’ and ‘inertial frame’.

Let \( V \) be a four-vector. It is easy to prove the following.

1. If \( V \) is future-pointing and timelike, then there is an inertial frame in which \( V \) has components

\[ V^0 = \sqrt{g(V,V)}, \quad V^1 = V^2 = V^3 = 0. \]

2. If \( V \) is spacelike, then there is an inertial frame in which \( V \) has components

\[ V^0 = 0, \quad V^1 = \sqrt{-g(V,V)}, \quad V^2 = V^3 = 0. \]

3. If \( V \) is null and future-pointing, then there is an inertial frame in which \( V \) has components

\[ V^0 = V^1 = 1, \quad V^2 = V^3 = 0. \]

\textbf{Example 3.4} In the third case, one first rotates the spatial axes to make \( V^1 > 0, V^2 = V^3 = 0 \). Then \( V^0 = V^1 \) since \( V \) is null and future-pointing. If we put

\[ \frac{v}{c} = \frac{1 - (V^0)^2}{1 + (V^0)^2}, \]

then

\[ \gamma(v) \left( \begin{array}{cc} 1 & v/c \\ v/c & 1 \end{array} \right) \left( \begin{array}{c} V^0 \\ V^1 \end{array} \right) = \left( \begin{array}{c} 1 \\ 1 \end{array} \right). \]

Hence, by a standard Lorentz transformation, the components of \( V \) can be transformed to \((1,1,0,0)\).

Let \( V \) be the displacement vector of \( B \) relative \( A \). Then, in the first case, \( A \) and \( B \) happen at the same point and are separated by a time interval \( \sqrt{g(V,V)}/c \), measured in
the inertial frame. In the second case, $A$ and $B$ happen at the same place in the inertial frame, and are separated by a distance $\sqrt{-g(V,V)}$. In the final case, $A$ and $B$ lie on the worldline of a photon travelling in the direction of the $x^1$-axis of the frame.

Exercise 3.5 Show that if the two inertial coordinate systems $x^a$ and $\tilde{x}^a$ are related by the inhomogeneous Lorentz transformation $x^a = L^a_b \tilde{x}^b + T^a$, then the corresponding coordinate bases $X_{(a)}$ and $\tilde{X}_{(a)}$ are related by $\tilde{X}_{(b)} = L^b_a X_{(a)}$.

### 3.5 Decomposition of 4-Vectors

Let $V$ be a future-pointing timelike vector and let $x^a$ be an inertial coordinate system in which $(V^a) = (k, 0, 0, 0)$, where $k = \sqrt{g(V,V)}$. Then the coordinates are determined by $V$ up to an inhomogeneous Lorentz transformation

$$ x^a = L^a_b \tilde{x}^b + T^a, \quad (3.3) $$

where the $T^a$s are arbitrary and the matrix $L = (L^a_b)$ is of the form

$$ L = \begin{pmatrix} 1 & 0 \\ 0 & H \end{pmatrix} $$

for some $3 \times 3$ orthogonal matrix $H$. Equation (3.3) represents a rotation of the spatial coordinate axes, followed by a translation of the space-time origin.

We can think of $x^1$, $x^2$, and $x^3$ as Cartesian coordinates in the Euclidean space $E_V$ of an inertial observer at rest relative to the coordinate system $x^a$. Equation (3.3) is a transformation of the Cartesian coordinates in $E_V$.

Now let $U$ be any 4-vector with components $U^a$ in the coordinate system $x^a$. We can write $U = (U^0, u)$, where $u$ is the 3-vector in $E_V$ with components $(U^1, U^2, U^3)$. Under (3.3), $U^0$ is unchanged, while

$$ U^i = \sum_{j=1}^{3} H_{ij} u^j $$

$(i = 1, 2, 3)$. Thus the $U^i$s transform as the components of a 3-vector in Euclidean space. The decomposition of $U$ into its temporal part $U^0$ (a scalar) and its spatial part $u$ (a 3-vector) depends only on $V$ and not on the particular choice of the $x^a$s.

Note that

$$ U^0 = \frac{g(V,U)}{\sqrt{g(V,V)}} \quad \text{and} \quad g(U,U) = (U^0)^2 - u \cdot u. \quad (3.4) $$

It is important to understand that the decomposition relative to another timelike vector $\tilde{V}$ is different, unless $\tilde{V}$ happens to be proportional to $V$. In general, there is no natural way to identify the Euclidean spaces $E_V$ and $E_{\tilde{V}}$. 
3.6 Vector Fields

A *vector field* is a map \( V \) that assigns a 4-vector to each event \( A \). We usually assume that \( V \) is smooth in the sense that its components \( V^a \) are infinitely differentiable functions of the space-time coordinates.

If \( U \) and \( V \) are vector fields, then the *derivative* of \( U \) along \( V \) is the vector field \( \nabla_V U \) with components

\[
\nabla_V U^a = V^b \frac{\partial U^a}{\partial x^b}
\]

in any inertial coordinate system. By the chain rule, these transform as the components of a 4-vector field, but only because the coefficients \( L^a_b \) and \( T^a \) in an inhomogeneous Lorentz transformation are constants: it is not so easy to write the components of \( \nabla_V U \) in a general curvilinear coordinate system.

If \( U \) is defined only at events on some worldline, then \( \nabla_V U \) is still well defined, provided that \( V \) is a tangent to the worldline: if the worldline is given parametrically by \( x^a = x^a(s) \) and if \( V^a = dx^a/ds \), then \( \nabla_V U \) has components \( dU^a/ds \).

3.7 4-Velocity and Proper Time

Let \( W \) be a curve in space-time, with the parametric representation \( x^a = x^a(s) \) in some inertial coordinate system \( x^a \), and let \( U \) be the tangent 4-vector. If \( U \) is everywhere timelike, then the curve is said to be timelike. A timelike curve represents the worldline of an accelerating particle moving at less than the speed of light.

The *proper time* \( \tau \) along a timelike curve \( W \) is the parameter defined by

\[
ct = \int \sqrt{g_{ab} \frac{dx^a}{ds} \frac{dx^b}{ds}} ds = \int \sqrt{g(U,U)} ds.
\]

The integral is taken from any fixed event on \( W \) and the sign of the square root is chosen to make \( \tau \) increasing into the future. Proper time is an invariant: up to an additive constant, which depends on the choice of the event \( \tau = 0 \), it is independent of the inertial coordinates. It is analogous to arc length along a curve in Euclidean space and it is interpreted as the time measured by a standard clock moving with the particle. If the particle is free, then \( \tau \) is the same as the proper time that we have met already.

If \( s \) is replaced by \( \tau \) as the parameter along \( W \), then the new tangent vector, which has components \( V^a = dx^a/d\tau \), satisfies

\[
g_{ab}V^aV^b = g(V,V) = c^2.
\]

Since \( \tau \) is increasing into the future, \( V \) is a future-pointing timelike vector. It is called the *4-velocity*.

Let \( X \) be the displacement vector of an event on \( W \) from the origin of the coordinates. Then \( X = (ct, \mathbf{r}) \), where \( \mathbf{r} \) is the 3-vector with components \( (x^1, x^2, x^3) \) and \( ct = x^0 \). Since \( dx^a/dt \) is tangent to \( W \), we have \( V = \lambda (\mathbf{e}, \mathbf{v}) \) for some \( \lambda \), where \( \mathbf{v} = d\mathbf{r}/dt \). But

\[
c^2 = g(V,V) = \lambda^2 (c^2 - \mathbf{v} \cdot \mathbf{v}).
\]
Therefore $V = \gamma(v)(c, v)$, where $\gamma(v) = 1/\sqrt{1 - v^2/c^2}$, with $v = |v|$. The spatial part of the 4-velocity is the velocity $v$ relative to the coordinate system, multiplied by $\gamma(v)$.

We deduce from this that the proper time $\tau$ along $W$ is related to the coordinate time $t$ by

$$\frac{dt}{d\tau} = c^{-1} \frac{dx^0}{d\tau} = c^{-1} V^0 = \gamma(v).$$

(3.5)

Note that $c\gamma(v) = g(V, X(0))$, where $X(0)$ is the 4-vector with components $(1, 0, 0, 0)$.

The 4-velocity of an inertial observer is given by $V = (c, 0)$ in the observer's rest frame. It is constant along his worldline, so that $\nabla_V V = 0$.

**Exercise 3.6** Show that if $A$ and $B$ are events on the worldline of an inertial observer with 4-velocity $V$, then the displacement 4-vector from $A$ to $B$ is $\tau V$, where $\tau$ is the proper time from $A$ to $B$.

**Exercise 3.7** An inertial observer $O$ has 4-velocity $V$. Show that $O$ reckons that two events $A$ and $B$ are simultaneous if and only if $g(V, X) = 0$, where $X$ is the displacement 4-vector from $A$ to $B$.

**Exercise 3.8** Let $O$ and $\tilde{O}$ be inertial observers with different 4-velocities. Show that there is a unique pair of events $A$ and $\tilde{A}$, with $A$ on $O$'s worldline and $\tilde{A}$ on $\tilde{O}$'s worldline, that the observers agree are simultaneous. Show that the observers agree that at these events, the distance separating them is minimal. (Hint: look at the problem in the rest frame of one of the observers, and use the result of the previous exercise.)

### 3.8 4-Acceleration

Let $W$ be the worldline of an accelerating particle and let $V$ be its 4-velocity. Since $V$ is tangent to $W$, the 4-vector $\dot{V} = \nabla_V V$ is well defined. It is called the 4-acceleration of $W$. In inertial coordinates, the 4-acceleration has components $\dot{V}^a = dV^a/d\tau$. By differentiating $g_{ab} V^a V^b = c^2$ with respect to $\tau$,

$$g_{ab} V^a \dot{V}^b = 0.$$

Therefore $\dot{V}$ is orthogonal to $V$.

Let $A$ be an event on $W$ and let $x^a$ be an inertial coordinate system in which the components of $V$ at $A$ are $(c, 0, 0, 0)$: we call such a system an instantaneous rest frame of the particle. Since $g(\dot{V}, V) = 0$, $V$ is of the form $(0, a)$ at $A$ in these coordinates. At proper time $\delta\tau$ after $A$, the 4-velocity is

$$(c, \delta v) = (c, 0) + \delta\tau(0, a) + O(\delta\tau^2).$$

Since $\gamma(0) = 1$, $\delta\tau = \delta t$ to the first order in $\delta\tau$, where $t = x^0/c$. Therefore $a = dv/dt$. In the instantaneous rest frame, the spatial part of $\dot{V}$ is the acceleration in the ordinary sense. The scalar $a = \sqrt{-g(\dot{V}, V)} = \sqrt{a \cdot a}$ measures the acceleration felt by an accelerating observer with worldline $W$. It is called the proper acceleration of the particle.
As an example, consider a particle moving along the $x$-axis of a fixed inertial coordinate system. Then the coordinates of the events on $W$ are of the form $(ct, x, 0, 0)$. The 4-velocity and 4-acceleration are given by

$$V = \gamma(v)(c, vi) \quad \text{and} \quad \dot{V} = \frac{d}{d\tau}\left(\gamma(v)(c, vi)\right),$$

where $v = \frac{dx}{dt}$ and $i$ is a unit 3-vector along the $x$-axis. By using (3.5),

$$\dot{V} = \frac{v}{c^2(1 - v^2/c^2)^{3/2}} \frac{dv}{d\tau}(c, vi) + \frac{1}{(1 - v^2/c^2)^{1/2}} \frac{dv}{d\tau}(0, i)$$

$$= \frac{1}{c(1 - v^2/c^2)^{3/2}} \frac{dv}{d\tau}(v, ci)$$

and

$$a = \frac{1}{1 - v^2/c^2} \frac{dv}{d\tau} = \frac{1}{(1 - v^2/c^2)^{3/2}} \frac{dv}{dt}.$$

If the proper acceleration $a$ remains bounded, then $dv/dt \to 0$ as $v \to c$.

Suppose that $a$ is constant. The components of $V$ and $\dot{V}$ are $(ci, \dot{x}, 0, 0)$ and $(ci, \ddot{x}, 0, 0)$, where the dot denotes the derivative with respect to $\tau$. Hence

$$c^2i^2 - \dot{x}^2 = c^2 \quad \text{and} \quad a^2 = \ddot{x} - c^2i^2.$$

If we make an appropriate choice for the zero of $\tau$, then after a short calculation,

$$i = \cosh(a\tau/c) \quad \text{and} \quad \dot{x} = c\sinh(a\tau/c).$$

With a suitable choice of origin for the inertial coordinates,

$$t = \frac{c}{a} \sinh(a\tau/c) \quad \text{and} \quad x = \frac{c^2}{a} \cosh(a\tau/c).$$

It follows that $W$ is a hyperbola in space-time, with null lines through the origin as asymptotes.

Let $A$ and $B$ be the events on $W$ at which $\tau = -T$ and $\tau = T$. Relative to the fixed inertial frame, $A$ and $B$ happen at the same place (they have the same $x$ coordinate), but they are separated by the coordinate time interval $(2c/a)\sinh(aT/c)$. A clock which is at rest relative to the inertial frame measures time $(2c/a)\sinh(aT/c)$ between $A$ and $B$, while a clock moving along the accelerating worldline measures time $2T$. This is an example of the *clock paradox*, although it is no more paradoxical than the fact that the distance you travel in driving from one town to another depends on the route you take.\(^8\)

It is amusing to put in some numbers. If $a \simeq g = 10\text{m/s}^2$, $T = 35\text{y}$, and $c = 3 \times 10^8\text{m/s}$, then $aT/c \simeq 35$ and

$$(2c/a)\sinh(aT/c) \simeq e^{35}\text{y}.$$  

\(^8\) In Euclidean geometry, the shortest route between two points is the straight one. In Minkowski space, it is the other way round. The nonaccelerating worldline from $A$ to $B$ is the one with the longest proper time: the curved worldline $W$ is 'shorter' than the straight line from $A$ to $B$. 
If you travel at one $g$, the acceleration due to gravity, for a lifetime (70 years), then to an observer in the rest frame of your 35th birthday, you appear to live for $1.6 \times 10^{15}$ years.

**Exercise 3.9** An observer travels round a circular path of radius $D$ with constant angular speed $\omega$, relative to some inertial coordinate system. What acceleration does the observer feel? What proper time elapses during one complete circuit? (Answers: $\gamma^2 D \omega^2$ and $2\pi/\gamma \omega$, where $\gamma = 1/\sqrt{1 - D^2 \omega^2/c^2}$.)

**Exercise 3.10** Consider the constant acceleration worldline in this section. Let $E$ be the event $(0, 0, 0, 0)$, let $F$ be any event on the worldline, and let $X$ be the displacement vector from $E$ to $F$. Show that $g(V, X) = 0$ and that $g(X, X)$ is the same for all choices of $F$. Deduce that, as perceived by an observer with this worldline, the event $E$ is simultaneous with every event in the observer's history, and is always the same distance away. Comment on this apparent absurdity. (What is the distance when $a = g$? See Exercise 2.8.) Show that it is possible for two observers accelerating in opposite directions along the $x$-axis to remain a fixed distance apart (as measured by both of them in their instantaneous rest frames).

**Exercise 3.11†** Show that for the constant acceleration worldline in this section, the proper acceleration is $cd\theta/d\tau$, where $\theta$ is the rapidity (Exercise 2.14).

**Exercise 3.12†** A spaceship sets out from earth to visit a distant planet, which is at rest relative to the earth. The spaceship accelerates away from the earth with acceleration
for 10 years (the time and the acceleration are both measured in the spaceship), and then decelerates for 10 years, coming to rest relative to the earth at the planet. After spending a few minutes looking around, the crew return to earth in the same way. How far away was the planet (in the earth’s frame)? Why does the fact that this distance is much greater than 20 light years not violate the prohibition on faster than light travel? How much time elapses on earth before the return?

3.9 Relative Speed

Consider two free particles with 4-velocities $U$ and $U'$. Their relative speed $v$ is defined to be the speed of one measured in the rest frame of the other. It follows from (3.4) that

$$c^2 \gamma(v) = \frac{c^2}{\sqrt{1 - v^2/c^2}} = g(U, U'),$$

since $g(U', U') = c^2$. (Note that the right-hand side is symmetric in $U$ and $U'$.)

If the particles have velocities $u$ and $u'$ relative to a general inertial frame, then

$$U = \gamma(u)(c, u) \quad \text{and} \quad U' = \gamma(u')(c, u').$$

Hence $\gamma(v) = \gamma(u)\gamma(u')(1 - u \cdot u'/c^2)$. It follows after a short calculation that

$$v^2 = \frac{c^2(u - u') \cdot (u - u') - c^2u^2u'^2 + c^2(u \cdot u')^2}{(c^2 - u \cdot u')^2}.$$ 

When $u$ and $u'$ are antiparallel, this reduces to (2.12).

Exercise 3.13 A solid is moving with 4-velocity $U$. Measured in its rest frame, it has volume $\Delta$. Show that an observer with 4-velocity $V$ will measure the solid’s volume to be $\Delta c^2/g(U, V)$ (see §2.6).

3.10 The 4-Gradient

Let $f$ be a function on space-time. Under the inhomogeneous Lorentz transformation $x^a = L^a_b \tilde{x}^b + T^a$, the partial derivatives of $f$ transform by

$$\frac{\partial f}{\partial \tilde{x}^a} = L^b_a \frac{\partial f}{\partial x^b},$$

which is the wrong way round to define the components of a 4-vector field. However we can write this

$$\frac{\partial f}{\partial x^a} = M^b_a \frac{\partial f}{\partial \tilde{x}^b},$$

where $(M^a_b)$ is the inverse matrix of $(L^a_b)$. We saw in §2.2 that $M^b_a = g^{bd}g_{ae}L^e_d$. Therefore,
$$\frac{\partial f}{\partial x^a} = g_{ae} L^e_d g^{bd} \frac{\partial f}{\partial x^b}.$$ 

Since $g^{ak} g_{ae} = \delta^k_e$, it follows that $\nabla^k f = L^k_b \tilde{\nabla}^b f$, where

$$\nabla^a f = g^{ab} \frac{\partial f}{\partial x^b} \text{ and } \tilde{\nabla}^a f = g^{ab} \frac{\partial f}{\partial x^b},$$

and hence that the $\nabla^a f$s are the components of a 4-vector field.

The operators $\nabla^a$, where

$$(\nabla^0, \nabla^1, \nabla^2, \nabla^3) = \left( \frac{\partial}{\partial x^0}, -\frac{\partial}{\partial x^1}, -\frac{\partial}{\partial x^2}, -\frac{\partial}{\partial x^3} \right),$$

are the components of the 4-gradient, which is a 4-vector operator, in the sense that $\nabla^a = L^a_b \tilde{\nabla}^b$. The 4-vector field with components $\nabla^a f$ is the gradient 4-vector field of $f$.

We can also use $\nabla^a$ to define the 4-divergence of a 4-vector field $X$: it is the scalar

$$g_{ab} \nabla^a X^b = \frac{\partial X^0}{\partial x^0} + \frac{\partial X^1}{\partial x^1} + \frac{\partial X^2}{\partial x^2} + \frac{\partial X^3}{\partial x^3}.$$ 

It is independent of the choice of coordinates.

**Example 3.5** The wave operator is invariant under change of inertial coordinates since

$$g_{ab} \nabla^a \nabla^b = \frac{1}{c^2} \frac{\partial^2}{\partial \tau^2} - \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial z^2}.$$ 

Its action on a function $f$ is the 4-divergence of the 4-gradient of $f$ (see Exercise 2.9).

### 3.11 The Frequency 4-Vector

A (scalar) plane wave is a solution of the wave equation of the form $u = \Re(\phi)$, where

$$\phi = A \exp(i \omega (t - c^{-1} e \cdot r)),$$

where $r = (x, y, z)$, $A$ is a complex constant, $\omega$ is a real constant, and $e$ is a unit 3-vector: $A$ is the complex amplitude, $\omega$ is the angular frequency, and $e$ gives the direction of propagation (see Exercise 1.2).

We define the frequency 4-vector $K$ by

$$K^a = -i \phi^{-1} \nabla^a \phi.$$ 

It is a 4-vector because $\nabla^a$ is a 4-vector operator. Its components are constant: the temporal part is $c^{-1} \omega$ and the spatial part is $c^{-1} \omega e$.

**Example 3.6** A plane wave in one inertial coordinate system is a plane wave in every inertial coordinate system. An observer with 4-velocity $V$ measures the frequency of the wave to be $\omega = g(K, V)$ since in the observer's rest frame, $V = (c, 0)$ and $K = c^{-1} \omega (1, e)$. 
If an observer at rest in some frame sees a plane wave propagating in the direction \( e \) with angular frequency \( \omega \), then a second observer with velocity \( u \) will see a plane wave with frequency \( \tilde{\omega} = g(K,U) \), where \( K = c^{-1}\omega(1,e) \) and \( U = \gamma(u)(c,u) \). That is,

\[
\tilde{\omega} = c^{-1}\omega\gamma(u)(c - e \cdot u).
\]

This is the Doppler effect. Note that, in contrast to the classical theory, \( \tilde{\omega} \neq \omega \) when \( e \cdot u = 0, u \neq 0 \) (why?).

Exercise 3.14  Scalar plane waves of frequencies \( p^{-1}, q^{-1}, r^{-1}, s^{-1} \) propagate in the directions \((1,1,1), (1,-1,-1), (-1,1,-1), \) and \((-1,-1,1)\), respectively. Show that there exists an inertial frame in which all four have the same frequency if and only if

\[
3(p + q - r - s)^2 + 3(p - q + r - s)^2 + 3(p - q - r + s)^2 < (p + q + r + s)^2.
\]

(Hint: look for a timelike vector \( V \) such that \( g(V,K) = g(V,L) = g(V,M) = g(V,N) \), where \( K, L, M, N \) are the frequency 4-vectors.)

3.12 The Appearance of Moving Objects

An observer \( O \) looks at a solid object. Suppose that at some event \( A \) on the observer's worldline, the object appears to have a circular outline. Then the photons that reach the observer at \( A \) travel along the generators of a right circular cone in the observer's 3-space, with vertex at the observer. That is, they travel in the direction of the vectors \( k \) such that

\[
k \cdot x = -\sqrt{x \cdot x \sqrt{k \cdot k}} \cos \alpha,
\]

where \( x \) is a 3-vector in the direction from the observer to the centre of the circular outline and \( 2\alpha \) is the angle subtended by the outline.

Let \( K \) and \( X \) be the 4-vectors such that \( K = (\sqrt{k \cdot k}, k) \) and \( X = (-\cos \alpha \sqrt{x \cdot x}, x) \) in the observer's frame (\( K \) is null and \( X \) is spacelike). Then (3.6) can be rewritten

\[
g(K,X) = 0.
\]

That is, the tangents to the worldlines of the photons that reach the observer at \( A \) from the outline are orthogonal to \( X \).

Conversely, if there exists a 4-vector \( X = (X^0, x) \) such that (3.7) holds for every photon travelling from the apparent outline of a body to \( A \), then we can recover (3.6) by defining

\[
\cos \alpha = -X^0/|x|,
\]

and we can conclude that the outline appears to the observer at \( A \) to be circular.

Equation (3.7) is invariant, and does not involve the 4-velocity of the observer. We conclude that a body that appears to one observer at some event to have a circular outline

---

9 This 'transverse Doppler shift' has been verified to one part in \( 2.5 \times 10^4 \) by M. Kaivola, O. Poulsen, E. Riis, and S. A. Lee: Measurement of the relativistic Doppler shift in neon Phys. Rev. Lett. 54 255-8 (1985).
Fig. 3.5. The cone in the observer's 3-space

appears to every observer at that event to have a circular outline. Thus we have Penrose's remarkable result that a moving sphere appears circular. There is no contradiction with the Lorentz contraction because photons that reach the observer at $A$ do not necessarily leave the body simultaneously in the observer's frame.

We shall call $X$ the outline 4-vector: it encodes the apparent size and direction of the outline, and is unique up to multiplication by a positive scalar.

**Example 3.7** A sphere subtends an angle $2\theta$ when seen from a point $P$ in its rest frame. An observer $O$ passes through $P$ with velocity $v$. Show that to $O$, the sphere appears to subtend an angle $2\alpha$ such that

$$c \cot \alpha = \gamma(v)(c \cot \theta + e \cdot v \csc \theta),$$

where $e$ is a unit 3-vector in the frame of the sphere in the direction from $P$ to the centre of the sphere.

**Solution.** To an observer at $P$ at rest relative to the sphere, the outline subtends a right circular cone of semivertical angle $\theta$. Therefore we can take the outline 4-vector to be $X = (\cos \theta, e)$ in the rest frame of the sphere.

Let $O$'s 4-velocity be $V$. In $O$'s rest frame, $V = (c, 0)$ and $X = (-\cos \alpha \sqrt{x \cdot x}, x)$, where $x$ points to the apparent centre of the circular outline. Therefore,

$$c \cot \alpha = -\frac{g(X, V)}{\sqrt{-g(X, X)}}.$$

In the rest frame of the sphere, $V = \gamma(v)(c, v)$ and $X = (\cos \theta, e)$. Therefore,

---

\[ g(X, V) = \gamma(v)(-c \cos \theta - e \cdot v), \quad g(X, X) = \cos^2 \theta - 1 = -\sin^2 \theta. \]

The formula follows.

A disconcerting consequence of the formula in this example is that \( \alpha \to \pi \) as \( v \to -ce \). If \( O \) accelerates instantaneously directly away from the sphere, and increases his speed towards the velocity of light, then the area of the sky filled by the sphere grows. When his speed is \( c \cos \theta \), \( \alpha = \pi/2 \), and the sphere occupies half the sky. At higher speeds, it occupies more than half the sky, and in the limit, only a small hole in the observer's forward direction is left uncovered.

![Diagram of a sphere with a shaded portion representing the sky](image)

**Fig. 3.6.** The outline of the sphere as \( O \) accelerates away from it to the left. The spheres in the figure represent the sky, with the observer at the centre. The shaded portions are the areas covered by the images of the solid sphere.

If an astronaut sees a space monster approaching, with jaws open ready to swallow his spaceship, and if he attempts to accelerate sharply away from the danger, then, when his speed exceeds \( c \cos \theta \), where \( 2\theta \) is the angle originally subtended by the jaws, he appears to be inside the monster’s mouth. As he accelerates further, the jaws close around him. He is eventually killed by the glint on the monster’s teeth, which is blue-shifted to high-frequency gamma radiation (exercise).

### 3.13 Example and Exercises

Notice the technique used throughout this chapter: express the quantity you want to calculate in terms of invariants, and then evaluate them in different frames.

**Example 3.8** A rod is of length \( L_0 \) in its rest frame. In a second inertial frame, it is oriented in the direction of the unit vector \( \mathbf{i} \) and is moving with constant velocity \( v \). What is its length in the second inertial frame?

**Solution.** Let \( A \) be an event at one end of the rod and let \( B \) and \( C \) be the events at the other end that are simultaneous with \( A \) in the given frame and in the rest frame,
respectively. Let \( V \) be the 4-velocity of the rod and let \( X \) and \( Y \) be the displacement 4-vectors from \( A \) to \( B \) and from \( A \) to \( C \), respectively (Fig. 3.7). In the second frame, 
\[
V = \gamma(v)(c,v) \quad \text{and} \quad X = (0, Li),
\]
where \( L \) is the length we are required to find.

![Fig. 3.7. The events A, B and C]

We know that \( g(Y, Y) = -L_0^2 \) and that \( g(Y, V) = 0 \), since \( A \) and \( C \) are simultaneous in the rest frame. Also, \( Y - X = \tau V \), where \( \tau \) is the proper time from \( B \) to \( C \) along the worldline of the end of the rod. By taking the scalar product with \( V \), 
\[
c^2 \tau = -g(X, V).
\]

It follows that
\[
L^2 = -g(X, X) = -g(Y - \tau V, Y - \tau V) = L_0^2 - c^{-2}g(X, V)^2 = L_0^2 - c^{-2}L^2\gamma(v)^2(i \cdot v)^2
\]

Hence, if \( \theta \) is the angle between \( v \) and \( i \),
\[
L = \frac{L_0\sqrt{c^2 - v^2}}{\sqrt{c^2 - v^2 \sin^2 \theta}}.
\]

Note that \( L = L_0 \) when \( \theta = \pi/2 \) and that \( L = L_0/\gamma(v) \) when \( i \) is parallel to \( v \). \( \square \)

**Exercise 3.15** An inertial observer \( O \) looks at the standard clock of a second inertial observer \( \tilde{O} \), who is moving directly away from him with speed \( v \). Show that \( \tilde{O}'s \) clock appears *visually* to run slow by a factor \( \sqrt{c - v}/\sqrt{c + v} \). (Hint: find the coordinates of the pairs of events \( A, \tilde{A} \), with \( A \) on \( O's \) worldline and \( \tilde{A} \) on \( \tilde{O}'s \) worldline, such that the displacement vector from \( \tilde{A} \) to \( A \) is future-pointing null.)

**Exercise 3.16** Two thin straight rods \( R \) and \( R' \) are rigidly bolted together so that, in their rest frame, the angle between them is \( \theta \). Relative to a second inertial frame, the rods are moving with velocity \( v \). In the second frame, the angle between \( R \) and \( v \) is \( \alpha \), the angle between \( R' \) and \( v \) is \( \alpha' \), and the angle between \( R \) and \( R' \) is \( \phi \). Show that
\[
\cos \phi = \frac{\cos \theta \sqrt{c^2 - v^2} \sin^2 \alpha \sqrt{c^2 - v^2} \sin^2 \alpha' - c^2 \cos \alpha \cos \alpha'}{c^2 - v^2}.
\]

**Exercise 3.17** Relative to an inertial coordinate system \(ct, x, y, z\), a particle travels with varying speed \(v\) along a curve with curvature \(\kappa\). Show that the acceleration \(a\) measured in the particle’s instantaneous rest frame is given by

\[
a^2 = \left(\frac{c^2}{c^2 - v^2}\right)^3 \left[ \left(\frac{dv}{dt}\right)^2 + \left(1 - \frac{v^2}{c^2}\right)v^4 \kappa^2 \right].
\]

**Exercise 3.18** Two spaceships \(S\) and \(T\) set off in the same direction but with different speeds from their home planet \(H\). After the elapse of proper time \(\tau\) on \(H\), two further spaceships \(S'\) and \(T'\) simultaneously set off from \(H\) to meet up with \(S\) and \(T\), respectively, which they do after the passage of proper time \(\tau\) on \(S'\) and \(T'\). On the assumption that the motions of \(S, T, S', T'\) and \(H\) are uniform, show that the rapidity of \(S'\) relative to \(T'\) is twice the rapidity of \(S\) relative to \(T\). (Hint: let \(U, V, U', V'\) and \(W\) be the 4-velocities of \(S, T, S', T'\) and \(H\), respectively. Show that \(tU = \tau(W + U')\), where \(t\) is the proper time on \(S\) from its leaving \(H\) to its being caught up by \(S'\). Show that

\[
2 \cosh^2 \psi = 1 + \cosh \theta,
\]

where \(\psi\) is the rapidity of \(S\) relative to \(H\) and \(\theta\) is the rapidity of \(S'\) relative to \(H\), by using \(c^2 \cosh \psi = g(U, W)\), and so on.)

**Exercise 3.19** An inertial observer \(\tilde{O}\) is moving directly away from an inertial observer \(O\) with velocity \(ve\), where \(e \cdot e = 1\). The observer \(O\) sees a distant star in a direction which makes an angle \(\theta\) with \(e\). Show that if \(\tilde{e}\) is the unit vector in \(\tilde{O}\)'s 3-space in the direction from \(O\) to \(\tilde{O}\), then the direction of the star appears to \(\tilde{O}\) to make an angle \(\alpha\) with \(\tilde{e}\), where

\[
\sin \alpha = \sin \theta \sqrt{\frac{c^2 - v^2}{c + v \cos \theta}}
\]

(this is a version of the aberration formula; you can derive it from Example 3.7).
4 Relativistic Particles

4.1 Conservation of 4-Momentum

Mass enters Newtonian mechanics in two ways—as inertial mass, the constant \( m \) in the second law \( F = ma \); and as gravitational mass, the constants \( m, m' \) in the inverse-square law \( F = Gmm'/r^2 \). In this chapter, we shall look at how the idea of inertial mass must be modified in the context of special relativity. We shall not consider gravitational mass, which would take us into general relativity.

Consider a closed system of particles which undergo collisions, but which are otherwise free. 'Collisions' can be collisions in the ordinary sense, or events at which particles break up or coalesce, or any other type of interaction between coincident particles at a single event (it is much harder to deal with interactions between spatially separated particles because of the conflict between the classical idea of 'action at a distance' and the basic principle of relativity that no influence can be transmitted from one particle to another at a speed faster than light). In Newtonian mechanics, collisions are governed by the following conservation laws.

N1 *Conservation of mass.* If the masses of the incoming particles are \( m_1, m_2, \ldots, m_k \) and those of the outgoing particles are \( m_{k+1}, m_{k+2}, \ldots, m_n \), then

\[
\sum_{i=1}^{k} m_i = \sum_{i=k+1}^{n} m_i
\]

(there is no assumption that the number of incoming particles is equal to the number of outgoing particles).

N2 *Conservation of momentum.* If the velocities of the incoming particles are \( v_1, v_2, \ldots, v_k \) and those of the outgoing particles are \( v_{k+1}, v_{k+2}, \ldots, v_n \), then

\[
\sum_{i=1}^{k} m_i v_i = \sum_{i=k+1}^{n} m_i v_i.
\]

N3 *Conservation of energy.* If the collisions are elastic, then kinetic energy is conserved. In a general collision, however, there is an exchange between kinetic energy and other forms of exchange (heat, chemical energy, and so on), and kinetic energy by itself is not conserved.

These cannot be taken over into special relativity without modification since (N2) and (N3) are not invariant under Lorentz transformations. We need to find relativistic conservation laws which reduce to the Newtonian laws in the limit of low velocities and which are invariant under Lorentz transformations.

The first step is to adopt an operational definition of mass. Given a standard mass \( M \), an observer can assign a mass \( m \) to any other particle by colliding it at low speeds
with the standard mass, measuring the resulting velocities, and applying the Newtonian law of conservation of momentum. Since this becomes exact as the velocities go to zero, the observer can use a limiting procedure to measure $m$ when the particle is at rest. This quantity is called the \textit{rest mass} of the particle. The 'limiting procedure' is needed to get over the awkward point that the particle is required to be simultaneously at rest and involved in collisions.

Each particle has a rest mass $m$ (a scalar) and a 4-velocity $V$ (a 4-vector). The 4-vector $P = mV$ is called the 4-\textit{momentum} of the particle. It has temporal and spatial parts $m\gamma(v)c$ and $m\gamma(v)v$, where $v$ is the velocity. As $v \to 0$, $\gamma(v) = 1 + O(v^2/c^2)$ and

$$P = (mc, mv) + O(v^2/c^2).$$

If all the velocities of the particles in a collision are so small that terms in $v^2/c^2$ can be neglected, then the Newtonian laws of conservation of mass and momentum are equivalent to the conservation of the spatial and temporal parts of 4-momentum.

This suggests the law of \textit{conservation of 4-momentum}, which asserts that in \textit{any} collision,

$$\sum_{i=1}^{k} P_i = \sum_{i=k+1}^{n} P_i,$$

(4.1)

where $P_1, P_2, \ldots, P_k$ are the 4-momenta of the incoming particles and $P_{k+1}, P_{k+2}, \ldots, P_n$ are the 4-momenta of the outgoing particles. The five classical conserved quantities (mass, energy, and the three components of momentum) are replaced by four relativistic ones (the four components of $P$); but nothing is lost because mass and energy are equivalent in special relativity, as we shall see shortly. Because (4.1) is a 4-vector equation, it is consistent with Lorentz transformations: if it holds in one inertial coordinate system, then it holds in every inertial coordinate system. It is a very plausible relativistic extension of the Newtonian laws, but its justification is derived, as always, from experiment.

When the particles are moving at much less than the speed of light, we recover (N1) and (N2) as the first order approximation in $v/c$. By expanding $\gamma(v) = 1/\sqrt{1 - v^2/c^2}$ in powers of $v/c$,

$$P = (mc(1 + \frac{1}{2}v^2/c^2 + \ldots), m(1 + \frac{1}{2}v^2/c^2 + \ldots)v).$$

When we include the terms of order $v^2/c^2$, the temporal part of $P$ becomes

$$P^0 = c^{-1}(mc^2 + \frac{1}{2}mv^2) + O(v^3/c^3),$$

and the spatial part becomes

$$m\gamma(v)v = (m + \frac{1}{2}mv^2/c^2)v + O(v^3/c^3).$$

In the second-order approximation, $cP^0$ is equal to the sum of the Newtonian kinetic energy $\frac{1}{2}mv^2$ and a much larger term $mc^2$. The kinetic energy also contributes a small correction $\frac{1}{2}mv^2/c^2$ to the mass measured by high speed collisions.

In a general frame, we write $P = (E/c, p)$. We call $p$ the 3-momentum and $E$ the total energy relative to the frame. These are related to the rest mass $m$ and the velocity $v$ relative to the frame by

$$E = mc^2\gamma(v) \quad \text{and} \quad p = m\gamma(v)v.$$

(4.2)
The terminology is justified, first, by the form of \( E \) and \( p \) in the limit of small \( v/c \) and, second, by the fact that it is \( E \) and \( p \) that are exactly conserved in all collisions.

The quantity \( mc^2 \) in (4.2) is called the rest energy since it is equal to the total energy for a particle at rest. The factor of proportionality \( m\gamma(v) \) between \( p \) and \( v \) is called the inertial mass: it is a particle's apparent dynamical mass, as measured by high speed collisions. It is denoted by \( m_I \).

Note that rest mass is a scalar (all observers agree on its value), but inertial mass is different in different inertial coordinate systems. Rest mass and inertial mass are equal for a particle at rest. The new feature of the relativistic theory is that a particle's inertial mass increases with its velocity, albeit only very slightly for velocities much less than that of light.\(^{11}\)

### 4.2 Equivalence of Mass and Energy

The definition of total energy can be rewritten \( E = m_I c^2 \). One consequence of the relativistic theory is that a particle's inertial mass increases in proportion to its total energy. Since \( E \to \infty \) as \( v \to c \), this is consistent with the impossibility of accelerating a particle with nonzero rest mass to the velocity of light. A more striking consequence is that it is possible to exchange rest mass for kinetic energy in a collision. For example, a particle of rest mass \( m \) can break up into a number of particles with large velocities, but total rest mass less than \( m \) (Example 4.1).

The equivalence of inertial mass and total energy expressed by the formula \( E = m_I c^2 \) is a fundamental principle of relativistic physics. It is not restricted to particles. In general relativity, it is extended to an equivalence of energy and gravitational mass.

In classical mechanics, the energy that can be stored in a given mass (for example, in the form of heat or nuclear or chemical energy) is, in principle, unlimited. In relativity, by contrast, any increase in energy is accompanied by an increase in mass. If a body is heated, for example, then its rest mass increases (usually only by an undetectable amount). The maximum energy that can be released from a body of rest mass \( m \) is its rest energy \( mc^2 \). This upper limit is very large: even in an atomic explosion, only about 0.1% of the available energy is released.

### 4.3 The 4-Momentum of a Photon

According to quantum theory, a photon of angular frequency \( \omega \) has energy \( E = h\omega \). The relativistic version of this formula assigns a 4-momentum \( P = hK \) to a photon with frequency 4-vector \( K \). For a particle of rest mass \( m \), we have \( m^2 c^2 = g(P, P) \). For a photon, \( g(P, P) = 0 \), since \( K \) is null. Consequently photons are 'zero-rest-mass particles'. (We shall look more closely at the frequency 4-vector of a photon in Chap. 6: all we need for the present is that \( K \) is a null vector tangent to the photon's worldline.)

The law of conservation of 4-momentum also holds for collisions involving photons.

\(^{11}\)In the modern terminology of particle physics, 'rest mass' is simply 'mass' and 'inertial mass' has disappeared since it is equivalent to (total) energy.
4.4 4-Force

If a particle has an accelerating worldline, then we again define the 4-momentum by $P = mV$, where $V$ is the 4-velocity and $m$ is the rest mass. The 4-force acting on the particle is the 4-vector $F = \nabla_V P = dP/d\tau$. In a general inertial frame,

$$F = \gamma(v) \left( \frac{1}{c} \frac{dE}{dt}, \mathbf{f} \right).$$

Here $u$ is the 3-velocity and $\mathbf{f}$ is the 3-vector defined by

$$\mathbf{f} = \frac{dp}{dt},$$

where $p = m\gamma(v)v$ is the 3-momentum; $\mathbf{f}$ is called the 3-force. If the rest mass is constant, then $g(F, V) = 0$ and

$$\mathbf{f}.v = \frac{dE}{dt},$$

which is the same as the Newtonian formula.

4.5 Examples and Exercises

The following examples illustrate the use of the law of conservation of 4-momentum. Note the very useful identities:

$$m^2c^2 = g(P, P) \quad \text{and} \quad v^2\gamma(v)^2 = c^2(\gamma(v)^2 - 1).$$

Note also that an observer with 4-velocity $U$ measures the total energy of a particle with 4-momentum $P$ to be $g(P, U)$, since in the rest frame of the observer, $U = (c, 0)$ and $P = (E/c, p)$.

Example 4.1 A particle of rest mass $M$ is at rest when it splits into two particles, each of rest mass $m$, which have velocities $v$ and $-v$. Show that $M = 2m\gamma(v)$, where $v = |v|$.

Solution. By conservation of 4-momentum,

$$M(c, 0) = m\gamma(v)(c, v) + m\gamma(v)(c, -v).$$

Therefore $M = 2m\gamma(v)$. Note that if $v \neq 0$, then $M > 2m$; and that $m \to 0$ as $v \to c$ (for fixed $M$).

Example 4.2 Elastic collisions. A particle of rest mass $m$ is moving with velocity $u$ relative to an inertial frame when it collides elastically with a second particle, also of rest mass $m$, which is at rest. After the collision, the particles have velocities $v$ and $w$. Show that if $\theta$ is the angle between $v$ and $w$, then

$$\cos \theta = \frac{c^2}{vw} \left( 1 - \sqrt{1 - v^2/c^2} \right) \left( 1 - \sqrt{1 - w^2/c^2} \right).$$
(In relativity, an elastic collision is one in which the rest masses of the particles involved are unchanged.)

**Solution.** By conservation of 4-momentum,

\[ m(c, 0) + m\gamma(u)(c, u) = m\gamma(v)(c, v) + m\gamma(w)(c, w). \]

Therefore,

\[ 1 + \gamma(u) = \gamma(v) + \gamma(w) \quad \text{and} \quad \gamma(u)u = \gamma(v)v + \gamma(w)w. \]  \hspace{1cm} (4.3)

By taking the modulus squared of both sides of the second of these,

\[ \gamma(u)^2u^2 = \gamma(v)^2v^2 + \gamma(w)^2w^2 + 2\gamma(v)\gamma(w)v \cdot w. \]

Hence

\[ 2\gamma(v)\gamma(w)v \cdot w = c^2(\gamma(u)^2 - \gamma(v)^2 - \gamma(w)^2 + 1). \]

But, from the first equation in (4.3),

\[ \gamma(u)^2 - \gamma(v)^2 - \gamma(w)^2 + 1 = (\gamma(v) + \gamma(w) - 1)^2 - \gamma(v)^2 - \gamma(w)^2 + 1 = 2(\gamma(v) - 1)(\gamma(w) - 1). \]

Therefore,

\[ v \cdot w = \frac{c^2(\gamma(v) - 1)(\gamma(w) - 1)}{\gamma(v)\gamma(w)}, \]

from which the required result follows. \(\square\)

Note that \(\cos \theta > 0\) whenever \(v, w > 0\), so the angle between \(v\) and \(w\) is always acute. Also, as \(v, w \to c\), \(\cos \theta \to 1\), and so \(\theta \to 0\).

In the analogous problem in Newtonian theory, the kinetic energy and momentum are both conserved, from which it follows that \(u = v + w\) and \(u^2 = v^2 + w^2\). These imply that \(v \cdot w = 0\) and hence that \(v\) and \(w\) are always orthogonal.

**Example 4.3†** A rocket accelerates along the \(x\)-axis of an inertial frame by firing out a stream of particles in the negative \(x\)-direction. As it does so, its rest mass \(m\) decreases. Show that

\[ m \frac{dv}{dm} + u \left( 1 - \frac{v^2}{c^2} \right) = 0, \]

where \(v\) is the speed of the rocket relative to the frame and \(u\) is the speed of the particles relative to the rocket.

**Solution.** Consider an event \(A\) on the worldline of the rocket at which it fires out a particle of rest mass \(\delta M\) and 4-velocity \(U\). Suppose that immediately before \(A\), the rocket has rest mass \(m\) and speed \(v\) relative to the inertial frame, and that immediately afterwards, it has rest mass \(m + \delta m\) (\(\delta m < 0\)) and speed \(v + \delta v\). By conservation of 4-momentum,

\[ m\gamma(v)(c, v, 0, 0) = (m + \delta m)\gamma(v + \delta v)(c, v + \delta v, 0, 0) + \delta M(U^0, U^1, 0, 0), \]

where \((U^0, U^1, 0, 0)\) are the components of \(U\) in the inertial frame. However,
\[ g(V, U) = \gamma(v)(cU^0 - vU^1) = c^2 \gamma(u)^2, \]

where \( V \) is the 4-velocity of the rocket (see §3.9). Therefore, \( U^0 = \gamma(v)\gamma(u)(c - vu/c) \) and \( U^1 = \gamma(v)\gamma(u)(v - u) \), since \( g(U, U) = c^2 \).

If we assume that \( \delta m, \delta M \) and \( \delta v \) are all small, and ignore second-order quantities, then by taking the scalar product of (4.4) first with \( \gamma(v)(c, v, 0, 0) \) and then with \( \gamma(v)(c, 0, 0, 0) \), we have

\[ c^2 \delta m + \gamma(u)c^2 \delta M = 0 \quad \text{and} \quad \gamma(v)^2 mc \delta v - \gamma(u)uc \delta M = 0. \]

By eliminating \( \gamma(u)\delta M \), therefore,

\[ u \delta m + m\gamma(v)^2 \delta v = 0, \]

and the result follows in the limit \( \delta m \to 0 \). \( \square \)

**Example 4.4† Compton scattering.** A photon of angular frequency \( \omega \) collides with an electron of rest mass \( m \), which is at rest. After the collision, the photon has angular frequency \( \omega' \). Show that

\[ \hbar \omega\omega'(1 - \cos \theta) = mc^2(\omega - \omega'), \]

where \( \theta \) is the angle between the initial and final trajectories of the photon.

**Solution.** In the frame in which the electron is initially at rest, the 4-momenta of the electron before and after the collision are

\[ P = m(c, 0) \quad \text{and} \quad P' = m\gamma(v)(c, v), \]

where \( v \) is the velocity of the electron after the collision. The 4-momenta of the photon before and after the collision are

\[ Q = \frac{\hbar \omega}{c}(1, e) \quad \text{and} \quad Q' = \frac{\hbar \omega'}{c}(1, e'), \]

where \( e \) and \( e' \) are unit vectors along the initial and final trajectories. In terms of these 4-vectors, the angle \( \theta \) and the change in the frequency of the photon are given by

\[ g(Q, Q') = \frac{\hbar^2 \omega \omega'}{c^2}(1 - \cos \theta) \quad \text{and} \quad g(P, Q - Q') = m\hbar(\omega - \omega'). \]

By conservation of 4-momentum,

\[ P + Q = P' + Q'. \]

Now \( g(P, P) = g(P', P') = m^2 c^2 \), and \( g(Q, Q) = g(Q', Q') = 0 \). Hence

\[ g(P, P') - m^2 c^2 = g(Q, Q') \]

since \( g(P - P', P - P') = g(Q' - Q, Q' - Q) \). Therefore,

\[ g(P, Q - Q') = g(P, P' - P) = g(P, P') - m^2 c^2 = g(Q', Q). \]

Consequently, \( g(Q, Q') = g(P, Q - Q') \), and the result follows. \( \square \)
Exercise 4.1 A particle of rest mass $M$ breaks up into particles of rest masses $m_1$, $m_2$, ..., $m_n$. Show that $M \geq \sum_1^n m_i$. When is this an equality? Show that two particles with nonzero rest mass cannot coalesce into a single photon.

Exercise 4.2 A $\pi^0$ meson decays into two photons of frequencies $\omega_1$ and $\omega_2$, which travel in opposite directions. Find the rest mass of the meson and its speed before the decay in terms of $\omega_1$ and $\omega_2$.

Exercise 4.3† A particle of rest mass $M$ decays from rest into a particle of rest mass $m$ and a photon. Find the energies of the products in terms of $m$ and $M$.

Exercise 4.4 Two particles of equal rest mass collide and annihilate each other, producing two photons. Initially one particle is at rest and the other is moving with velocity $v$. Show that if the angles between $v$ and the trajectories of the photons are $\theta$ and $\phi$, respectively, then

$$\frac{1 + \cos(\theta + \phi)}{\cos \theta + \cos \phi} = \sqrt{\frac{\gamma(v) - 1}{\gamma(v) + 1}}.$$

(Hint: take the inner product of the 4-momentum conservation equation with (i) the 4-velocity of the particle at rest, (ii) the 4-momentum of the first photon, and (iii) the 4-momentum of the second photon. Remember that $v\gamma(v) = c\sqrt{\gamma(v)^2 - 1}$.)

Exercise 4.5† The rocket in Example 4.3 has constant acceleration $a$, measured in its instantaneous rest frame at each event. Show that

$$\frac{m_0}{m} = \frac{(c + v)}{(c - v)}^{c/2u} = e^{a\tau/u},$$

where $m_0$ is its initial rest mass and $\tau$ is the proper time along its worldline. Estimate the total mass of fuel required for the expedition in Exercise 3.12 on the assumption that $u = \frac{1}{2}c$ and that the only filling station is on earth.

Exercise 4.6 A particle of rest mass $M$ and energy $E$ collides with a particle of rest mass $m$ at rest. Show that if $E'$ is the total energy of the two particles in the frame in which their centre of mass is at rest, then

$$E'^2 = (M^2 + m^2)c^4 + 2Emc^2.$$

(The 4-velocity $V$ of the centre of mass is defined to be the 4-velocity proportional to the total 4-momentum of the particles. Express $E'V$ in terms of the total 4-momentum of the particles.)

Exercise 4.7 In Example 4.2, show that if $\alpha$ and $\beta$ are, respectively, the angles between $u$ and $v$, and between $u$ and $w$, then

$$\cos^2 \alpha = \frac{(\gamma(u) + 1)(\gamma(v) - 1)}{(\gamma(u) - 1)(\gamma(v) + 1)},$$
with a similar formula for $\cos^2 \beta$. Deduce that $\tan \alpha \tan \beta = 2/(1 + \gamma(u))$. What is the corresponding classical result? (Hint: express $\gamma(w)^2 w \cdot w$ in terms of $u \cdot v$, $\gamma(u)$, and $\gamma(v)$.)

**Exercise 4.8** A particle accelerates by absorbing photons from a parallel beam of light. Relative to an inertial frame, the particle's velocity is $v$. Show that in this frame,

$$m\gamma(v)(c, v) = m_0\gamma(v_0)(c, v_0) + \lambda(1, k),$$

where $m$ is the particle's rest mass, $k$ is a unit 3-vector in the direction of the beam, $\lambda$ is a function of $m$, and $m_0$ and $v_0$ are the initial values of $m$ and $v$. Hence show that

$$c - v \cdot k = \frac{A}{m^2 + B} \quad \text{and} \quad c^2 - v^2 = \frac{2Ac m^2}{(m^2 + B)^2},$$

where $A$ and $B$ are constants that should be determined from the initial conditions. Show that $m \to \infty$ as $v \to c$. (Hint: (i) take the inner product of the conservation equation with itself, (ii) take the inner product of both sides with $(1, k)$, and (iii) pick out the temporal part of the conservation equation.)
5 Tensors in Space-Time

5.1 Covectors

We have seen that space-time vectors and the invariants constructed from them provide powerful tools for solving problems in relativity. Tensors are more general objects with components that also transform in a linear way under change of inertial coordinates. We shall use them in the next chapter to recast Maxwell’s equations in a Lorentz-invariant form.

We can express the relationship between 4-vector components in different inertial coordinate systems by writing the components as column vectors and multiplying on the left by the Lorentz transformation matrix. The next simplest example of a tensor is a ‘covector’, which follows the ‘dual’ transformation rule: when the four components are written in a row vector, the relationship is given multiplying on the right by the inverse of the Lorentz transformation matrix. The formal definition is as follows.

As in §3.4, let $\mathbb{T}^1$ denote the space of 4-vectors in Minkowski space. A covector or covariant vector is an element of the dual space $\mathbb{T}_1$. In other words, a covector is a linear function $\alpha : \mathbb{T}^1 \to \mathbb{R}$. Each 4-vector $X$ determines a covector $g(X) = g(X; Y) = g(X, Y)$.

Let $x^a$ be an inertial coordinate system and let $X^{(a)}$ be the corresponding coordinate basis in the space of 4-vectors. The dual basis $\alpha^{(a)}$ in $\mathbb{T}_1$ is characterized by

$$\alpha^{(a)}(X^{(b)}) = \delta^a_b.$$

If $\beta \in \mathbb{T}_1$, then $\beta = \beta^a \alpha^{(a)}$, where the components $\beta_a$ are the four real numbers $\beta_a = \beta(X^{(a)})$. The components of a 4-vector $Y$ are similarly given by $Y^a = \alpha^{(a)}(Y)$, and, in terms of these, and $\beta(Y) = \beta_a Y^a$.

Under a translation of the coordinates, the components of vectors and covectors are unchanged. Under a Lorentz transformation $x^a = L^a_b \tilde{x}^b$, vector components transform by $Y^a = L^a_b \tilde{Y}^b$ and the bases $X^{(a)}$ and $\tilde{X}^{(a)}$ of the inertial coordinate systems $x^a$ and $\tilde{x}^a$ are related by

$$X^{(a)} = M^b_a \tilde{X}^{(b)},$$

where $M^b_a = g_{ac}g^{bd}L^c_d$ (Exercise 3.5). Therefore covector components transform according to the rule $\beta_a = M^b_a \tilde{\beta}_b$. Since $M^a_b$ is the inverse of $L^a_b$ (§2.2), it follows that

$$\beta_a Y^a = \tilde{\beta}_a \tilde{Y}^a = \beta(Y).$$

We could equally well define covectors by the transformation rule for their components and use the invariance of $\beta_a Y^a$ to prove that covectors are dual to vectors.
5.2 Raising and Lowering of Indices

If \( X \) has components \( X^a \), then the covector \( g(X) \) has components \( g_{ab}X^b \). It is conventional to write the components of \( g(X) \) as

\[
X_a = g_{ab}X^b,
\]

so that mapping \( X \) to \( g(X) \) is represented by ‘lowering the index’ on \( X^a \). In the same way, we represent the inverse mapping by ‘raising the index’: if \( \beta \) is a covector with components \( \beta_a \), then the vector \( X = g^{-1}(\beta) \), determined by

\[
g(X,Y) = \beta(Y),
\]

\((Y \in \mathbb{T}^1)\), has components \( X^a = g^{ab}\beta_b \). Again it is conventional to write \( \beta^a \) for \( X^a \). The notation is consistent since \( g_{ac}g^{bc} = \delta^c_a \), so if we raise an index and then lower it again, we get back to where we started.

5.3 Tensors

A tensor \( T \) of type \((p,q)\) is a multilinear map

\[
T : \mathbb{T}_1 \times \ldots \times \mathbb{T}_1 \times \mathbb{T}^1 \times \ldots \times \mathbb{T}^1 \rightarrow \mathbb{R},
\]

That is, \( T \) assigns a real number \( T(\alpha, \ldots, \beta, X, \ldots, Y) \) to each collection of \( p \) covectors \( \alpha, \ldots, \beta \) and \( q \) vectors \( X, \ldots, Y \), and is linear in each argument, so that, for example,

\[
T(r\alpha + r'\alpha', \ldots) = rT(\alpha, \ldots) + r'T(\alpha', \ldots)
\]

\[
T(\ldots, rX + r'X', \ldots) = rT(\ldots, X, \ldots) + r'T(\ldots, X', \ldots),
\]

and so on. Here \( r, r' \in \mathbb{R}, \alpha, \alpha' \in \mathbb{T}_1 \), \( X, X' \in \mathbb{T}^1 \). If \( p = 0 \), then \( T \) is a covariant tensor of rank \( q \). If \( q = 0 \), then \( T \) is a contravariant tensor of rank \( p \). Otherwise, \( T \) is a mixed tensor. The set of tensors of type \((p,q)\) is denoted by \( \mathbb{T}^p_q \). In this language, a covector is a covariant tensor of rank one and a vector is a contravariant tensor of rank one (since \( \mathbb{T}^1_1 = (\mathbb{T}^1)^{**} = \mathbb{T}^1 \)).

The components of \( T \) in the inertial coordinate system \( x^a \) are the real numbers

\[
T^{ab \ldots de \ldots} = T(\alpha^a, \alpha^b, \ldots, X^d, X^e, \ldots).
\]

The components have \( p \) upper indices and \( q \) lower indices. Note that if \( T \) is a vector, then \( T = T^aX_a \); and if \( T \) is a covector, then \( T = T_a\alpha^a \).

In terms of the components of \( T \),

\[
T(\alpha, \beta, \ldots, X, Y, \ldots) = T^{ab \ldots de \ldots} \alpha_a \beta_b \ldots X^d Y^e \ldots \tag{5.1}
\]

The components of \( T \) are unchanged by translations of the coordinates. Under the Lorentz transformation \( x^a \mapsto \tilde{x}^a \) where \( x^a = L^a_b \tilde{x}^b \), \( T^{ab \ldots de \ldots} \mapsto \tilde{T}^{ab \ldots de \ldots} \), where

\[
T^{ab \ldots de \ldots} = L^a_b T^b \ldots M_d^l M_e^m \ldots \tilde{T}^{hk \ldots lm \ldots}.
\]
Again, we could equally well take this as the definition of a tensor, and use (5.1) to define \( T(\alpha, \beta, \ldots, X, Y, \ldots) \). In fact we shall often specify tensors by giving their components, and then checking that the components have this transformation property.

**Example 5.1** The metric tensor. The metric \( g \) is a second-rank covariant tensor since \( g(X, Y) = g_{ab} X^a Y^b \) is linear in \( X \) and \( Y \). Its components in any inertial coordinate system are the metric coefficients \( g_{ab} \).

**Example 5.2** The contravariant metric. The coefficients \( g^{ab} \) are the components of a second-rank contravariant tensor (which again has the same components in every inertial coordinate system).

**Example 5.3** The Kronecker delta. This is a mixed tensor of type \((1,1)\) with components \( \delta^a_b \) is any inertial coordinate system. As a bilinear map \( T^1 \times T^1 \to \mathbb{R} \), it is defined by \( \delta(\alpha, X) = \alpha(X) \).

**Example 5.4** The inertia tensor. The same formalism can be used to define tensors in three-dimensional Euclidean space. A familiar example is the inertia tensor \( I \) of a rigid body. This is a second-rank covariant tensor. If \( \mathbf{i} \) is a unit 3-vector, then \( I(\mathbf{i}, \mathbf{i}) \) is the moment of inertia of the body about an axis through the origin in the direction of the unit vector \( \mathbf{i} \). The components of \( I \) in a Cartesian coordinate system are the entries in the inertia matrix (the moments and products of inertia, with appropriate signs).

### 5.4 Operations on Tensors

The following operations on tensors are 'natural'. That is, they are independent of the choice of coordinates.

1. **Scalar multiplication.** If \( T \) is a tensor of type \((p,q)\) and \( r \in \mathbb{R} \), then \( rT \) is the tensor of type \((p,q)\) defined by

   \[
   (rT)(\alpha, \beta, \ldots, X, Y, \ldots) = r(T(\alpha, \beta, \ldots, X, Y, \ldots)).
   \]

   It has components \( rT_{ab\ldots de\ldots} \).

2. **Addition.** If \( T \) and \( T' \) are tensors of the same type, then their sum \( T + T' \) is also a tensor, with components

   \[
   T_{ab\ldots de\ldots} + T'_{ab\ldots de\ldots}.
   \]

3. **Tensor multiplication.** If \( T \) is of type \((p,q)\) and \( T' \) is of type \((p', q')\), then the tensor product \( T \otimes T' \) is the tensor of type \((p+p', q+q')\), defined by

   \[
   T \otimes T'(\alpha, \beta, \ldots, \alpha', \beta', \ldots, X, Y, \ldots, X', Y', \ldots) \\
   = T(\alpha, \beta, \ldots, X, Y, \ldots)T'(\alpha', \beta', \ldots, X', Y', \ldots).
   \]

   It has components \( T_{ab\ldots ef\ldots} T'_{cd\ldots h\ldots} \). Note that \( T \otimes T' \) and \( T' \otimes T \) are not necessarily the same.
4 *Contraction.* Let \( T \) be a tensor of type \((p, q)\). The contraction \( T' \) of \( T \) on the first and the \((p + 1)\)th indices is the tensor of type \((p - 1, q - 1)\) with components

\[
T'_{ab\ldots de\ldots} = T_{ab\ldots ade\ldots}.
\]

These have the right transformation property to define a tensor because \( L^a_k M^k_c = \delta^a_c \). One can also contract on other pairs of indices (one upper and one lower).

5 *Raising and Lowering.* If \( T \) is a tensor of type \((p, q)\), then we can define a tensor \( T' \) of type \((p - 1, q + 1)\) by

\[
T'_{\alpha, \beta, \ldots, X, Y, \ldots} = T(g(X), \alpha, \beta, \ldots, Y, \ldots).
\]

The components of \( T' \) are

\[
T'^{ab\ldots ef\ldots} = g_{ak} T^{kbc\ldots ef\ldots}.
\]

It is conventional to write \( T^{ab\ldots ef\ldots} \) for \( g_{ak} T^{kbc\ldots ef\ldots} \). The operation is called 'lowering the first index'. Put another way, lowering the first index means first taking the tensor product \( g \otimes T \) and then contracting over the second index on \( g \) and the first index on \( T \). One can also raise indices by using the contravariant metric. However, one must be very careful to keep track of the order of the indices. For example, if \( T \) is of type \((2, 0)\) and has components \( T^{ab} \), then \( T^{00} = T^{00}, T^{10} = T^{10}, T^{01} = T^{01}, \) and so on; but \( T^{00} = T^{00}, T^{10} = T^{10}, T^{01} = T^{01} \).

6 *Symmetrization and skew-symmetrization.* Let \( T \) be a covariant tensor of rank \( q \). Then \( T \) is said to be *(totally) symmetric* if its value is unaltered when any two of its arguments are interchanged. That is

\[
T(\ldots, X, \ldots, Y, \ldots) = T(\ldots, Y, \ldots, X, \ldots).
\]

It is said to be *(totally) anti-symmetric* or *skew-symmetric* if it changes sign when any two of its arguments are interchanged. That is

\[
T(\ldots, X, \ldots, Y, \ldots) = -T(\ldots, Y, \ldots, X, \ldots).
\]

There are similar definitions for contravariant tensors. Given any covariant tensor \( T \) of rank \( q \), we can construct a symmetric tensor \( S(T) \) and an anti-symmetric tensor \( A(T) \) by

\[
S(T)(X_1, X_2, \ldots, X_q) = \frac{1}{q!} \sum T(X_{\sigma(1)}, X_{\sigma(2)}, \ldots, X_{\sigma(q)}),
\]

\[
A(T)(X_1, X_2, \ldots, X_q) = \frac{1}{q!} \sum \text{sign}(\sigma) T(X_{\sigma(1)}, X_{\sigma(2)}, \ldots, X_{\sigma(q)}),
\]

where the sums are over \( \sigma \in S_q \) (the group of permutations of \( q \) symbols). The corresponding operations on the components of \( T \) are denoted by round and square brackets. Thus the components of \( S(T) \) are written \( T_{(abc\ldots)} \) and the components of \( A(T) \) are written \( T_{[abc\ldots]} \). For example,

\[
T_{(ab)} = \frac{1}{2}(T_{ab} + T_{ba}), \quad T_{[ab]} = \frac{1}{2}(T_{ab} - T_{ba}),
\]

\[
T_{(abc)} = \frac{1}{6}(T_{abc} + T_{bca} + T_{cab} + T_{acb} + T_{bac} + T_{cba}),
\]

\[
T_{[abc]} = \frac{1}{6}(T_{abc} + T_{bca} + T_{cab} - T_{acb} - T_{bac} - T_{cda}).
\]
We also write, for example,
\[ T_{a[b} = \frac{1}{2} (T_{abc} - T_{acb}). \]

All skew-symmetric tensors of rank four are proportional and all skew-symmetric
tensors of rank greater than four are zero.

**Exercise 5.1** Show that for any 4-vectors \( X \) and \( Y \), \( g(X, Y) = X_a Y^a = X^a Y_a. \)

### 5.5 The Alternating Tensor

The *alternating tensor* is the fourth-rank covariant tensor with components \( e_{abcd} \) is any
inertial coordinate system. Here \( e_{abcd} \) is the alternating symbol, which is defined by
\[ e_{0123} = 1, \quad e_{abcd} = e_{[abcd]} . \]

Thus
\[ e_{abcd} = \begin{cases} 
1, & \text{if } abcd \text{ is an even permutation of 0123;} \\
-1, & \text{if } abcd \text{ is an odd permutation of 0123;} \\
0, & \text{otherwise.}
\end{cases} \]

This defines a tensor since for any proper orthochronous Lorentz transformation \( L^a_b \),
\[ e_{abcd} L^a_f L^b_h L^c_k L^d_l = (\det L) e_{fhkl} = e_{fhkl}. \]
(It is important here that we do not allow Lorentz transformations with \( \det L = -1 \).)

The contravariant alternating tensor is defined by raising all four indices on \( e \). Take care to note that
\[ e^{0123} = g^{0a} g^{1b} g^{2c} g^{3d} e_{abcd} = -e_{0123} = -1. \]

The alternating tensor determines a ‘scalar quadruple product’ \( e(T, X, Y, Z) \) of four vectors \( T, X, Y, Z \). It is closely related to the analogous scalar triple product of vectors in
3-space. In fact, if \( T = (1, 0), X = (X^0, x), Y = (Y^0, y), Z = (Z^0, z) \), then
\[ e(T, X, Y, Z) = e_{abcd} T^a X^b Y^c Z^d = x \cdot (y \wedge z). \]

### 5.6 Tensor Fields

A tensor field of type \((p, q)\) is a map that assigns a tensor \( T \) of type \((p, q)\) to each event. As
in the case of vector fields, we usually require that the components \( T^{ab\ldots de\ldots} \) be smooth functions of the coordinates.

If we make a Lorentz transformation \( x^a = L^a_b x^b \), then
\[ \frac{\partial}{\partial x^b} = M^a_b \frac{\partial}{\partial x^a}, \]
where, as usual, \( M \) is the inverse of \( L \) (\( L^a_b M^b_c = \delta^a_c \)) (§3.10). It follows that the operators
\( \nabla_a = \partial/\partial x^a \) transform like the components of a covector.
The covariant derivative $\nabla T$ of a tensor field $T$ of type $(p, q)$ is the tensor field of type $(p, q + 1)$ with components

$$\nabla_a T^{b...}_{c...d...}.$$ 

In particular, if $f$ is a scalar field (a function on space-time), then $\nabla f$ is a covector field with components $\partial f / \partial x^a$: it is called the gradient covector field of $f$. The 4-gradient is obtained by raising the index: it has components $\nabla^a f = g^{ab} \nabla_b f$.

Note that $\nabla_a g_{bc} = 0$, so that covariant differentiation can be interchanged with raising and lowering of indices. Also $\nabla_a e_{bcde} = 0$.

### 5.7 Example and Exercises

**Example 5.5** For a covariant tensor $T$ of type $(0, 3)$, $\hat{T}$ is defined to be the tensor with components $\hat{T}_{abc} = \hat{T}_{[ac]b}$, where $\hat{T}_{abc} = T_{(ac)b}$. Show that

$$\hat{T}_{abc} = \frac{3}{4} T_{abc}.$$

**Solution.** Since $\hat{T}_{abc} = \frac{1}{2} (T_{acb} + T_{cac})$,

$$\hat{T}_{abc} = \frac{1}{2} (T_{acb} - T_{cab})$$

$$= \frac{1}{4} (T_{abc} + T_{bac} - T_{cbe} - T_{bca}).$$

Therefore

$$\hat{T}_{abc} = \frac{1}{16} (T_{abc} + T_{bac} - T_{cbe} - T_{bca} + T_{bca} - T_{cbe} + T_{abc} + T_{bac} - T_{bca} + T_{cbe} + T_{abc} + T_{bac})$$

$$= \frac{3}{16} (T_{abc} + T_{bac} - T_{cbe} - T_{bca})$$

$$= \frac{3}{4} \hat{T}_{abc}.$$ 

\[ \square \]

**Exercise 5.2** Show that for any $T \in \mathbf{T}_4^0$,

$$T_{[a[bc]d]} = T_{[abcd]}$$

$$T_{(a(bc)d)} = T_{(abcd)}$$

$$T_{(a[bcd]} = 0$$

$$T_{[a(bc]d]} = 0.$$ 

(Note the potential ambiguity in the notation on the left-hand sides: the convention is that the inner brackets go together and the outer ones go together.) Show that in general the operations of symmetrization and skew-symmetrization are idempotent.

**Exercise 5.3** Show that for any tensor $T$,
\[ T = T_{de...}^{ab...} X(a) \otimes X(b) \otimes \cdots \otimes \alpha^{(d)} \otimes \alpha^{(e)} \otimes \cdots. \]

**Exercise 5.4** Let \( T \in \mathfrak{T}_1 \). Show that the mapping \( L : \mathfrak{T}^1 \rightarrow \mathfrak{T}^1 \) determined by
\[ \alpha(L(X)) = T(\alpha, X), \]
\( \alpha \in \mathfrak{T}_1 \) and \( X \in \mathfrak{T}^1 \), is linear. Show that its trace is the contraction of \( T \).

**Exercise 5.5** Let \( X \) be a vector field and let \( \nabla X \) be its covariant derivative; \( \nabla X \) is a tensor field of type \((1, 1)\). Show that
\[ (\nabla X)(\alpha, Y) = \alpha(\nabla_Y X), \]
where \( \alpha \) is a covector field and \( Y \) is a vector field.

**Exercise 5.6** Let \( X \) be a vector field. For any tensor field \( T \), let \( \nabla_X T \) denote the tensor field of the same type with components \( X^a \nabla_a T^{bc...} \). Show that
\[ \nabla_X (T \otimes T') = (\nabla_X T) \otimes T' + T \otimes (\nabla_X T') \]
for any tensor fields \( T \) and \( T' \).

**Exercise 5.7** Show that
\[
\begin{align*}
(i) \quad & e_{abcd}e^{abcd} = -24 \\
(ii) \quad & e_{abcd}e^{abcdef} = -6\delta^f_i \\
(iii) \quad & e_{abcd}e^{abfh} = -4\delta^f_i \delta^h_c \\
(iv) \quad & e_{abcd}e^{afhk} = -6\delta^f_i \delta^h_c \delta^k_d \\
v) \quad & e_{abcd}e^{fhkl} = -24\delta^f_i \delta^h_c \delta^k_d \delta^l_a.
\end{align*}
\]

**Exercise 5.8** Show that if \( T \in \mathfrak{T}_0^4 \) is totally skew-symmetric, then
\[ T^{abcd} = -\frac{1}{24} T^{fhkl} e_{fhk} e^{abcd}. \]
(Hint: first convince yourself that all fourth-rank skew-symmetric tensors are proportional.)

**Exercise 5.9** Show that for any \( T \in \mathfrak{T}_2^0 \),
\[ T_{ab} = T_{(ab)} + T_{[ab]}. \]
Show that if \( T_{ab} X^a X^b = 0 \) for every 4-vector \( X \), then \( T_{ab} = T_{[ab]} \). Show that if \( T_{ab} X^a X^b = 0 \) for every null 4-vector \( X \), then \( T_{ab} = T_{[ab]} + \frac{1}{4} T_c T_c g_{ab} \).

**Exercise 5.10** Consider the cube in §2.6. Let \( U \) be the 4-velocity of an arbitrary observer. Let \( A_i (i = 1, 2, 3, 4) \) be any four events such that \( A_i \in W_i \) and such that the \( A_i \)'s are simultaneous relative to the observer. Show that the observer measures the volume to be \( c^{-1} \epsilon(U, X, Y, Z) \) where \( X, Y, Z \) are the displacement 4-vectors from \( A_1 \) to \( A_2, A_3, \) and
A_4$, respectively. Show that the volume measured in the rest frame is $c^{-1} \varepsilon(V, X, Y, Z)$, where $V$ is the 4-velocity of the cube. Hence show that the observed volume is $D^3/\gamma(u)$, where $u$ is the speed of the observer relative to the cube. (Hint: the scalar quadruple product of four linearly dependent 4-vectors vanishes.)
6 Electrodynamics

6.1 Tensor Formulation of Maxwell’s Equations

Although Maxwell’s equations do not need relativistic correction, we have yet to write them in a form in which their invariance under Lorentz transformations is explicit.

In 3-vector notation, Maxwell’s equations relate the electric and magnetic fields $E$ and $B$ to the charge density $\rho$ and the current density $j$. They are

\[
\text{curl } E = -\frac{\partial B}{\partial t} \tag{6.1}
\]

\[
\text{curl } B = \frac{1}{c^2 \varepsilon_0} j + \frac{1}{c^2} \frac{\partial E}{\partial t} \tag{6.2}
\]

\[
\text{div } B = 0 \tag{6.3}
\]

\[
\text{div } E = \frac{\rho}{\varepsilon_0}. \tag{6.4}
\]

The first task is to write down the transformation rules for $\rho$ and $j$; or, from a less observer-dependent point of view, to identify $\rho$ and $j$ with the components of a tensorial object in space-time.

It follows from Maxwell’s equations that $\rho$ and $j$ satisfy the continuity equation

\[
\text{div } j + \frac{\partial \rho}{\partial t} = 0.
\]

That is, $\partial J^a/\partial x^a = 0$, where $(J^a) = (\rho c, j_1, j_2, j_3)$. The simplest way to ensure that this holds in every inertial frame is to assume that the $J^a$s are the components of a 4-vector (called the 4-current). Then the continuity equation becomes $\nabla_a J^a = 0$.

The physical content of the assumption is that the charge of a particle is an invariant. To see this, suppose that $J$ is a 4-vector. Then, by assumption, the charge density seen by an observer with 4-velocity $V$ is $\rho = g(J, V)/c^2$, since in the observer’s rest frame, $V = (c, 0)$ and $J = (\rho c, j)$. If the current is due to a stream of charged particles with 4-velocity $U$, then $J = enU$, where $e$ and $n$ are the charge per particle and the number of particles per unit volume, both measured in the instantaneous rest frame of the particles.

The observer reckons that there are $n g(U, V)/c^2$ particles per unit volume (see Exercise 3.13) and that the charge density is $g(V, J)/c^2 = ne g(U, V)/c^2$. In other words, the observer also reckons that each particle has charge $e$. The overall neutrality of matter is strong evidence for the invariance of charge: if it were not invariant, then an atom would carry a net charge as the result of the motion of its electrons.

It follows from the first and third of Maxwell’s equations that there is a scalar $\phi$ and 3-vector $A$ such that

\[
B = \text{curl } A \quad \text{and} \quad E = -\frac{\partial A}{\partial t} - \text{grad } \phi.
\]
The potentials $A$ and $\phi$ are not uniquely determined by $B$ and $E$: we can make gauge transformations $A \mapsto A + \text{grad } f$, $\phi \mapsto \phi - \partial f / \partial t$, where $f$ is any function of position and time. We can use this freedom to impose the Lorenz gauge condition\(^\text{12}\)

$$c^2 \text{div } A + \frac{\partial \phi}{\partial t} = 0.$$ 

With $\phi^0 = \phi$, $\phi^1 = cA_1$, $\phi^2 = cA_2$, $\phi^3 = cA_3$, the gauge condition and Maxwell’s equations become

$$\frac{1}{c} \frac{\partial \phi^0}{\partial t} + \frac{\partial \phi^1}{\partial x} + \frac{\partial \phi^2}{\partial y} + \frac{\partial \phi^3}{\partial z} = 0 \quad (6.5)$$

$$\left( \frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial z^2} \right) \phi^a = \frac{1}{\epsilon_0 c} J^a. \quad (6.6)$$

Let us now assume that the $\phi^a$s transform as the components of a 4-vector $\Phi$ (we shall justify this later). Then (6.5) and (6.6) take the explicitly invariant form

$$\nabla_a \phi^a = 0 \quad \text{and} \quad \nabla_b \nabla^b \phi^a = \frac{1}{\epsilon_0 c} J^a. \quad (6.7)$$

The vector field $\Phi$ is called the 4-potential and the operator

$$\nabla_a \nabla^a = \frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial z^2}$$

is the wave operator ($\S 3.10$).

Sometimes it is more convenient not to impose the gauge condition, and to use a general pair of potentials $\phi$ and $A$. Then, with $\Phi = (\phi, cA)$, Maxwell’s equations are equivalent to

$$\nabla_b \nabla^b \phi^a - \nabla^a (\nabla_b \phi^b) = \frac{1}{\epsilon_0 c} J^a. \quad (6.8)$$

We define the electromagnetic field tensor to be the second rank covariant tensor $F$ with components $F_{ab} = 2\nabla_{[a} \phi_{b]}$. Explicitly,

$$\begin{pmatrix}
0 & E_1 & E_2 & E_3 \\
-E_1 & 0 & -cB_3 & cB_2 \\
-E_2 & cB_3 & 0 & -cB_1 \\
-E_3 & -cB_2 & cB_1 & 0
\end{pmatrix}.$$

In terms of $F$, Maxwell’s equations are

$$\nabla_a F^{ab} = \frac{1}{\epsilon_0 c} J^b \quad (6.9)$$

$$\nabla_{[a} F_{bc]} = 0. \quad (6.10)$$

The first is equivalent to (6.2) and (6.4); the second to (6.1) and (6.3). For example, by taking $b = 1$ in (6.9), one obtains the first component of (6.2). By taking $a = 1$, $b = 2$, $c = 3$ in (6.10), one obtains (6.3).

---

\(^{12}\)After L. Lorenz, not H. A. Lorentz.
The requirement that the $\Phi^a$s should transform as the components of a 4-vector fixes the transformation law for $E$ and $B$. The tensorial form of (6.9) and (6.10) then ensures that Maxwell's equations take the same form in all inertial frames.\footnote{One can also treat in a similar way the general form of Maxwell's equations, with arbitrary $\mu$ and $\epsilon$. The transformations rules for $E$, $B$, $D$, and $H$ are perhaps less fundamental from a theoretical point of view, but are of considerable importance as far experimental predictions are concerned.}

**Exercise 6.1** Show that if $x^a = L^a_b x^b$ is the standard Lorentz transformation (§2.4), then the components of $E$ and $B$ in the two coordinate systems are related by

\[
E_1 = \tilde{E}_1 \\
E_2 = \gamma(v)(\tilde{E}_2 + v\tilde{B}_3) \\
E_3 = \gamma(\tilde{E}_3 - v\tilde{B}_2) \\
B_1 = \tilde{B}_1 \\
B_2 = \gamma(v)(\tilde{B}_2 - vc^{-2}\tilde{E}_3) \\
B_3 = \gamma(v)(\tilde{B}_3 + vc^{-2}\tilde{E}_2).
\]

### 6.2 Gauge Transformations

The 4-potential is not determined uniquely by $F_{ab}$ since $\nabla_a \nabla_b u = 0$ for any function $u$ on space-time. Therefore $\Phi^a$ and $\Phi^a + \nabla^a u$ generate the same field tensor $F_{ab}$. The gauge transformation $\Phi^a \mapsto \tilde{\Phi}^a = \Phi^a + \nabla^a u$ is equivalent to a gauge transformation of $\phi$ and $A$, with $u = -cf$. If $\Phi^a$ satisfies the Lorenz condition, then so does $\tilde{\Phi}^a$ provided that $u$ is a solution of the wave equation $\nabla_a \nabla^a u = 0$.

### 6.3 The Dual Electromagnetic Field Tensor

Let $F^*$ denote the tensor with components

\[
F^*_{ab} = \frac{1}{2} \varepsilon_{abcd} F^{cd}.
\]

That is,

\[
(F^*_{ab}) = \begin{pmatrix}
0 & -cB_1 & -cB_2 & -cB_3 \\
cB_1 & 0 & -E_3 & E_2 \\
cB_2 & E_3 & 0 & -E_1 \\
cB_3 & -E_2 & E_1 & 0
\end{pmatrix}.
\]

The tensor $F^*$ is called the dual of the electromagnetic field. When $J = 0$, we can use the dual tensor to write Maxwell's equations in the symmetric form

\[
\nabla_a F_{bc} = 0, \quad \nabla_a F^*_{bc} = 0.
\]
The contractions \( F_{ab}F^{*ab} \) and \( F_{ab}F^{ab} \) are invariant under Lorentz transformations. They are given by
\[
F_{ab}F^{*ab} = 4cE \cdot B \quad \text{and} \quad F_{ab}F^{ab} = 2(c^2 B \cdot B - E \cdot E).
\]
If \( E \cdot B = 0 \) and \( c^2 B \cdot B - E \cdot E > 0 \) (or \( < 0 \)), then \( F \) is purely magnetic (purely electric) in the sense that there exists an inertial frame in which \( E = 0 \) (or \( B = 0 \)); see §6.9.

6.4 Advanced and Retarded Solutions

The 4-potential is not uniquely determined by the 4-current. Given one solution \( \Phi \) of (6.7), we are free to add \( K \), where \( K \) is any 4-vector field satisfying the source-free equations
\[
\nabla_a K^a = 0, \quad \nabla_b \nabla^b K^a = 0.
\]
If \( K^a = \nabla^a u \), where \( u \) is a solution of the wave equation, then this is nothing more than a gauge transformation. In general, however, \( \Phi \) and \( \Phi + K \) are potentials for physically distinct electromagnetic fields with the same source.

Suppose, for example, that the source consists of a collection of charges in some bounded region of space, and that the charges are at rest relative to some inertial coordinate system \( ct, x, y, z \) except between times \( t_1 \) and \( t_2 \). We know what the corresponding electromagnetic field should look like. For \( t < t_1 \), we expect to see only the electrostatic field of the charges. In this part of space-time, we can take \( \Phi = (\phi, 0) \), where \( \phi \) is independent of \( t \) and
\[
\phi = O(r^{-1}), \quad \text{grad} \phi = O(r^{-2}) \tag{6.11}
\]
as \( r \to \infty \), where \( r \) is the spatial distance from the origin. When the charges move, they radiate electromagnetic waves, which travel outwards from the source along future-directed null lines (the worldlines of photons). After the radiation has dispersed, the field settles down again to the electrostatic solution.

It follows that as we go out to infinity in past null directions,
\[
\Phi_a = O(r^{-1}), \quad \text{and} \quad F_{ab} = O(r^{-2}) \tag{6.12}
\]
since, if we go out to infinity along a past-directed null line, then we must eventually enter the region of space-time in which \( \Phi = (\phi, 0) \) and (6.11) holds. If we go to infinity in a future null direction, then we may stay in the radiation region, and (6.12) need not hold (see Fig. 6.1).

There is nothing in Maxwell's equations, however, that singles out a particular arrow of time. There will therefore be an equally good solution in which the space-time diagram is the other way up and the radiation emitted by the moving charges travels backwards in time; or, from another point of view, radiation focuses onto the source with exactly the right phase and frequency to interfere destructively with the radiation emitted by the charges and to leave a static field after time \( t_2 \). In this solution \( \Phi \) falls off like (6.12) in future null directions.

This suggests that we should interpret the fall-off condition (6.12) in past null directions as the condition that there is no incoming radiation (NIR). This can, in fact, be justified under more realistic conditions.
Proposition 6.1 Suppose that \( J \) is the 4-current of a spatially bounded source and that \( \Phi \) satisfies (6.7) and the NIR condition. Then

\[
\Phi^a(ct', r') = \frac{1}{4\pi \epsilon_0 c} \int J^a(ct' - R, r) \frac{dx dy dz}{R},
\]

where \( r = (x, y, z) \), \( R = \sqrt{(r - r') \cdot (r - r')} \), and the integral is over all space.

Proof. For simplicity, take \( (ct', r') = 0 \), so that \( R = r = |r| \). We shall take \( a = 0 \) (the other components are dealt with in the same way).

Put \( \phi = \Phi^0 \) and \( f = J^0 / \epsilon_0 c \). Define \( \psi \) and \( \chi \) as functions of \( r \) by putting \( ct = -r \) in \( \phi(ct, r) \) and its time derivative. That is

\[
\psi(r) = \phi(-r, r) \quad \text{and} \quad \chi(r) = \phi_t(-r, r),
\]

where the subscript denotes the partial derivative with respect to \( t \). Then

\[
\frac{\partial \psi}{\partial x} = \frac{\partial \phi}{\partial x} - \frac{x}{cr} \frac{\partial \phi}{\partial t},
\]

\[
\frac{\partial^2 \psi}{\partial x^2} = \frac{\partial^2 \phi}{\partial x^2} - \frac{2x}{cr} \frac{\partial^2 \phi}{\partial x \partial t} + \frac{x^2}{c^2 r^2 \partial t^2} - \frac{1}{cr} \frac{\partial \phi}{\partial t} + \frac{x^2}{cr^3} \frac{\partial \phi}{\partial t},
\]

\[
\text{div} \left( \frac{\chi}{r^2} \frac{r}{r} \right) = \frac{1}{r^2} \frac{\partial \phi}{\partial t} + \frac{1}{r^2} r : \text{grad} \left( \frac{\partial \phi}{\partial t} \right) - \frac{1}{cr} \frac{\partial^2 \phi}{\partial t^2},
\]
where the right-hand sides are evaluated with \( ct = -r \). Therefore
\[
\left[ \frac{1}{r} \nabla^2 \psi + \frac{2}{c} \text{div} \left( \frac{\chi}{r^2} r \right) \right]_r = -\frac{1}{r} \left( \frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2} - \nabla^2 \phi \right) = -\frac{f(-r, r)}{r},
\]
by using (6.7).

Let \( S_1 \) be a small sphere of radius \( r_1 \) with its centre at the origin \( r = 0 \) and let \( S_2 \) be a large concentric sphere of radius \( r_2 \). By applying the divergence theorem to the volume \( V \) between \( S_1 \) and \( S_2 \), and letting \( r_1 \to 0 \), we find
\[
4\pi \psi(0) = \int_V f(-r, r) \frac{dxdydz}{r} + \int_{S_2} \left[ 2\chi c^{-1} r^2 - 2r + r^{-1} \text{grad} \psi - \psi \text{grad} (r^{-1}) \right] \cdot dS.
\]
The surface integral on the right-hand side goes to zero as \( r_2 \to \infty \) because of the NIR condition (as \( r \to \infty \), the event \( (ct, r) = (-r, r) \) goes to infinity in a past null direction). The proposition follows.

It can also be shown, by a long but straightforward calculation, that if \( J \) satisfies certain plausible conditions, and if \( \Phi \) is defined by (6.13), then \( \Phi \) satisfies (6.7), together with the NIR condition. By a shorter calculation (exercise) one can also show that the conservation law \( \nabla_a J^a = 0 \) implies that \( \nabla_a \Phi^a = 0 \). Finally, it is easy to see that if the components of a 4-vector \( \Phi \) satisfy the NIR condition in one inertial frame, then they satisfy it in every inertial frame. Thus Maxwell's equations, supplemented by the NIR condition, imply that the 4-potential is given uniquely by (6.13).

The integral in (6.13) can be put into a Lorentz-invariant form. Let \( A' \) be the event \((ct', r')\) and let \( A \) be the event \((ct - R, r)\). Then \( X = (-R, r - r') \) is the displacement 4-vector from \( A' \) to \( A \), and \( g(X, X) = R^2 - R^2 = 0 \). Thus \( X \) is null, and, since \( X^0 < 0 \), \( A \) lies on the past light-cone \( N_- \) of \( A' \). Therefore, (6.13) can be written as an integral over \( N_- \):
\[
\Phi^a(A') = \frac{1}{4\pi\varepsilon_0 c} \int_{A \in N_-} J^a(A) \frac{dX^1dX^2dX^3}{|X^0|}.
\]
Here we think of \( X^1, X^2, X^3 \) as coordinates on \( N_- \), with \( X^0 \) defined as a function of \( X^1, X^2, X^3 \) by \( X^0 = -\sqrt{(X^1)^2 + (X^2)^2 + (X^3)^2} \). We have the following proposition (an alternative proof is outlined in Exercise 6.7).

**Proposition 6.2**  The volume element \( dN = dX^1dX^2dX^3/|X^0| \) on \( N_- \) is invariant under Lorentz transformations.

**Proof.** Let \( A \) be the event on \( N_- \) with displacement 4-vector \( X \) relative to \( A' \) and let \( B, C, D \) be nearby events in \( N_- \) with displacement 4-vectors \( \epsilon U, \epsilon V, \epsilon W \) relative to \( A \), where \( \epsilon \) is small. If we ignore terms of order \( \epsilon^2 \), then
\[
g(X, U) = g(X, V) = g(X, W) = 0,
\]
since \( X, X + \epsilon U, X + \epsilon V, \) and \( X + \epsilon W \) are null. The four events \( A, B, C, \) and \( D \) are the neighbouring vertices of a small parallelepiped in \( N_- \). With respect to \( dN \), this has volume

\[ \Lambda = \frac{e^3 u \cdot (v \wedge w)}{|X^0|}, \]

where \(U = (U^0, u), V = (V^0, v), W = (W^0, w)\). We have to show that \(\Lambda\) is independent of the choice of inertial frame. (We assume that \(u \cdot (v \wedge w) > 0\), so that \(\Lambda \neq 0\).)

Consider the 4-vector with components \(e^a_{bcd} U^b V^c W^d\). This is orthogonal to \(U, V\) and \(W\). Consequently, it must be proportional to \(X\), since the space of 4-vectors orthogonal to \(U, V\), and \(W\) is one-dimensional and contains \(X\). Therefore, there exists a scalar \(\lambda\), which is independent of the choice of frame, such that

\[ e^a_{bcd} U^b V^c W^d + \lambda X^a = 0. \]

By taking \(a = 0\),

\[ \lambda = \frac{u \cdot (v \wedge w)}{|X^0|}. \]

Hence \(\Lambda = e^3 \lambda\). It follows that \(\Lambda\) is invariant.

We can now write (6.13) in the invariant form

\[ \Phi^a(A') = \frac{1}{4\pi \epsilon_0 c} \int_{N_-} J^a \, dN. \]  \hspace{1cm} (6.14)

This is called the \textit{retarded solution} of Maxwell's equations. If we drop the NIR condition, then we have an equally good solution,

\[ \Phi^a(A') = \frac{1}{4\pi \epsilon_0 c} \int_{N_+} J^a \, dN, \]

where \(N_+\) is the future half of the light-cone of \(A'\). It is called the \textit{advanced} solution.

The retarded solution corresponds to the first space-time diagram. The potential at an event \(A'\) is determined by the behaviour of the source on the past light-cone of \(A'\): an observer at \(A'\) 'sees' the 4-current at events on his past light-cone and the radiation emitted by the source propagates forwards in time along future-directed null lines. The advanced solution corresponds to the second diagram: here the radiation propagates backwards in time or, in the alternative interpretation, there is incoming radiation focusing in on the source.

Remark One can use these results to justify the transformation rule for \(\Phi\). If we assume that a solution of Maxwell's equations given by a retarded potential in one frame should be given by a retarded potential in every frame, then it follows that if \(\Phi\) is given by (6.14) in one frame, then it is given by (6.14) in every frame. Since the \(J^a\)'s are the components of a 4-vector, this implies that the \(\Phi^a\)'s must be as well. Thus the transformation law for \(\Phi\) is a consequence of causality: it follows from the assumption that if radiation propagates forwards in time in one frame, then it must do so in every frame.
6.5 The Field of a Uniformly Moving Charge

Consider the electromagnetic field of a uniformly moving particle with charge $e$ and 4-velocity $V$. In the rest frame of the charge, the magnetic field vanishes and the electric field is the electrostatic field

$$E = \frac{er}{4\pi\varepsilon_0 D^3},$$

where $r$ is the position 3-vector from the charge and $D^2 = r \cdot r$. Let $X$ be the displacement 4-vector from an origin on the worldline of the charge to an event at which the field is measured. In the rest frame of the charge, $X = (ct, r)$ and $V = (c, 0)$. Therefore

$$D^2 = c^{-2}(V_a X^a)^2 - X_a X^a$$ (6.15)

and

$$F^{[ab]} = -\frac{2eV^a X^b}{4\pi\varepsilon_0 c D^3}.$$ (6.16)

(For example, with $a = 0$, $b = 1$, this is $F^{01} = -E_1 = -ex/4\pi\varepsilon_0 D^3$, where $x$ is the first component of $r$.) If we use (6.15) to define $D$ in terms of $V$ and $X$, then (6.16) is an equation relating the components of two tensors. Since it is valid in one frame (the rest frame) it must be valid in every frame. Note that $D$ and $F^{ab}$ are unchanged by adding a multiple of $V$ to $X$ (Fig. 6.3).

In a general frame, $V = \gamma(v)(c, v)$, where $v$ is the velocity of the charge. We can use the freedom in $X$ to set its temporal component in the frame to zero. Then $X = (0, R)$ where $R$ is a 3-vector from the charge at some time to a simultaneous event (in the frame) at which the field is to be measured (Fig. 6.4). By evaluating the components of $F$, the electric and magnetic fields in the general frame are

$$E = kR \quad \text{and} \quad B = kc^{-2}v \wedge R,$$

where $k = e\gamma(v)/4\pi\varepsilon_0 D^3$.

By evaluating the right-hand side of (6.15) in the general frame

$$D^2 = c^{-2}\gamma(v)^2(v \cdot R)^2 + R \cdot R = \frac{R^2(c^2 - v^2 \sin^2 \theta)}{c^2 - v^2},$$

where $\theta$ is the angle between $v$ and $R$. Hence, by evaluating the right-hand side of (6.16), the electric and magnetic fields in the general frame are given by

$$E = \frac{e\gamma(v)}{4\pi\varepsilon_0} \left(\frac{c^2 - v^2}{c^2 - v^2 \sin^2 \theta}\right)^{3/2} \frac{R}{R^3},$$ (6.17)

and $c^2 B = v \wedge E$.

When $v << c$, the expression in large brackets in (6.17) is close to unity, $E$ is the usual electrostatic field, and $B$ is given by the nonrelativistic formula for the field of a moving charge. The field of a fast charge is reduced by a factor $1/\gamma(v)^2$ on $\theta = 0$ (the trajectory of the charge), and enhanced by a factor $\gamma(v)$ on $\theta = \pi/2$, the plane through the charge orthogonal to its velocity. In the limit $v \to c$, the field is entirely concentrated on this plane.
Fig. 6.3. The worldline of the charge and the displacement vector X (space-time diagram)

Fig. 6.4. The calculation of the field in a general frame (space diagram)

6.6 The Equation of Motion of a Test Charge

Consider a test charge \( e \) of rest mass \( m \) moving through an electromagnetic field \( F \) (by a ‘test charge’ we mean one whose own field be neglected). In its instantaneous rest frame, the particle is subject to the electrostatic force \( eE \) and has acceleration \( \mathbf{a} \), where \( ma = eE \), by Newton’s second law. In this frame, the 4-velocity and the 4-acceleration are \( V = (c, 0) \) and \( \dot{V} = (0, \mathbf{a}) \). If we define \( G \) to be the 4-vector with components

\[
G^a = \frac{e}{c} F^{ab} V_b,
\]

then \( G = (0, eE) \) in the rest frame. It follows that in the instantaneous rest frame, \( m\dot{V} = G \). This equates two 4-vectors. Since it holds in one frame, it must hold in every frame. Therefore, if the rest mass is constant, the equation of motion of the charge is

\[
\frac{dP^a}{d\tau} = \frac{e}{c} F^{ab} V_b
\]

in any frame, where \( P = mV \) is the 4-momentum.

Written out in a general frame, this is

\[
\frac{d}{dt} \left( m\gamma c^2 \right) = e\mathbf{E} \cdot \mathbf{v}, \quad \frac{d}{dt} \left( m\gamma v \right) = e(\mathbf{E} + \mathbf{v} \wedge \mathbf{B}),
\]

(6.18)

where \( \gamma = \gamma(v) = dt/d\tau \). The second equation is the Lorentz force law. The first, which follows from the second, gives the time derivative of the energy.
6.7 Plane Waves

When there is no source \((J = 0)\), the 4-potential in the Lorenz gauge satisfies

\[
\nabla_a \nabla^a \Phi^b = 0 \quad \text{and} \quad \nabla_a \Phi^a = 0.
\]

The four components of \(\Phi\) satisfy the wave equation.

A *monochromatic plane wave* is a solution of the form

\[
\Phi^a = \text{Re} \left( P^a e^{iK_a x^b} \right),
\]

where \(P\) is a constant complex 4-vector and \(K\) is a constant real 4-vector. Since

\[
\nabla_a \nabla^a \Phi^b = -\text{Re} \left( K_a K^a P^b e^{iK_c x^c} \right) \quad \text{and} \quad \nabla_a \Phi^a = \text{Re} \left( iK_a P^a e^{iK_c x^c} \right),
\]

\(\Phi\) is a potential for a source-free electromagnetic field if \(K_a K^a = 0\) and \(P_a K^a = 0\). The null vector \(K\) is called the *frequency 4-vector*, by analogy with the scalar case (§3.11).

The field tensor is

\[
F^{ab} = \text{Re} \left( 2iK^{[a} P^{b]} e^{iK_c x^c} \right).
\]

Neither the value of \(F\) nor the Lorenz gauge condition are upset by replacing \(P\) by \(P + zK\) for some constant \(z \in \mathbb{C}\). We can use this freedom to set \(P^0 = 0\). Then \(P = (0, \mathbf{p})\) for some complex 3-vector \(\mathbf{p}\). If we write \(K = c^{-1} \omega(\mathbf{1}, \mathbf{e})\), where \(\mathbf{e} \cdot \mathbf{e} = 1\), then

\[
\mathbf{E} = -\text{Re} \left( iK^0 e^{iK_c x^c} \mathbf{p} \right) = -\text{Re} \left( ic^{-1} \omega \mathbf{p} \exp(i\omega(t - \mathbf{e} \cdot \mathbf{r}/c)) \right)
\]

and \(c \mathbf{B} = \mathbf{e} \wedge \mathbf{E}\). The condition \(K_a P^a = 0\) implies that \(\mathbf{e} \cdot \mathbf{p} = 0\). Therefore

1. the electric and magnetic fields are plane waves with angular frequency \(\omega\) travelling in the direction of the unit vector \(\mathbf{e}\);
2. the electric field is at all times orthogonal to the magnetic field and both are orthogonal to the direction of propagation.

To understand the significance of \(\mathbf{p}\), consider the behaviour of \(\mathbf{E}\) at a fixed point in space. In general, \(\mathbf{E}\) traces out an ellipse in the plane orthogonal to the direction of propagation of the wave. The parameters of the ellipse are determined by \(\mathbf{p}\). Two special cases are

1. when \(\mathbf{p}\) is real, the wave is said to be *linearly polarized*: in this case, \(\mathbf{E}\) oscillates along a fixed line;
2. when \(\mathbf{p} = \mathbf{x} + iy\), where \(\mathbf{x}\) and \(\mathbf{y}\) are orthogonal real vectors of the same length, the wave is said to be *circularly polarized*: in this case, \(\mathbf{E}\) traces out a circle in the plane orthogonal to the direction of propagation.
6.8 Geometric Optics

We now consider how Maxwell's equations describe the propagation of photons through space-time. We shall look for approximate solutions such that

\[ J^a = 0 \quad \text{and} \quad \Phi^a = \Re \left( P^a e^{i\omega / \lambda} \right) \]

where \( \lambda \) is a small parameter, and \( P^a \) and \( \omega \) depend on the space-time coordinates. As \( \lambda \to 0 \), the oscillations in \( \Phi \) become more and more rapid. We shall show that in this high frequency limit, we can find approximate solutions which are concentrated around single null lines in space-time, and which can be regarded as representing 'particles' of light.

We substitute the complex potential \( P^a e^{i\omega / \lambda} \) into the left-hand side of the Lorenz gauge condition. The result is

\[ (i\lambda^{-1} P^a \nabla_a \omega + \nabla_a P^a) e^{i\omega / \lambda} . \]

The dominant term here is the one in \( \lambda^{-1} \), and so we can impose an approximate form of the Lorenz condition by requiring that \( P^a K_a = 0 \), where \( K_a = \lambda^{-1} \nabla_a \omega \). We now substitute the complex potential into the left-hand side of (6.8) (we have to use this form of Maxwell's equations because the Lorenz gauge condition does not hold exactly). If we keep only the terms in \( \lambda^{-2} \) and \( \lambda^{-1} \), then the result is

\[ (-\lambda^{-2} P^a (\nabla_b \omega) \nabla_b \omega + 2i\lambda^{-1} (\nabla_b \omega) \nabla_b P^a + i\lambda^{-1} P^a \nabla_b \omega - i\lambda^{-1} (\nabla^a \omega) \nabla_b P^b) e^{i\omega / \lambda} . \]

By equating to zero the terms in \( \lambda^{-1} \) and \( \lambda^{-2} \),

\[ K_a K^a = 0, \quad 2P^b \nabla_a P^a + P^a \nabla_b K^b - K^a \nabla_b P^b = 0, \quad (6.19) \]

This is the geometric optics approximation: if we solve these for \( \omega \) and \( P \), with \( K_a P^a = 0 \), then we obtain an electromagnetic field that, near each event, looks like a high-frequency plane wave with frequency 4-vector \( K \), but with the amplitude and phase varying slowly from event to event.

The first step in constructing such a field is to solve the first-order partial differential equation \( (\nabla_a \omega) \nabla^a \omega = 0 \) (some particular solutions are given in Exercise 6.2). Then, with \( K_a = \lambda^{-1} \nabla_a \omega \),

\[ \nabla_a K_b - \nabla_b K_a = \lambda^{-1} \left( \frac{\partial^2 \omega}{\partial x^a \partial x^b} - \frac{\partial^2 \omega}{\partial x^b \partial x^a} \right) = 0. \]

Also from (6.19), \( K^a \nabla_b K_a = \frac{1}{2} \nabla_b (K_a K^a) = 0 \). Therefore

\[ K^a \nabla_a K^b = 0. \]

(6.20)

The rays are defined to be the curves \( x^a = x^a(s) \) in space-time such that

\[ \frac{dx^a}{ds} = K^a. \]

It follows from (6.20) that \( d^2 x^a / ds^2 = 0 \) and hence that the rays are straight lines. Since they are also null \( (K^a K_a = 0) \), they are photon worldlines.
We then turn to the second equation in (6.19), which is called the transport equation. It determines the behaviour of \( P \) along the rays as follows. We can write \( P = fK + Q \), where \( Q = (0, q) \) has zero temporal part and \( f \) is a complex function of the space-time coordinates. Then, since \( K^a \nabla_a = d/ds \) and

\[
\nabla_a P^a = f \Theta + \frac{df}{ds} + \text{div} \ q, \quad \text{where} \quad \Theta = \nabla_a K^a,
\]

the transport equation is equivalent to

\[
\frac{dq}{ds} = -\frac{1}{2} \Theta q, \quad \frac{df}{ds} = \text{div} \ q
\]

(exercise). We can specify the components of \( P \) as arbitrary complex functions of \( x, y, z \) at \( t = 0 \), subject to the constraint \( P^a K_a = 0 \). The first equation then determines \( q \) along the rays, and the second equation determines \( f \).

The idea is to represent a photon by a high-frequency solution of Maxwell’s equations, which is localized, as near as is possible, on a single null line in space-time: such a solution would naturally represent a single particle of light, with the null line as its worldline. It would be tempting to try to do this by taking delta-functions for the initial values of the \( P^a \) s: the initial values would then be localized at a single point \( (x, y, z) \) and the solution itself would be localized on the ray through this point. However, the approximation we have made only makes sense if the derivatives of \( P \) are much smaller than \( \lambda^{-1} P \). This condition is violated if we try to localize the nonzero initial values of \( P \) in too small a region. In fact we see that \( P \) must be nonzero over a spatial region several times larger than the wavelength \( 2\pi/K^0 \) (Fig. 6.5). For ordinary light, with a wavelength of a few times \( 10^{-7} \) m, we can take the \( P^a \) s to be zero outside a region \( 10^{-5} \) m across. Provided that we only do measurements on a scale much larger than this, it is legitimate to think of the corresponding high frequency solution as representing a point particle of light. On smaller scales, however, the particle is fuzzy; it fills out a tube in space-time and cannot be localized to a single null line.

To combine the particle and wave pictures in a more satisfactory way, one must turn to the quantum theory of the electromagnetic field.

**Exercise 6.2** Show that the following are solutions of the equation \((\nabla_a w)\nabla^a w = 0\).

(i) \( w = ct - x \),

(ii) \( w = ct - \sqrt{x^2 + y^2 + z^2} \),

(iii) \( w = \sqrt{c^2 t^2 - x^2} - \sqrt{y^2 + z^2} \),

(iv) \( w = \sqrt{c^2 t^2 - x^2} - y^2 - z \).

In each case, describe the rays.
Fig. 6.5. The geometric optics approximation

6.9 Examples and Exercises

Example 6.1 Pure Electric and Magnetic Fields. Let $F$ be a constant electromagnetic field. Find the conditions on $F$ for there to exist a frame in which the electric or magnetic field vanishes.

Solution. An observer with 4-velocity $V$ sees no electric field if $F^{ab}V_b = 0$ (to prove this, write out the components of $F^{ab}V_b$ in the observer's rest frame). In a general frame, $V = \gamma(v)(c, v)$, and this condition is

$$E \cdot v = 0 \quad \text{and} \quad E + v \wedge B = 0. \quad (6.21)$$

Equations (6.21) have no solution if $E \cdot B \neq 0$. When $E \cdot B = 0$, the general solution is

$$v = B^{-2}(E \wedge B + \lambda B), \quad (6.22)$$

where $B^2 = B \cdot B$ and $\lambda \in \mathbb{R}$ is arbitrary.

If $v$ is to represent a physical 3-velocity, then we must have $v \cdot v < c^2$. That is, $E^2 < c^2B^2$ and $\lambda^2 < c^2B^2 - E^2$, where $E^2 = E \cdot E$. It follows that there are no observers for whom the electric field vanishes unless $E \cdot B = 0$ and $E^2 < c^2B^2$. When these conditions hold, the electric field vanishes for all observers with velocity $v$ given by (6.22), with $\lambda^2 < c^2B^2 - E^2$.

When $E \cdot B = 0$ and $E^2 - c^2B^2 > 0$, the same argument applied to the dual tensor $F^*$ gives the frames in which the magnetic field vanishes.

Example 6.2 Reflection from a moving mirror. An electromagnetic plane wave of frequency $\omega$ is travelling in the direction of the unit vector $e$ relative to an inertial frame.
The wave is reflected at a plane mirror moving with velocity \( \mathbf{v} \). The normal to the mirror is in the direction of the unit vector \( \mathbf{k} \). Find the frequency \( \omega' \) and direction \( \mathbf{e}' \) of the reflected wave.

**Solution.** Let the frequency 4-vectors of the incident and reflected waves be \( L \) and \( L' \). In the given frame, \( cL = \omega(1, e) \) and \( cL' = \omega'(1, e') \); and the 4-velocity of the mirror is \( V = \gamma(v)(c, v) \).

Let \( x^a \) be an inertial coordinate system in which the mirror is at rest in the \( x^2, x^3 \) plane. In these coordinates, the frequencies of the incident and reflected waves are the same, and the angle of incidence is equal to the angle of reflection. Therefore, if \( L \) has components \( L^a \) in this rest frame, then \( L' \) has components \( L'^a \), where

\[
(L'^0, L'^1, L'^2, L'^3) = (L^0, -L^1, L^2, L^3).
\]

We can write this as the 4-vector equation

\[
L' = L + 2g(L, N)N,
\]

where \( N \) is the 4-vector with components \((0, 1, 0, 0)\) in the rest frame. We can characterize \( N \), up to sign, by \( g(N, N) = -1 \), together with the condition that \( g(N, X) = 0 \) whenever \( X \) is the displacement 4-vector between two events in the history of the mirror (i.e. two events with zero \( x^1 \) coordinate in the rest frame). In particular, \( g(N, V) = 0 \).

Now look at (6.23) in the given frame in which the mirror is moving. Let \( N = (N^0, \mathbf{n}) \) in this frame and let \( X \) be the displacement vector between two simultaneous events in the history of the mirror. Then in the given frame \( X = (0, \mathbf{x}) \), where \( \mathbf{x} \cdot \mathbf{n} = 0 \). We have

\[
\begin{align*}
g(N, N) &= (N^0)^2 - \mathbf{n} \cdot \mathbf{n} = -1 \\
g(N, V) &= \gamma(v)(cN^0 - \mathbf{n} \cdot \mathbf{v}) = 0 \\
g(N, X) &= \mathbf{n} \cdot \mathbf{x} = 0 \quad \text{whenever} \quad \mathbf{x} \cdot \mathbf{k} = 0.
\end{align*}
\]

It follows that

\[
N^0 = \frac{v \cdot k}{\sqrt{c^2 - (v \cdot k)^2}} \quad \text{and} \quad \mathbf{n} = \frac{cN^0 \mathbf{k}}{v \cdot k},
\]

(the positive sign of \( N^0 \) is chosen to make the direction of \( \mathbf{n} \) match the direction of \( \mathbf{k} \)). We can write (6.23) as

\[
\omega'(1, e') = \omega(1, e) + 2\omega(N^0 - \mathbf{e} \cdot \mathbf{n})(N^0, \mathbf{n}).
\]

By substituting from (6.24),

\[
\begin{align*}
\omega' &= \omega \left( 1 + \frac{2(v \cdot k) k \cdot (v - ce)}{c^2 - (v \cdot k)^2} \right) \\
\omega'e' &= \omega \left[ e + 2c \left( \frac{(v - ce) \cdot k}{c^2 - (v \cdot k)^2} \right) k \right].
\end{align*}
\]

If the mirror is moving parallel to itself \((v \cdot k = 0)\), then the laws of reflection are the same as for a stationary mirror. \( \Box \)

**Exercise 6.3** Derive (6.9) and (6.10) directly from (6.7).
Exercise 6.4 Derive the first equation in (6.18) from the second. (Remember that \( v\gamma(v) = c\sqrt{\gamma(v)^2 - 1} \).)

Exercise 6.5 In an inertial frame, the electric field \( \mathbf{E} \) is uniform and constant, and the magnetic field vanishes. A test charge \( e \) of rest mass \( m \) has initial velocity \( v_0 \) perpendicular to \( \mathbf{E} \). Show that

\[
\mathbf{E} \cdot \mathbf{v} = cE \tanh \left( \frac{eE\tau}{mc} \right) \quad \text{and} \quad \gamma(v) = \gamma(v_0) \cosh \left( \frac{eE\tau}{mc} \right),
\]

where \( \mathbf{v} \) is the velocity relative to the frame, \( \tau \) is the proper time, and \( E = |\mathbf{E}| \). Hence find the trajectory of the particle. In the classical theory, the speed of the particle increases without limit. What happens here?

Exercise 6.6 In an inertial frame, the magnetic field \( \mathbf{B} \) is uniform and constant, and the electric field vanishes. A test charge \( e \) of rest mass \( m \) has initial velocity \( \mathbf{v}_0 \) perpendicular to \( \mathbf{B} \). Show that the particle moves in a circle and that the proper time that elapses on each circuit is \( 2\pi m/eB \).

A test charge moves in a constant uniform electromagnetic field such that \( F_{ab}^{\text{field}} > 0 \) and \( F_{ab}^{\text{field}} F^{ab} = 0 \). Show that after the elapse of proper time \( 2\sqrt{2} \pi mc/e\sqrt{F_{ab}^{\text{field}} F^{ab}} \), measured along the worldline of the charge, its 4-velocity is the same as its initial value.

Exercise 6.7 Let \( ct, x, y, z \) and \( c\tilde{t}, \tilde{x}, \tilde{y}, \tilde{z} \) be two inertial coordinate systems related by a standard Lorentz transformation. An event on the future light-cone of the origin has coordinates

\[
(\sqrt{x^2 + y^2 + z^2}, x, y, z) \quad \text{and} \quad (\sqrt{\tilde{x}^2 + \tilde{y}^2 + \tilde{z}^2}, \tilde{x}, \tilde{y}, \tilde{z})
\]

in the respective systems. Express \( x, y, z \) in terms of \( \tilde{x}, \tilde{y}, \tilde{z} \) and find the Jacobian \( \partial(x, y, z)/\partial(\tilde{x}, \tilde{y}, \tilde{z}) \). Hence show that

\[
\frac{dx dy dz}{\sqrt{x^2 + y^2 + z^2}} = \frac{d\tilde{x} d\tilde{y} d\tilde{z}}{\sqrt{\tilde{x}^2 + \tilde{y}^2 + \tilde{z}^2}}.
\]

Exercise 6.8 Let \( F^{ab} \) be an electromagnetic field tensor. Show that if \( K \) is an eigenvector of \( F \) with nonzero eigenvalue \( \lambda \) (that is, \( F^a_b K^b = \lambda K^a \)), then \( K \) is null. By writing \( K \) as a multiple of \((1, k)\), where \( k \cdot k = 1 \), show that the eigenvector equation is equivalent to

\[
\mathbf{E} \cdot \mathbf{k} = \lambda \quad \text{and} \quad \mathbf{E} + c\mathbf{k} \wedge \mathbf{B} = \lambda \mathbf{k}.
\]

Show that, in general, \( F \) has two independent real null eigenvectors. What are the special cases?

An observer has 4-velocity which is the sum of two null eigenvectors of \( F \). Show that the electric and magnetic fields seen by the observer are proportional.
7 Energy-Momentum Tensors

7.1 Dust

The energy-momentum tensor of a continuous distribution carries information about the total energy density measured by an arbitrary inertial observer. The examples we have in mind are continuous media, such as fluids and gases, and fields, such as the electromagnetic field. The 'energy density' includes the matter density as well as heat and other forms of energy.

We shall start by considering a dust cloud. By this, we mean a stream of particles with no interactions. Such a system is characterized by a 4-vector field $V$ (the 4-velocity) and a scalar function $\rho$ (the rest-mass per unit volume measured in the rest frame at each event).

Consider the particles that occupy a small volume $\Delta$ in the instantaneous rest frame at an event. These have total 4-momentum $P = \rho V \Delta$. An observer with 4-velocity $U$ will measure total energy of the particles to be

$$g(P, U) = \rho U_a V^a,$$

and will reckon that they occupy a volume

$$\frac{\Delta}{\gamma(u)} = \frac{c^2 \Delta}{U_a V^a},$$

where $u$ is the observer's speed in the instantaneous rest frame. Therefore the observer will measure the energy density to be $\rho c^{-2} (U_a V^a)^2 = c^{-2} T^{ab} U_a U_b$, where $T^{ab} = \rho V^a V^b$. The tensor $T^{ab}$ is called the energy-momentum tensor of the dust. The other components of $T$ also have physical significance.

1. The 4-vector field with components $c^{-2} T^{ab} U_a = \rho c^{-2} U_a V^a V^b$ is the total 4-momentum per unit volume measured by the observer. In the observer's rest frame, this 4-vector is $\rho \gamma(v)^2(c, v)$, where $v$ is the velocity of the particles.

2. The component $T^{11}$ in the observer's frame is half the normal force per unit area exerted by the particles striking a screen orthogonal to the observer's $x$-axis. This can be seen as follows. Let $v$ be the velocity of the particles relative to the observer and let $\mathbf{i}$ be a unit vector along the observer's $x$ axis. The particles that strike a small area $A$ of the screen in time $\delta t$ occupied a volume $\mathbf{i} \cdot v A \delta t$ just before hitting the screen. The $x$-component of their 3-momentum before striking the screen is $\rho A \delta t \gamma(v)^2(i, v)^2$. Immediately afterwards, it is minus this. Therefore the $x$-component of the force is $2T^{11}$, which is the change in 3-momentum. (This is clear if $v << c$, in which case 'force' simply means force in the sense of Newtonian mechanics. If $v$ is comparable to $c$, then we must take the 'change in 3-momentum per unit time' as the definition of force.)
The expression for the observed energy density is linear in $T^{ab}$. Therefore the energy density of two superimposed dust clouds with different velocities will again be given by $c^{-2}T^{ab}U_aU_b$, where the total energy-momentum tensor $T^{ab}$ is the sum of the individual energy-momentum tensors for the two clouds.

### 7.2 The Ideal Gas

We consider first an oversimplified model of an ideal gas made up of a large number of superimposed clouds of particles, moving with different velocities, but without interaction. We assume that there is a rest frame, in which the gas is isotropic. This means that there is no preferred direction and that the distribution of velocities in this frame is spherically symmetric.

The energy-momentum tensor $T^{ab}$ of the gas is the sum of the energy-momentum tensors of the individual streams of particles. In a general inertial frame, the component $T^{00}$ is the energy density (both the rest energy and the kinetic energy of the particles contribute).

In the rest frame, the diagonal components $T^{11}$, $T^{22}$, and $T^{33}$ at an event are equal, by symmetry, and all the off diagonal components of $T$ vanish. The components $T^{0a}$, $a = 1, 2, 3$, vanish since otherwise they would be the components of a nonzero 3-vector which would pick out a preferred direction. The $3 \times 3$ matrix with entries $T^{ab}$, $a, b = 1, 2, 3$, must be a multiple of the identity matrix since otherwise its eigenvectors would pick out preferred directions.

The diagonal components $T^{aa}$, $a = 1, 2, 3$, in the rest frame are equal to the pressure: $T^{11}$ is the total normal force per unit area exerted on one side of a small screen orthogonal to the $x$-axis and bombarded elastically by the particles. This follows by the same argument as before: the factor of 2 does not appear here because only the particles with a negative $x$-component of velocity strike the screen from the positive $x$-direction, and these make up half the total.

In the rest-frame, therefore,

$$
(T^{ab}) = \begin{pmatrix}
\rho c^2 & 0 & 0 & 0 \\
0 & p & 0 & 0 \\
0 & 0 & p & 0 \\
0 & 0 & 0 & p
\end{pmatrix},
$$

where $\rho c^2$ is the energy density and $p$ is the pressure. We can write this in the tensor form

$$
T^{ab} = (\rho + c^{-2}p)V^aV^b - pg^{ab},
$$

(7.1)

where $V^a$ is the 4-velocity of the rest frame. This holds in the rest frame, by comparing components, and therefore in any frame.

A perfect fluid is modelled as a collection of particles for which the total energy-momentum tensor has the form (7.1), where $p$, $\rho$, and $V$ are functions of the space-time coordinates (be careful to note that when this formula is used in a general frame, the $p$ and $\rho c^2$ that appear are the pressure and energy density measured in the frame with 4-velocity $V$ at each event, and not in the general frame).
7.3 Conservation of Energy

What property of the energy-momentum tensor encodes the conservation of energy? First consider a dust cloud. Conservation of energy here means conservation of the rest masses of the individual particles, since there are no interactions. Each particle moves in a straight line, and so

\[ V^a \nabla_a V^b = 0 \quad (7.2) \]

in any frame.

Now pick out one of the particles and choose an inertial coordinate system in which it is at rest at the origin. The nearby particle at \( r \) has velocity \( v(r) \), with \( v = |v| < c \). Its 4-velocity is \( V = \gamma(v)(c, v) \).

Consider the particles in a small region containing the origin. Suppose that the volume occupied by the particles at time \( t \) is \( \Delta(t) \). Then their total rest mass is \( \rho(t, 0) \Delta(t) \). At time \( t + \delta t \), they occupy a volume \( \Delta(t + \delta t) \). To the first order in \( \delta t \),

\[ \Delta(t + \delta t) - \Delta(t) = \delta t \int_S v \cdot dS = \delta t \Delta(t) \text{ div } v, \]

where \( S \) is the surface of the region (the surface moves with the particles and in time \( \delta t \), the surface element \( dS \) sweeps out a volume \( \delta t \ dS \cdot v \)). Since the total rest mass of the particles is constant,\[
\Delta(t) \rho(t + \delta t, 0)(1 + \delta t \text{ div } v) = \rho(t, 0) \Delta(t). \]

We conclude that the continuity equation

\[ \frac{\partial \rho}{\partial t} + \text{div} (\rho v) = 0 \]

holds at the origin, at which point \( v = 0 \).

Now consider the 4-divergence \( \nabla_a (\rho V^a) \). At the origin, \( \gamma(v) = 1 \) and the partial derivatives of \( \gamma(v) \) vanish, since \( v = 0 \). It follows that, at \( r = 0 \),

\[ \nabla_a (\rho V^a) = \frac{\partial \rho}{\partial t} + \text{div} (\rho v) = 0. \]

But \( \nabla_a (\rho V^a) \) is a scalar: it takes the same value in every inertial coordinate system. By applying the same argument to the other particles, we conclude that \( \nabla_a (\rho V^a) = 0 \) at all events, in all inertial frames. By combining this with (7.2), we deduce that

\[ \nabla_a T^{ab} = \nabla_a (\rho V^a V^b) = \nabla_a (\rho V^a) V^b + \rho V^a \nabla_a V^b = 0. \]

This equation encodes both the equation of motion and the conservation of rest mass: by contracting with \( V_b \), we obtain first the continuity equation \( \nabla_a (\rho V^a) = 0 \), and then by substituting back, (7.2).

By linearity we deduce that \( \nabla_a T^{ab} = 0 \) also holds when we combine streams of particles, without allowing interactions. By using conservation of 4-momentum, it is not hard to show that this is still true when the different streams are allowed to interact through collisions, although it no longer holds for each stream individually because of the exchange of energy in collisions.
7.4 Another Form of Energy Conservation

The conservation law $\nabla_a T^{ab} = 0$ can be understood in another way. First consider a dust cloud. Pick an inertial coordinate system and let $\Omega$ now be a fixed region. Let $f = \rho c^2 \gamma(v)^2 v$ be the 3-vector field with components $c(T^{01}, T^{02}, T^{03})$. It represents the total energy flow: the particles that cross a small surface element $dS$ between time $t$ and time $t + dt$ have total energy $\delta t \mathbf{f} \cdot dS$ relative to the coordinate system: this is because the total energy density is $T^{00} = \gamma(v)^2 \rho c^2$ and the particles that cross the surface element in the time interval occupy a volume $\delta t \mathbf{v} \cdot dS$ at time $t$.

The change in total energy in $\Omega$ in time $\delta t$ must be balanced by a flow of energy across the boundary $S$. Therefore

$$0 = \frac{d}{dt} \int_\Omega T^{00} dV + \int_S \mathbf{f} \cdot dS$$

$$= \int_\Omega \left( \frac{\partial T^{00}}{\partial t} + \text{div } \mathbf{f} \right) dV$$

$$= c \int_\Omega \left( \nabla_0 T^{00} + \nabla_1 T^{01} + \nabla_2 T^{02} + \nabla_3 T^{03} \right) dV.$$ 

Since this holds for any $\Omega$, we again deduce that the equation $\nabla_a T^{ab} = 0$ encodes the conservation of total energy for a dust cloud, and hence, by linearity, for an ideal gas.

7.5 Fluids

The next step is to extend the analysis to a general fluid, which we think of as a continuous limit of a large number of superimposed streams of particles, without any special assumptions about the distribution of velocities. The energy-momentum tensor $T^{ab}$ is the sum of the energy-momentum tensors of the individual streams.

It is hard to pass to the limit of a continuous medium in a mathematically coherent way. One could instead start with the classical equations for a fluid with a general stress tensor, and construct a Lorentz-invariant theory by introducing appropriate relativistic modifications. This is not easy to make physically convincing. A more honest approach is to start with the conservation law $\nabla_a T^{ab} = 0$ as a basic assumption (in the absence of external forces), and then to work back to the equations of motion: we have seen already in the case of a single stream (a dust cloud) that this equation encodes both conservation of energy and the equations of motion (it is remarkable that the requirement that conservation of energy should hold for all observers is sufficient to determine the equations of the motion). The key is the following proposition.

Proposition 7.1 Let $T$ be a symmetric second-rank contravariant tensor with components $T^{ab}$. If $T^{ab} U_a U_b > 0$ for every timelike and null 4-vector $U$, then there exists a unique future-pointing timelike 4-vector $V$ such that

$$T^{ab} V_b = \rho V^a \quad \text{and} \quad V^a V_a = c^2 \quad .$$

for some $\rho > 0$. 
Proof. Pick an inertial coordinate system and put \( U = \gamma(u)(c,u) \). Define the function \( f(u) \) on the open ball \( u \cdot u < c^2 \) by

\[
f(u) = T(U,U) = \frac{c^2T^{00} - 2cT^{01}u_1 - \cdots + T^{11}u_1^2 + \cdots}{1 - u^2/c^2}.
\]

Then \( f \) is positive and \( f(u) \to \infty \) as \( u \to ce \), where \( e \cdot e = 1 \). This is because the numerator on the right-hand side remains positive in the limit since it is equal to \( T(K,K) \), where \( K \) is the null 4-vector \( c(1,e) \).

Since the closed ball is compact, \( f \) must achieve its minimum at some point \( v \) with \( v < c \). Put \( V = \gamma(v)(c,v) \). Because \( f \) is Lorentz invariant, there is no loss of generality in assuming that the inertial frame has been chosen so that \( v = 0 \). Since the partial derivatives of \( f \) vanish at its minimum, \( T^{01} = T^{02} = T^{03} = 0 \) in this frame. Therefore

\[
T^{ab}V_a = \rho c^2 V^b, \tag{7.3}
\]

where \( \rho c^4 = T(V,V) > 0 \), since \( V \) is timelike. Equation (7.3) is a tensor equation, so it must hold in every frame.

This proves that \( V \) exists. To show that it is unique, suppose there exists a second such 4-vector \( V' \). Then

\[
T^{ab}V_b = \rho c^2 V^a \quad \text{and} \quad T^{ab}V'_b = \rho' c^2 V'^a,
\]

with \( \rho > 0 \) and \( \rho' > 0 \). Since \( V \) and \( V' \) are both timelike, \( g(V,V') \neq 0 \). By contracting the first equation with \( V_a \) and the second with \( V'_a \), and subtracting, we deduce that \( \rho = \rho' \). It follows that \( T^{ab}L_b = \rho c^2 L^a \) for every linear combination \( L \) of \( V \) and \( V' \). But either \( V \) and \( V' \) are proportional, in which case they must be equal since they are both future-pointing and \( g(V,V) = g(V',V') = c^2 \), or \( L \) is null for some linear combination. In the latter case \( T^{ab}L_aL_b = \rho c^2 L^aL_a = 0 \), a contradiction. Therefore \( V = V' \). \( \square \)

If \( T^{ab} \) is the energy-momentum tensor field of a fluid, then the 4-vector field \( V \) is called the 4-\textit{velocity} of the fluid. An inertial frame in which \( V = (c,0) \) at some event is called an \textit{instantaneous rest frame} of the fluid at that event. A rest frame is a frame in which the energy density is minimal.

The condition \( T^{ab}U_aU_b > 0 \) for every timelike \( U \) is the condition that every inertial observer sees a positive energy density. This should always hold. For reasonable fluids, the strict inequality also holds for null \( U \), but it is not true of the electromagnetic energy-momentum tensor (Exercise 7.3). A fluid has a well-defined rest frame at each event and a well-defined 4-\textit{velocity}, but, as is to be expected, this is not true of an electromagnetic field.

For a fluid, we can write

\[
T^{ab} = \rho V^aV^b + t^{ab},
\]

where \( t \) has components

\[
(t^{ab}) = \begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & t^{11} & t^{12} & t^{13} \\
0 & t^{21} & t^{22} & t^{23} \\
0 & t^{31} & t^{32} & t^{33}
\end{pmatrix}
\]

in the instantaneous rest frame; \( t \) is called the \textit{stress tensor} of the fluid. We think of the solution curves of
\[ \frac{dx^a}{d\tau} = V^a \]

as the worldlines of small elements of the fluid.

Consider such a fluid element. Suppose that it occupies volume \( \Delta \), measured in its instantaneous rest frame. Then, in the instantaneous rest frame,

\[ \frac{d\Delta}{dt} = \Delta \text{div } \mathbf{v} \]

by the same argument as in §7.3. Since \( \gamma(v) = 1 \) and its derivatives vanish at the fluid element in its instantaneous rest frame, we can deduce that along the worldline

\[ \frac{d\Delta}{d\tau} = \Delta \nabla_a V^a. \]

The 4-momentum of the fluid element is \( P = \Delta \rho V \). From the conservation law,

\[
V^a \nabla_a P^b = V^a \nabla_a (\Delta \rho V^b) \\
= \Delta (V^a \nabla_a (\rho V^b) + \rho V^b \nabla_a V^a) \\
= \Delta \nabla_a (\rho V^a V^b) \\
= -\Delta \nabla_a \epsilon^{ab}.
\]

Therefore the fluid element behaves as a particle subject to a 4-force \( F \) with components \( F^a = -\nabla_b (\epsilon^{ab}) \). Hence the term 'stress tensor', by analogy with classical fluid dynamics. In fact the spatial components \( \epsilon^{ab} \) in the instantaneous rest frame are precisely the components of the classical stress tensor, and for \( v \ll c \), (7.4) reduces to the classical equation of motion.

### 7.6 The Electromagnetic Energy-Momentum Tensor

We shall now consider the conservation of energy in a cloud of charged particles which interact electromagnetically, but which are not subject to other external forces. If each particle has rest mass \( m \) and charge \( e \), then the current 4-vector is \( J = \lambda \rho V \), where \( \lambda = e/m \) and \( \rho \) and \( V \) are as in §7.1.

Consider the particles that occupy a small volume \( \Delta \) in the instantaneous rest frame at some event. These have 4-momentum \( P = \Delta \rho V \) and total charge \( \lambda \rho \Delta \). As in (7.4),

\[ V^b \nabla_b P^a = \Delta \nabla_b (\rho V^b V^a) \].

The equation of motion of the charges is

\[ V^b \nabla_b P^a = \frac{\lambda \rho \Delta}{c} F^{ab} V_b = \frac{\Delta}{c} F^{ab} J_b \]

(see §6.6). Also, by (6.9),

\[ F^{ab} J_b = c e_0 F^{ab} \nabla^k F_{kb} = c e_0 (\nabla^k (F^{ab} F_{kb}) - (\nabla^k F^{ab}) F_{kb}). \]

By (6.10),

\[ \nabla^k F^{ab} + \nabla^b F^{ka} + \nabla^a F^{bk} = 0. \]

Therefore, by contracting with \( F_{kb} \),
\[ 2F_{kb} \nabla^k F^{ab} = F_{kb} \nabla^k F^{ab} + F_{kb} \nabla^b F^{ka} = -F_{kb} \nabla^a F^{bk} = \frac{1}{2} \nabla^a (F_{dk} F^{dk}), \]

since \( F_{bk} = -F_{kb} \). It follows that
\[ F^{ab} J_b = \epsilon_0 c \nabla^b \left( F^{ak} F_{bk} - \frac{1}{4} \delta^a_b F_{kd} F^{kd} \right), \]

and hence that the equation of motion can be written
\[ \nabla_b \left( \rho V^a V^b + K^{ab} \right) = 0, \]

where
\[ K^{ab} = \epsilon_0 \left( F^{ak} F_k^b + \frac{1}{2} g^{ab} F_{dk} F^{dk} \right). \]

The energy of the particles themselves is not conserved. However the total energy of the particles and the electromagnetic field is conserved if we interpret \( K^{ab} \) as the electromagnetic energy-momentum tensor.

With this interpretation, \( K^{00} = \frac{1}{2} \epsilon_0 (c^2 B^2 + E^2) \) is the energy density of the field, and the vector field \( \epsilon_0 c^2 E \wedge B \) with components \( \epsilon_0 (K^{01}, K^{02}, K^{03}) \) is the energy flow: it is called the Poynting vector.

**Exercise 7.1** Show that the electromagnetic energy-momentum tensor can be written
\[ K^{ab} = \frac{1}{2} \epsilon_0 \left( F^{ac} F_{cb} + F^{*bc} F_{ca} \right). \]

(Hint: it is enough to consider \( K^{00} \). Why?)

**Exercise 7.2** Find the energy density and energy flow of a monochromatic plane wave.

**Exercise 7.3** Let \( K \) be the energy-momentum tensor of an electromagnetic field \( F \). Show that, except when \( F_{ab} F^{ab} = F_{ab} F^{*ab} = 0 \), there are two independent real null 4-vectors \( L \) such that \( K^{ab} L_b = \lambda L^a \) for some \( \lambda \) (they are the principal null vectors). Deduce that the electromagnetic energy-momentum tensor does not satisfy the hypotheses of Prop. 7.1. How many principal null vectors are there when \( F_{ab} F^{ab} = F_{ab} F^{*ab} = 0 \)? How are they related to the Poynting vector? (Hint: see Exercise 6.8.)

**Exercise 7.4** Show that for a perfect fluid (§7.2), the conservation equation \( \nabla_a T^{ab} = 0 \) is equivalent to
\[ c^2 \nabla_a (\rho V^a) + p \nabla_a V^a = 0 \]
\[ (c^2 \rho + p) \frac{dV^a}{d\tau} + (V^a V^b - c^2 g^{ab}) \nabla_b p = 0, \]

where \( \tau \) is the proper time along the worldlines of the fluid elements. Why does \( \nabla_a (\rho V^a) \) not vanish? To which equations in classical fluid dynamics do these reduce when \( V = \gamma(v)(c, v) \), with \( v \ll c \)?
8.1 Isometries

The world function is the symmetric two-point function on Minkowski space defined by \( \Phi(A, B) = g(X, X) \), where \( A \) and \( B \) are events and \( X \) is the displacement vector from \( A \) to \( B \). If \( \Phi(A, B) > 0 \), then \( \Phi(A, B) = c^2t^2 \), where \( t \) is the time between \( A \) and \( B \) in a frame in which they happen at the same place. If \( \Phi(A, B) = 0 \), then \( A \) and \( B \) lie on the worldline of a photon. If \( \Phi(A, B) < 0 \), then \( \Phi(A, B) = -d^2 \), where \( d \) is the distance from \( A \) to \( B \) in a frame in which they happen at the same time.

An isometry of Minkowski space is a map \( \iota : M \to M \) which preserves the world function in the sense that \( \Phi(\iota(A), \iota(B)) = \Phi(A, B) \) for every pair of events \( A \) and \( B \). Thus an isometry is a map that leaves invariant the spatial and temporal separation of events. We use this terminology by analogy with Euclidean geometry, where an isometry is a map that preserves distance.

In this chapter, we shall find the infinitesimal isometries of Minkowski space and explore their relationship with conservation laws.

8.2 Killing Vectors

An infinitesimal mapping \( \iota : M \to M \) can be represented by the 4-vector field \( V \) such that \( \iota \) sends the event with inertial coordinates \( x^a + sV^a \), where \( s \) is some small parameter (we ignore terms of order \( s^2 \)). We call \( V \) the generator of the mapping.

For \( \iota \) to be an infinitesimal isometry, we must have \( \Phi(\iota(A), \iota(B)) = \Phi(A, B) + O(s^2) \) for every pair of events \( A \) and \( B \). Now if the displacement vector from \( A \) to \( B \) is \( X \), then the displacement vector form \( \iota(A) \) to \( \iota(B) \) is \( X - sV(A) + sV(B) \) (Fig. 8.1). It follows that if \( \iota \) is an infinitesimal isometry, then

\[
g_{ab}(X^a - sV^a(A) + sV^a(B))(X^b - sV^b(A) + sV^b(B)) = g_{ab}X^aX^b + O(s^2), \tag{8.1}
\]

and hence that

\[
X_a(V^a(B) - V^a(A)) = 0. \tag{8.2}
\]

In particular, this must be true when \( A \) is close to \( B \), in which case

\[
V^a(B) = V^a(A) + X^b(\nabla_b V^a)_A \tag{8.3}
\]

to the first order in \( X \). By substituting (8.3) into (8.2), \( X^aX^b\nabla_a V_b = 0 \). This must hold for all small \( X^a \). Therefore

\[
\nabla_a V_b = 0. \tag{8.4}
\]

Equation (8.4) is the Killing equation. Its solutions are called Killing vectors (strictly, we should say 'Killing vector fields', but that is a bit clumsy). We have shown that every infinitesimal isometry is generated by a Killing vector.
Fig. 8.1. An infinitesimal isometry

The converse is also true. We shall prove this by finding the general solution of the Killing equation. By differentiating with respect to $x^c$,

$$\nabla_c \nabla_a V_b + \nabla_c \nabla_b V_a = 0. \quad (8.5)$$

By permuting the indices,

$$\nabla_b \nabla_c V_a + \nabla_b \nabla_a V_c = 0 \quad (8.6)$$

$$\nabla_a \nabla_b V_c + \nabla_a \nabla_c V_b = 0. \quad (8.7)$$

By adding (8.6) and (8.7) and subtracting (8.5), and by using $\nabla_a \nabla_b V_c = \nabla_b \nabla_a V_c$,

$$\nabla_a \nabla_b V_c = \frac{\partial^2 V_c}{\partial x^a \partial x^b} = 0.$$

It follows that $V_a = L_{ab} x^b + T_a$, where $L_{ab}$ and $T_a$ are constants. By substituting back into the Killing equation, we see that a vector field of this form is a Killing vector if and only if $L_{(ab)} = 0$. Therefore the general Killing vector is

$$V^a = L^a_b x^b + T^a, \quad (8.8)$$

where $L_{ab} = -L_{ba}$, with $L_{ab}$ and $T^a$ constant. It follows by substituting into (8.1) that any Killing vector generates an infinitesimal isometry.

Note that $L_{ab} = \nabla_b V_a$ and that $T^a = V^a(0)$, from which it follows that the $L_{ab}$s are the components of a tensor, but that the $T^a$s depend on the choice of origin.
8.3 Translations and Lorentz Rotations

We say that \( V \) is a translation if \( L_{ab} = 0 \) or a Lorentz rotation about \( A \) if \( V(A) = 0 \)—that is, if the corresponding infinitesimal isometry fixes \( A \). Every Killing vector is a combination of a translation and a Lorentz rotation about the origin.

The translations are the constant vector fields, and the corresponding isometries simply translate each event by the same constant displacement vector. The classification of the Lorentz rotations is a slightly harder problem. We shall tackle it by finding the invariant lines (in Euclidean space, every infinitesimal rotation has one invariant line—its axis).

Let \( V \) be the generator of a Lorentz rotation about \( A \). Choose an inertial coordinate system with \( A \) as origin. Then \( V^a = L^a_b x^b \).

Consider the line through \( A \) given by \( x^a = uX^a \), where \( u \) is a parameter along the line and \( X \) is a 4-vector. This line is mapped to itself by the infinitesimal isometry if \( X \) is an eigenvector of \( L \). That is
\[
L^a_b X^b = \lambda X^a
\]
for some \( \lambda \in \mathbb{R} \). By contracting with \( X_a \), it follows that either \( X_a X^a = 0 \) or \( \lambda = 0 \) since \( L_{ab} X^a X^b = 0 \) by the skew-symmetry of \( L_{ab} \).

Consider the first possibility, \( X^a X_a = 0 \). There is no loss of generality in taking \( X^0 = 1 \). So we can write \( X = (1, x) \) where \( x \cdot x = 1 \), and, since \( L_{ab} = -L_{ba} \),
\[
(L_{ab}) = \begin{pmatrix}
0 & e_1 & e_2 & e_3 \\
-e_1 & 0 & -b_3 & b_2 \\
-e_2 & b_3 & 0 & -b_1 \\
e_3 & -b_2 & b_1 & 0
\end{pmatrix}
\]
(8.10)
where \( e \) and \( b \) are 3-vectors (note the similarity to the electromagnetic field tensor). Then (8.9) becomes
\[
e \cdot x = \lambda, \quad e + x \wedge b = \lambda x,
\]
(8.11)
from which it follows that
\[
e \cdot b = \lambda x \cdot b \quad \text{and} \quad e \wedge b + (x \wedge b) \wedge b = \lambda x \wedge b.
\]
When \( \lambda \neq 0 \), these imply that
\[
(\lambda^2 + b^2) x = e \wedge b + k \lambda^{-1} b + \lambda e,
\]
where \( k = e \cdot b \) and \( b^2 = b \cdot b \). By substituting back into (8.11), we find that \( X = (1, x) \) is a solution provided that
\[
\lambda^4 + \lambda^2 (b^2 - e^2) - k^2 = 0,
\]
(8.12)
where \( e^2 = e \cdot e \). When \( \lambda = 0 \) and \( b \neq 0 \), the solution is \( x = b^{-2} e \wedge b \), provided that \( k = 0 \).

In general, (8.12) has two distinct nonzero real roots in \( \lambda \). For a general Lorentz rotation, there are two invariant null lines. Also in the general case, there are no other invariant lines since there are no solutions of (8.9) with \( \lambda = 0 \). There are three special cases (apart from \( V = 0 \)).
1. $k = 0$ and $b^2 < e^2$. In this case, (8.12) has zero as a repeated root, and two other nonzero real roots. Equation (8.9) has solutions with $\lambda = 0$. These are of the form $X = (X^0, x)$, where

$$e \cdot x = 0 \quad \text{and} \quad X^0 e + x \wedge b = 0. \quad (8.13)$$

That is,

$$X = (X^0, x) = (X^0, b^{-2}(X^0 e \wedge b + \mu b)),$$

with $X^0, \mu \in \mathbb{R}$ arbitrary. Since

$$x \cdot x = \frac{(X^0)^2 e^2 + \mu^2}{b^2} \geq (X^0)^2,$$

$X$ is spacelike for all values of $X^0 \neq 0$ and $\mu$. Consequently there is a 2-dimensional space $\Lambda$ of spacelike vectors such that $L_a^b X_b = 0$. The events with displacement vectors from the origin in $\Lambda$ are fixed by the infinitesimal transformation generated by $V$. The two nonzero roots of (8.12) give two null vectors $N$ such that

$$L_a^b N^b = \pm \sqrt{e^2 - b^2} N^a.$$  

The corresponding null lines $x^a = s N^a$ are also invariant. Killing vectors of this type are called boosts. For example, if $b = 0$ and $e$ is a unit vector along the $z$-axis, then $V$ is a boost: the invariant events are those in the $x, y$ plane, and the invariant null lines are along the two null vectors $(1, 0, 0, 1)$ and $(1, 0, 0, -1)$. The Killing vector has components $(z, 0, 0, ct)$ (Fig. 8.2).

2. $k = 0$ and $b^2 > e^2$. In this case, $\lambda = 0$ is a double root of (8.12) and the other roots are complex. Again the solutions of (8.13) form a 2-dimensional space $\Lambda$, but it now contains spacelike, timelike and null vectors. The corresponding events are fixed, as are all the lines through the origin with tangents in $\Lambda$. There are no other fixed lines. Killing vectors of this type are called rotations (Fig. 8.3). For example, if $e = 0$ and $b$ is a unit vector along the $z$-axis, then the isometry is an infinitesimal rotation about the $z$-axis. The Killing vector has components $(0, y, -z, 0)$ and the fixed events are those in the $t, z$-plane.

3. $k = 0$ and $e^2 = b^2$. This is the same as the previous case, except that $\Lambda$ contains only spacelike and null vectors (it is tangent to the light-cone). Killing vectors of this type are called null rotations. An example is the Killing vector with components $(z, 0, z, ct - y)$, for which the fixed events are those with coordinates $(ct, x, ct, 0)$.

Exercise 8.1  Show that $\det (L_{ab}) = (e \cdot b)^2$.

Exercise 8.2  Show that if $L_{ab}$ is a skew-symmetric tensor of the form (8.10), then $e$ and $b$ behave as 3-vectors under rotations of the spatial axes. (Hint: it is easy to deal with $e$. Now consider the dual tensor $\frac{1}{2} e_{abcd} L^{cd}$.)
8.4 Active and Passive Transformations

Equation (8.8) looks familiar: it looks like an inhomogeneous Lorentz transformation. In fact the infinitesimal inhomogeneous Lorentz transformations are precisely the transformations of the form

\[ \tilde{x}^a = x^a + s \left( L^a_{\ b} x^b + T^a \right) + O(s^2), \]

(8.14)

where \( g_{a(\ b} L^a_{\ c)} = 0 \). The difference between this and the infinitesimal isometry in (8.8) is that (8.14) is a passive transformation, in which we think of the events as fixed and the coordinates changing, while (8.8) is an active transformation: the events are moved while the coordinate system is held fixed.

8.5 Conservation of Momentum and Angular Momentum

Consider a free particle with 4-momentum \( P \) and rest mass \( m \). In any inertial coordinate system,

\[ P^a \nabla_a P^b = 0 \]

since \( P^a \nabla_a = m \, d/d\tau = 0 \), where \( \tau \) is the proper time along the particle’s worldline. It follows that if \( V \) is a Killing vector, then

\[ P^a \nabla_a (P_b V^b) = P^a P^b \nabla_a V_b = P^a P^b \nabla_{(a} V_{b)} = 0. \]

Therefore \( P^a V^a = g(P, V) \) is constant along the worldline. If \( V \) is a translation, then this is simply the statement that one of the components of \( P \) is constant.

If \( V^a = L^a_{\ b} x^b \), so that \( V \) is a Lorentz rotation about the origin, then the conservation law is
\[ L_{ab} P^a x^b = L_{ab} P^{[a} x^{b]} = \text{constant}. \]  

(8.15)

Let \( M^{ab} \) be the tensor defined at events of the particle's worldline by \( M^{ab} = 2x^{[a} P^{b]} \). Since (8.15) holds for any skew-symmetric \( L_{ab} \), we have

\[
\frac{dM^{ab}}{dr} = 0.
\]

The tensor \( M \) is called the \textit{angular-momentum tensor} of the particle. Equation (8.15) is the relativistic form of the law of conservation angular momentum. The \( 3 \times 3 \) matrix of spatial components of \( M \) is

\[
\begin{pmatrix}
0 & L_3 & -L_2 \\
-L_3 & 0 & L_1 \\
L_2 & -L_1 & 0
\end{pmatrix},
\]

where \( \mathbf{L} = \mathbf{r} \wedge \mathbf{p} \) is the angular momentum 3-vector (\( \mathbf{p} \) is the 3-momentum). Thus the relativistic law implies the conservation of angular momentum in the traditional sense.

### 8.6 Exercises

**Exercise 8.3** Suppose that \( X \) is a vector field that vanishes for large \( x, y, z \) in some inertial coordinate system and that \( \nabla_a X^a = 0 \). Show that

\[
\frac{d}{dt} \int X^0(ct, x, y, z) \, dx \, dy \, dz = 0
\]

by differentiating under the integral sign and by applying the divergence theorem.

Let \( V \) be a Killing vector and let \( T \) be the energy-momentum tensor of some distribution of matter or fields. Show that, if \( \nabla_a T^{ab} = 0 \), then \( \nabla_a X^a = 0 \), where \( X^a = T^{ab} \mathbf{V}_b \). Write down the components of \( X \) when \( T \) is the energy-momentum tensor of a dust cloud (occupying a finite volume) and interpret the conserved quantity in (8.16) in the following cases.

(i) \( V \) has components \((1, 0, 0, 0)\).

(ii) \( V \) is a Lorentz rotation about the origin.

**Exercise 8.4** Let \( \Phi \) be the 4-potential of an electromagnetic field \( F \). Let \( V \) be a Killing vector such that \( V^a \nabla_a \Phi_b + \Phi_a \nabla_b V^a = 0 \). Show that \( V^a (U_a + \kappa \Phi_a) \) is constant along the worldline of a test charge, where \( U \) is the charge's 4-velocity and \( \kappa \) is a constant depending on the charge and rest mass, which should be determined.

In an inertial coordinate system \( ct, x, y, z \), the vector fields \( T, X, Y, Z \) have components \((1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0) \), and \((0, 0, 0, 1)\), respectively, and \( F^{ab} = \alpha X^{[a} T^{b]} + \beta Z^{[a} Y^{b]} \), where \( \alpha \) and \( \beta \) are constants. Show that \( \Phi^a = -\alpha x T^a - \beta z Y^a \) is a 4-potential. Hence show that \( U^0 - \kappa \alpha x \) and \( U^2 - \kappa \beta z \) are constants of the motion for the test charge. Write down another 4-potential and hence derive two more constants of the motion. Use your results to give a brief qualitative description of the motion.

**Exercise 8.5** Show that if
\[ X = \begin{pmatrix} ct + z & x + iy \\ x - iy & ct - z \end{pmatrix}, \]

then \( \det X = c^2t^2 - x^2 - y^2 - z^2 \). Hence show that if \( S \) is a \( 2 \times 2 \) complex matrix with unit determinant, then \( X \mapsto SX\overline{S}^t \) determines a Lorentz transformation (\( \overline{S}^t \) is the complex conjugate of the transposed matrix). Show that every proper orthochronous Lorentz transformation can be obtained in this way (this bit is not easy: if you cannot do it, assume that it is true and carry on with the question).

Show that the 4-vector with components \((ct, x, y, z)\) is null if and only if \( X = \alpha \overline{\alpha}^t \), where

\[ \alpha = \begin{pmatrix} a \\ b \end{pmatrix} \]

for some \( a, b \in \mathbb{C} \). Given \( X \), what is the freedom in the choice of \( a \) and \( b \)?

By considering the eigenvectors of \( S \), show that a general Lorentz transformation leaves invariant two null lines through the origin. What are the special cases?
absolute distance 3
absolute time 3
4-acceleration 26
active transformation 81
advanced solution 56–9
alternating tensor 49
angular momentum tensor 82
boost 80
clock paradox 27
clock
  standard 5
collision 37
  elastic 40
Compton scattering 42
coordinate basis 23
covariant derivative 50
covector 45
4-current 53
4-divergence 30
Doppler effect 31
electromagnetic field
dual tensor 55
tensor 54
energy 37, 38
energy conservation 71, 72
energy-momentum tensor
dust 69
  electromagnetic 74–5
fluid 70, 72–3
  ideal gas 70
ether 2
4-force 40
free particle 5
frequency 4-vector 30–31, 39
Galilean transformation 1, 4
gauge transformation 54, 55
geometric optics 63–5
4-gradient 29, 50
gradient covector field 50
inertia tensor 47
inertial coordinates 8
inertial frame 2, 8, 23
inertial mass 39
inertial observer 5
inner product 20
instantaneous rest frame 73
isometry 77
Killing vector 77
Kronecker delta 8, 47
light-cone 6, 8
Lorentz contraction 2, 14, 33
Lorentz force law 61
Lorentz rotation 79
Lorentz transformation 10
  orthochronous 10
  proper 10
  standard 12–13, 16
Lorenz gauge 54
Maxwell's equations 1, 53
metric 20
  coefficients 8
tensor 47
  contravariant 47
Michelson-Morley experiment 2, 17
Minkowski space 20
3-momentum 38
4-momentum 38

Newton’s first law 6
null rotation 80

operational definitions 4
orthonormal frame 22-4
outline 4-vector 32

passive transformation 81
photon 5
  4-momentum 39
plane wave 4, 30
  circularly polarized 62
  linearly polarized 62
  monochromatic 62
Poincaré transformation 10
4-potential 54
Poynting vector 75
pressure 70
principal null vector 75
principle of relativity 5
proper acceleration 26
proper time 5, 25
pseudo-velocity 17
pure electric field 65
pure magnetic field 65

radar method 6
range convention 8
rapidity 17
relative speed 29
rest energy 39
rest mass 38
retarded solution 56–9
rotation 11

scalar product 20

scalar quadruple product 49
simultaneity 3, 6–7
skew-symmetrization 48
space-time 5
summation convention 8
symmetrization 48
tensor 46
  contravariant 46
  covariant 46
  type (p,q) 46
tensor field 49
time dilatation 11–12
translation 10, 79
3-vector 19
4-vector 19
vector field 25
derivative 25

vector
  covariant 45
displacement 19
future-pointing 20
null 20
past-pointing 20
spacelike 20
spatial part 24
tangent 19
temporal part 24
timelike 20
4-velocity 25
  of a fluid 73
velocity, addition formula 16, 29
volume element
  on the light-cone 58
volumes
  contraction of 15

wave equation 1, 4
wave operator 13, 30, 54
world function 77