Introduction to GENERAL RELATIVITY

Lewis Ryder

CAMBRIDGE www.cambridge.org/9780521845632

This page intentionally left blank

Introduction to General Relativity

A student-friendly style, over 100 illustrations, and numerous exercises are brought together in this textbook for advanced undergraduate and beginning graduate students in physics and mathematics.

Lewis Ryder develops the theory of General Relativity in detail. Covering the core topics of black holes, gravitational radiation and cosmology, he provides an overview of General Relativity and its modern ramifications. The book contains a chapter on the connections between General Relativity and the fundamental physics of the microworld, explains the geometry of curved spaces and contains key solutions of Einstein's equations – the Schwarzschild and Kerr solutions.

Mathematical calculations are worked out in detail, so students can develop an intuitive understanding of the subject, as well as learn how to perform calculations. Password-protected solutions for instructors are available at www.cambridge.org/Ryder.

Lewis Ryder is an Honorary Senior Lecturer in Physics at the University of Kent, UK. His research interests are in geometrical aspects of particle theory and its parallels with General Relativity.

Introduction to General Relativity

Lewis Ryder
University of Kent, UK

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,
São Paulo, Delhi, Dubai, Tokyo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521845632

First published in print format 2009

ISBN-13 978-0-511-57746-8 eBook (Adobe Reader)

ISBN-13 978-0-521-84563-2 Hardback

For Mildred Elizabeth Ryder

It is always a source of pleasure when a great and beautiful idea proves to be correct in actual fact. Albert Einstein [letter to Sigmund Freud]

The answer to all these questions may not be simple. I know there are some scientists who go about preaching that Nature always takes on the simplest solutions. Yet the simplest by far would be nothing, that there would be nothing at all in the universe. Nature is far more interesting than that, so I refuse to go along thinking it always has to be simple.

Richard Feynman

Contents

Preface

This book is designed for final year undergraduates or beginning graduate students in physics or theoretical physics. It assumes an acquaintance with Special Relativity and electromagnetism, but beyond that my aim has been to provide a pedagogical introduction to General Relativity, a subject which is now – at last – part of mainstream physics. The coverage is fairly conventional; after outlining the need for a theory of gravity to replace Newton's, there are two chapters devoted to differential geometry, including its modern formulation in terms of differential forms and coordinate-free vectors, then the Einstein field equations, the Schwarzschild solution, the Lense–Thirring effect (recently confirmed observationally), black holes, the Kerr solution, gravitational radiation and cosmology. The book ends with a chapter on field theory, describing similarities between General Relativity and gauge theories of particle physics, the Dirac equation in Riemannian space-time, and Kaluza–Klein theory.

As a research student I was lucky enough to attend the Les Houches summer school in 1963 and there, in the magnificent surroundings of the French alps, began an acquaintance with many of the then new aspects of this subject, just as it was entering the domain of physics proper, eight years after Einstein's death. A notable feature was John Wheeler's course on gravitational collapse, before he had coined the phrase 'black hole'. In part I like to think of this book as passing on to the community of young physicists, after a gap of more than 40 years, some of the excitement generated at that school.

I am very grateful to the staff at Cambridge University Press, Tamsin van Essen, Lindsay Barnes and particularly Simon Capelin for their unfailing help and guidance, and generosity over my failure to meet deadlines. I also gratefully acknowledge helpful conversations and correspondence with Robin Tucker, Bahram Mashhoon, Alexander Shannon, the late Jeeva Anandan, Brian Steadman, Daniel Ryder and especially Andy Hone, who have all helped to improve my understanding. Finally I particularly want to thank my wife, who has supported me throughout this long project, with constant good humour and generous and selfless encouragement. To her the book is dedicated.

Notation, important formulae and physical constants

Latin indices i, j, k, and so on run over the three spatial coordinates 1, 2, 3 or x, y, z or r, θ, φ

Greek indices $\alpha, \beta, \gamma, \ldots \kappa, \lambda, \mu, \ldots$ and so on run over the four space-time coordinates 0, 1, 2, 3 or ct, x, y, z or ct, r, θ, ϕ

Minkowski space-time: metric tensor is $\eta_{\mu\nu} = \mathrm{diag}\,(-1, 1, 1, 1)$, $\mathrm{d}s^2 = -c^2\,\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2$ in Cartesian coordinates

Riemannian space-time: $\mathrm{d}s^2 = g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu = -\,c^2\,\mathrm{d}\tau^2$

The Levi-Cività totally antisymmetric symbol (in Minkowski space) is

$$\varepsilon^{0123} = -\varepsilon_{0123} = 1$$

Connection coefficients: $\Gamma^\nu_{\mu\kappa} = {}^1\!/\!_2\, g^{\nu\rho}(g_{\mu\rho,\kappa} + g_{\kappa\rho,\mu} - g_{\mu\kappa,\rho})$

Riemann tensor: $R^\kappa{}_{\lambda\mu\nu} = \Gamma^\kappa{}_{\lambda\nu,\mu} - \Gamma^\kappa{}_{\lambda\mu,\nu} + \Gamma^\kappa{}_{\rho\mu}\Gamma^\rho{}_{\lambda\nu} - \Gamma^\kappa{}_{\rho\nu}\Gamma^\rho{}_{\lambda\mu}$

Ricci tensor: $R_{\mu\nu} = R^\rho{}_{\mu\rho\nu}$

Curvature scalar: $R = g^{\mu\nu}R_{\mu\nu}$

Field equations: $G_{\mu\nu} = R_{\mu\nu} - {}^1\!/\!_2\, g_{\mu\nu}R = \dfrac{8\pi G}{c^2}T_{\mu\nu}$

Covariant derivatives:

$$\frac{\mathrm{D}V^\mu}{\mathrm{d}x^\nu} = \frac{\partial V^\mu}{\partial x^\nu} + \Gamma^\mu{}_{\lambda\nu}V^\lambda \quad \text{or} \quad V^\mu{}_{;\nu} = V^\mu{}_{,\nu} + \Gamma^\mu{}_{\lambda\nu}V^\lambda$$

$$\frac{\mathrm{D}W_\mu}{\mathrm{d}x^\nu} = \frac{\partial W_\mu}{\partial x^\nu} - \Gamma^\lambda{}_{\mu\nu}W_\lambda \quad \text{or} \quad W_{\mu;\nu} = W_{\mu,\nu} - \Gamma^\lambda{}_{\mu\nu}W_\lambda$$

Speed of light	$c = 3.00 \times 10^8\ \mathrm{m\,s^{-1}}$
Gravitational constant	$G = 6.67 \times 10^{-11}\ \mathrm{N\,m^2\,kg^{-1}}$
Planck's constant	$\hbar = 1.05 \times 10^{-34}\ \mathrm{J\,s}$
	$= 6.58 \times 10^{-22}\ \mathrm{MeV\,s}$
Electron mass	$m_\mathrm{e} = 9.11 \times 10^{-31}\ \mathrm{kg}$
	$m_\mathrm{e}c^2 = 0.51\ \mathrm{MeV}$
Proton mass	$m_\mathrm{p} = 1.672 \times 10^{-27}\ \mathrm{kg}$
	$m_\mathrm{p}c^2 = 938.3\ \mathrm{MeV}$
Neutron mass	$m_\mathrm{n} = 1.675 \times 10^{-27}\ \mathrm{kg}$
	$m_\mathrm{n}c^2 = 939.6\ \mathrm{MeV}$
Boltzmann constant	$k = 1.4 \times 10^{-23}\ \mathrm{J\,K^{-1}}$
	$= 8.6 \times 10^{-11}\ \mathrm{MeV\,K^{-1}}$
Solar mass	$M_\mathrm{S} = 1.99 \times 10^{30}\ \mathrm{kg}$

Solar radius	$R_S = 6.96 \times 10^8$ m
Earth mass	$M_E = 5.98 \times 10^{24}$ kg
Earth equatorial radius	$R_E = 6.38 \times 10^6$ m
Mean Earth–Sun distance	$R = 1.50 \times 10^{11}$ m = 1 AU
Schwarzschild radius of Sun	$2m = \dfrac{2M_S G}{c^2} = 2.96$ km
Stefan–Boltzmann constant	$\sigma = 5.67 \times 10^{-8}$ W m^{-2} K^{-4}

1 light year (ly) = 9.46×10^{15} m
1 pc = 3.09×10^{16} m = 3.26 ly
1 radian = 2.06×10^5 seconds of arc

1 Introduction

Einstein's General Theory of Relativity, proposed in 1916, is a theory of gravity. It is also, as its name suggests, a generalisation of Special Relativity, which had been proposed in 1905. This immediately suggests two questions. Firstly, why was a new theory of gravity needed? Newton's theory was, to put it mildly, perfectly good enough. Secondly, why is it that a generalisation of Special Relativity yields, of all things, a theory of *gravity*? Why doesn't it give a theory of electromagnetism, or the strong or weak nuclear forces? Or something even more exotic? What is so special about gravity, that generalising a theory of space and time (because that is what Special Relativity is) gives us an account of it? We begin this chapter by answering the first question first. By the end of the chapter we shall also have made a little bit of headway in the direction of answering the second one.

1.1 The need for a theory of gravity

Newton's theory of gravitation is a spectacularly successful theory. For centuries it has been used by astronomers to calculate the motions of the planets, with a staggering success rate. It has, however, the fatal flaw that it is inconsistent with Special Relativity. We begin by showing this.

As every reader of this book knows, Newton's law of gravitation states that the force exerted on a mass m by a mass M is

$$\mathbf{F} = -\frac{MmG}{r^3}\mathbf{r}. \tag{1.1}$$

Here M and m are not necessarily point masses; r is the distance between their centres of mass. The vector $\mathbf{r}$ has a direction from M to m. Now suppose that the mass M depends on time. The above formula will then become

$$\mathbf{F}(t) = -\frac{M(t)mG}{r^3}\mathbf{r}. \tag{1.2}$$

This means that the force felt by the mass m at a time t depends on the value of the mass M *at the same time* t. There is no allowance for time delay, as Special Relativity would require. From our experience of advanced and retarded potentials in electrodynamics, we can say that Special Relativity would be satisfied if, in the above equation, $M(t)$ were modified to $M(t - r/c)$. This would reflect the fact that the force felt by the small mass at time t depended on the value of the large mass at an *earlier* time $t - r/c$; assuming, that is, that the relevant gravitational 'information' travelled at the speed of light. But this would then not be Newton's law. Newton's law is Equation (1.2) which allows for no time delay, and therefore implicitly suggests that the *information* that the mass M is changing travels with infinite velocity, since the effect of a changing M is felt at the same instant by the mass m. Since Special Relativity implies that nothing can travel faster than light, Equations (1.1) and (1.2) are incompatible with it. If two theories are incompatible, at least one of them must be wrong. The only possible attitude to adopt is that Special Relativity must be kept intact, so Newton's law has to be changed.

Faced with such a dramatic situation – not to say crisis – the instinctive, and perfectly sensible, reaction of most physicists would be to try to 'tinker' with Newton's law; to change it slightly, in order to make it compatible with Special Relativity. And indeed many such attempts were made, but none were successful.[1] Einstein eventually concluded that nothing less than a complete 'new look' at the problem of gravitation had to be taken. We shall return to this in the next section, but before leaving this one it will be useful to rewrite the above equations in a slightly different form; it should be clear that, although Newton's equations are 'wrong', they are an extremely good approximation to whatever 'correct' theory is eventually found, so this theory should then give, as a first approximation, Newton's law. We have by no means finished with Newton!

Let us define $\mathbf{g} = \mathbf{F}/m$, the gravitational *field intensity*. This is a parallel equation to $\mathbf{E} = \mathbf{F}/q$ in electrostatics; the electric field is the force per unit charge and the gravitational field the force per unit mass. Mass is the 'source' of the gravitational field in the same way that electric charge is the source of an electric field. Then Equation (1.1) can be written

$$\mathbf{g} = -\frac{GM}{r^2}\hat{r}, \tag{1.3}$$

which gives an expression for the gravitational field intensity at a distance r from a mass M. This expression, however, is of a rather special form, since the right hand side is a *gradient*. We can write

$$\mathbf{g} = -\nabla\phi, \quad \phi(r) = -\frac{GM}{r}. \tag{1.4}$$

The function $\phi(r)$ is the gravitational *potential*, a scalar field. Newton's theory is then described simply by *one function*. (In contrast, as we shall see in due course, the gravitational field in General Relativity is described by *ten* functions, the ten components of the metric tensor. The non-relativistic limit of *one* of these components is, in essence, the Newtonian potential.) A mass, or a distribution of masses, gives rise to a scalar gravitational potential that completely determines the gravitational field. The potential ϕ in turn satisfies *field equations*. These are Laplace's and Poisson's equations, relevant, respectively, to the cases where there is a vacuum, or a matter density ρ:

$$\text{(Laplace)} \qquad \nabla^2\phi = 0 \qquad \text{(vacuum)}, \tag{1.5}$$

$$\text{(Poisson)} \qquad \nabla^2\phi = 4\pi G\rho \qquad \text{(matter)}. \tag{1.6}$$

[1] For references to these see 'Further reading' at the end of the chapter.

In the case of a point mass, of course, we have $\rho(r) = M\delta^3(\mathbf{r})$, and by virtue of the identity

$$\nabla^2(1/r) = -4\pi\delta^3(\mathbf{r}) \tag{1.7}$$

Equations (1.4) and (1.6) are in accord.

This completes our account of Newtonian gravitational theory. The field **g** depends on r but not on t. Such a field is incompatible with Special Relativity. It is not a Lorentz covariant field; such a field would be a *four*-vector rather than a three-vector and would depend on t as well as on r, so that the equations of gravity looked the same in all frames of reference related by Lorentz transformations. This is not the case here. Since Newton's theory is inconsistent with Special Relativity it must be abandoned. This is both a horrifying prospect and a slightly encouraging one; horrifying because we are having to abandon one of the best theories in physics, and encouraging because Newton's theory is so precise and so successful that any new theory of gravity will immediately have to fulfil the very stringent requirement that in the non-relativistic limit it should yield Newton's theory. This will provide an immediate test for a new theory.

1.2 Gravitation and inertia: the Equivalence Principle in mechanics

Einstein's new approach to gravity sprang from the work of Galileo (1564–1642; he was born in the same year as Shakespeare and died the year Newton was born). Galileo conducted a series of experiments rolling spheres down ramps. He varied the angle of inclination of the ramp and timed the spheres with a water clock. Physicists commonly portray Galileo as dropping masses from the Leaning Tower of Pisa and timing their descent to the ground. Historians cast doubt on whether this happened, but for our purposes it hardly matters whether it did or didn't; what matters is the conclusion Galileo drew. By extrapolating to the limit in which the ramps down which the spheres rolled became vertical, and therefore that the spheres fell freely, he concluded that *all bodies fall at the same rate in a gravitational field*. This, for Einstein, was a crucially important finding. To investigate it further consider the following 'thought-experiment', which I refer to as 'Einstein's box'. A box is placed in a gravitational field, say on the Earth's surface (Fig. 1.1(a)). An experimenter in the box releases two objects, made of different materials, from the same height, and measures the times of their fall in the gravitational field **g**. He finds, as Galileo found, that they reach the floor of the box at the same time. Now consider the box in free space, completely out of the reach of any gravitational influences of planets or stars, but subject to an *acceleration* **a** (Fig. 1.1(b)). Suppose an experimenter in this box also releases two objects at the same time and measures the time which elapses before they reach the floor. He will find, of course, that they take the same time to reach the floor; he *must* find this, because when the two objects are released, they are then subject to no force, because no acceleration, and it is the floor of the box that accelerates up to meet them. It clearly reaches them at the same time. We conclude that this experimenter, by releasing objects and timing their fall, *will not be able to tell whether he is in a gravitational field or being accelerated through*

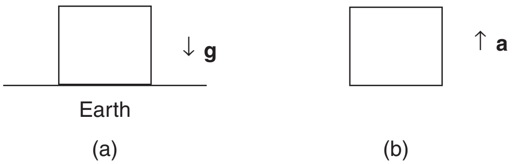


Fig. 1.1 The Einstein box: a comparison between a gravitational field and an accelerating frame of reference.

empty space. The experiments will give identical results. A gravitational field is therefore *equivalent* to an accelerating frame of reference – at least, as measured in this experiment. This, according to Einstein, is the significance of Galileo's experiments, and it is known as the *Equivalence Principle*. Stated in a more general way, the Equivalence Principle says that *no experiment in mechanics can distinguish between a gravitational field and an accelerating frame of reference*. This formulation, the reader will note, already goes beyond Galileo's experiments; the claim is made that *all* experiments in mechanics will yield the same results in an accelerating frame and in a gravitational field. Let us now analyse the consequences of this.

We begin by considering a particle subject to an acceleration **a**. According to Newton's second law of motion, in order to make a particle accelerate it is necessary to apply a force to it. We write

$$\mathbf{F} = m_{\mathrm{i}}\mathbf{a}. \tag{1.8}$$

Here m_{i} is the *inertial mass* of the particle. The above law states that the reason a particle needs a force to accelerate it is that the particles possesses *inertia*. A very closely related idea is that acceleration is *absolute*; (constant) velocity, on the other hand, is *relative*. Now consider a particle falling in a gravitational field **g**. It will experience a force (see (1.2) and (1.3) above) given by

$$\mathbf{F} = m_{\mathrm{g}}\mathbf{g}. \tag{1.9}$$

Here m_{g} is the *gravitational mass* of the particle. It measures the response of a particle to a *gravitational field*. It is very important to appreciate that gravitational mass and inertial mass are conceptually *entirely distinct*. Acceleration in free space is an entirely different thing from a gravitational field, and we make this distinction clear by distinguishing gravitational and inertial mass, as in the two equations above. Now, however, consider a particle falling freely in a gravitational field, as in the Einstein box experiments. Both equations above apply. Because the particle is in a gravitational field it will experience a force, given by (1.9); and because a force is acting on the particle it will accelerate, the acceleration being given by (1.8). These two equations then give

$$\mathbf{a} = \frac{\mathbf{F}}{m_{\mathrm{i}}} = \frac{m_{\mathrm{g}}}{m_{\mathrm{i}}}\mathbf{g}; \tag{1.10}$$

the acceleration of a particle in a gravitational field **g** is the ratio of its gravitational and inertial masses times **g**. Galileo's experiments therefore imply that $m_{\mathrm{g}}/m_{\mathrm{i}}$ *is the same for all* *materials*. Without loss of generality we may put $m_g = m_i$ for all materials; this is because the formula for **g** contains G (see (1.3)), so by scaling G, m_g/m_i can be made equal to unity. (In fact, of course, historically G was found by *assuming* that $m_g = m_i$; no distinction was made between gravitational and inertial masses. We are now 'undoing' history.) We conclude that the Equivalence Principle states that

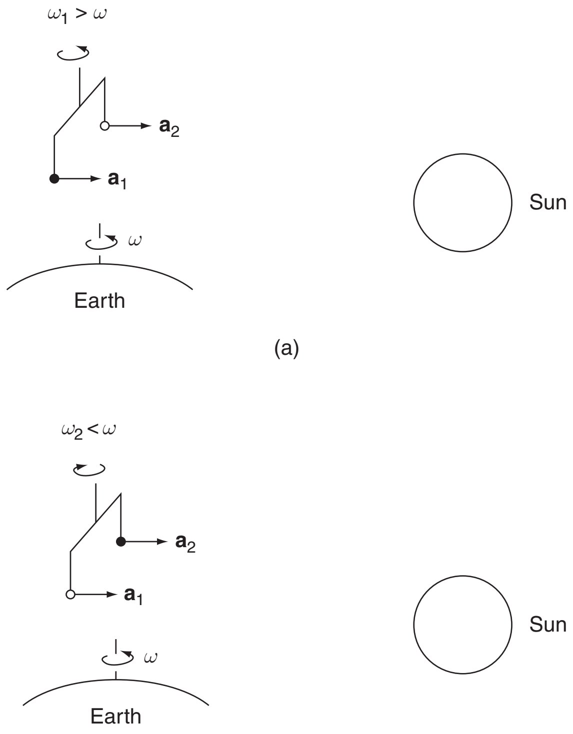
$$m_g = m_i. \tag{1.11}$$

Gravitational mass is the *same as* inertial mass for *all materials*. This is an interesting and non-trivial result. Some very sensitive experiments have been performed, and continue to be performed, to test this equality to higher and higher standards of accuracy. After Galileo, the most interesting experiment was done by Eötvös and will be described below. Before that, however, it is worth devoting a few minutes' thought to the significance of the equality (1.11) above.

The inertial mass of a piece of matter has contributions from two sources; the mass of the 'constituents' and the binding energy, expressed in mass units ($m = E/c^2$). This is the case no matter what the type of binding. So for example the mass of an atom is the sum of the masses of its constituent protons and neutrons *minus the nuclear binding energy* (divided by c^2). In the case of nuclei, the binding energy makes a contribution of the order of 10^{-3} to the total mass. Atoms are bound together by electromagnetic forces and stars and planets are bound by gravitational forces. In all of these cases, the binding energy, as well as the inertial mass of the constituents, contributes to the overall inertial mass of the sample. The statement (1.11) above then implies that the binding energy of a body will *also* contribute to its *gravitational* mass, so binding energy (in fact, energy in general) has a *gravitational effect* since its mass equivalent will in turn give rise to a gravitational field. The gravitational force itself, by virtue of the binding it gives rise to, also gives rise to further gravitational effects. In this sense gravity is *non-linear*. Electromagnetism, on the other hand, is linear; electromagnetic forces give rise to (binding) energy, which acts as a source of gravity, but *not* as a source of further electromagnetic fields, since electromagnetic energy possesses *no charge*. Gravitational energy, however, possesses an effective mass and therefore gives rise to further gravitational fields.

Now let us turn to experiments to test the Equivalence Principle. The simplest one to imagine is simply the measurement of the displacement from the vertical with which a large mass hangs, in the gravitational field of the (rotating) Earth. From Problem 1.1 we see that this displacement is (in Budapest) of the order of 6 minutes of arc multiplied by m_g/m_i. To see whether m_g/m_i is the same for all substances, then, involves looking for tiny variations in this angle, for masses made of differing materials. This is a very difficult measurement to make, not least because it is *static*.

A better test for the constancy of m_g/m_i relies on the gravitational attraction of the *Sun*, whose position relative to the Earth varies with a 24 hour period. We are therefore looking for a periodic signal, which stands more chance of being observed above the noise than does a static one. The simplest version of this is the *Eötvös* or *torsion balance*; the original torsion balance was invented by Coulomb and by Mitchell, and was used by Cavendish to verify the inverse square law of gravity. For the purposes of this experiment the torsion balance takes the form shown in Fig. 1.2.



(a) (b)

Fig. 1.2 A torsion balance at the North Pole. (a) and (b) represent two situations with a 12 hour time separation. The Earth is rotating with angular velocity ω and $\mathbf{a}_1$ and $\mathbf{a}_2$ are the accelerations of the gold and aluminium masses towards the Sun. Assuming that $\mathbf{a}_1 > \mathbf{a}_2$ the resulting torques are of opposite sign.

Two masses, one of gold (shaded) one of aluminium (not shaded), hang from opposite ends of an arm suspended by a thread in the gravitational field of the Earth. Consider such a balance at the North Pole, with the Sun in some assigned position to the right of the diagram. Then at 6 a.m., say, the situation is as shown in (a), the Earth rotating with angular velocity ω. The force exerted by the Sun on the gold mass is (M is the mass of the Sun and r the Earth–Sun distance)

$$F_{\mathrm{Au}} = \frac{GM(m_{\mathrm{g}})_{\mathrm{Au}}}{r^2} \tag{1.12}$$

and hence its acceleration towards the Sun is

$$a_{\mathrm{Au}} = \frac{GM}{r^2}\left(\frac{m_{\mathrm{g}}}{m_{\mathrm{i}}}\right)_{\mathrm{Au}}. \tag{1.13}$$

A similar formula holds for the aluminium mass. Putting

$$\frac{m_{\mathrm{g}}}{m_{\mathrm{i}}} = 1 + \delta, \tag{1.14}$$

then if $\delta_{\mathrm{Au}} \neq \delta_{\mathrm{Al}}$ a *torque* is exerted on the balance, of magnitude ($2l$ is the length of the arm)

$$T = \frac{GMl}{r^2}[(m_g)_{\mathrm{Au}} - (m_g)_{\mathrm{Al}}]. \tag{1.15}$$

This results in an angular acceleration α given by $T = I\alpha$, with I, the moment of inertia, given by $I = m_i l^2$, so we have, at 6 a.m.,

$$(\alpha)_{6\mathrm{am}} = \frac{GM}{lr^2}(\delta_{\mathrm{Au}} - \delta_{\mathrm{Al}}) \equiv \frac{GM}{lr^2}\Delta, \tag{1.16}$$

where $\Delta = \delta_{\mathrm{Au}} - \delta_{\mathrm{Al}}$. In diagram (a) we suppose that $\Delta > 0$, i.e. the acceleration of the gold mass is greater than that of the aluminium mass. This in effect causes the torsion balance to rotate with angular velocity $\omega_1 > \omega$. At 6 p.m., however, the situation is reversed (Fig. 1.2(b)) so the direction of the torque will be reversed, and

$$(\alpha)_{6\mathrm{pm}} = -\frac{GM}{lr^2}\Delta. \tag{1.17}$$

Thus there would be a periodic variation in the torque, with a period of 24 hours. No such variation has been observed,[2] allowing the conclusion that

$$\delta < 10^{-11}; \tag{1.18}$$

gravitational mass and inertial mass are equal to one part in 10^{11} – at least as measured using gold and aluminium.

1.1.1 A remark on inertial mass

The Equivalence Principle states the equality of gravitational and inertial mass, as we have just seen above. It is worthwhile, however, making the following remark. The inertial mass of a particle refers to its mass (deduced, for example, from its behaviour analysed according to Newton's laws) when it undergoes non-uniform, or non-inertial, motion. There are, however, two different types of such motion; it may for instance be acceleration in a straight line, or circular motion with constant speed. In the first case the magnitude of the velocity vector changes but its direction remains constant, while in the second case the magnitude is constant but the direction changes. In each of these cases the motion is non-inertial, but there is a conceptual distinction to be made. To be precise we should observe this distinction and denote the two types of mass $m_{\mathrm{i,acc}}$ and $m_{\mathrm{i,rot}}$. We believe, without, as far as I know, proper evidence, that they are equal

$$m_{\mathrm{i,acc}} = m_{\mathrm{i,rot}}. \tag{1.19}$$

The interesting thing is that Einstein's formulation of the Equivalence Principle referred to inertial mass measured in an accelerating frame, $m_{\mathrm{i,acc}}$, whereas the Eötvös experiment, described above, establishes the equality (to within the stated bounds) of $m_{\mathrm{i,rot}}$ and the

[2] Roll *et al*. (1964), Braginsky & Panov (1972).

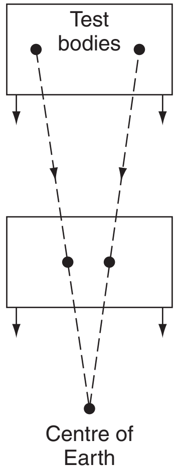


Fig. 1.3 **Test bodies falling to the centre of the Earth.**

gravitational mass. The question is: can an experiment be devised to test the equality of $m_{i,acc}$ and m_g? Or even to test (1.19)?

1.1.2 Tidal forces

The Principle of Equivalence is a *local* principle. To see this, consider the Einstein box in the gravitational field of the Earth, as in Fig. 1.3. If the box descends over a large distance towards the centre of the Earth, it is clear that two test bodies in the box will approach one another, so over this *extended* journey it is clear that they are in a genuine gravitational field, and *not* in an accelerating frame (in which they would stay the same distance apart). In other words, the Equivalence Principle has broken down. We conclude that this principle is only valid as a *local* principle. Over small distances a gravitational field is equivalent to an acceleration, but over larger distances this equivalence breaks down. The effect is known as a *tidal effect*, and ultimately is due to the *curvature* produced by a real gravitational field.

Another way of stating the situation is to note that an object in free fall is in an inertial frame. The effect of the gravitational field has been *cancelled* by the acceleration of the elevator (the 'acceleration due to gravity'). The accelerations required to annul the gravitational fields of the two test bodies, however, are slightly different, because they are directed along the radius vectors. So the inertial frames of the two bodies differ slightly. The frames are 'locally inertial'. The Equivalence Principle treats a gravitational field *at a single point* as equivalent to an acceleration, but it is clear that no gravitational fields encountered in nature give rise to a *uniform* acceleration. Most real gravitational fields are produced by more or less spherical objects like the Earth, so the equivalence in question is only a local one.

We may find an expression for the tidal forces which result from this non-locality. Figure 1.4 shows the forces exerted on the two test bodies – call them A and B – in the gravitational field of a body at O. They both experience a force towards O of magnitude

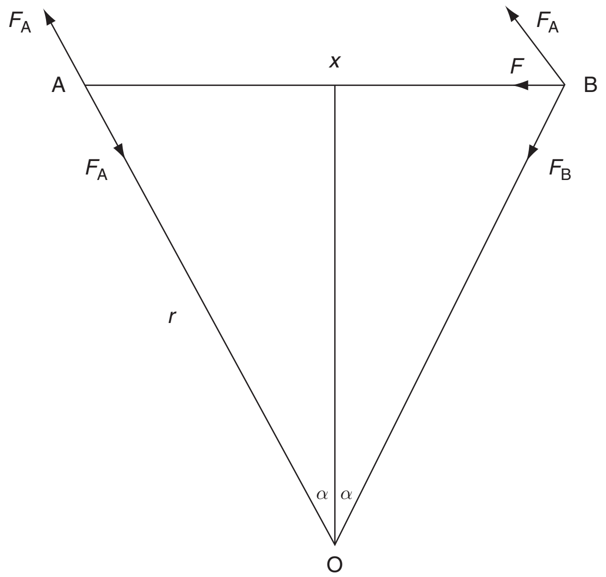


Fig. 1.4 Tidal effect: forces on test bodies A and B.

$$F_A = F_B = \frac{mMG}{r^2}$$

where m is the mass of A and B, M is the mass of the Earth and r the distance of A and B from its centre. In addition, let the distance between A and B be x. Consider the frame in which A is at rest. This frame is realised by applying a force equal and opposite to F_A, to both A and B, as shown in Fig. 1.4. In this frame, B experiences a force F, directed towards A, which is the vector sum of F_B and $-F_A$:

$$F = 2F_A \sin\alpha = 2F_A \cdot \frac{x}{2r} = \frac{mMG}{r^3}x.$$

A then observes B to be accelerating towards him with an acceleration given by $F = -m\,d^2x/dt^2$, i.e.

$$\frac{d^2x}{dt^2} = -\frac{MG}{r^3}x. \tag{1.20}$$

The $1/r^3$ behaviour is characteristic of tidal forces.

1.3 The Equivalence Principle and optics

The Equivalence Principle is a principle of *indistinguishability*; it is impossible, using any experiment in mechanics, to *distinguish* between a gravitational field and an accelerating frame of reference. To this extent it is a *symmetry principle*. If a symmetry of nature is *exact*,

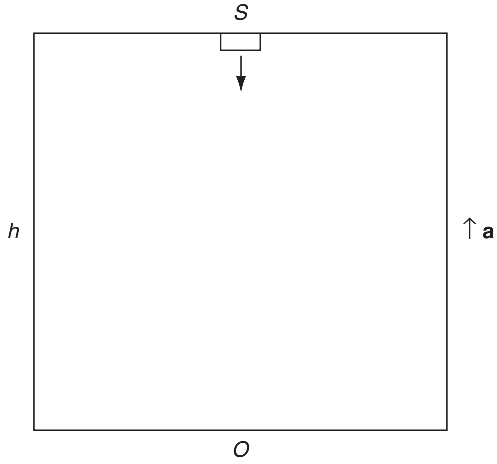


Fig. 1.5 Light propagating downwards in a box accelerating upwards.

this means that various situations are experimentally indistinguishable. If, for example, parity were an exact symmetry of the world (which it is not, because of beta decay), it would be impossible to distinguish left from right. The fact that it *is* possible to distinguish them is a direct indication of the breaking of the symmetry.

No experiment in mechanics, then, can distinguish a gravitational field from an accelerating frame. What about other areas of physics? Let us generalise the Equivalence Principle to *optics*, and consider the idea that no experiment in optics could distinguish a gravitational field from an accelerating frame.[3] To make this concrete, return to the Einstein box and consider the following simple two experiments. The first one is to release monochromatic light (of frequency ν) from the ceiling of the accelerating box, and receive it on the floor (Fig. 1.5). The light is released from the source S at $t = 0$ towards the observer O. At the same instant $t = 0$ the box begins to accelerate upwards with acceleration a. The box is of height h. Light from S reaches O after a time interval $t = h/c$, at which time O is moving upwards with speed $u = at = ah/c$.

Now consider the emission of two successive crests of light from S. Let the time interval between the emission of these crests be $\mathrm{d}t$ in the frame of S. Then

$$\mathrm{d}t = \frac{1}{\nu} \quad \text{in frame } S, \tag{1.21}$$

where ν is the frequency of the light in frame S. Arguing *non-relativistically*, the time interval between the *reception* of these crests at O is

$$\mathrm{d}t' = \mathrm{d}t - \Delta t = \mathrm{d}t - u\frac{\mathrm{d}t}{c} = \mathrm{d}t\left(1 - \frac{u}{c}\right) = \frac{1}{\nu'},$$

[3] This generalisation is sometimes characterised as a progression from a Weak Equivalence Principle (which is the statement $m_\mathrm{i} = m_\mathrm{g}$) to a Strong Equivalence Principle, according to which all the laws of nature (not just those of freely falling bodies) are affected in the same way by a gravitational field and a constant acceleration.

hence

$$\frac{\nu'}{\nu} = \frac{1}{1 - u/c} = 1 + \frac{u}{c} + \mathrm{O}\left(\frac{u}{c}\right)^2 > 1; \tag{1.22}$$

the light is Doppler (blue) shifted. With $\nu' = \nu + \Delta\nu$ we have

$$\frac{\Delta\nu}{\nu} = \frac{u}{c} + \mathrm{O}\left(\frac{u}{c}\right)^2 = \frac{ah}{c^2} + \mathrm{O}\left(\frac{ah}{c^2}\right)^2. \tag{1.23}$$

Arguing *relativistically*, the above result is unchanged to order $(ah/c^2)^2$; the equation above becomes (with $\gamma = (1 - u^2/c^2)^{-1/2}$)

$$\mathrm{d}t' = \gamma(\mathrm{d}t - \Delta t) = \gamma\,\mathrm{d}t(1 - u/c) = \frac{1}{\nu'},$$

hence

$$\frac{\nu'}{\nu} = \frac{1}{\gamma\left(1 - \frac{u}{c}\right)} = \sqrt{\frac{1 + \frac{u}{c}}{1 - \frac{u}{c}}} = 1 + \frac{u}{c} + \mathrm{O}\left(\frac{u}{c}\right)^2,$$

which is the same as (1.22), to the given order. The Equivalence Principle then implies that this is the relativistic frequency shift of light *in a gravitational field*. That is to say, if light is emitted at a point S in a gravitational field and observed at a point O closer to the source of the field, the measured frequency of the light at O is greater than that at S; light 'falling into' a gravitational field is *blue-shifted*. By the same token, if light moves 'out of' a gravitational field its frequency is decreased – it is red-shifted. To get an order of magnitude estimate for this effect, it follows directly from (1.23) that for light travelling 10 metres vertically downwards in the Earth's gravitational field, $h = 10$ m, $a = 10\ \mathrm{m\,s^{-2}}$, we have

$$\frac{\Delta\nu}{\nu} \approx 10^{-15}. \tag{1.24}$$

Of course in the gravitational case the frequency shift described above is *not* a Doppler shift. It is a purely gravitational effect, in which the source and the detector are *not* in relative motion. The formula was, however, derived from the hypothesis that the physical consequences of observing light frequency in a gravitational field are the same as those of observing it in an accelerating frame; and this *is* a Doppler shift, because in this case the source and detection point are in relative motion. This concludes the first thought-experiment on the Equivalence Principle and optics.

The second such thought-experiment is also concerned with light propagation; this time the light travels from left to right across the Einstein box. Consider the situations drawn schematically in Fig. 1.6. In (a) a beam of laser light travels *in an inertial frame* (that is, in neither a gravitational field nor an accelerating frame) across the box. It leaves the laser on the left hand wall and is detected on the right hand wall, after travelling in a straight line. In (b) the box is accelerating upwards with an acceleration a; this acceleration commences at the same time that the light leaves the laser. After a time Δt the light has travelled in the x direction a distance $\Delta x = c\,\Delta t$, while the box has moved upwards a distance $\Delta y = \frac{1}{2}a(\Delta t)^2$, from which

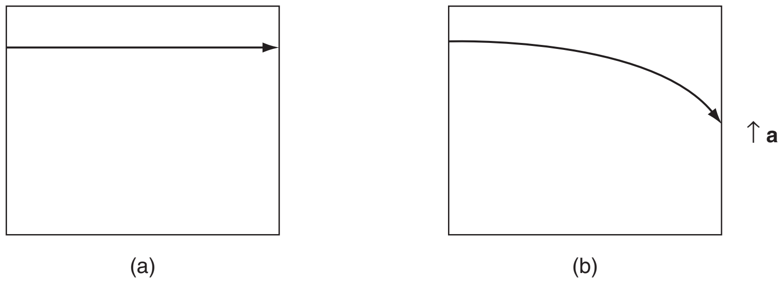


Fig. 1.6 **Light travelling across a box (a) in an inertial frame, (b) in an accelerating frame, or equivalently a gravitational field.**

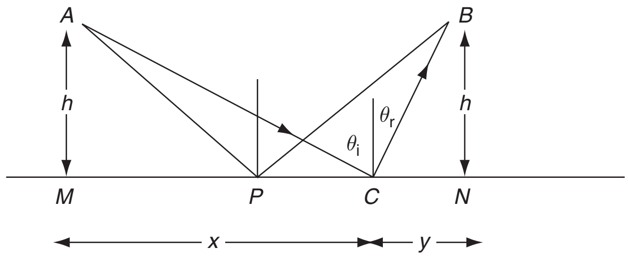


Fig. 1.7 **Fermat's Principle: light reaches *A* from *B* after reflection at a mirror surface.**

$$\Delta y = \frac{a}{2c^2}(\Delta x)^2. \tag{1.25}$$

Since Δy and Δx are the coordinates of the light as measured in the box, it follows that the light describes a *parabolic path* if the box is accelerated. It will therefore be detected at a detector nearer the floor of the box than the laser is. The Equivalence Principle then implies that *light follows a curved path in a gravitational field*, since it does so in an accelerating frame.

This conclusion is extremely far-reaching; even more so than the prediction of a gravitational frequency shift. Fermat (1601–1665) postulated that light takes a *minimum time* to travel from one point to another. For example, consider (Fig. 1.7) the passage of light from A to B, after reflection in a mirror. Let an arbitrary path be ACB, where C is the point where the light beam strikes the mirror, and let the angles of incidence and reflection be θ_i and θ_r, as marked. For simplicity, let A and B each be a perpendicular distance h from the mirror ($AM = BN = h$) and let x and y be the horizontal distances MC and CN respectively. Then, since A and B are fixed points, $x + y = d$ (fixed). The distance s travelled by the light is

$$s = AC + CB = \sqrt{x^2 + h^2} + \sqrt{y^2 + h^2} = \sqrt{x^2 + h^2} + \sqrt{(d - x)^2 + h^2}.$$

The *time* taken to travel a distance s is then s/c, with c the speed of light; more generally, the time taken to travel over a given path is $\int \mathrm{d}s/c$. The requirement that the time taken be a minimum is, since c is constant, clearly the same as the requirement that the distance travelled be a minimum. In the example of reflection of light at the mirror above, if s is a minimum then $\mathrm{d}s/\mathrm{d}x = 0$, which is easily seen to give

$$\mathrm{d}s/\mathrm{d}x = 0 \Rightarrow x = y = \frac{d}{2};$$

in other words, the light beam strikes the mirror at P, at which $\theta_\mathrm{i} = \theta_\mathrm{r}$. Fermat's requirement of *least time* yields Snell's law, that the angles of incidence and reflection are equal.

Fermat's principle is, however, much more far-reaching than this. To begin with, in a general sense, the demand that light propagation takes a minimum time is the requirement that $\int \mathrm{d}s$ be a minimum; or that, in the sense of the *variational principle*,[4]

$$\delta \int \mathrm{d}s = 0. \tag{1.26}$$

The *bending* of light in a gravitational field then implies, if we take Fermat's principle seriously, that the shortest path between two points in a gravitational field is not a straight line. In any flat (Euclidean) space, however, the shortest path between two points *is* a straight line. We therefore conclude that the effect of the gravitational field is to make space *curved*. This is Einstein's conclusion: to study gravity we have to study *curved spaces*. The motion of particles in a gravitational field is to be formulated as the motion of particles in curved spaces. And more generally we can then learn to formulate *any* laws of physics in a gravitational field; for example, the study of electrodynamical effects in a gravitational field is 'simply' arrived at by writing Maxwell's equations in a curved space. The study of curved spaces is, however, not easy, and this is precisely why General Relativity is so difficult. On the other hand, in a *qualitative* sense some results become immediately 'comprehensible' in this new language. For example, the reason that planetary orbits are curves (ellipses, in general) is that planets travel in *free-fall* motion, so they trace out the shortest path they can. In a flat space this would be a straight line, but the effect of the Sun's gravity is to make the space surrounding it curved, making the planetary orbits curved paths (there are no straight lines in a curved space.). Newton's account of gravity, involving a *force*, becomes replaced by an entirely different account, involving a curved space. This is an absolutely totally different vision! It is, however, worth remarking again that the effects of a curved space are not going to be easy to detect on Earth; the deflection of a light beam on Earth, travelling over a distance of 100 km (an order of magnitude larger than, for example, SLAC, the particle accelerator at Stanford), is, from (1.25), with $a = g = 10\,\mathrm{m\,s^{-2}}$, about 10^{-3} mm.

The reader who has followed the logic so far will agree that the plan of action is now, in principle, clear. We have to learn about curved spaces; and this includes learning how to describe vectors in curved spaces, and how to *differentiate* them, which we must do if we are to carry over ideas such as Gauss's theorem and Stokes' theorem into curved spaces. The task is large, not to say daunting, but thanks to the efforts of differential geometers and theoretical relativists over a long period of time (from before Einstein's birth to after his

[4] The variational principle continues to play a crucial role in the formulation of fundamental theories in physics, from classical mechanics and quantum mechanics to General Relativity and gauge field theory. For an introduction to the central role of the variational principle see Yourgrau & Mandelstam (1968).

death) it is not impossible; and, like the task of climbing a mountain, great efforts are rewarded with excellent views. The chapters ahead chart, I hope, a sensible way through this rather complex material, but we close this chapter by making some simple observations and calculations about curved surfaces.

1.4 Curved surfaces

A surface is a 2-dimensional space. It has the distinct advantage that we can imagine it easily, because we see it (as the mathematicians say) 'embedded' in a 3-dimensional space – which also happens to be flat (I mean, of course, Euclidean 3-space). What I want to demonstrate, however, is that there are measurements *intrinsic to a surface* that may be performed to see whether it is flat or not. It is not necessary to embed a surface in a 3-dimensional space in order to see whether or not the surface is curved; we can tell just by performing measurements on the surface itself. The reader will appreciate that this is a necessary exercise; for if we are to make the statement that *3-dimensional space* is curved, this statement must have an *intrinsic* meaning. There is no fourth dimension into which our 3-dimensional space may be embedded (time does not count here).

To begin, consider the three surfaces illustrated in Fig. 1.8. They are a plane, a sphere and a saddle. On each surface draw a circle of radius a and measure its circumference C and area A. On the plane, of course, $C = 2\pi a$ and $A = \pi a^2$, but our claim is that these relations do not hold on the curved surfaces. In fact we have

$$\begin{array}{llll} \text{Plane:} & C = 2\pi a & A = \pi a^2 & \text{flat (zero curvature),} \\ \text{Sphere:} & C < 2\pi a & A < \pi a^2 & \text{curved (positive curvature),} \\ \text{Saddle:} & C > 2\pi a & A > \pi a^2 & \text{curved (negative curvature).} \end{array} \tag{1.27}$$

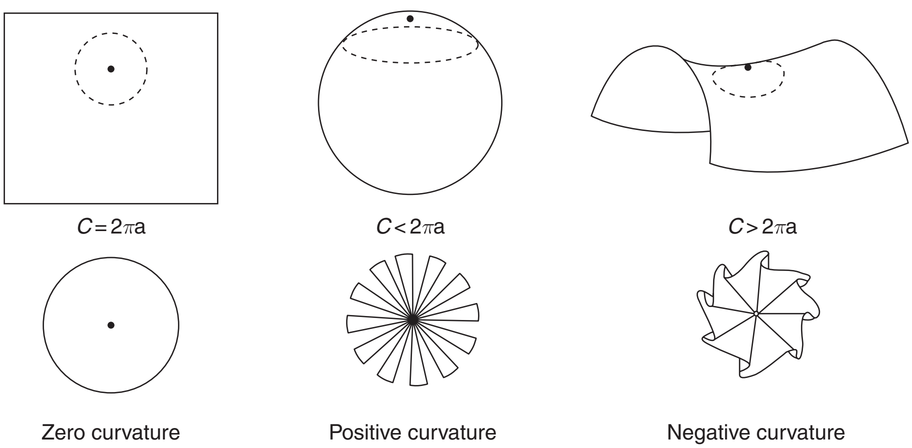


Fig. 1.8 Circles inscribed on a plane, on a sphere and on a saddle.

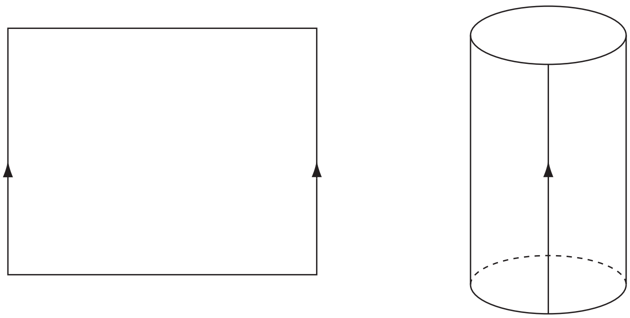


Fig. 1.9 **A cylinder is made by joining the edges of a plane (those marked with arrows). No cutting or stretching is involved.**

In the case of the sphere, for example, to see that $C < 2\pi a$ imagine cutting out the circular shape which acts as a 'cap' to the sphere. This shape cannot be pressed flat. In order to make it flat some radial incisions must be inserted, but this then has the consequence that the *total* circumference C of the dotted circle, which is equal to the sum of the arcs of all the incisions in the diagram below, is less than $2\pi a$. In the case of the saddle the opposite thing happens; in order to get the circular area to lie flat, we have to *fold* parts of it back on itself, so that the true circumference C is greater than $2\pi a$.

It is important to bear in mind in the above reasoning that a, the radius of the circle, is the actual distance from the centre to the perimeter, *as measured in the space*. For simplicity imagine drawing a circle of radius 1000 km on the surface of the Earth, with the centre of the circle at the North Pole. This could be done by having a piece of wire 1000 km long, fixing one end at the North Pole and walking round in a circle, with the wire kept taught. The distance travelled before returning to one's starting point is the circumference C of the circle. The radius of this circle, 1000 km, is the length of the wire, which is laid out along the curved surface. One might feel tempted to point out that one could define the radius of the circle as the 'straight' distance between a point on the circumference and the North Pole, measured by tunnelling through the Earth. But this would be cheating, because it would involve *leaving the space*. We are to imagine the surface as being a world in itself, which we do not leave; we are insisting, in other words, on making measurements *intrinsic* to the space. It should now be clear that the statements (1.27) above constitute a way of telling whether a space – in this case a 2-dimensional one – is flat or curved (and if curved, whether open or closed), and this by means of measurements made entirely within the space. A corollary of this is that, on this definition, a cylinder is *flat*; for a cylinder can be made by joining together the edges of a flat piece of paper, without stretching or tearing (see Fig. 1.9, where the edges with arrows are joined together). Since $C = 2\pi a$ before joining the edges, the same relation holds after joining them, so a cylinder is not *intrinsically* curved. It is said that a cylinder has zero intrinsic curvature but non-zero *extrinsic* curvature.

It is interesting to make one final observation about the exercise of drawing circles on spheres. As the circle S^1 is lowered over the sphere, becoming further and further south, its circumference increases as its radius increases (with $C < 2\pi a$ always holding).[5] This happens until the circle becomes the Equator. This is the circle with a *maximum* circumference in S^2; beyond the Equator, when the circle enters the southern hemisphere, its radius continues to increase, but its circumference actually *decreases*. This continues to be the case until the circle itself approaches the South Pole, at which its circumference tends to zero. This is the limit of a circle with maximum radius (which is the maximum distance attainable in the space) but with a circumference approaching zero. These observations are, in a sense, obvious, but they become interesting and physically relevant in a particular cosmological model, in which the geometry of 3-dimensional space is S^3, the 3-sphere. The above exercise can then be rehashed, increasing the dimension of everything by 1; that is, to discuss surfaces S^2 in S^3, rather than lines S^1 in S^2. This model describes a 'closed' universe, and is described further in Section 10.2 below.

Further reading

Accounts of the various attempts to construct relativistic theories of gravity (other than General Relativity) are outlined in Pauli (1958) pp. 142–5, Mehra (1973), Pais (1982) Chapter 13, Torretti (1996) Chapter 5 and Cao (1997) Chapter 3.

For details of the Eötvös experiment on the torsion balance, see Dicke (1964) and Nieto *et al*. (1989). A modern assessment of the experimental evidence for the Equivalence Principle is contained in Will (2001). The reference to Einstein's seminal paper on the Equivalence Principle and optics is Einstein (1911).

Good introductory accounts of General Relativity and curved spaces are to be found in Hoffmann (1983) Chapter 6, and in Harrison (2000) Chapters 10 and 12. See also, for a slightly more advanced treatment, Ellis & Williams (1988).

Problems

1.1 Find an expression for the angle of displacement from the vertical with which a mass hangs in the gravitational field of the Earth as a function of latitude λ, and calculate its value at Budapest (latitude 47.5° N).

1.2 Suppose that mass, like electric charge, can take on both positive and negative values, but with Newton's laws continuing to hold. Consider two masses, m_1 and m_2, a distance r apart. Describe their motion in the cases (i) $m_1 = m_2 = m$ ($m > 0$), (ii) $m_1 = m_2 = -m$, (iii) $m_1 = m$, $m_2 = -m$. Is momentum conserved in all these cases?

[5] S^n is an n-dimensional subspace of the $(n+1)$-dimensional Euclidean space, given by the formula ${x_1}^2 + {x_2}^2 + \cdots + {x_{n+1}}^2 = \text{const}$. So S^1 is a circle, S^2 (the surface of) a sphere, etc.

1.3 The usual formula for the period T of a simple pendulum of length l is

$$T = 2\pi\sqrt{\frac{l}{g}},$$

where g is the acceleration due to gravity. Denoting the inertial mass of the pendulum bob by m_i and its gravitational mass by m_g, derive an alternative expression for T in terms of these masses, the radius R of the Earth and its mass M_g.

1.4 By employing spherical polar coordinates show that the circumference C of a circle of radius R inscribed on a sphere S^2 (as in Fig. 1.8) obeys the inequality $C < 2\pi R$.

2 Special Relativity, non-inertial effects and electromagnetism

General Relativity is a generalisation of Special Relativity, and this chapter begins with a brief summary of the special theory, with which the reader is assumed already to have some familiarity. After an account of the famous ‘Einstein train’ thought experiment, the more formal matters of Minkowski space-time and Lorentz transformations are discussed. We then consider some non-inertial effects in the shape of the twin paradox and the Sagnac effect. Mach’s Principle, which concerns itself with the origin of inertia, is considered, and this is followed by a section on Thomas precession; an effect derivable from Special Relativity alone, but associated with forces, and therefore with non-inertial frames. The chapter finishes with a brief treatment of electrodynamics – which was Einstein’s starting point for Special Relativity.

2.1 Special Relativity: Einstein’s train

We are concerned with the laws of transformation of coordinates between frames of reference in (uniform) relative motion. Two frames, S and S', both inertial, move relative to one another with (constant) speed v, which we may take to be along their common x axis. The space-time coordinates in each frame are then

$$S : (x, y, z, t); \quad S' : (x', y', z', t').$$

What is the relation between these? In the physics of Galileo and Newton it is

$$x' = x - vt, \quad y' = y, \quad z' = z, \quad t' = t \tag{2.1}$$

whose inverse is

$$x = x' + vt', \quad y = y', \quad z = z', \quad t = t'; \tag{2.2}$$

S and S' have a common origin at $t = 0$. There is an infinite number of inertial frames and the laws of Newtonian mechanics *are the same in all of them*. There is no such thing as absolute velocity; we can only meaningfully talk about the *relative velocity* of one inertial frame relative to another one. This is the Newtonian–Galilean *Principle of Relativity*. Under the above transformations the laws of Newtonian mechanics are *covariant* (of the same form). These transformations form a group – the Galileo group – which is the symmetry group of Newtonian mechanics. Its actions take one from one frame of reference S to another one S', in which the laws of mechanics are the same. If there is a frame S'', moving relative to S' with speed u along their common x axis, then the speed of S'' relative to S is

$$w = u + v. \tag{2.3}$$

This is the law of addition of velocities in the Newtonian-Galilean Principle of Relativity.

Can this Principle of Relativity be generalised from mechanics to all of physics? This is surely a worthy aim, but a strong hint of trouble came when Maxwell, in his theory of electromagnetism, showed that the speed of light (electromagnetic waves) was given by the formula

$$c = (\varepsilon_0 \mu_0)^{-1/2}, \tag{2.4}$$

where ε_0 is the electric permitivity and μ_0 the magnetic permeability of free space. When the values are inserted this gives $c \approx 3 \times 10^8\,\mathrm{m\,s^{-1}}$ – the observed speed of light. So in Maxwell's electrodynamics the speed of light (in a vacuum) depends only on electric and magnetic properties of the vacuum, and is therefore *absolute*; this clearly contradicts the Principle of Relativity above. It must be the same in all frames of reference and Equation (2.3) must therefore break down (at least when applied to light).

The most famous demonstration of this is the Michelson–Morley experiment, which showed that the speed of light is indeed the same in different frames of reference. It is therefore clear that Equations (2.1) and (2.2) must be revised. It was in fact already known that the transformations which left Maxwell's equations invariant were the Lorentz transformations, which for relative motion along the x axis take the form

$$x' = \gamma(x - vt), \quad y' = y, \quad z' = z, \quad t' = \gamma(t - vx/c^2) \tag{2.5}$$

with inverse

$$x = \gamma(x' + vt'), \quad y = y', \quad z = z', \quad t = \gamma(t' + vx'/c^2), \tag{2.6}$$

where

$$\gamma = (1 - v^2/c^2)^{-1/2}. \tag{2.7}$$

Einstein interpreted these equations not just as a mathematical curiosity, but as a demonstration that time, like space, is relative: $x' \neq x$, $t' \neq t$. Let us illustrate this by considering the 'Einstein train'.

Trains A and B, with the same length L, pass one another with relative speed v in the x direction. How long does this take? Let us consider two events:

Event 1 : front of train B passes front of train A
Event 2 : rear of train B passes front of train A

These are illustrated in Fig. 2.1.

Let us adopt the coordinates

$$\begin{aligned} (x', t') &: \text{ coordinates in moving frame} \\ (x, t) &: \text{ coordinates in stationary frame} \end{aligned} \tag{2.8}$$

What is the time interval between these events, as measured in the two coordinate systems? To be definite, let us consider train A as stationary and train B moving. We take the origins ($x = 0$, $x' = 0$) at the right hand ends of the trains and synchronise the clocks so that event 1 happens at $t = 0$, $t' = 0$. Then, for event 1

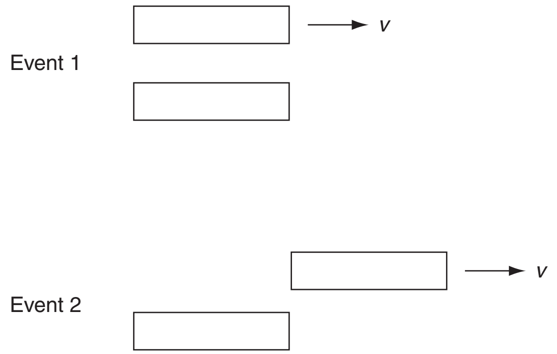


Fig. 2.1 **Event 1: front of train B passes front of train A; Event 2: rear of train B passes front of train A.**

$$({x_1}', {t_1}') = (0, 0), \quad (x_1,\ t_1) = (0, 0). \tag{2.9}$$

If event 2 happens after a time interval T in the stationary frame (train A) and after an interval T' in the moving frame (train B) then we have

$$({x_2}', {t_2}') = (-L, T'), \quad (x_2, t) = (0, T). \tag{2.10}$$

The Lorentz transformation (2.5) applied to event 2 gives $-L = \gamma\,(-v\,T)$, $T' = \gamma\,T$, or

$$L = \gamma v T, \quad T' = \gamma T. \tag{2.11}$$

Since $\gamma > 1$, then $T' > T$; the time interval between events 1 and 2 in the moving frame is greater than in the stationary frame – 'time goes slower in moving frames'. So when Andrei (in train A) looks at Bianca's clock (in train B), he sees it goes slower than his own. It is also true that when Bianca looks at Andrei's clock, she sees it goes slower than her own (since of course the whole sequence of events can be considered in the frame in which B is at rest). One is tempted to ask the question, whose clock is *really* going slower? But this is a bit like asking, when walking along a road, is the house on the left or the right hand side of the road? It all depends in which direction you are walking; and in our case it all depends who is looking at the two clocks: if Andrei is looking, Bianca's clock is going slower, and if Bianca is looking, Andrei's clock is lagging. This is, after all, a theory of *relativity* – only relative motion has physical significance. It has no meaning to say that A is at rest and B is moving, any more than it has to say that B is at rest and A is moving. Since only relative motion has significance, anything observed by A must also be observed by B; the situation is symmetrical. Einstein's train gives a neat demonstration of the relativity of time – to be precise, of time intervals.

There is, as the reader will know, a similar result for space intervals: what is the length of train B as viewed from train A? Call it L'. It is of course Tv:

$$L' = T\,v = L/\gamma = L\,(1 - v^2/c^2)^{1/2} < L. \tag{2.12}$$

A measures B's train as being shorter than his own. Similarly, B measures A's train as being shorter than her own: moving objects appear contracted. This is the Fitzgerald–Lorentz contraction.

We see that time intervals and lengths are not invariant under Lorentz transformations. Infinitesimally, $dx^2+dy^2+dz^2$ is not invariant, and neither is dt^2. The quantity which *is* invariant between the events (x, y, z, t) and $(x+dx, y+dy, z+dz, t+dt)$ is

$$ds^2 = -c^2\,dt^2 + dx^2 + dy^2 + dz^2. \tag{2.13}$$

In the present case of the train, $dy=dz=0$, so $ds^2=-c^2\,dt^2+dx^2$. This should be the same in all frames of reference, where dt and dx refer to the time and space separation of the two events above. We then have, in the rest frame S with coordinates (x, t)

$$ds^2 = -c^2\,dt^2 + dx^2 = -c^2(t_2-t_1)^2 + (x_2-x_1)^2 = -c^2T^2, \tag{2.14}$$

while in the moving frame S', with coordinates (x', t')

$$\begin{aligned} ds^2 &= -c^2\,dt'^2 + dx'^2 = c^2(t_2'-t_1')^2 + (x_2'-x_1')^2 \\ &= -c^2T'^2 + L^2 \\ &= -c^2\gamma^2T^2 + \gamma^2v^2T^2 = -c^2\gamma^2T^2(1-v^2/c^2) \\ &= -c^2T^2, \end{aligned} \tag{2.15}$$

where (2.11) and (2.12) have been used. We see that ds^2 is the same in the two frames. We also see the force of Minkowski's remark,[1] 'Henceforth space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality.'

2.1.1 Minkowski space-time

We now formalise Special Relativity as follows. Space and time become a 4-dimensional manifold, Minkowski space-time. Points in this space-time ('events') have coordinates x^μ ($\mu=0, 1, 2, 3$), with, in Cartesian coordinates $(x^0, x^1, x^2, x^3)=(ct, x, y, z)$, and in spherical polars $(x^0, x^1, x^2, x^3)=(ct, r, \theta, \phi)$. We also adopt the notation that while Greek suffices take on the values (0, 1, 2, 3), Latin suffices take on the values (1, 2, 3) for space variables only; $x^\mu=(x^0, x^i)$. The invariant distance, or 'separation' between two events (in Cartesian coordinates), $ds^2=-c^2\,dt^2+dx^2+dy^2+dz^2$, is written in the form

$$ds^2 = \eta_{\mu\nu}\,dx^\mu\,dx^\nu, \tag{2.16}$$

where the *summation convention* has been used: repeated indices are summed over the values 0, 1, 2, 3. Thus (2.16) is short-hand for

$$ds^2 = \eta_{00}(dx^0)^2 + \eta_{01}\,dx^0\,dx^1 + \eta_{02}\,dx^0\,dx^2 + \cdots \;\; (16\,\text{terms}),$$

and $\eta_{\mu\nu}$ has the following values, in *Cartesian coordinates*:

$$ds^2 = -c^2\,dt^2 + dx^2 + dy^2 + dz^2, \tag{2.17}$$

[1] In Lorentz *et al.*, 1952, p. 75.

hence

$$\eta_{00} = -1, \quad \eta_{11} = \eta_{22} = \eta_{33} = 1, \quad \eta_{\mu\nu} = 0, \quad \mu \neq \nu;$$

or in matrix form

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}; \tag{2.18}$$

and in *spherical polar coordinates*:

$$ds^2 = -c^2\, dt^2 + dr^2 + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\phi^2 \tag{2.19}$$

hence

$$\eta_{00} = -1, \quad \eta_{11} = 1, \quad \eta_{22} = r^2, \quad \eta_{33} = r^2 \sin^2\theta, \quad \eta_{\mu\nu} = 0, \quad \mu \neq \nu;$$

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2 \sin^2\theta \end{pmatrix}. \tag{2.20}$$

The object $\eta_{\mu\nu}$ is the *metric tensor* of Minkowski space; it makes the space a *metric space*, one in which distance is defined.

A useful concept in Special Relativity is that of *proper time* τ. It is defined by

$$ds^2 = -c^2\, d\tau^2. \tag{2.21}$$

In a particle's own rest frame – in which, of course, $dx = dy = dz = 0$, τ coincides with t, so proper time is simply time as measured in the rest frame, or time recorded on one's own clock.

2.1.2 Lorentz transformations

Lorentz transformations are transformations between coordinates labelling space-time events recorded by two inertial observers in uniform relative motion. They take a system from one inertial frame to another one, and consist of rotations and Lorentz 'boost' transformations.[2] Under a general Lorentz transformation

[2] The *maximal* set of transformations leaving ds^2 invariant includes, in addition to rotations and boosts, also translations in space and time, $x^i \to x^i + a^i$, $t \to t + t_0$ (or simply $x^\mu \to x^\mu + a^\mu$). These are *inhomogeneous* transformations and, corresponding to the philosophy outlined above, their inclusion represents the fact that the laws of physics are invariant under space and time translations; there is no absolute origin in space, nor in time (the Big Bang is not relevant here; firstly, we are not considering cosmology, and secondly, we are concerned with the laws of physics themselves, not with whatever state the Universe happens to be in). Enlarging the group of Lorentz transformations to include these translations produces the *inhomogeneous Lorentz group*, or *Poincaré group*. The importance of the Poincaré group as the maximal invariance group in Minkowski space was emphasised particularly by Wigner, whose analysis remains of fundamental importance in particle physics. For more details, see Wigner (1939, 1964), Wightman (1960), Sexl & Urbantke (1976), Tung (1985), Doughty (1990), Ryder (1996), Cao (1997).

$$x^\mu \to x'^\mu = \Lambda^\mu{}_\nu x^\nu \tag{2.22}$$

so $dx'^\mu = \Lambda^\mu{}_\nu \, dx^\nu$ and the invariance of ds^2 gives

$$\eta_{\mu\nu}\, dx'^\mu\, dx'^\nu = \eta_{\mu\nu}\, dx^\mu dx^\nu,$$

hence

$$\eta_{\mu\nu}\Lambda^\mu{}_\rho\Lambda^\nu{}_\sigma\, dx^\rho\, dx^\sigma = \eta_{\rho\sigma}\, dx^\rho\, dx^\sigma$$

or

$$\eta_{\mu\nu}\Lambda^\mu{}_\rho\Lambda^\nu{}_\sigma = \eta_{\rho\sigma}. \tag{2.23}$$

Let us now check that this holds for some specific Lorentz transformations. First consider a rotation about the z axis through an angle θ:

$$x' = x\cos\theta + y\sin\theta, \quad y' = -x\sin\theta + y\cos\theta.$$

The corresponding matrix is

$$\Lambda^\mu{}_\nu = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & \sin\theta & 0 \\ 0 & -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{2.24}$$

Equation (2.23) with $\rho = \sigma = 1$ then gives $\eta_{\mu\nu}\, \Lambda^\mu{}_1\Lambda^\nu{}_1 = 1$, i.e. (summation convention!)

$$-\Lambda^0{}_1\Lambda^0{}_1 + \Lambda^1{}_1\Lambda^1{}_1 + \Lambda^2{}_1\Lambda^2{}_1 + \Lambda^3{}_1\Lambda^3{}_1 = 1,$$

or $\cos^2\theta + \sin^2\theta = 1$, which is correct. Taking different values for ρ and σ also gives consistency with (2.23), as may easily be checked.

Now consider a Lorentz boost along the x direction. With $x^0 = ct$ (and replacing $-v$ by $+v$), Equation (2.5) corresponds to the Lorentz matrix

$$\Lambda^\mu{}_\nu = \begin{pmatrix} \gamma & \gamma v/c & 0 & 0 \\ \gamma v/c & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{2.25}$$

Now put, for example, $\rho = \sigma = 0$. With $\Lambda^0{}_0 = \gamma$, $\Lambda^1{}_0 = \gamma v/c$, $\Lambda^2{}_0 = \Lambda^3{}_0 = 0$, we have

$$\eta_{00}(\Lambda^0{}_0)^2 + \eta_{11}(\Lambda^1{}_0)^2 = -1,$$

or $\gamma^2(1 - v^2/c^2) = 1$, which is correct.

For future reference it is convenient to give the most general form of a Lorentz (boost) transformation, from frame S to frame S' moving with relative velocity $\mathbf{v}$:

$$\mathbf{x}' = \mathbf{x} + (\gamma - 1)\frac{\mathbf{x}\cdot\mathbf{v}}{v^2}\mathbf{v} - \gamma\mathbf{v}t; \quad t' = \gamma\left(t - \frac{\mathbf{x}\cdot\mathbf{v}}{c^2}\right), \tag{2.26}$$

with inverse

$$\mathbf{x} = \mathbf{x}' + (\gamma - 1)\frac{\mathbf{x}' \cdot \mathbf{v}}{v^2}\mathbf{v} + \gamma\mathbf{v}t'; \quad t = \gamma\left(t' - \frac{\mathbf{x}' \cdot \mathbf{v}}{c^2}\right), \tag{2.27}$$

and, as usual, $\gamma = (1 - v^2/c^2)^{-1/2}$.

The matrix (2.25) may be written in a 'trigonometric' form, similar to (2.24). Defining the hyperbolic angle ϕ by ($\beta = v/c$)

$$\gamma = \cosh\phi, \quad \gamma\beta = \sinh\phi, \tag{2.28}$$

the Lorentz transformation given by (2.25) may be written

$$\begin{pmatrix} x'^0 \\ x'^1 \\ x'^2 \\ x'^3 \end{pmatrix} = \begin{pmatrix} \cosh\phi & \sinh\phi & 0 & 0 \\ \sinh\phi & \cosh\phi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix}. \tag{2.29}$$

We now define the *generator* of Lorentz boosts along the x axis by

$$K_x = \frac{1}{\mathrm{i}}\frac{\partial\Lambda}{\partial\phi}\bigg|_{\phi=0} = -\mathrm{i}\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{2.30}$$

where Λ is the matrix in (2.29). It may then easily be checked that

$$\exp(\mathrm{i}K_x\phi) = \begin{pmatrix} \cosh\phi & \sinh\phi & 0 & 0 \\ \sinh\phi & \cosh\phi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{2.31}$$

The generators of boosts along the y and z axes are defined analogously and turn out as

$$K_y = -\mathrm{i}\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad K_z = -\mathrm{i}\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}. \tag{2.32}$$

Generators of rotations may be defined similarly. The matrix (2.24) represents a rotation about the z axis, whose generator is defined as $J_z = \frac{1}{\mathrm{i}}\frac{\partial\Lambda}{\partial\theta}\Big|_{\theta=0}$. This and analogous definitions for J_x and J_y yield

$$J_x = -\mathrm{i}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}, \quad J_y = -\mathrm{i}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad J_z = -\mathrm{i}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{2.33}$$

These six generators obey the commutation relations ($[A,B] \equiv AB - BA$)

$$\begin{aligned}
[J_x, J_y] &= \mathrm{i}\,J_z \text{ and cyclic perms} \\
[K_x, K_y] &= -\mathrm{i}\,J_z \text{ and cyclic perms} \\
[J_x, K_y] &= \mathrm{i}\,K_z \text{ and cyclic perms} \\
[J_x, K_x] &= 0 \text{ etc.}
\end{aligned} \tag{2.34}$$

Equivalently, relabelling the subscripts x, y, z as 1, 2, 3,

$$[J_i, J_k] = \mathrm{i}\,\varepsilon_{ikm} J_m, \tag{2.35a}$$

$$[J_i, K_k] = \mathrm{i}\,\varepsilon_{ikm} K_m, \tag{2.35b}$$

$$[K_i, K_k] = \mathrm{i}\,\varepsilon_{ikm} J_m, \tag{2.35c}$$

where ε_{ikm} is the totally antisymmetric symbol

$$\varepsilon_{ikm} = \begin{cases} 1\ (ikm) \text{ even permutation of } (123), \\ -1\ (ikm) \text{ odd permutation of } (123), \\ 0 \text{ otherwise.} \end{cases} \tag{2.36}$$

In terms of these six generators a general Lorentz boost transformation is

$$\Lambda(\boldsymbol{\phi}) = \exp(\mathrm{i}\mathbf{K} \cdot \boldsymbol{\phi}); \tag{2.37}$$

a general rotation is represented by

$$\Lambda(\boldsymbol{\theta}) = \exp(\mathrm{i}\mathbf{J} \cdot \boldsymbol{\theta}); \tag{2.38}$$

while a general Lorentz transformation, comprising both a boost and a rotation is given by

$$\Lambda(\boldsymbol{\phi}, \boldsymbol{\theta}) = \exp(\mathrm{i}\mathbf{K} \cdot \boldsymbol{\phi} + \mathrm{i}\mathbf{J} \cdot \boldsymbol{\theta}). \tag{2.39}$$

The relations (2.35) define the *Lie algebra* of the Lorentz group, involving three generators K_i of Lorentz 'boosts' (or 'pure' Lorentz transformations) and three generators J_i of rotations in space. The algebra is closed, corresponding to the fact that Lorentz transformations form a group. Rotations in space form a subgroup of the Lorentz group, as may be seen from the fact that the generators J_i form by themselves a closed algebra. The boost generators K_i however do not generate a closed system, as is seen from (2.35c); pure Lorentz transformations do not form a group. As a simple consequence of this, the product of two Lorentz boosts in different directions is *not* a single Lorentz boost, but also involves a rotation. It is this fact which is responsible for Thomas precession (see Section 2.5 below) – and which, as far as I can tell, seems to have been unknown to Einstein.

We finish this section with an additional remark about notation. In Equation (2.16),

$$\mathrm{d}s^2 = \eta_{\mu\nu}\, \mathrm{d}x^\mu\, \mathrm{d}x^\nu,$$

it was pointed out that the summation convention is understood. To be more precise, indices to be summed over appear twice, *once in a lower and once in an upper position*. We may write (2.16), however, in an alternative way. Defining

$$x_\mu = \eta_{\mu\nu} x^\nu,$$

we may put

$$ds^2 = dx^\mu \, dx_\mu, \tag{2.40}$$

where still the summation convention holds, and the repeated index – only one index now – appears once in a lower and once in an upper position. Note that the components x_μ and x^μ are quite different: in Cartesian coordinates we have

$$(x^0, x^1, x^2, x^3) = (ct, x, y, z); \quad (x_0, x_1, x_2, x_3) = (-ct, x, y, z);$$

and in spherical polar coordinates

$$(x^0, x^1, x^2, x^3) = (ct, r, \theta, \phi); \quad (x_0, x_1, x_2, x_3) = (-ct, r, r^2, \theta, \{r^2 \sin^2\theta\}\phi).$$

In General Relativity the position of indices on vectors is important. Vectors with an *upper* index, V^μ, are called *contravariant vectors*, and those with a lower index, V_μ, *covariant vectors*. In the modern mathematical formulation, these vectors actually arise in conceptually different ways, as will be explained in the next chapter.

2.2 Twin paradox: accelerations

The so-called twin paradox is not a paradox. It is the following statement: if A and B are twins and A remains on Earth while B goes on a long trip, say to a distant star and back again, then on return B is younger than A. Suppose the star is a distance l away and B travels with speed v there and back. Then, as measured in A's frame, B is away for a time $2l/v$, and that is how much A has aged when B returns. When A looks at B's clock, however, there is a time dilation factor of $\gamma = (1 - v^2/c^2)^{-1/2}$, so B's clock – including her biological clock – has only registered a passage of time $2l/\gamma v = (2l/v)(1 - v^2/c^2)^{1/2}$; on return, therefore, she is younger than A. This is the true situation. It *appears* paradoxical because one is tempted to think that 'time is relative', so that while A reckons B to be younger on return, as argued above, B should also reckon A to be younger; so in actual fact, one might think, they are the same age after the trip, just as before it. This, however, is wrong, and the reason is that while A remains in an inertial frame (or at least the *approximately* inertial frame of the Earth), B does not, since B has to reverse her velocity for the return trip, and that means she undergoes an *acceleration*. There is no reason why the twins should be the same age after B's space trip, and they are not.

It may be useful to consider some numbers. Suppose the star is 15 light years away and B (Bianca) travels at speed $v = (3/5)c$. Then, measured by A (Andrei), Bianca reaches the star in $\dfrac{15}{3/5} = 25$ years, so Andrei is 50 years older when Bianca returns (see Fig. 2.2). The time dilation factor is $1/\gamma = (1 - v^2/c^2)^{1/2} = 4/5$, so, as seen by Andrei, Bianca takes a time $25 \times (4/5) = 20$ years to reach the star, and will therefore be 40 years older when she returns. She will therefore be 10 years younger than Andrei after the trip. Of course, this is an approximation, since we have ignored the time taken for Bianca to change her velocity from

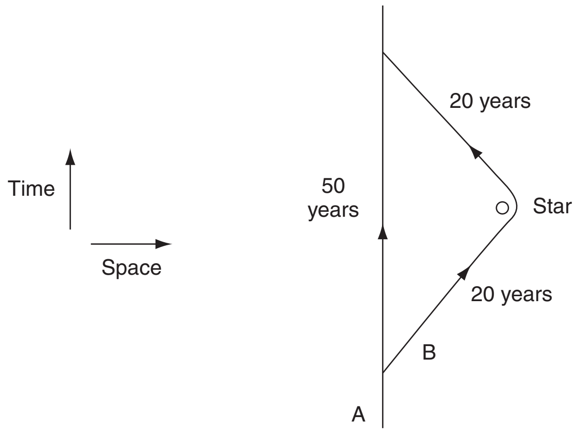


Fig. 2.2 A stays on Earth while B travels to a star and back.

$+v$ to $-v$; this is indicated by B's 'smoothed out' world-line near the star in Fig. 2.2. We now show, however, that this time may be as short as desired, if B is subjected to a large enough acceleration. We must therefore consider the treatment of accelerations in Minkowski space-time.

First define the 4-velocity u^μ:

$$u^\mu = \frac{dx^\mu}{d\tau} = \left(c\frac{dt}{d\tau}, \frac{dx}{d\tau}, \frac{dy}{d\tau}, \frac{dz}{d\tau} \right). \tag{2.41}$$

In view of (2.16) and (2.17) we have

$$\eta_{\mu\nu} u^\mu u^\nu = u^\mu u_\mu = -c^2; \tag{2.42}$$

the 4-velocity has constant length. Differentiating this gives $\left(\text{with } \dot{u}^\mu = \frac{du^\mu}{d\tau}\right)$

$$\frac{d}{d\tau}(u^\mu u_\mu) = 0 = 2\dot{u}^\mu u_\mu,$$

or, defining the *acceleration four-vector* $a^\mu = \dot{u}^\mu$,

$$\eta_{\mu\nu} a^\mu u^\nu = a^\mu u_\mu = 0. \tag{2.43}$$

Now consider a particle moving in the x^1 direction with constant acceleration g. The velocity and acceleration 4-vectors are

$$c\frac{dt}{d\tau} = u^0, \quad \frac{dx^1}{d\tau} = u^1; \quad \frac{du^0}{d\tau} = a^0, \quad \frac{du^1}{d\tau} = a^1;$$

(both vectors have vanishing 2- and 3-components). Equations (2.42) and (2.43) give

$$-(u^0)^2 + (u^1)^2 = -c^2; \quad -u^0 a^0 + u^1 a^1 = 0. \tag{2.44}$$

In addition

$$a^\mu a_\mu = -(a^0)^2 + (a^1)^2 = g^2; \tag{2.45}$$

this last equation *defines* the constant acceleration g. These two equations have the solutions

$$a^0 = \frac{g}{c} u^1, \quad a^1 = \frac{g}{c} u^0, \tag{2.46}$$

from which $\dfrac{\mathrm{d}a^0}{\mathrm{d}\tau} = \dfrac{g}{c}\dfrac{\mathrm{d}u^1}{\mathrm{d}\tau} = \dfrac{g}{c} a^1 = \dfrac{g^2}{c^2} u^0$ and hence

$$\frac{\mathrm{d}^2 u^0}{\mathrm{d}\tau^2} = \frac{g^2}{c^2} u^0. \tag{2.47}$$

Similarly,

$$\frac{\mathrm{d}^2 u^1}{\mathrm{d}\tau^2} = \frac{g^2}{c^2} u^1. \tag{2.48}$$

The solution to (2.48) is

$$u^1 = \mathrm{A}\mathrm{e}^{g\tau/c} + \mathrm{B}\mathrm{e}^{-g\tau/c},$$

hence

$$\frac{\mathrm{d}u^1}{\mathrm{d}\tau} = \frac{g}{c}(\mathrm{A}\mathrm{e}^{g\tau/c} - \mathrm{B}\mathrm{e}^{-g\tau/c}).$$

With the boundary conditions $t=0$, $\tau=0$; $u^1=0$, $\dfrac{\mathrm{d}u^1}{\mathrm{d}\tau} = a^1 = g$ we find $\mathrm{A} = -\,\mathrm{B} = c/2$ and hence $u^1 = \dfrac{\mathrm{d}x}{\mathrm{d}\tau} = c \sinh(g\tau/c)$. Equation (2.46) then gives

$$a^0 = \frac{\mathrm{d}u^0}{\mathrm{d}\tau} = g \sinh(g\tau/c),$$

hence $u^0 = c\dfrac{\mathrm{d}t}{\mathrm{d}\tau} = c \cosh(g\tau/c)$, and finally

$$x = \frac{c^2}{g}\cosh(g\tau/c), \quad ct = \frac{c^2}{g}\sinh(g\tau/c). \tag{2.49}$$

The space and time coordinates then fall on the hyperbola

$$x^2 - c^2t^2 = \frac{c^4}{g^2} \tag{2.50}$$

sketched in Fig. 2.3. The non-relativistic limits (i.e. $g\tau / c \ll 1$) of x and t above are

$$t = \tau, \quad x = c^2/g + {}^1\!/\!{}_2\, gt^2.$$

We may now return to the twin paradox. We saw that the amount of proper time elapsing for B, travelling at a constant speed $v=3c/5$, was 20 years for each of the journeys to and from the star. The remaining question was, how much proper time elapses while B reverses her velocity from $+\,v$ to $-v$? If this is achieved with a constant acceleration a, then we have from (2.49) $\dfrac{\mathrm{d}x}{\mathrm{d}\tau} = c \sinh(a\tau/c) = \dfrac{3}{5}c$, hence

$$\tau = \frac{c}{a}\sinh^{-1} 0.6 \approx \frac{0.55c}{a}.$$