COSMOLOGICAL PHYSICS

J. A. PEACOCK

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011–4211, USA
10 Stamford Road, Oakleigh, VIC 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 1999
Reprinted 1999 (with corrections), 2000, 2001 (with corrections), 2002, 2003 (with corrections)

Printed in the United Kingdom at the University Press, Cambridge

Typeface Monotype Times $9\frac{1}{2}$/12pt System TEX [UPH]

A catalogue record for this book is available from the British Library

Library of Congress Cataloguing in Publication data
Peacock, John A.
Cosmological physics / J.A. Peacock.
p. cm.
Includes bibliographical references and index.
ISBN 0 521 41072 X (hardcover)
1. Cosmology. 2. Astrophysics. I. Title.
QB981.P37 1999
523.1–dc21 98-29460 CIP

ISBN 0 521 41072 X hardback
ISBN 0 521 42270 1 paperback

Contents

Part 4: The early universe

Part 5: Observational cosmology

1 Essentials of general relativity

1.1 The concepts of general relativity

SPECIAL RELATIVITY To understand the issues involved in general relativity, it is helpful to begin with a brief summary of the way space and time are treated in special relativity. The latter theory is an elaboration of the intuitive point of view that the properties of empty space should be the same throughout the universe. This is just a generalization of everyday experience: the world in our vicinity looks much the same whether we are stationary or in motion (leaving aside the inertial forces experienced by accelerated observers, to which we will return shortly).

The immediate consequence of this assumption is that any process that depends only on the properties of empty space must appear the same to all observers: the velocity of light or gravitational radiation should be a constant. The development of special relativity can of course proceed from the experimental constancy of c, as revealed by the Michelson-Morley experiment, but it is worth noting that Einstein considered the result of this experiment to be inevitable on intuitive grounds (see Pais 1982 for a detailed account of the conceptual development of relativity). Despite the mathematical complexity that can result, general relativity is at heart a highly intuitive theory; the way in which our everyday experience can be generalized to deduce the large-scale structure of the universe is one of the most magical parts of physics. The most important concepts of the theory can be dealt with without requiring much mathematical sophistication, and we begin with these physical fundamentals.

4-VECTORS From the constancy of c, it is simple to show that the only possible linear transformation relating the coordinates measured by different observers is the Lorentz transformation:

$$\boxed{\begin{aligned} dx' &= \gamma\left(dx - \frac{v}{c}c\,dt\right) \\ c\,dt' &= \gamma\left(c\,dt - \frac{v}{c}dx\right). \end{aligned}} \tag{1.1}$$

Note that this is written in a form that makes it explicit that x and ct are treated in the same way. To reflect this interchangeability of space and time, and the absence of any preferred frame, we say that special relativity requires all true physical relations to be written in terms of **4-vectors**. An equation valid for one observer will then apply to all others because the quantities on either side of the equation will transform in the same way. We ensure that this is so by constructing physical 4-vectors out of the fundamental interval

$$dx^\mu = (c\,dt, dx, dy, dz) \quad \mu = 0, 1, 2, 3, \tag{1.2}$$

by manipulations with relativistic invariants such as rest mass m and proper time $d\tau$, where

$$(c\,d\tau)^2 = (c\,dt)^2 - (dx^2 + dy^2 + dz^2). \tag{1.3}$$

Thus, defining the 4-momentum $P^\mu = m\,dx^\mu/d\tau$ allows an immediate relativistic generalization of conservation of mass and momentum, since the equation $\Delta P^\mu = 0$ reduces to these laws for an observer who sees a set of slowly moving particles. This is a very powerful principle, as it allows us to reject 'obviously wrong' physical laws at sight. For example, Newton's second law $\mathbf{F} = m\,d\mathbf{u}/dt$ is not a relation between the spatial components of two 4-vectors. The obvious way to define 4-force is $F^\mu = dP^\mu/d\tau$, but where does the 3-force $\mathbf{F}$ sit in F^μ? Force will still be defined as rate of change of momentum, $\mathbf{F} = d\mathbf{P}/dt$; the required components of F^μ are $\gamma(\dot{E}, \mathbf{F})$, and the correct relativistic force–acceleration relation is

$$\mathbf{F} = m\frac{d}{dt}(\gamma\mathbf{u}). \tag{1.4}$$

Note again that the symbol m denotes the **rest mass** of the particle, which is one of the invariant scalar quantities of special relativity. The whole ethos of special relativity is that, in the frame in which a particle is at rest, its intrinsic properties such as mass are always the same, independently of how fast it is moving. The general way in which quantities are calculated in relativity is to evaluate them in the rest frame where things are simple, and then to transform out into the lab frame.

GENERAL RELATIVITY Nothing that has been said so far seems to depend on whether or not observers move at constant velocity. We have in fact already dealt with the main principle of general relativity, which states that the only valid physical laws are those that equate two quantities that transform in the same way under any arbitrary change of coordinates.

Before getting too pleased with ourselves, we should ask how we are going to construct general analogues of 4-vectors. Consider how the components of dx^μ transform under the adoption of a new set of coordinates x'^μ, which are functions of x^ν:

$$\boxed{dx'^\mu = \frac{\partial x'^\mu}{\partial x^\nu}dx^\nu.} \tag{1.5}$$

This apparently trivial equation (which assumes, as usual, the summation convention on repeated indices) may be divided by $d\tau$ on either side to obtain a similar transformation law for 4-velocity, U^μ; so U^μ is a general 4-vector. Things unfortunately go wrong at the next level, when we try to differentiate this new equation to form the 4-acceleration $A^\mu = dU^\mu/d\tau$:

$$A'^\mu = \frac{\partial x'^\mu}{\partial x^\nu}A^\nu + \frac{\partial^2 x'^\mu}{\partial\tau\,\partial x^\nu}U^\nu. \tag{1.6}$$

The second term on the right-hand side (rhs) is zero only when the transformation coefficients are constants. This is so for the Lorentz transformation, but not in general. The conclusion is therefore that $F^\mu = dP^\mu/d\tau$ cannot be a general law of physics, since $dP^\mu/d\tau$ is not a general 4-vector.

INERTIAL FRAMES AND MACH'S PRINCIPLE We have just deduced in a rather cumbersome fashion the familiar fact that $\mathbf{F} = m\mathbf{a}$ only applies in inertial frames of reference. What exactly are these? There is a well-known circularity in Newtonian mechanics, in that inertial frames are effectively defined as being those sets of observers for whom $\mathbf{F} = m\mathbf{a}$ applies. The circularity is only broken by supplying some independent information about $\mathbf{F}$ – for example, the Lorentz force $\mathbf{F} = e(\mathbf{E} + \mathbf{v}\wedge\mathbf{B})$ in the case of a charged particle. This leaves us in a rather unsatisfactory situation: $\mathbf{F} = m\mathbf{a}$ is really only a statement about cause and effect, so the existence of non-inertial frames comes down to saying that there can be a motion with no apparent cause. Now, it is well known that $\mathbf{F} = m\mathbf{a}$ can be made to apply in all frames if certain 'fictitious' forces are allowed to operate. In respectively uniformly accelerating and rotating frames, we would write

$$\begin{aligned} \mathbf{F} &= m\mathbf{a} + m\mathbf{g} \\ \mathbf{F} &= m\mathbf{a} + m\boldsymbol{\Omega}\wedge(\boldsymbol{\Omega}\wedge\mathbf{r}) - 2m(\mathbf{v}\wedge\boldsymbol{\Omega}) + m\dot{\boldsymbol{\Omega}}\wedge\mathbf{r}. \end{aligned} \tag{1.7}$$

The fact that these 'forces' have simple expressions is tantalizing: it suggests that they should have a direct explanation, rather than taking the Newtonian view that they arise from an incorrect choice of reference frame. The relativist's attitude will be that if our physical laws are correct, they should account for what observers see from any arbitrary point of view – however perverse.

The mystery of inertial frames is deepened by a fact of which Newton was well aware, but did not explain: an inertial frame is one in which the bulk of matter in the universe is at rest. This observation was taken up in 1872 by Ernst Mach. He argued that since the acceleration of particles can only be measured relative to other matter in the universe, the existence of inertia for a particle must depend on the existence of other matter. This idea has become known as **Mach's principle**, and was a strong influence on Einstein in formulating general relativity. In fact, Mach's ideas ended up very much in conflict with Einstein's eventual theory – most crucially, the rest mass of a particle is a relativistic invariant, independent of the gravitational environment in which a particle finds itself. However, controversy still arises in debating whether general relativity is truly a 'Machian' theory – i.e. one in which the rest frame of the large-scale matter distribution is inevitably an inertial frame (e.g. Raine & Heller 1981).

A hint at the answer to this question comes by returning to the expressions for the inertial forces. The most satisfactory outcome would be to dispose of the notion of inertial frames altogether, and to find a direct physical mechanism for generating 'fictitious' forces. Following this route in fact leads us to conclude that Newtonian gravitation cannot be correct, and that the inertial forces can be effectively attributed to gravitational radiation. Since we cannot at this stage give a correct relativistic argument, consider the analogy with electromagnetism. At large distances, an accelerating charge produces an electric field given by

$$\mathbf{E} = \frac{e}{4\pi\epsilon_0 r c^2}(\hat{\mathbf{r}}\wedge[\mathbf{a}])\wedge\hat{\mathbf{r}}, \tag{1.8}$$

i.e. with components parallel to the retarded acceleration $[\mathbf{a}]$ and perpendicular to the acceleration axis. A charge distribution symmetric about a given point will then generate a net force on a particle at that point in the direction of **a**. It is highly plausible that something similar goes on in the generation of inertial forces via gravity, and we can guess the magnitude by letting $e/(4\pi\epsilon_0) \to Gm$. This argument was proposed by Dennis Sciama, and is known as **inertial induction**. Integrating such a force over all mass in a spherically symmetric universe, we get a total of

$$\frac{F_{\rm tot}}{m} = 2\pi \frac{Ga}{c^2} \int_0^{c/H_0} \int_0^{\pi} \rho\, r \, \sin^3\theta \; d\theta \; dr \; = a \, \frac{\pi^2 G\rho}{2H_0^2}. \tag{1.9}$$

This calculation is rough in many respects. The main deficiency is the failure to include the expansion of the universe: objects at a vector distance **r** appear to recede from us at a velocity $\mathbf{v} = H_0\mathbf{r}$, where H_0 is known as Hubble's constant (and is not constant at all, as will become apparent later). This law is only strictly valid at small distances, of course, but it does tell us that objects with $r \simeq c/H_0$ recede at a speed approaching that of light. This is why it seems reasonable to use this as an upper cutoff in the radial part of the above integral. Having done this, we obtain a total acceleration induced by gravitational radiation that is roughly equal to the acceleration we first thought of (the dimensionless factor on the rhs of the above equation is known experimentally to be unity to within a factor 10 or so). Thus, it does seem qualitatively valid to think of inertial forces as arising from gravitational radiation. Apart from being a startlingly different view of what is going on in non-inertial frames, this argument also sheds light on Mach's principle: for a symmetric universe, inertial forces clearly vanish in the average rest frame of the matter distribution. Frames in constant relative motion are allowed because (in this analogy) a uniformly moving charge does not radiate.

It is not worth trying to make this calculation more precise, as the approach is not really even close to being a correct relativistic treatment. Nevertheless, it does illustrate very well the prime characteristic of relativistic thought: we must be able to explain what we see from any point of view.

THE EQUIVALENCE PRINCIPLE In the previous subsection, we were trying to understand the non-inertial effects that are seen in accelerating reference frames as being gravitational in origin. In fact, it is more conventional to state this equivalence the other way around, saying that gravitational effects are identical in nature to those arising through acceleration. The seed for this idea goes back to the observation by Galileo that bodies fall at a rate independent of mass. In Newtonian terms, the acceleration of a body in a gravitational field **g** is

$$m_{\rm I}\, \mathbf{a} = m_{\rm G}\, \mathbf{g}, \tag{1.10}$$

and no experiment has ever been able to detect a difference between the inertial and gravitational masses $m_{\rm I}$ and $m_{\rm G}$ (the equality holds to better than 1 part in 10^{11}: Will 1993). This equality is trivially obvious in the case of inertial forces, and the apparent gravitational acceleration **g** becomes simply the acceleration of the frame **a**. These considerations led Einstein to suggest that inertial and gravitational forces were indeed one and the same. Formally, this leads us to the equivalence principle, which comes in two forms.

The **weak equivalence principle** is a statement only about space and time. It says that in any gravitational field, however strong, a freely falling observer will experience no gravitational effects – with the important exception of tidal forces in non-uniform fields. The spacetime will be that of special relativity (known as **Minkowski spacetime**).

The **strong equivalence principle** takes this a stage further and asserts that not only is the spacetime as in special relativity, but all the laws of physics take the same form in the freely falling frame as they would in the absence of gravity. This form of the equivalence principle is crucial in that it will allow us to deduce the generally valid laws governing physics once the special-relativistic forms are known. Note however that it is less easy to design experiments that can *test* the strong equivalence principle (see chapter 8 of Will 1993).

It may seem that we have actually returned to something like the Newtonian viewpoint: gravitation is merely an artifact of looking at things from the 'wrong' point of view. This is not really so; rather, the important aspects of gravitation are not so much to do with first-order effects as second-order tidal forces: these cannot be transformed away and are the true signature of gravitating mass. However, it is certainly true in one sense to say that gravity is *not* a real force: the gravitational acceleration is not derived from a 4-force F^μ and transforms differently.

GRAVITATIONAL TIME DILATION Many of the important features of general relativity can be obtained via rather simple arguments that use the equivalence principle. The most famous of these is the thought experiment that leads to gravitational time dilation, illustrated in figure 1.1. Consider an accelerating frame, which is conventionally a rocket of height h, with a clock mounted on the roof that regularly disgorges photons towards the floor. If the rocket accelerates upwards at g, the floor acquires a speed $v = gh/c$ in the time taken for a photon to travel from roof to floor. There will thus be a blueshift in the frequency of received photons, given by $\Delta\nu/\nu = gh/c^2$, and it is easy to see that the rate of reception of photons will increase by the same factor.

Now, since the rocket can be kept accelerating for as long as we like, and since photons cannot be stockpiled anywhere, the conclusion of an observer on the floor of the rocket is that in a real sense the clock on the roof is running fast. When the rocket stops accelerating, the clock on the roof will have gained a time Δt by comparison with an identical clock kept on the floor. Finally, the equivalence principle can be brought in to conclude that gravity must cause the same effect. Noting that $\Delta\phi = gh$ is the difference in potential between roof and floor, it is simple to generalize this to

$$\boxed{\frac{\Delta t}{t} = \frac{\Delta\phi}{c^2}.} \tag{1.11}$$

The same thought experiment can also be used to show that light must be deflected in a gravitational field: consider a ray that crosses the rocket cabin horizontally when stationary. This track will appear curved when the rocket accelerates.

The experimental demonstration of the gravitational redshift by Pound & Rebka (1960) was one of the main pieces of evidence for the essential correctness of the above reasoning, and provides a test (although not the most powerful one) of the equivalence principle.

THE TWIN PARADOX One of the neatest illustrations of gravitational time dilation is in resolving the twin paradox. This involves twins A and B, each equipped with a clock.

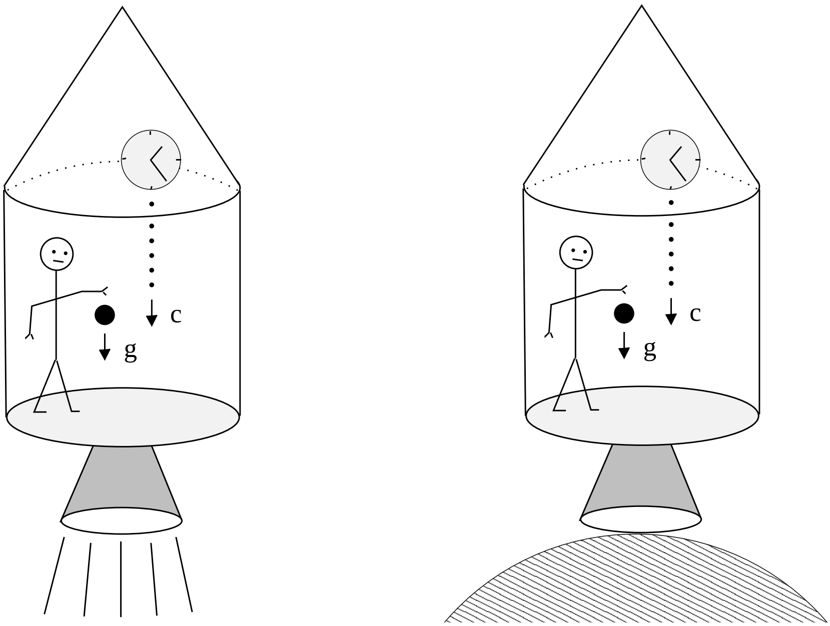

Figure 1.1. Imagine you are in a box in free space far from any source of gravitation. If the box is made to accelerate 'upwards' and has a clock that emits a photon every second mounted on its roof, it is easy to see that you will receive photons more rapidly once the box accelerates (imagine yourself running into the line of oncoming photons). Now, according to the equivalence principle, the situation is exactly equivalent to the second picture in which the box sits at rest on the surface of the Earth. Since there is nowhere for the excess photons to accumulate, the conclusion has to be that clocks above us in a gravitational field run fast.

A remains on Earth, while B travels a distance d on a rocket at velocity v, fires the engines briefly to reverse the rocket's velocity, and returns. The standard analysis of this situation in special relativity concludes, correctly, that A's clock will indicate a longer time for the journey than B's:

$$t_{\rm A} = \gamma\, t_{\rm B}. \tag{1.12}$$

The so-called paradox lies in the broken symmetry between the twins. There are various resolutions of this puzzle, but these generally refuse to meet the problem head-on by analysing things from B's point of view. However, at least for small v, it is easy to do this using the equivalence principle. There are three stages to consider:

(1) Outward trip. According to B, in special relativity A's clock runs slow: $t_{\rm A} = \gamma^{-1} t_{\rm B} \simeq [1 - v^2/(2c^2)](d/v)$.

(2) Return trip. Similarly, A's clock runs slow, resulting in a total lag with respect to B's of $(v^2/c^2)(d/v) = vd/c^2$.

(3) In between comes the crucial phase of turning. During this time, B's frame is non-inertial; there is an apparent gravitational field causing A to halt and start to return to B (at least, what else is B to conclude? There is obviously a force acting on the Earth, but the Earth is clearly not equipped with rockets). If an acceleration g operates for a time $t_{\rm turn}$, then A's clock will run fast by a fractional amount gd/c^2, leading to a total time step of $gdt_{\rm turn}/c^2 = 2vd/c^2$ (since $gt_{\rm turn} = 2v$).

Thus, in total, B returns to find A's clock in advance of B's by an amount

$$t_{\rm A} - t_{\rm B} = -\frac{vd}{c^2} + \frac{2vd}{c^2} \simeq (\gamma - 1)t_{\rm B}, \tag{1.13}$$

exactly (for small v) in accordance with A's entirely special relativity calculation.

1.2 The equation of motion

It was mentioned above that the equivalence principle allows us to bootstrap our way from physics in Minkowski spacetime to general laws. We can in fact obtain the full equations of general relativity in this way, in an approach pioneered by Weinberg (1972). In what follows, note the following conventions: Greek indices run from 0 to 3 (spacetime), Roman from 1 to 3 (spatial). The summation convention on repeated indices of either type is assumed.

Consider freely falling observers, who erect a special-relativity coordinate frame ξ^μ in their neighbourhood. The equation of motion for nearby particles is simple:

$$\frac{d^2\xi^\mu}{d\tau^2} = 0; \qquad \xi^\mu = (ct, x, y, z), \tag{1.14}$$

i.e. they have zero acceleration, and we have Minkowski spacetime

$$c^2 d\tau^2 = \eta_{\alpha\beta}\, d\xi^\alpha d\xi^\beta, \tag{1.15}$$

where $\eta_{\alpha\beta}$ is just a diagonal matrix $\eta_{\alpha\beta} = \mathrm{diag}(1, -1, -1, -1)$. Now suppose the observers make a transformation to some other set of coordinates x^μ. What results is the perfectly general relation

$$d\xi^\mu = \frac{\partial \xi^\mu}{\partial x^\nu} dx^\nu, \tag{1.16}$$

which on substitution leads to the two principal equations of dynamics in general relativity:

$$\boxed{\begin{aligned} &\frac{d^2x^\mu}{d\tau^2} + \Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau} = 0 \\ &\qquad\qquad c^2 d\tau^2 = g_{\alpha\beta}\, dx^\alpha\, dx^\beta. \end{aligned}} \tag{1.17}$$

At this stage, the new quantities appearing in these equations are defined only in terms of our transformation coefficients:

$$\begin{aligned} \Gamma^\mu_{\alpha\beta} &= \frac{\partial x^\mu}{\partial \xi^\nu}\frac{\partial^2 \xi^\nu}{\partial x^\alpha \partial x^\beta} \\ g_{\mu\nu} &= \frac{\partial \xi^\alpha}{\partial x^\mu}\frac{\partial \xi^\beta}{\partial x^\nu}\,\eta_{\alpha\beta}. \end{aligned} \tag{1.18}$$

COORDINATE TRANSFORMATIONS What is the physical meaning of this analysis? We have taken the special relativity equations for motion and the structure of spacetime and looked at the effects of a general coordinate transformation. One example of such a transformation is a Lorentz boost to some other inertial frame. However, this is not very interesting since we know in advance that the equations retain their form in this case (it is easy to show that $\Gamma^{\mu}_{\alpha\beta} = 0$ and $g_{\mu\nu} = \eta_{\mu\nu}$). A more general transformation could be one to the frame of an accelerating observer, but the transformation might have *no* direct physical interpretation at all. It is important to realize that general relativity makes no distinction between coordinate transformations associated with motion of the observer and a simple change of variable. For example, we might decide that henceforth we will write down coordinates in the order (x, y, z, ct) rather than (ct, x, y, z) (as is indeed the case in some formalisms). General relativity can cope with these changes automatically. Indeed, this flexibility of the theory is something of a problem: it can sometimes be hard to see when some feature of a problem is 'real', or just an artifact of the coordinates adopted. People attempt to distinguish this second type of coordinate change by distinguishing between 'active' and 'passive' Lorentz transformations; a more common term for the latter class is **gauge transformation**. The term gauge will occur often throughout this book: it always refers to some freedom within a theory that has no observable consequence (e.g. the arbitrary value of $\boldsymbol{\nabla} \cdot \mathbf{A}$, where $\mathbf{A}$ is the vector potential in electrodynamics).

METRIC AND CONNECTION The matrix $g_{\mu\nu}$ is known as the **metric tensor**. It expresses (in the sense of special relativity) a notion of distance between spacetime points. Although this is a feature of many spaces commonly used in physics, it is easy to think of cases where such a measure does not exist (for example, in a plot of particle masses against charges, there is no physical meaning to the distance between points). The fact that spacetime *is* endowed with a metric is in fact something that has been *deduced*, as a consequence of special relativity and the equivalence principle. Given a metric, Minkowski spacetime appears as an inevitable special case: if the matrix $g_{\mu\nu}$ is symmetric, we know that there must exist a coordinate transformation that makes the matrix diagonal:

$$\tilde{\mathbf{\Lambda}}\mathbf{g}\mathbf{\Lambda} = \mathrm{diag}(\lambda_0, \ldots, \lambda_3), \tag{1.19}$$

where $\mathbf{\Lambda}$ is the matrix of transformation coefficients, and λ_i are the eigenvalues of this matrix.

The object $g_{\mu\nu}$ is called a **tensor**, since it occurs in an equation $c^2 d\tau^2 = g_{\mu\nu} dx^\mu dx^\nu$ that must be valid in all frames. In order for this to be so, the components of the matrix $\mathbf{g}$ must obey certain transformation relations under a change of coordinates. This is one way of defining a tensor, an issue that is discussed in detail below.

So much for the metric tensor, what is the meaning of the coefficients $\Gamma^{\mu}_{\alpha\beta}$? These are known as components of the **affine connection** or as **Christoffel symbols** (and are sometimes written in the alternative notation $\left\{ {\mu \atop \alpha\beta} \right\}$). These quantities obviously correspond roughly to the gravitational force – but what determines whether such a force exists? The answer is that gravitational acceleration depends on spatial change in the metric. For a simple example, consider gravitational time dilation in a weak field: for events at the same spatial position, there must be a separation in proper time of

$$d\tau \simeq dt \left(1 + \frac{\Delta\phi}{c^2} \right). \tag{1.20}$$

This suggests that the gravitational acceleration should be obtained via

$$\mathbf{a} = -\frac{c^2}{2}\boldsymbol{\nabla} g_{00}. \tag{1.21}$$

More generally, we can differentiate the equation for $g_{\mu\nu}$ to get

$$\frac{\partial g_{\mu\nu}}{\partial x^\lambda} = \Gamma^\alpha_{\lambda\mu} g_{\alpha\nu} + \Gamma^\beta_{\lambda\nu} g_{\beta\mu}. \tag{1.22}$$

Using the symmetry of the Γ's in their lower indices, and defining $g^{\mu\nu}$ to be the matrix inverse to $g_{\mu\nu}$, we can find an equation for the Γ's directly in terms of the metric tensor:

$$\boxed{\Gamma^\alpha_{\lambda\mu} = \tfrac{1}{2} g^{\alpha\nu}\left(\frac{\partial g_{\mu\nu}}{\partial x^\lambda} + \frac{\partial g_{\lambda\nu}}{\partial x^\mu} - \frac{\partial g_{\mu\lambda}}{\partial x^\nu}\right).} \tag{1.23}$$

Thus, the metric tensor is the crucial object in general relativity: given it, we know both the structure of spacetime and how particles will move.

1.3 Tensors and relativity

Before proceeding further, the above rather intuitive treatment should be set on a slightly firmer mathematical foundation. There are a variety of possible approaches one can take, which differ sufficiently that general relativity texts for physicists and mathematicians sometimes scarcely seem to refer to the same subject. For now, we stick with a rather old-fashioned approach, which has the virtue that it is likely to be familiar. Amends will be made later.

COVARIANT AND CONTRAVARIANT COMPONENTS So far, tensors have been met in their role as quantities that provide generally valid relations between different 4-vectors. If such relations are to be physically useful, they must apply in different frames of reference, and so the components of tensors have to change to compensate for the fact that the components of 4-vectors alter under a coordinate transformation. The transformation law for tensors is obtained from that for 4-vectors. For example, consider $c^2 d\tau^2 = g_{\alpha\beta}\, dx^\alpha dx^\beta$: substitute for dx^μ in terms of dx'^α and require that the resulting equation must have the form $c^2 d\tau^2 = g'_{\alpha\beta}\, dx'^\alpha dx'^\beta$. We then deduce the tensor transformation law

$$\boxed{g'_{\alpha\beta} = \frac{\partial x^\mu}{\partial x'^\alpha}\,\frac{\partial x^\nu}{\partial x'^\beta}\; g_{\mu\nu},} \tag{1.24}$$

of which law our above definition of $g_{\mu\nu}$ in terms of $\eta_{\alpha\beta}$ is an example.

Note that this transformation law looks rather like a generalization of that for a single 4-vector (with one transformation coefficient per index), but with the important difference that the coefficients are upside down in the tensor relation. For Cartesian coordinates, this would make no difference:

$$\frac{\partial x^\mu}{\partial x'^\alpha} = \frac{\partial x'^\alpha}{\partial x^\mu} = \cos\theta, \tag{1.25}$$

where θ is the angle of rotation between the two coordinate axes. In general, though, the transformations are not the same. To illustrate this, consider a set of non-orthogonal basis vectors $\mathbf{e}_i$: there are two ways to define the components of a vector $\mathbf{a}$:

$$\begin{aligned} &(1) \quad \mathbf{a} = \sum a^i \mathbf{e}_i \\ &(2) \quad a_i = \mathbf{a} \cdot \mathbf{e}_i. \end{aligned} \tag{1.26}$$

These clearly differ in general if $\mathbf{e}_i \cdot \mathbf{e}_j \neq 0$, and they are distinguished by writing an index 'upstairs' on one and 'downstairs' on the other.

Something very similar goes on in general relativity. If we define a new vector

$$dx_\nu \equiv g_{\alpha\nu} dx^\alpha, \tag{1.27}$$

then our metric is given by

$$c^2 d\tau^2 = dx_\nu dx^\nu. \tag{1.28}$$

In special relativity, this would yield just $x_\mu = (ct, -x, -y, -z)$. In general relativity, it defines the relations between the **contravariant components** of a 4-vector A^μ and the **covariant components** A_μ. These names reflect that the components transform either in the *same* way as basis vectors (covariant) or oppositely (contravariant). The relevant transformation laws are

$$\boxed{\begin{aligned} A'^\mu &= \frac{\partial x'^\mu}{\partial x^\nu} A^\nu \\ A'_\mu &= \frac{\partial x^\nu}{\partial x'^\mu} A_\nu. \end{aligned}} \tag{1.29}$$

This generalizes to any tensor, by multiplying by the appropriate factor for each index.

INVARIANTS To summarize the above arguments, one can only construct an invariant quantity in general relativity (i.e. one that is the same for all observers) by **contracting** vector or tensor indices in pairs: $A^\mu A_\mu$ is the invariant 'size' or **norm** of the vector A^μ. However, $A^\mu A^\mu$ would not be a constant, since the effects of arbitrary coordinate changes do not cancel out unless upstairs and downstairs indices contract with each other.

This sounds like a tedious complication, but it can be turned to advantage. Suppose we are given an equation such as $A^\mu B_\mu = 1$, and that A^μ is known to be a 4-vector. Clearly, the right-hand side of the equation is invariant, and so the only way in which this can happen in general is if B_μ is also a 4-vector. This trick of deducing the nature of quantities in a relativistic equation is called the principle of **manifest covariance**. As an example, consider the coordinate derivative, for which there exists the common shorthand

$$\boxed{\partial_\mu \equiv \frac{\partial}{\partial x^\mu}.} \tag{1.30}$$

The index must be 'downstairs' since the derivative operates only on one coordinate:

$$\partial_\mu x^\nu = \delta^\nu_\mu = \mathrm{diag}(1,1,1,1). \tag{1.31}$$

Thus, $\partial_\mu x^\mu = 4$ is an invariant, justifying the use of a downstairs index for ∂_μ. More generally, the tensor δ^ν_μ must be an **isotropic tensor** (meaning one whose components are the same in all frames), since it must exist in order to define the inverse matrix to some tensor $T^{\mu\nu}$. Indeed, we have already met an example of this in writing the inverse matrix to $g_{\mu\nu}$ as a contravariant tensor:

$$\boxed{g^{\mu\alpha} g_{\mu\beta} = \delta^\alpha_\beta.} \tag{1.32}$$

The metric tensor is therefore the tool that is used to raise and lower indices, so that

$$\boxed{A_\mu \equiv g_{\mu\nu} A^\nu.} \tag{1.33}$$

For example, in special relativity, the 4-derivatives are therefore

$$\begin{aligned} \partial_\mu &= \left(\frac{\partial}{\partial ct}, \boldsymbol{\nabla} \right) \\ \partial^\mu &= \left(\frac{\partial}{\partial ct}, -\boldsymbol{\nabla} \right). \end{aligned} \tag{1.34}$$

Manifest covariance allows quantities like the **4-current** $J^\mu = (c\rho, \mathbf{j})$ to be recognized as 4-vectors, since they allow the conservation law to be written relativistically: $\partial^\mu J_\mu = 0$.

To summarize, tensor equations with indices in the same relative positions on either side of the expression must be generally valid. An unfortunate term is used for this: the equations are said to be **generally covariant** – i.e. to have the same form for all observers. This has nothing to do with the usage of the term when referring to covariant vectors; it is a historical accident with which one simply has to live.

PSEUDOTENSORS AND TENSOR DENSITIES If we regard a second-rank tensor as a matrix, there is another familiar way of forming a number. In addition to tensor contraction, we can also take the determinant:

$$\boxed{g \equiv -\det g_{\mu\nu}.} \tag{1.35}$$

This is *not* an invariant scalar; thinking of tensor transformations in matrix terms ($\mathbf{g}' = \tilde{\mathbf{\Lambda}}\mathbf{g}\mathbf{\Lambda}$) shows that g' depends on the Jacobian of the coordinate transformation:

$$g' = \left| \frac{\partial x'^\mu}{\partial x^\nu} \right|^{-2} g. \tag{1.36}$$

The reason why this quantity arises in relativity comes from volume elements: under a general coordinate transformation, the hypervolume element behaves as

$$d^4 x'^\mu = \left\| \frac{\partial x'^\alpha}{\partial x^\beta} \right\| d^4 x^\mu, \tag{1.37}$$

so that an invariant normalization of some scalar ρ can only be constructed via

$$\int \sqrt{-g}\, \rho \, d^4 x^\mu = \text{constant}. \tag{1.38}$$

The quantity $\sqrt{-g}\,\rho$ is referred to as a **scalar density**. More generally, an object formed from a tensor and n powers of $\sqrt{-g}$ is called a **tensor density** of weight n.

PROPER AND IMPROPER TRANSFORMATIONS One important consequence of the existence of tensor densities arises when considering coordinate transformations that involve a spatial reflection. It is usual to distinguish between different classes of Lorentz transformations according to the sign of their corresponding Jacobians: **proper Lorentz transformations** have $J > 0$, whereas those with negative Jacobians are termed **improper**. In special relativity, where $g = -1$ always, there are two possibilities: $J = \pm 1$. Thus, a tensor density will in special relativity transform like a tensor if we restrict ourselves to proper transformations. However, on spatial inversion, densities of odd weight will change sign. Such quantities are referred to as **pseudotensors** (or, in special cases pseudovectors or pseudoscalars). The most famous example of this is the totally antisymmetric **Levi–Civita pseudotensor** $\epsilon^{\alpha\beta\gamma\delta}$, which has components $+1$ when $\alpha\beta\gamma\delta$ is an even permutation of 0123, -1 for odd permutations and zero otherwise. One can show explicitly that this frame-independent component definition produces a tensor density of weight -1 by applying the transformation law for such a quantity and verifying the invariance of the components (not too hard since ϵ enters into the definition of the determinant). Lowering indices with the metric tensor produces a covariant density of weight -1:

$$\epsilon_{\alpha\beta\gamma\delta} = g\epsilon^{\alpha\beta\gamma\delta}. \tag{1.39}$$

In special relativity, $\epsilon^{\alpha\beta\gamma\delta}$ is therefore of opposite sign to $\epsilon_{\alpha\beta\gamma\delta}$.

PHYSICS IN GENERAL RELATIVITY So far, we have dealt with how to generalize gravitational dynamics, but how are other parts of physics incorporated into general relativity? A hint at the answer is obtained by looking again at the equation of motion $d^2x^\mu/d\tau^2 + \Gamma^\mu_{\alpha\beta}(dx^\alpha/d\tau)(dx^\beta/d\tau) = 0$. Remembering that $d^2x^\mu/d\tau^2$ is not a general 4-vector, this equation must have added two non-vectors in such a way that the 'errors' in their transformation properties have cancelled to yield a covariant answer. We may say that the addition of the term containing the affine connection has made the equation **gauge invariant**. The term 'gauge' means that there are hidden degrees of freedom (coordinate transformations in this case) that do not affect physical observables.

In fact, we have been dealing with a special case of

$$\boxed{DA^\mu \equiv dA^\mu + \Gamma^\mu_{\alpha\beta}A^\alpha dx^\beta,} \tag{1.40}$$

which is known as the **covariant derivative**. The equation of motion under gravity is then most simply expressed by saying that the covariant derivative of 4-velocity vanishes: $DU^\mu/d\tau = 0$. One can show directly from the definition of the affine connection that the covariant derivative transforms as a 4-vector, but this is a rather messy exercise. It is simpler to use manifest covariance: the form of DU^μ was deduced by transforming the relation $dU^\mu/d\tau = 0$ from the local freely falling frame to a general frame. If $DU^\mu/d\tau$ vanishes in all frames, it must be a general 4-vector. It is immediately clear how to generalize other equations: simply replace ordinary derivatives by covariant ones. Thus, in the presence of non-gravitational forces, the equation of motion for a particle would become

$$m\frac{DU^\mu}{d\tau} = F^\mu. \tag{1.41}$$

There is a frequently used notation to simplify such substitutions. Coordinate partial derivatives may be represented by indices following a comma, and covariant derivatives by a semicolon:

$$\boxed{\begin{aligned} V^{\mu}{}_{,\nu} &\equiv \frac{\partial V^{\mu}}{\partial x^{\nu}} \equiv \partial_{\nu} V^{\mu} \\ V^{\mu}{}_{;\nu} &\equiv \frac{D V^{\mu}}{\partial x^{\nu}} \equiv V^{\mu}{}_{,\nu} + \Gamma^{\mu}_{\alpha\nu} V^{\alpha}, \end{aligned}} \tag{1.42}$$

and the notation extends in an obvious way: $V^{\mu}{}_{;\beta}^{\;\alpha} \equiv D^2 V^{\mu}/\partial x_{\alpha} \partial x^{\beta}$. General relativity thus introduces a simple 'comma goes to semicolon' rule for the bootstrapping of special relativity laws to generally covariant ones. The meaning and origin of the covariant derivative is discussed further below, where two generalizations are proved. The analogous result for the derivatives of covariant vectors is

$$DA_{\mu} \equiv dA_{\mu} - \Gamma^{\alpha}_{\mu\beta} A_{\alpha} dx^{\beta} \tag{1.43}$$

(note the opposite sign and different index arrangements).

This procedure is not completely foolproof, unfortunately. A manifestly covariant equation containing only covariant derivatives is clearly a possible generalization of the corresponding special relativity law, but it may not be unique. For example, since the Ricci tensor discussed below vanishes in special relativity, it can be introduced into almost any equation without destroying the special relativity limit: $T^{\mu\nu}_{;\nu} = R^{\mu\nu}_{;\nu}$ has the same limit as $T^{\mu\nu}_{;\nu} = 0$. The best to be said here is that the more complex law is inelegant and lacking in physical motivation; such more complex alternatives can only be ruled out with certainty by experiment. According to the **principle of minimal coupling** we leave out any such extra terms.

A more difficult case arises with equations containing more than one derivative: partial derivatives commute, but covariant derivatives do not:

$$V^{\mu}_{;\alpha\beta} - V^{\mu}_{;\beta\alpha} = R^{\mu}_{\nu\beta\alpha} V^{\nu}, \tag{1.44}$$

where $R^{\mu}_{\nu\beta\alpha}$ is the Riemann curvature tensor discussed below. There is no general solution to this problem, which is analogous to the ambiguity of operator orderings in quantum mechanics encountered when going from a classical equation to a quantum one. Sometimes the difficulty can be resolved by dealing with the derivatives one at a time. For example, the natural generalization of Maxwell's equations is clear in terms of the field tensor:

$$\begin{aligned} F^{\mu\nu}_{;\nu} &= -\mu_0 J^{\mu} \\ F_{\mu\nu} &= A_{\mu;\nu} - A_{\nu;\mu}, \end{aligned} \tag{1.45}$$

even though the combined equation for A^{μ} is ambiguous.

GEODESICS There is an important way of visualizing the meaning of the general relativity equation of motion. In special relativity a free particle travels along a straight line in space; in general relativity an analogous statement applies to the paths of particles in spacetime. These are *geodesics*: paths whose 'length' in spacetime is stationary with respect to small variations about them in the same way as small perturbations about a straight line in space will increase its total length. This is familiar from special relativity, where particles travel along paths of maximum proper time, as is easily shown by considering a particle that propagates from event A to event B at constant velocity. This motion is most simply viewed from the point of view of the rest frame of this particle. Now consider another path, which corresponds to motion in the frame of the first particle. Suppose the second particle reaches position x at time $t_{\rm C}$, and then returns to the origin, both legs of the journey being at constant speed. The proper time for this path is the sum of the proper times for each leg:

$$\tau' = \sqrt{(t_{\rm C} - t_{\rm A})^2 - x^2} + \sqrt{(t_{\rm B} - t_{\rm C})^2 - x^2} < t_{\rm B} - t_{\rm A}. \tag{1.46}$$

An arbitrary path can be made out of excursions of this sort, so the unaccelerated path has the maximum proper time. This is the special relativity 'solution' of the twin paradox, although it is actually an evasion, since it refuses to analyse things from the point of view of the accelerated observer. In any case, returning to general relativity, the equivalence principle says that a general path is locally a trajectory in Minkowski spacetime, so it is not surprising that the general path is also one in which the proper time is stationary with respect to variations of the path.

A slightly more formal approach is to express the particle dynamics in terms of an **action principle** and use the calculus of variations. Consider the equation

$$\delta \int L\, dp = 0, \tag{1.47}$$

where p is any parameter describing the path of the particle (t, τ etc.), and δ represents the effects of small variations of the path $x^\mu(p)$. The function L is the **Lagrangian** and Newtonian mechanics can be represented in this form with $L = T - V$, i.e. the difference of kinetic and potential energies for the particle. The principal result of variational calculus is obtained by assuming some arbitrary perturbation $\Delta x^\mu(p)$, expanding L in a Taylor series and integrating by parts for a path with fixed endpoints. This produces the **Euler equation**:

$$\frac{d}{dp}\left(\frac{\partial L}{\partial \dot{x}^\mu}\right) - \frac{\partial L}{\partial x^\mu} = 0, \tag{1.48}$$

where $\dot{x}^\mu$ denotes dx^μ/dp. In special relativity, this equation clearly yields the correct equation of motion for $p = t$, $L = \eta^{\mu\nu} U_\mu U_\nu$, which suggests a manifestly covariant generalization of the action principle:

$$\boxed{\delta \int g^{\mu\nu} U_\mu U_\nu \, d\tau = 0.} \tag{1.49}$$

If this intuitive derivation does not appeal, one can show directly, as an exercise in the calculus of variations, that this yields the correct equation of motion. Now, this formulation of the equations of motion in terms of an action principle has a rather peculiar aspect: the integrand is the Lagrangian, but it must be a constant because it is the norm of the 4-velocity, which is always c^2. Hence the geodesic equation says that the total proper time for a particle to move under gravity between two points in spacetime must be stationary (not necessarily a global maximum, because there may be several possible paths). In short, particles in curved spacetimes do their best to travel in straight lines locally, and it is only the global curvature of spacetime that produces the appearance of a gravitational force.

CONFORMAL TRANSFORMATIONS A related concept is that of making a **conformal transformation** of the metric,

$$g^{\alpha\beta} \to f(x^\mu) g^{\alpha\beta}. \tag{1.50}$$

The name 'conformal' arises by analogy with complex number theory, since such a transformation clearly preserves the 'angle' between two 4-vectors, $A^\mu B_\mu / \sqrt{A^\mu A_\mu B^\mu B_\mu}$. The importance of conformal transformations lies in the structure of null geodesics: clearly the condition $d\tau = 0$ is not affected by a conformal transformation. The paths of light rays are thus independent of these transformations; this result will be useful later in discussing cosmological light propagation.

1.4 The energy–momentum tensor

The only ingredient now missing from a classical theory of relativistic gravitation is a field equation: the presence of mass must determine the gravitational field. To obtain some insight into how this can be achieved, it is helpful to consider first the weak-field limit and the analogy with electromagnetism. The simplest limit of the theory is that of a stationary particle in a stationary (i.e. time-independent) weak field. To first order in the field we can replace τ by t, and the spatial part of the equation of motion is then

$$\ddot{x}^i + c^2 \Gamma^i_{00} = 0, \tag{1.51}$$

where $\Gamma^i_{00} = g^{\nu i}(0 + 0 - \partial g_{00}/\partial x^\nu)/2$. So, as we guessed before via a rough argument from time-dilation considerations, the equation of motion in this limit is

$$\boxed{\ddot{\mathbf{x}} = -\frac{c^2}{2} \boldsymbol{\nabla} g_{00}.} \tag{1.52}$$

THE ELECTROMAGNETIC ANALOGY For a moving particle, it is clear there will be velocity-dependent forces. Before dealing with these in detail, suppose we guess that the weak-field form of gravitation will look like electromagnetism, i.e. that we will end up working with both a scalar potential ϕ and a vector potential $\mathbf{A}$ that together give a velocity-dependent acceleration $\mathbf{a} = -\boldsymbol{\nabla}\phi - \dot{\mathbf{A}} + \mathbf{v} \wedge (\boldsymbol{\nabla} \wedge \mathbf{A})$. Making the usual $e/4\pi\epsilon_0 \to Gm$ substitution would suggest the field equation

$$\partial^\nu \partial_\nu A^\mu \equiv \Box A^\mu = \frac{4\pi G}{c^2} J^\mu, \tag{1.53}$$

where $\Box$ is the **d'Alembertian wave operator**, $A^\mu = (\phi/c, \mathbf{A})$ is the 4-potential and $J^\mu = (\rho c, \mathbf{j})$ is a quantity that resembles a 4-current, whose components are a mass density and mass flux density. The solution to this equation is well known:

$$A^\mu(\mathbf{r}) = \frac{G}{c^2} \int \frac{[J^\mu(\mathbf{x})]}{|\mathbf{r} - \mathbf{x}|} \, d^3x, \tag{1.54}$$

where the square brackets denote retarded values.

Now, in fact this analogy can be discarded immediately as a theory of gravitation in the weak-field limit without any knowledge whatsoever of general relativity. The problem lies in the vector J^μ: what would the meaning of such a quantity be? In electromagnetism, it describes conservation of charge via

$$\partial_\mu J^\mu = \dot{\rho} + \boldsymbol{\nabla} \cdot \mathbf{j} = 0 \tag{1.55}$$

(notice how neatly such a conservation law can be expressed in 4-vector form). When dealing with mechanics, however, we have not one conserved quantity, but *four*: energy and vector momentum. So, although J^μ is a perfectly good 4-vector mathematically, it is not *physically* relevant for describing conservation laws involving mass. For example, conservation laws involving J^μ predict that density will change under Lorentz transformations as $\rho \to \gamma\rho$, whereas the correct law is clearly $\rho \to \gamma^2\rho$ (one power of γ for change in number density, one for relativistic mass increase).

The electromagnetic analogy is nevertheless useful, as it suggests that the source of gravitation might still be mass and momentum: what we need first is to find the object that will correctly express conservation of 4-momentum. Informally, what is needed is a way of writing four conservation laws for each component of P^μ. We can clearly write four equations of the above type in matrix form:

$$\boxed{\partial_\nu T^{\mu\nu} = 0.} \tag{1.56}$$

Now, if this equation is to be covariant, $T^{\mu\nu}$ must be a tensor and is known as the **energy–momentum tensor** (or sometimes as the stress–energy tensor). The meanings of its components in words are: $T^{00} = c^2 \times$ (mass density) = energy density; T^{12} = x-component of current of y-momentum etc. From these definitions, the tensor is readily seen to be symmetric. Both momentum density and energy flux density are the product of a mass density and a net velocity, so $T^{0\mu} = T^{\mu 0}$. The spatial stress tensor T^{ij} is also symmetric because any small volume element would otherwise suffer infinite angular acceleration: any asymmetric stress acting on a cube of side L gives a couple $\propto L^3$, whereas the moment of inertia is $\propto L^5$.

For example, a cold fluid with density ρ_0 in its rest frame only has one non-zero component for the energy–momentum tensor: $T^{00} = c^2\rho_0$. Carrying out the Lorentz transformation, we conclude that the tensor's components in another frame are

$$T^{\mu\nu} = c^2\rho_0 \begin{pmatrix} \gamma^2 & -\gamma^2\beta & 0 & 0 \\ -\gamma^2\beta & \gamma^2\beta^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{1.57}$$

Thus, we obtain in one step quantities such as momentum density $= \gamma^2\rho_0 v$, that would be derived at a more basic level by transforming mass and number density separately.

PERFECT FLUID This line of argument can be taken a little further to obtain a very important result: the energy–momentum tensor for a perfect fluid. In matrix form, the rest-frame $T^{\mu\nu}$ is given by just $\mathrm{diag}(c^2\rho, p, p, p)$ (using the fact that the meaning of the pressure p is just the flux density of x-momentum in the x-direction etc.). We can bypass the step of carrying out an explicit Lorentz transformation (which would be rather cumbersome in this case) by the powerful technique of manifest covariance. The following expression is clearly a tensor and reduces to the above rest-frame answer in special relativity:

$$\boxed{T^{\mu\nu} = (\rho + p/c^2)U^\mu U^\nu - p g^{\mu\nu};} \tag{1.58}$$

thus it must be the general expression for the energy–momentum tensor of a perfect fluid. This is all that is needed in order to derive all the equations of relativistic fluid mechanics (see chapter 2).

1.5 The field equations

Armed with the energy–momentum tensor, we can now return to the search for the relativistic field equations. These cannot be derived in any rigorous sense; all that can be done is to follow Einstein and start by thinking about the simplest form such an equation might take. Our experience with an attempted electromagnetic analogy is again helpful. Consider Maxwell's equations: $\Box A^\mu = \mu_0 J^\mu$. The way in which each component of the 4-current acts as a source in a wave equation for one component of the potential can be generalized to matter if (at least in the weak-field limit) we are dealing with a tensor potential $\phi^{\mu\nu}$:

$$\Box \phi^{\mu\nu} = \kappa T^{\mu\nu}, \tag{1.59}$$

where κ is some constant. The (0, 0) component of this equation will just be Poisson's equation for the Newtonian potential in the case of a stationary field. Since we have already shown that $-c^2 g_{00}/2$ may be identified with this potential, there is a strong suspicion that $\phi^{\mu\nu}$ will be closely related to the metric tensor.

In general, we are looking for a covariant equation that reduces to the above in special relativity. There are many possibilities, but the starting point is to find the simplest alternative that does the job. The reasoning so far suggests that we are looking for a tensor that contains second derivatives of the metric, so why not consider $\partial^2 g_{\mu\nu}/\partial x_\alpha \partial x_\beta$? There are six such creatures, corresponding to the distinct ways in which four indices can be split into two pairs. As usual, such derivatives are not general tensors, and it is not so easy to cure this problem. Normally, one would replace ordinary derivatives with covariant ones and consider $g_{\mu\nu\,;\alpha\beta}$, but the covariant derivatives of the metric vanish identically (see below), so this is no good. It is now not obvious that a tensor can be constructed from second derivatives at all, but there is in fact one combination of the six second-derivative matrices that does work (with the addition of appropriate Γ terms). It is possible (although immensely tedious – see p. 133 of Weinberg 1972) to prove that this is the unique choice for a tensor that is *linear* in second derivatives of the metric.

The tensor in question is the **Riemann tensor**:

$$\boxed{R^\mu{}_{\alpha\beta\gamma} = \frac{\partial \Gamma^\mu_{\alpha\gamma}}{\partial x^\beta} - \frac{\partial \Gamma^\mu_{\alpha\beta}}{\partial x^\gamma} + \Gamma^\mu_{\sigma\beta}\Gamma^\sigma_{\gamma\alpha} - \Gamma^\mu_{\sigma\gamma}\Gamma^\sigma_{\beta\alpha}.} \tag{1.60}$$

A discussion of the full significance of this tensor will be postponed briefly, but we should note that it is reasonable that some such tensor must exist. The existence of a general metric says that spacetime is curved in a way that is revealed by non-zero second derivatives of $g^{\mu\nu}$. There has to be some covariant description of this curvature, and this is exactly what the Riemann tensor provides.

The Riemann tensor is fourth order, but may be contracted to the **Ricci tensor** $R^{\mu\nu}$, or further to the **curvature scalar** R:

$$R_{\alpha\beta} = R^{\mu}{}_{\alpha\beta\mu}, \quad R = R_{\mu}{}^{\mu} = g^{\mu\nu} R_{\mu\nu}. \tag{1.61}$$

Unfortunately, these definitions are not universally agreed, and different signs can arise in the final equations according to which convention is adopted (see below). All authors, however, agree on the definition of the **Einstein tensor** $G^{\mu\nu}$:

$$G^{\mu\nu} = R^{\mu\nu} - \tfrac{1}{2} g^{\mu\nu} R. \tag{1.62}$$

This tensor is what is needed, because it has zero covariant divergence [problem 1.6]:

$$G^{\mu\nu}_{;\nu} = \frac{DG^{\mu\nu}}{\partial x^{\nu}} = \frac{\partial G^{\mu\nu}}{\partial x^{\nu}} + \Gamma^{\mu}_{\alpha\nu} G^{\alpha\nu} + \Gamma^{\nu}_{\alpha\nu} G^{\mu\alpha} = 0. \tag{1.63}$$

Since we know that $T^{\mu\nu}$ also has zero covariant divergence by virtue of the conservation laws it expresses, it therefore seems reasonable to guess that the two are proportional:

$$\boxed{G^{\mu\nu} = -\frac{8\pi G}{c^4} T^{\mu\nu}.} \tag{1.64}$$

These are Einstein's gravitational field equations, where the correct constant of proportionality has been inserted. This is obtained below by considering the weak-field limit, where Einstein's theory must go over to Newtonian gravity.

PARALLEL TRANSPORT AND THE RIEMANN TENSOR First, however, we ought to take a closer look at the meaning of the crucial Riemann tensor. This has considerable significance in that it describes the degree of curvature of a space: it is the fact that the Riemann tensor is non-zero in general that produces the general relativity interpretation of gravitation as concerned with curved spacetime.

How do we tell whether a space is curved in general? Even the simple case of a surface embedded in 3D space can be tricky. Most people would agree that the surface of a sphere is a curved 2D space, but what about a cylinder? In the sense we are concerned with here, the surface of a cylinder is *not* curved: it can be obtained from a flat plane by bending the plane without folding or distorting it. In other words, the geodesics on a cylinder are exactly those that would apply if the cylinder were unrolled to make a plane; this is a hint of how to proceed in making a less intuitive assessment of curvature.

Gauss was the first to realize that curvature can be measured without the aid of a higher-dimensional being, by making use of the intrinsic properties of a surface. This is a familiar idea: the curvature of a sphere can be measured by examining a (small) triangle whose sides are great circles, and using the relation

$$\text{sum of interior angles} = \pi + 4\pi \frac{\text{area of triangle}}{\text{area of sphere}}. \tag{1.65}$$

Very small triangles have a sum of angles equal to π, but triangles of size comparable to

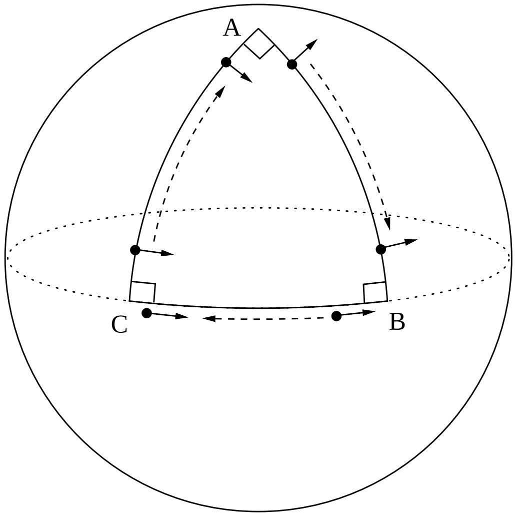

Figure 1.2. This figure illustrates the parallel transport of a vector around the closed loop ABC on the surface of a sphere. For the case of the spherical triangle with all angles equal to 90°, the vector rotates by 90° in one loop. This failure of vectors to realign under parallel transport is the fundamental signature of spatial curvature, and is used to define the affine connection and the Riemann tensor.

the radius of the sphere sample the curvature of the space, and the angular sum starts to differ from the Euclidean value.

To generalize this process, the concept of **parallel transport** is introduced. Here, imagine an observer travelling along some path, carrying with them some vector that is maintained parallel to itself as the observer moves. This is easy to imagine for small displacements, where a locally flat tangent frame can be used to apply the Euclidean concept of parallelism without difficulty. This is clearly a reversible process: a vector can be carried a large distance and back again along the same path, and will return to its original state. However, this need not be true in the case of a loop where the observer returns to the starting point along a different path, as illustrated in figure 1.2. In general, parallel transport around a loop will cause a change in a vector, and it is this that is the intrinsic signature of a curved space. We can think of the effect of parallel transport as producing a change in a vector proportional both to the vector itself (rotation), and to the distance along the loop (to first order), so that the total change in going once round a small loop can be written as

$$\delta V^\mu = -\oint \Gamma^\mu_{\alpha\beta}\, V^\alpha\, dx^\beta. \tag{1.66}$$

Why the minus sign? The reason for this is apparent when we consider the **covariant derivative** of a vector. To differentiate involves taking the limit of $[V^\mu(x+\delta x) - V^\mu(x)]/\delta x$, but the difference of V^μ at two different points is not meaningful in the face of general coordinate transformations. A more sensible procedure is to compare the value of the vector at the new point with the result of parallel-transporting it from the old point.

This is the 'observable' change in the vector, and gives the definition of the covariant derivative as

$$\boxed{V^\mu{}_{;\nu} \equiv V^\mu{}_{,\nu} + \Gamma^\mu_{\alpha\nu} V^\alpha} \tag{1.67}$$

(for a contravariant vector; hence the minus sign above). For a scalar, the covariant and coordinate derivatives are equal. Consider applying this fact to a scalar constructed from two vectors:

$$\begin{aligned} (U^\mu V_\mu)_{;\alpha} &= V_\mu U^\mu{}_{;\alpha} + U^\mu V_{\mu;\alpha} = V_\mu U^\mu{}_{,\alpha} + U^\mu V_{\mu,\alpha} \\ \Rightarrow \quad U^\mu V_{\mu;\alpha} &= U^\mu V_{\mu,\alpha} - V_\mu \left(U^\mu_{,\alpha} - U^\mu_{;\alpha} \right). \end{aligned} \tag{1.68}$$

Since U^μ is an arbitrary vector, this allows us to deduce the covariant derivative for V_μ (different by a sign and index placement in the second term):

$$\boxed{V_{\mu;\nu} \equiv V_{\mu,\nu} - \Gamma^\alpha_{\mu\nu} V_\alpha} \tag{1.69}$$

The covariant derivative of a tensor may be deduced by considering products of vectors, and requiring that the covariant derivative obeys the **Leibniz rule** for differentiation:

$$T^{\mu\nu}_{;\alpha} = (V^\mu U^\nu)_{;\alpha} = V^\mu U^\nu_{;\alpha} + V^\mu_{;\alpha} U^\nu = T^{\mu\nu}_{,\alpha} + \Gamma^\mu_{\beta\alpha} T^{\beta\nu} + \Gamma^\nu_{\beta\alpha} T^{\mu\beta}. \tag{1.70}$$

This introduces a separate Γ term for each index (the appropriate sign depending on whether the index is up or down).

The Riemann tensor arises in terms of the connection Γ when we consider again the change in a vector under parallel transport around a small loop. Since the connection Γ is a function of position in the loop, it cannot be taken outside the loop. What we can do, though, is to make first-order expansions of V^μ and Γ as functions of total displacement from the starting point, Δx^μ. For Γ, the expansion is just a first-order Taylor expansion; for V^μ, what matters is the first-order change in V^μ due to parallel transport. The expression for the total change in V^μ then contains a second-order contribution

$$\delta V^\mu = \left(\Gamma^\mu_{\nu\beta} \Gamma^\nu_{\lambda\alpha} - \Gamma^\mu_{\lambda\beta,\alpha} \right) V^\lambda \oint \Delta x^\alpha \, dx^\beta \tag{1.71}$$

(there is no first-order term because $\oint dx^\mu = 0$ around a loop). Writing this equation twice and permuting α and β allows us to obtain

$$\boxed{\delta V^\mu = \tfrac{1}{2} R^\mu_{\alpha\lambda\beta} \oint \left(\Delta x^\beta \, dx^\alpha - \Delta x^\alpha \, dx^\beta \right),} \tag{1.72}$$

where the tensor $R^\mu_{\alpha\lambda\beta}$ is the **Riemann tensor** encountered earlier:

$$R^\mu{}_{\alpha\beta\gamma} = \Gamma^\mu_{\alpha\gamma,\beta} - \Gamma^\mu_{\alpha\beta,\gamma} + \Gamma^\mu_{\sigma\beta} \Gamma^\sigma_{\gamma\alpha} - \Gamma^\mu_{\sigma\gamma} \Gamma^\sigma_{\beta\alpha}. \tag{1.73}$$

The antisymmetric integral on the right of the expression for δV^μ is clearly a measure of the loop area, so the change in a vector under parallel transport around a loop is proportional both to the loop area and to the Riemann tensor, which is therefore the