Four-Vectors, Tensors
&
The General Theory of Relativity

Dr. J. J. Binney

Trinity Term 1991

Books

(i) *Principles of Cosmology and Gravitation* by M. V. Berry (Adam Hilger). Good on the physical ideas but leaves out most of the mathematical machinery and with a strong slant towards cosmology. (£10)

(ii) *Introduction to Tensor Calculus and Relativity* by D. F. Lawden (Wiley). An adequate introduction. Employs an imaginary time coordinate. Perhaps the mathematics and the physics are kept too much apart. (£11.95)

(iii) *Gravitation* by C. W. Misner, K. S. Thorne & J. A. Wheeler (Freeman). An immense book (1279 pages) notable for its emphasis on geometry. To benefit from its many valuable insights you have to learn modern coordinate-free differential geometry (which the book will obligingly teach you). At £49.95 (about three times original publication price) one of the most expensive paperbacks ever printed.

(iv) *Gravitation & Cosmology* by S. Weinberg (Wiley). An immensely scholarly work, notable for its clear derivations and pathalogical distaste for geometrical interpretations. Makes no use of coordinate-free methods. At a criminal £51.50 (five times original publication price) this book is strictly for reference in your college library.

1 Space, Time & Spacetime

Physics like betting is about predicting the future. From a study of the past (the “form”) and some general principles gleaned from experience, we endeavour to predict what will happen somewhere at some future time. Thus it behoves us to be clear about the concepts of “space” and “time”.

If you think this is a simple matter, hardly worthy of further thought, think again! There is nothing inevitable or absolute about our notions of time and space; they are merely one stage in a continuing process of development and refinement and are even now being challenged by theories in which time and space are merely crude reductions of an underlying 10- or 26-dimensional world. So let us look closely at our current notions of space-time before progress sweeps them aside.

We owe our current concepts of space and time to the “Special Theory of Relativity”. Nothing is so absurd about this theory as its name—it ought properly to be called the “Theory of Absolutivity”. For the theory’s central thesis is that there exists an absolute continuum called spacetime and that spacetime carries a definite and unvarying structure, the Minkowski metric, which firmly divides it into “space” and “time”. A few paragraphs about the more familiar concept of absolute space will help to elucidate this point.

We master the concept of absolute space before we can even talk: we are all familiar with the objects parents hang across their babys’ cots and with the games of “peek-a-boo” with which they seek to develop in their offspring a sense for spatial relationships. Soon a parent is spending hours looking with his or her child at picture books, pointing out the houses, cars, cats etc. shown there, and the child automatically thinks of these items as solid objects firmly seated in an imaginary three-dimensional world. He learns to speak by learning to name these objects. Thus an essential preliminary to the acquisition of language is the the development of the conviction that the two-dimensional patterns sensed by the retina are caused by three-dimensional objects; a chair looks quite different when seen from the front and when viewed from below. But every toddler sees the power of the unifying abstraction “chair”: the chair is the thing that stays the same when his point of view changes.

Having grasped the power of abstrations such as “box” and “car” to organize and explain his sense-experience, he is soon exploring the relations between these abstractions—discovering which toy cars can be put into a box this way or that. The relationships he is exploring have nothing whatever to do with his point of view. This car won’t go into that box irrespective of the angle from which he views the two objects. We most of us become pretty good at this kind of relationship-building, so that even the most innumerate humanist can hit a good shot on a squash court or solve a jigsaw puzzle.

However, lamentably few people ever learn to place their ability to manipulate spatial relationships on a quantitative basis, that is, to do Euclidean geometry. Yet it is perhaps significant that the latter was the first branch of mathematics to be brought to something like its present state of completeness—by late classical times the Greek-speaking world knew an astonishing amount about how planar and even solid figures can be disected and superposed.

Very much later, with the invention towards the end of the $17^{\rm th}$ century of coordinate geometry, it was discovered how geometry can be reduced to algebra and arithmetic. This reduction is a great boon computationally, but it suffers from one notable disadvantage: we reduce a series of points, curves and surfaces to algebraic form only by choosing a coordinate system, and there is no unique way of doing that. So later still (in fact, not until the dawn of the $20^{\rm th}$ century) vectors were invented as a device which enables geometrical entities to be manipulated algebraically without the distraction of picking out one particular coordinate system. Thus instead of indicating a displacement by giving its components a_x, a_y and a_z parallel to given coordinate directions, we write simply $\boldsymbol{a}$ for the entire displacement. Similarly denoting some other displacement by $\boldsymbol{b}$ we form the compound displacement $\boldsymbol{c} = \boldsymbol{a} + \boldsymbol{b}$ and write the angle between $\boldsymbol{a}$ and $\boldsymbol{b}$ as $\phi_{ab} = \arccos(\boldsymbol{a}\cdot\boldsymbol{b}/|\boldsymbol{a}||\boldsymbol{b}|)$. In short, vector notation enables us to perform the largest possible number of calculations without polluting our intellectual environment with any particular coordinate system.

We see then that both for an individual human and for the species as a whole, the line of development has been

$$\text{spatial relationships} \quad \rightarrow \quad \text{coordinate geometry} \quad \rightarrow \quad \text{vectors.}$$

By contrast, the crucial concept of spacetime is approached from the middle; we first meet spacetime in coordinate form and only rather later, and with difficulty, perceive its intrinsic geometrical structure. Thus if I say some event occurred at the spacetime point (t, x, y, z) you will know exactly what I mean, while if I started to explain the nature of the spacetime two-surface described by the electromagnetic field tensor $\mathbf{F}$, I would encounter only blank stares. The first goal of this course is to indicate how the discipline of organizing relations between physically interesting quantities into "covariant" equations between tensors (in effect, making the transition: coordinate geometry $\rightarrow$ vectors) leads us to introduce spacetime objects $\mathbf{p}$, $\mathbf{F}$, $\mathbf{T}$ etc, which are the spacetime analogues of boxes and cars; it is $\mathbf{p}$, $\mathbf{F}$ and $\mathbf{T}$ which stay the same when we change our "point of view" by boosting to a speeding frame of reference.

In infancy we discovered the relations between things by manipulating boxes and cars in our hands. Unfortunately, our hands do not permit us to manipulate $\mathbf{p}$, $\mathbf{F}$ and $\mathbf{T}$. So we inform ourselves about spacetime relationships by developing a calculus, tensor calculus, which, like vector algebra, is designed enable us to deal as far as possible in the underlying entities $\mathbf{p}$, $\mathbf{F}$, $\mathbf{T}$ etc rather than in coordinate-dependent expressions. Experience in the manipulation of spacetime objects with this calculus helps us gradually to develop that feel for spacetime relationships which we were sadly unable to develop in our playpens.[1]

1.1 Inertial Coordinates & Lorentz Transformations

Special relativity teaches us to think of experience as being made up of numbers of events, each with a definite location in the four-dimensional continuum of spacetime. Any given observer assigns to each event a unique 4-tuple of numbers (t, x, y, z). Of course he can do this in many, many ways. But special relativity claims that there are certain specially favoured systems for assigning coordinates to events, the so-called inertial coordinate systems. An observer O can set up an inertial coordinate system by assigning the coordinates $(t, -1, 0, 0)$ and $(t, 1, 0, 0)$ to the events that occur at the ends of a ruler which drifts force-free and non-rotating through space, and the coordinates $(t, 0, 1, 0)$ to events local to a particle that likewise moves freely and is always equidistant from the ruler's ends and as far from its middle as are the ends. The times t should be marked off by a good clock located at the origin. Another observer O′ may set up an entirely different inertial coordinate system, but according to special relativity the coordinates (t', x', y', z') he assigns to any event can be related to O's coordinates (t, x, y, z) of the same event by

$$\begin{pmatrix} ict' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} ict_0 \\ x_0 \\ y_0 \\ z_0 \end{pmatrix} + \mathbf{L} \cdot \begin{pmatrix} ict \\ x \\ y \\ z \end{pmatrix},$$

where c is the speed of light and (t_0, x_0, y_0, z_0) is a set of numbers characteristic of the two observers, as is the 4×4 matrix $\mathbf{L}$. Clearly, (t_0, x_0, y_0, z_0) are the coordinates O′ assigns

[1] Why do not children develop an instinct for spacetime structures? It would be rash to offer a definitive answer to this question, but undoubtedly a contributing factor is the circumstance that for all everyday purposes a particular direction for the time-axis is picked out for us—the temporal direction of our rest frame. Furthermore, we tend to look out into spacetime along either this particular temporal direction or one of the associated spatial directions: "at that time" or "over there". We simply don't have opportunity to ramble at will through through spacetime viewing its contents from a variety of different perspectives.

to the event that marks the origin of O's coordinates. For simplicity we shall assume that $(t_0, x_0, y_0, z_0) = \mathbf{0}$. In general $\mathbf{L}$ can be represented as the product of matrices generating a rotation, a boost parallel to a coordinate direction and a second rotation: $\mathbf{L} = \mathbf{R}' \cdot \mathbf{L}_0 \cdot \mathbf{R}$, where $\mathbf{R}$ rotates the coordinate axes so as to align the boost direction with a coordinate direction, $\mathbf{L}_0$ effects the boost along the given axis and $\mathbf{R}'$ rotates the coordinates to any desired final orientation. If $\mathbf{R}$ is chosen such that the x-axis becomes the boost direction, $\mathbf{L}_0$ has the form

$$\mathbf{L}_0 = \begin{pmatrix} \gamma & -i\beta\gamma & 0 & 0 \\ i\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{where} \quad \begin{array}{c} \beta \equiv v/c \\ \gamma \equiv 1/\sqrt{1-\beta^2} \end{array}. \tag{1.1}$$

Defining

$$\psi \equiv \text{arctanh}(\beta) = \text{arccosh}(\gamma), \tag{1.2a}$$

we can rewrite equation (1.1) as

$$\mathbf{L}_0 = \begin{pmatrix} \cosh\psi & -i\sinh\psi & 0 & 0 \\ i\sinh\psi & \cosh\psi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{1.2b}$$

Recalling that the matrix of rotation by angle θ about the z-axis is

$$\mathbf{R} = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

and bearing mind the relations $\cos i\psi = \cosh\psi$ and $\sin i\psi = i\sinh\psi$, we see from (1.2b) that $\mathbf{L}_0$ is identical with the matrix for a rotation in the (t, x) plane through the imaginary angle $\theta = -i\psi$.

Special relativity claims that the fundamental laws of physics must look the same when written in terms of O's coordinates as they do when expressed in terms of the coordinates of O′; these gentlemen are simply observing the world from two different but equally valid points of view. The dynamics of what they observe, being independent of any observer, can depend only on those relations between the objects that control the dynamics which are unaffected by pseudo-rotations of the coordinates (1.2b). Thus every physical quantity should be associated with a geometrical entity in spacetime such as a line, or a directed area or an ellipsoid etc, and physics should consist of statements such as "the length of the line associated with a proton is $(1836)^2$ times the length of an electron's line" and "when particles interact, the sum of their lines is invariant". We should, in fact, be able to suppose that everything in which we are interested in physics is associated with a spacetime object as real as any box or car.

Sadly for most of us the reality must fall far short of this Platonic ideal because, not having had adequate spacetime toys in our playpens, we experience difficulty picturing and manipulating geometrical entities in spacetime. Hence we fall back on algebra and calculus as a blind man reaches for his stick. But by hard work and diligent thought *is* possible to extend one's geometrical intuition to abstract spaces, and the development of theoretical physics in the 20^{th} century has time and again shown the power of geometrical formulations of fundamental theories. In the next section we show how the most important physical quantities are represented geometrically. But first we must deal with a tiresome technicality.

1.2 Real & Imaginary Coordinates

The key difference between a true rotation and the pseudo-rotation (1.2b) is that while a true rotation would change the coordinates so as to leave invariant the squared distance $d^2 =$

$(ct)^2 + x^2 + y^2 + z^2$, (1.2b) holds constant the quantity $s^2 = -(ct)^2 + x^2 + y^2 + z^2$. Now there are two ways in which we can generate that irritating minus sign in s^2: either we work with pure imaginary time coordinates, or we stick to real coordinates and introduce a metric tensor

$$\boldsymbol{\eta} \equiv \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

With $\boldsymbol{\eta}$ we can write s^2 really:

$$s^2 = (\, ct, x, y, z \,) \cdot \boldsymbol{\eta} \cdot \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix}.$$

If one is doing only special relativity, it is arguable that it is better to put up with the inconvenience of using an imaginary time coordinate rather than to carry $\boldsymbol{\eta}$ through all one's calculations. But for calculations in general relativity one is without question better off with a real time coordinate and a generalized $\boldsymbol{\eta}$. The reason for this is simply that in general relativity one must use coordinates in which s^2 *cannot* be expressed simply as a sum or difference of the squares of coordinates, but necessarily involves some 4×4 matrix. We may as well take advantage of this matrix to banish the unaesthetic factor of i from our time coordinates.

1.3 Summary

Physics is about relations between geometrical objects in space-time. To quantify these objects we have to choose a coordinate system. But space-time objects enjoy an existence independent of the point of view imposed by any particular system of coordinates.

2 Tensors in Special Relativity

Observers who move relative to one another do not always agree about the values of quantities, such as mass, energy, momentum etc, associated with the same physical system. The special theory of relativity tells us how we may predict the values measured by any observer once we know the values assigned by one particular observer, for example ourselves. For simplicity we confine ourselves to observers who assign the space-time coordinates $(0, \mathbf{0})$ to the same event, whose spatial coordinate systems are aligned and whose relative motion lies along their (mutually parallel) x-axes.

According to special relativity, all quantities of physical interest can be grouped into n-tuples. Each n-tuple specifies some geometrical quantity.

2.1 1-tuples (4-scalars)

On some things all observers agree, for example the charge and total spin of the an electron. These quantities are called **4-scalars** or relativistic invariants. A scalar describes an intrinsic shapeless property of an object, such as a colour or a temperature.

2.2 4-tuples (4-vectors)

Two observers O and O′ use Cartesian axes that are parallel to one another and at $t = 0$ their coordinate origins coincide. But O′ moves at speed v with respect to O down their mutual x-axis. Then coordinates assigned by O and O′ to the same event are related by

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \tag{2.1}$$

Moreover, if O measures the wave-vector and frequency of a photon to be $\boldsymbol{k}$ and ω, then an observer O′ who moves at speed v along O's x-axis measures wave-vector $\boldsymbol{k}'$ and frequency ω' given by

$$\begin{pmatrix} \omega'/c \\ k'_x \\ k'_y \\ k'_z \end{pmatrix} = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \omega/c \\ k_x \\ k_y \\ k_z \end{pmatrix}. \tag{2.2}$$

In matrix notatioin these equations are written

$$\begin{pmatrix} ct' \\ \boldsymbol{x}' \end{pmatrix} = \mathbf{\Lambda} \cdot \begin{pmatrix} ct \\ \boldsymbol{k} \end{pmatrix} \quad ; \quad \begin{pmatrix} \omega'/c \\ \boldsymbol{k}' \end{pmatrix} = \mathbf{\Lambda} \cdot \begin{pmatrix} \omega/c \\ \boldsymbol{k} \end{pmatrix} \tag{2.3a}$$

where

$$\mathbf{\Lambda} \equiv \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{2.3b}$$

Notes:

(i) Vectors written in italic boldface, $\boldsymbol{x}$ or ($\boldsymbol{k}$), are 3-vectors, while those written in Roman boldface, $\mathbf{x}$ or ($\mathbf{k}$), are 4-vectors.

(ii) The Lorentz transformation matrix $\mathbf{\Lambda}$ is dimensionless, so t has to be multiplied by c and ω has to be divided by c to give the same dimensions as $\boldsymbol{x}$ and $\boldsymbol{k}$, respectively, before being put into the first place of a 4-vector with $\boldsymbol{x}$ or $\boldsymbol{k}$.

If we define $x^0 \equiv ct$, $x^1 \equiv x$, $x^2 \equiv y$, $x^3 \equiv z$, $k^0 \equiv \frac{\omega}{c}$, $k^1 \equiv k_x$ etc then

$$x'^\mu = \sum_{\nu=0}^{3} \Lambda^\mu{}_\nu x^\nu \equiv \Lambda^\mu{}_\nu x^\nu \quad ; \quad k'^\mu = \sum_{\nu=0}^{3} \Lambda^\mu{}_\nu k^\nu \equiv \Lambda^\mu{}_\nu k^\nu. \tag{2.4}$$

Here we introduce two additional conventions:

(i) by "$k^\mu =$" we mean to imply the four equations for the four components of the 4-vector k^μ—thus k^μ is usually synonymous with $\mathbf{k}$;

(ii) by the **Einstein summation convention** we omit the summation sign when summing over repeated indices in a product. (Note that in $B^{\mu\nu} + k^\mu k^\nu$ there is *no* implied summation.)

The reason for writing some indices up and other down will emerge shortly.

Another familiar 4-tuple: if observer O measures energy E and momentum $\mathbf{p}$ for some particle, then O′ will measure E' and $\boldsymbol{p}'$ given by

$$\begin{pmatrix} E'/c \\ \boldsymbol{p}' \end{pmatrix} = \mathbf{\Lambda} \cdot \begin{pmatrix} E/c \\ \boldsymbol{p} \end{pmatrix}, \tag{2.5}$$

or setting $p^0 \equiv \frac{E}{c}$, we have $p'^\mu = \Lambda^\mu{}_\nu p^\nu$.

Each 4-tuple corresponds to an arrow in spacetime.

Lengths of 4-vectors With every 4-vector $\mathbf{v}$ there comes a free 4-scalar, called the **length** of the vector, written $|\mathbf{v}|^2$. (Notwithstanding this notation, $|\mathbf{v}|^2$ is not necessarily positive, and therefore we do not take its square root, as we would to find the length of a 3-vector $\boldsymbol{v}$.) For example, the length of a photon's 4-vector is the scalar

$$|\mathbf{k}|^2 \equiv (k^0)^2 + (k^1)^2 + (k^2)^2 + (k^3)^2 = -\frac{\omega^2}{c^2} + |\boldsymbol{k}|^2 = 0.$$

One can prove that this really is a scalar by brute force:

$$\begin{aligned}
|\mathbf{k}'|^2 &= -(k'^0)^2 + (k'^1)^2 + (k'^2)^2 + (k'^3)^2 \\
&= -\Big(\gamma\frac{\omega}{c} - \beta\gamma k^1\Big)^2 + \Big(-\beta\gamma\frac{\omega}{c} + \gamma k^1\Big)^2 + (k^2)^2 + (k^3)^2 \\
&= -\gamma^2\Big(1-\beta^2\Big)\frac{\omega^2}{c^2} + \gamma^2\Big(1-\beta^2\Big)(k^1)^2 + (k^2)^2 + (k^3)^2 \\
&= -(k^0)^2 + (k^1)^2 + (k^2)^2 + (k^3)^2.
\end{aligned}$$

The length of the momentum-energy 4-vector of a particle of rest mass $m_0 \neq 0$ is just $-c^2$ times the square of its rest mass m_0. We show this by arguing that it doesn't matter in whose frame we evaluate a scalar. We choose the particle's rest frame. Then $\boldsymbol{p} = 0$ and $E = m_0c^2$, so

$$-p^0p^0 + p^1p^1 + p^2p^2 + p^3p^3 = -m_0^2c^2.$$

Four-vectors that have negative lengths are called **time-like**, while those with positive lengths are **space-like**. Vectors with zero length are said to be **null**.

Note:

Every book on relativity uses a different convention. The sign of the lengths of space-like vectors is called the "signature of the metric".

With $\boldsymbol{\eta}$ we can write the length of the vector neatly:

$$-m_0^2c^2 = \mathbf{p}\cdot\boldsymbol{\eta}\cdot\mathbf{p} = p^\mu\eta_{\mu\nu}p^\nu.$$

(Notice that we are summing over both μ and ν from 0 to 3.)

Covariant and contravariant vectors We write the result of matrix multiplication of $\mathbf{p}$ by $\boldsymbol{\eta}$ as

$$p_\mu \equiv \eta_{\mu\nu}p^\nu.$$

We have $p_0 = -p^0$, $p_1 = p^1$, $p_2 = p^2$ and $p_3 = p^3$. Thus the length of p^μ is

$$p^\mu p_\mu = -m_0^2c^2.$$

Notice that here as everywhere else, we are summing over *one up and one down index.* In order to stick rigidly to this rule, we define

$$\eta^{\mu\nu} \equiv \eta_{\mu\nu} \equiv \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{2.6}$$

Note:

Later, when $\eta_{\mu\nu}$ goes over into the general relativistic metric tensor $g_{\mu\nu}$, the analogue of $\eta^{\mu\nu}$ will be the inverse $g^{\mu\nu}$ of the matrix $g_{\mu\nu}$. It so happens that $\eta_{\mu\nu}$ is its own inverse.

From p_μ we can recover p^μ;

$$p^\mu = \eta^{\mu\nu} p_\nu. \tag{2.7}$$

p_μ is a 4-vector, but of a slightly different type than p^μ, because under a Lorentz transformation we have

$$\begin{aligned} p'_\mu &= \eta_{\mu\nu} p'^\nu = \eta_{\mu\nu} \Lambda^\nu{}_\kappa p^\kappa = \eta_{\mu\nu} \Lambda^\nu{}_\kappa \eta^{\kappa\lambda} p_\lambda \\ &= \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix} \\ &= \begin{pmatrix} \gamma & \beta\gamma & 0 & 0 \\ \beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix} \equiv \Lambda_\mu{}^\nu p_\nu, \end{aligned} \tag{2.8}$$

where we have defined a new matrix

$$\Lambda_\mu{}^\lambda \equiv \eta_{\mu\nu} \Lambda^\nu{}_\kappa \eta^{\kappa\lambda}. \tag{2.9}$$

Notice that the transpose of $\Lambda_\mu{}^\nu$ is the inverse of $\Lambda^\mu{}_\nu$:

$$\Lambda^\mu{}_\kappa \Lambda_\mu{}^\nu = \delta^\nu_\kappa, \tag{2.10}$$

where we have written the 4×4 identity matrix as δ^ν_κ (see §2.4).

Vectors with their indices below are called **covariant** (p_μ). Vectors with indices above are called **contravariant** (p^μ). I shall call them down and up vectors. The operation of setting two indices equal and summing from 0 to 3 is called **contraction**. In a contraction one index must be up and one down. Quantities like $\sum_\mu p_\mu p_\mu$ are *not* 4-scalars and have nothing to do with physics. An important motivation for writing p^μ rather than $\mathbf{p}$ is to distinguish the up from the down form of $\mathbf{p}$. However, when there is no danger of confusion, I shall sometimes refer simply to $\mathbf{p}$.

Remember that the coordinates form an up-vector $x^\mu = (ct, x, y, z)$.

2.3 6-tuples (antisymmetric 2nd rank tensors)

If the electric and magnetic fields measured by O are arranged into the antisymmetric matrix $\mathbf{F}$,

$$F^{\mu\nu} \equiv \begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ -E_x/c & 0 & B_z & -B_y \\ -E_y/c & -B_z & 0 & B_x \\ -E_z/c & B_y & -B_x & 0 \end{pmatrix} \quad \text{(SI units)}, \tag{2.11}$$

then O′ will measure E' and B' as

$$\begin{pmatrix} 0 & E'_x/c & E'_y/c & E'_z/c \\ -E'_x/c & 0 & B'_z & -B'_y \\ -E'_y/c & -B'_z & 0 & B'_x \\ -E'_z/c & B'_y & -B'_x & 0 \end{pmatrix} \equiv F'^{\mu\nu} = \Lambda^\mu{}_\kappa \Lambda^\nu{}_\lambda F^{\kappa\lambda}. \tag{2.12}$$

Note that $F^{\mu\nu}$ transforms *as if* it were the product $p^\mu p^\nu$ of two down-vectors (which it isn't). Objects that transform in this way are called second-rank tensors.

$\mathbf{F}$ is called the **Maxwell field tensor**.[2]

[2] James Clerk-Maxwell was born, brought up and educated as plain James Clerk; he took the name Clerk-Maxwell only on marriage to an heiress, Miss Maxwell. Thus but for the power of Mamon, we would speak of Clerk's equations and Clerkians. Mr Maxwell got a better deal than I imagine he bargained for!

Exercise (1):

Transform $F^{\kappa\lambda}$ with the matrix $\Lambda^{\mu}{}_{\nu}$ defined by (2.2b) to show that an observer who moves at speed v down the x-axis of an observer who sees fields $\boldsymbol{E} = (E_x, E_y, 0)$ and $\boldsymbol{B} = 0$, perceives fields $\boldsymbol{E}' = (E_x, \gamma E_y, 0)$ and $\boldsymbol{B}' = (0, 0, \gamma v E_y/c)$. [Hint: since $\boldsymbol{\Lambda}$ is symmetric, we can write $\mathbf{F}' = \boldsymbol{\Lambda}\cdot\mathbf{F}\cdot\boldsymbol{\Lambda}$.] Hence deduce the general rules $\boldsymbol{E}'_{\|} = \boldsymbol{E}_{\|}$, $\boldsymbol{E}'_{\perp} = \gamma(\boldsymbol{E}_{\perp} + \boldsymbol{v}\times\boldsymbol{B})$, $\boldsymbol{B}'_{\|} = \boldsymbol{B}_{\|}$, $\boldsymbol{B}_{\perp} = \gamma(\boldsymbol{B}_{\perp} - \boldsymbol{v}\times\boldsymbol{E}/c^2)$. Verify that $(B^2 - E^2/c^2) = (B'^2 - E'^2/c^2)$.

Some 6-tuples correspond to elements of area. This correspondence works as follows. With any two displacements, say $\mathbf{u}$ and $\mathbf{v}$, we associate the parallelogram bounded by $\mathbf{u}$ and $\mathbf{v}$. Information about the size and orientation of this parallelogram is conveyed by the antisymmetric tensor $S^{\alpha\beta} \equiv u^{\alpha}v^{\beta} - u^{\beta}v^{\alpha}$; in particular, if $\mathbf{u} = \mathbf{v}$, then $\mathbf{S} = 0$. $\mathbf{S}$ has six degrees of freedom rather than the eight numbers involved in $\mathbf{u}$ and $\mathbf{v}$ because we can add to $\mathbf{u}$ any multiple of $\mathbf{v}$ without affecting $\mathbf{S}$, and vice versa for $\mathbf{v}$ and $\mathbf{u}$.

Exercise (2):

Prove the last statement and give a geometrical interpretation of this result.

In three-space the size and orientation of a parallelogram may be specified by giving the magnitude and direction of the normal. Hence in three-space full information about an antisymmetric 2nd rank tensor can be packed into the three components of the 3-vector which we call the cross-product of the parallelogram's sides. In four-dimensional spacetime each parallelogram has a magnitude and two mutually perpendicular normals, requiring six numbers for its full specification. Consequently there is no direct analogue of the cross product and we must represent areas directly with antisymmetric tensors.

Exercise (3):

Relate the above statements to the number of independent components of an antisymmetric $n \times n$ matrix for $n = 2, 3, 4$.

A physically interesting 6-tuple that describes an area is the tensor $(x^{\mu}p^{\nu} - x^{\nu}p^{\mu})$ formed from the space-time coordinate vector $x^{\mu} = (ct, x, y, z)$ and the 4-momentum of a particle. If the angular momentum about the origin is $\boldsymbol{L}$, we have

$$H^{\mu\nu} \equiv (x^{\mu}p^{\nu} - x^{\nu}p^{\mu}) = \begin{pmatrix} 0 & \ddots & \ddots & \\ c(xE/c^2 - tp_x) & 0 & \ddots & \ddots \\ c(yE/c^2 - tp_y) & L_z & 0 & \ddots \\ c(zE/c^2 - tp_z) & -L_y & L_x & 0 \end{pmatrix}, \tag{2.13}$$

where the diagonal dots stand for minus the quantities in the lower left triangle of the matrix. The numbers in the first column of this matrix give mc times the particle's initial position vector.

With every 6-tuple we get two free scalars. If the 6-tuple is of the form $(u^{\alpha}v^{\beta} - u^{\beta}v^{\alpha})$, then one of these is twice the squared magnitude of the corresponding parallelogram:

$$\begin{aligned} S^{\mu\nu}(\eta_{\mu\kappa}\eta_{\nu\lambda}S^{\kappa\lambda}) \equiv S^{\mu\nu}S_{\mu\nu} &= (u^{\mu}v^{\nu} - u^{\nu}v^{\mu})(u_{\mu}v_{\nu} - u_{\nu}v_{\mu}) \\ &= 2[|\mathbf{u}|^2|\mathbf{v}|^2 - (\mathbf{u}\cdot\mathbf{v})^2]. \end{aligned}$$

Evaluation in the particle's rest frame shows that the scalar $\frac{1}{2}H_{\mu\nu}H^{\mu\nu} = [|\mathbf{x}|^2|\mathbf{p}|^2 - (\mathbf{x}\cdot\mathbf{p})^2] = -(m_0 c r_0)^2$, where r_0 is the distance (in the rest frame) between the particle and the origin at $t = 0$.

It is interesting to evaluate this same scalar for the Maxwell field tensor. Straightforward matrix multiplication shows that the down-down shadow of $F^{\mu\nu}$ is[3]

$$F_{\mu\nu} \equiv \begin{pmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ E_x/c & 0 & B_z & -B_y \\ E_y/c & -B_z & 0 & B_x \\ E_z/c & B_y & -B_x & 0 \end{pmatrix} \quad \text{(SI units)}, \tag{2.14}$$

Multiplying each element of $F_{\mu\nu}$ by the corresponding element of $F^{\mu\nu}$ we find

$$\begin{aligned} m &\equiv \tfrac{1}{2} F_{\mu\nu} F^{\mu\nu} \\ &= (B^2 - E^2/c^2). \end{aligned} \tag{2.15}$$

We will discuss the other scalar associated with a 6-tuple in §2.5.

2.4 10-tuples (symmetric 2nd rank tensors)

We introduced the matrix $\eta_{\alpha\beta}$ in order to get a compact expression for the length of a 4-vector, and later used it to define a down vector. No matter what coordinate system we are considering, we want $\eta_{\alpha\beta}$ to be the matrix defined by (2.6). Is this compatible with $\eta_{\alpha\beta}$ being a down-down tensor? Let $b_{\alpha\beta}$ be the tensor that has the form (2.6) in the unprimed coordinate system. Then in the primed system by the rule for transforming down-down tensors,

$$b'_{\alpha\beta} = \Lambda_\alpha{}^\kappa \Lambda_\beta{}^\lambda \eta_{\kappa\lambda}.$$

But by the definition (2.9) of $\Lambda_\beta{}^\lambda$, this may be written

$$b'_{\alpha\beta} = \Lambda_\alpha{}^\kappa (\eta_{\beta\gamma} \Lambda^\gamma{}_\delta \eta^{\delta\lambda}) \eta_{\kappa\lambda}.$$

Now $\eta^{\delta\lambda}\eta_{\kappa\lambda}$ is just the $\delta\kappa$ component of the 4×4 identity matrix. So

$$b'_{\alpha\beta} = (\Lambda_\alpha{}^\kappa \Lambda^\gamma{}_\kappa) \eta_{\beta\gamma}.$$

Finally, the quantity in $(\ldots)$ is the matrix product of $\Lambda_\alpha{}^\kappa$ with the transpose of $\Lambda^\gamma{}_\kappa$, which is its inverse [see (2.10)]. Hence the bracket is the $\alpha\gamma$ component of the identity matrix, so

$$b'_{\alpha\beta} = \eta_{\beta\alpha} = \eta_{\alpha\beta}.$$

Consequently, if an up-up tensor has the form (2.6) in one coordinate system system, it has that form in all coordinate systems and we may treat $\eta_{\alpha\beta}$ as a tensor rather than as a matrix if we wish. This will be important and non-trivial in g.r.

In a very similar way one can show that the identity matrix can be regarded as either an up-down or as a down-up tensor. It is conventional to write it as

$$\delta^\beta_\alpha = \delta_\alpha{}^\beta = \delta^\alpha{}_\beta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.16}$$

It is hard to get excited about $\eta_{\alpha\beta}$ and δ^β_α as they don't contain any physics. Something meatier is the **energy momentum tensor** of a dust cloud: let there be at the event $\mathbf{x}$ $n_0(\mathbf{x})$

[3] It is worth remembering that in special relativity the lowering operation only *changes the sign of the mixed space-time components*; see Appendix A.

particles of rest-mass m_0 per unit volume in the particle's rest frame, and suppose all these particles are moving with 4-momentum p^μ. Then the energy-monentum tensor at $\mathbf{x}$ is

$$T^{\mu\nu} = \frac{n_0}{m_0} p^\mu p^\nu. \tag{2.17}$$

Examination of $\mathbf{T}$'s 00-component will show why it's of fundamental importance. Since $p^0 = \gamma m_0 c$, we have

$$T^{00} = (\gamma n_0)(\gamma m_0 c^2). \tag{2.18}$$

This is the product of the particle density γn_0 measured by an observer at rest in the given frame (the γ arises because the region occupied by the particles is Lorentz contracted parallel to their direction of motion), with the energy per particle in the given frame, $\gamma m_0 c^2$. So T^{00} is precisely the energy density of the particles and as such the thing which we expect to be the source of the particles' gravitational field. If T^{00} has to appear in the differential equation that governs a gravitational field, then the whole of the rest of $\mathbf{T}$ will have to appear in this equation too.

So let's take a look at the 3-vector formed by $\mathbf{T}$'s mixed space-time components $T^{i0} = (\gamma n_0 \gamma m_0 c^2)(p^i/\gamma m_0 c)$. This is the product of the energy density and a dimensionless vector which is proportional to the velocity of the particles. In fact, this vector is $(\mathrm{d}x^i/\mathrm{d}t)/c$ so T^{i0} is $1/c$ times the flux of energy associated with the dust cloud.

The form (2.17) provides a good approximation to the energy momentum tensors of a surprisingly wide range of systems. In fact, it is seriously in error only for systems in which vibrations such as sound waves propagate at a speed comparable to the speed of light; in all other systems the relative velocities of constituent particles are highly sub-relativistic and the interaction energies between particles are small compared to the particles' rest-mass energy $m_0 c^2$. So the energy-momentum tensor of an ordinary star is approximately of the form (2.17), but that of a neutron star is not.

A medium whose energy momentum tensor is most definitely not of the form (2.17) is the electromagnetic field in a vacuum. What does its energy-momentum tensor look like? We know that e.m. energy is proportional to $|\boldsymbol{E}|^2$ and $|\boldsymbol{B}|^2$ so it should involve a product of the Maxwell field tensor $\mathbf{F}$ (2.11) with itself. So a prime candidate is a multiple of the matrix product $\mathbf{F}\cdot\mathbf{F}$. Now the trace T^μ_μ of the energy-momentum tensor of a dust cloud is $(n_0/m_0)p_\mu p^\mu = -n_0 m_0 c^2$ and thus proportional to the rest-mass of the constituent particles. The e.m. field is made up of photons, which have zero rest mass. So it is a reasonable guess that the energy-momentum tensor of the e.m. field has vanishing trace. If we assume that its trace does vanish, then it must be a multiple of

$$T_\mu{}^\nu = -[F_{\mu\gamma}F^{\gamma\nu} + \tfrac{1}{4}(F_{\delta\gamma}F^{\delta\gamma}\delta^\nu_\mu)]. \tag{2.19}$$

$\mathbf{T}$ is an up-down tensor because it is constructed from tensors in the proper way. (Hence the importance of showing that δ^ν_μ is a tensor!) The first term in (2.19) is the matrix product of $F_{\mu\gamma}$ with $F^{\gamma\nu}$. The second is the invariant $m = B^2 - E^2/c^2$ evaluated above [eq. (2.15)]. By construction $\mathbf{T}$ is traceless.

A little slog shows that in terms of $\boldsymbol{E}$ and $\boldsymbol{B}$ the tensor $\mathbf{T}$ is

$$T_\mu{}^\nu = \begin{pmatrix} -\frac{1}{2}(B^2+E^2/c^2) & -N_x & -N_y & -N_z \\ N_x & & & \\ N_y & & P_{ij} & \\ N_z & & & \end{pmatrix}, \tag{2.20}$$

where $\boldsymbol{N} \equiv \boldsymbol{E}\times\boldsymbol{B}/c$ and

$$P_{ij} \equiv \tfrac{1}{2}\delta_{ij}\Big(B^2 + \frac{E^2}{c^2}\Big) - \Big(B_iB_j + \frac{E_iE_j}{c^2}\Big) \quad (i,j = 1,2,3). \tag{2.21}$$

We recognize $-T_0{}^0$ as the energy density in the e.m. field, and $T_i{}^0$ as the Poynting vector, which gives the energy flux. The 3×3 matrix P_{ij} was already known to Clerk-Maxwell; it describes the flux of the three kinds of momentum: P_{ix} = flux of x-momentum etc. The down-up form of $\mathbf{T}$ is not symmetric. But if we premultiply by $\eta^{\lambda\mu}$ to raise the first index, we find

$$T^{\lambda\nu} = \begin{pmatrix} \frac{1}{2}(B^2 + \frac{1}{c^2}E^2) & \boldsymbol{N} \\ \boldsymbol{N} & P_{ij} \end{pmatrix} \tag{2.22}$$

In g.r. $\mathbf{T}$ is what the e.m. field uses to generate a gravitational field.

Digression: A concrete example of T Suppose we have a plane e.m. wave running along $\hat{\mathbf{i}}$ polarized parallel to $\hat{\mathbf{j}}$. Then

$$\boldsymbol{E} = (0, E, 0)\cos(\omega t - kx)$$
$$\boldsymbol{B} = (0, 0, B)\cos(\omega t - kx).$$

E and B are related by $-\partial \boldsymbol{B}/\partial t = \nabla \times \boldsymbol{E} \quad \Rightarrow \quad B = kE/\omega = E/c$. Hence

$$\boldsymbol{N} = (E^2/c^2, 0, 0)\cos^2(\omega t - kx).$$

The first term in our expression (2.21) is non-zero only on the diagonal. The second term is non-zero only in the yy and zz slots and there cancels the first term. So $\boldsymbol{P}$ is

$$P_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \frac{E^2}{c^2}\cos^2(\omega t - kx),$$

and finally

$$T^{\mu\nu} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \frac{E^2}{c^2}\cos^2(\omega t - kx). \tag{2.23}$$

The stress tensor $\boldsymbol{P}$ has only an entry in the xx slot because our wave is engaged in the business of carrying x-type momentum in the x-direction; the wave would push back a mirror placed in a plane x = constant. Clearly the Poynting vector is also directed along the x axis, which accounts for the off-diagonal units in $\mathbf{T}$. In proper relativistic units the wave employs unit energy density ("capital employed") to carry unit fluxes of energy and momentum ("turnover"). Notice that the wave's phase is the scalar $-\mathbf{k} \cdot \mathbf{x}$.

We can use this example of the energy momentum tensor of a plane e.m. wave to caculate the energy-momentum tensor of a blackbody radiation field—which is interesting cosmologically, since in the beginning there was only heat. Any radiation field can be decomposed into plane waves, so we have that the $\mathbf{T}$ of a blackbody field is the sum[4] of contributions $\delta\mathbf{T}$ that in appropriately oriented axes may be written in the form (2.23). Now the trace T^μ_μ of the $\mathbf{T}$ given by (2.23) is zero, and the trace of a tensor is a scalar. So the trace δT^μ_μ of every contributor to the energy-momentum tensor $\widetilde{\mathbf{T}}$ of a black-body radiation field is zero, from which it follows that $\widetilde{T}^\mu_\mu = 0$. Furthermore, in the radiation-field's rest frame there is no net transport of energy, so in this frame $\widetilde{\mathbf{T}}$'s mixed space-time components T^{i0} vanish. On the other hand, the momentum flux doesn't vanish even in the radiation's rest frame; momentum is a signed quantity, so equal numbers of particles moving at the same speed to right and left both contribute positively to the momentum flux. But in the rest frame the only net flux across

[4] Recall from problems about waves on strings that the energies of normal modes add even though the kinetic and potential energies of the waves are proportional to the square of the amplitude.

a surface $x =$ constant is of x-type momentum, so $\boldsymbol{P}$ must be diagonal. Furthermore, in this frame the three space axes are equivalent, so $\widetilde{\mathbf{T}}$ must be of the form

$$\widetilde{T}_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{3} \end{pmatrix} \frac{\sigma}{c} T^4 \qquad \begin{pmatrix} \text{where } T \text{ is temperature and } \sigma \text{ is} \\ \text{the Stephan-Boltzmann constant} \end{pmatrix}. \tag{2.24}$$

The constant $\sigma T^4/c$ has been chosen so that $\widetilde{T}_{00}$ agrees with the usual expression for the energy density in a cavity at temperature T.

We may use (2.23) to find the $T^{\mu\nu}$ of an arbitrary plane wave by rotating our coordinate system so that the old x-axis lies along an arbitrary unit vector $\hat{\boldsymbol{n}}$. The $\boldsymbol{\Lambda}$ matrix of such a rotation is

$$\Lambda^\mu{}_\nu = \begin{pmatrix} 1 & \mathbf{o} \\ \mathbf{o} & R_{ij} \end{pmatrix},$$

where $\mathbf{o} \equiv (0, 0, 0)$ and $\boldsymbol{R}$ is a rotation matrix such that

$$\boldsymbol{R} \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \hat{\boldsymbol{n}} \qquad \text{(the } x\text{-axis is rotated into } \hat{\boldsymbol{n}}\text{)}.$$

Hence

$$\Lambda^\mu{}_\nu T^{\nu\gamma} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ n_x & n_x & 0 & 0 \\ n_y & n_y & 0 & 0 \\ n_z & n_z & 0 & 0 \end{pmatrix} \frac{E^2}{c^2} \cos^2(\mathbf{k} \cdot \mathbf{x}). \tag{2.25}$$

To transform the second index of $\mathbf{T}$ we transpose the matrix (2.25) and again premultiply by $\Lambda^\mu{}_\nu$:

$$\begin{aligned} T'^{\mu\nu} &= \Lambda^\mu{}_\gamma \Lambda^\nu{}_\delta T^{\gamma\delta} = \Lambda^\nu{}_\delta (\Lambda^\mu{}_\gamma T^{\gamma\delta}) \\ &= \begin{pmatrix} 1 & \mathbf{o} \\ \mathbf{o} & \boldsymbol{R} \end{pmatrix} \begin{pmatrix} 1 & n_x & n_y & n_z \\ 1 & n_x & n_y & n_z \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \frac{E^2}{c^2} \cos^2(\mathbf{k} \cdot \mathbf{x}) \\ &= \begin{pmatrix} 1 & n_x & n_y & n_z \\ n_x & n_x n_x & n_y n_x & n_z n_x \\ n_y & n_x n_y & n_y n_y & n_z n_y \\ n_z & n_x n_z & n_y n_z & n_z n_z \end{pmatrix} \frac{E^2}{c^2} \cos^2(\mathbf{k} \cdot \mathbf{x}). \end{aligned}$$

Multiplying this matrix through by $k^2 = \omega^2/c^2$ we get

$$T'^{\mu\nu} = k^\mu k^\nu (E/\omega)^2 \cos^2(\mathbf{k} \cdot \mathbf{x}), \tag{2.26}$$

where k^μ is the null up vector whose spatial part runs along $\boldsymbol{k} \equiv \frac{\omega}{c}\hat{\boldsymbol{n}}$. From this little exercise we learn the following: Since $k^\mu k^\nu$ is already a tensor, $T^{\mu\nu}$ can be a tensor only if the coefficient $(E/\omega)^2 \cos^2(\mathbf{k} \cdot \mathbf{x})$ is a scalar. The argument of the cosine is itself a scalar, so E^2/ω^2 must also be a scalar. Consequently, different observers of the same e.m. wave measure field strengths proportional to the perceived frequencies of the wave.

Exercise (4):

The Earth's velocity changes by about $60\,\mathrm{km\,s^{-1}}$ each six months. When pointed to a particular galaxy at angle θ from the direction of the Earth's motion, a telescope detects $I(\omega)\mathrm{d}\Omega\mathrm{d}\omega$ photons per unit time with frequencies in $(\omega + \mathrm{d}\omega, \omega)$ and wavevectors in the solid angle $\mathrm{d}\Omega$. Use equation (2.25) to show that six months later the telescope measures $(\omega'/\omega)^2 I(\omega)$ photons per unit frequency, time and steradian, where $\omega' = \gamma(1 + \beta\cos\theta)\omega$ ($\beta \simeq 2 \times 10^{-4}$, $\gamma = 1/\sqrt{1-\beta^2}$). [Hint: you'll probably want to prove the relations (i) $I(\omega)\mathrm{d}\ln\omega\mathrm{d}\Omega = I'(\omega')\mathrm{d}\ln\omega'\mathrm{d}\Omega'$, (ii) $\cos\theta' = (\cos\theta + \beta)/(1 + \beta\cos\theta)$ and (iii) $\mathrm{d}\Omega' = \mathrm{d}\Omega/[\gamma^2(1 + \beta\cos\theta)^2]$.]

2.5 Geometrical Interpretation of 2nd Rank Tensors

10-tuples From matrix algebra we know that any symmetric (or even Hermitian) matrix $\mathbf{M}$ can be diagonalized by moving to coordinates in which the axial directions coincide with $\mathbf{M}$'s eigenvectors: if $\mathbf{B}$ is the unitary matrix formed by lining up in columns the properly normalized eigenvectors, then $\mathbf{M}' \equiv \mathbf{B}^\dagger \cdot \mathbf{M} \cdot \mathbf{B}$ is diagonal with $\mathbf{M}$'s eigenvalues λ_i for its non-zero elements. Hence the quadratic form $Q \equiv \mathbf{x}' \cdot \mathbf{M}' \cdot \mathbf{x}'$ is just $\lambda_1 {x'}_1^2 + \lambda_2 {x'}_2^2 + \ldots$. Consequently the equation $Q =$ constant defines either an ellipsoid (if all the λ_i have the same sign) or a hyperboloid.

In the case of a symmetric spacetime tensor we consider the quadratic form $Q = x_\alpha M^{\alpha\beta} x_\beta = x^\alpha M_{\alpha\beta} x^\beta$. It is easy to see that in the case $\mathbf{M} = \boldsymbol{\eta}$ this defines a hyperboloid:

Surfaces of constant $\mathbf{x} \cdot \boldsymbol{\eta} \cdot \mathbf{x}$

The hyperboloids obtained when $Q < 0$ are called **mass shells** because they contain all the points reachable by the momentum vector of a particle of given rest mass. The singular hyperboloid $Q = 0$ is the **light cone**.

In the case in which $\mathbf{M}$ is the energy momentum tensor $\widetilde{\mathbf{T}}$ of a black-body radiation field, equation (2.24) shows that $Q =$ constant defines an ellipsoid with axes in the ratios $1 : \sqrt{3} : \sqrt{3} : \sqrt{3}$, i.e. a 4-sphere that has been squashed along the t-axis.

Algebraically a 10-tuple $\mathbf{M}$ is a machine for turning one vector into another: $\mathbf{b} = \mathbf{M} \cdot \mathbf{a}$. How is the ellipsoid/hyperboloid of $\mathbf{M}$ related to this machine? Consider the gradient with respect to $\mathbf{a}$ of the quadratic form $Q(\mathbf{a}) = a_\alpha M^{\alpha\beta} a_\beta$; we have

$$\begin{aligned} \frac{\partial Q}{\partial a_\gamma} &= M^{\gamma\beta} a_\beta + a_\alpha M^{\alpha\gamma} = 2M^{\gamma\beta} a_\beta \qquad \text{(by the symmetry of } \mathbf{M}\text{)} \\ &= 2b^\gamma. \end{aligned} \tag{2.27}$$

Thus the output vector $\mathbf{b}$ is half the gradient of the ellipsoid/hyperboloid represented by Q at the position of the input vector $\mathbf{a}$:

What are the physical interpretations of the input and output vectors appropriate to an energy-momentum tensor **T**? Every 3-dimensional element $\mathrm{d}V$ of 4-dimensional spacetime may be characterized by the 4-vector $\mathrm{d}\mathbf{V}$ whose magnitude gives $\mathrm{d}V$'s volume and whose direction is normal to every vector contained in $\mathrm{d}V$. If $\mathrm{d}V$ is a cell with sides parallel to the vectors **u**, **v** and **w**, then we have

$$\mathrm{d}V^\alpha = \epsilon^{\alpha\beta\gamma\delta} u_\beta v_\gamma w_\delta, \tag{2.28}$$

where the **Levi-Civita** symbol $\epsilon^{\alpha\beta\gamma\delta}$ is the object which changes sign on every exchange of indices and has $\epsilon^{0123} = 1$; in particular it vanishes if any two indices are the same (see Appendix A). Now suppose that **u**, **v** and **w** join events that all occur at one instant on an observer O's clock. Then $T_{\alpha\beta}\mathrm{d}V^\beta$ is the amount of energy-momentum at that instant in the cell formed by **u**, **v** and **w**. If on the other hand **u** and **v** join simultaneous events while **w** joins two events that happen one after another at the same place (from O's point of view), then $T_{\alpha\beta}\mathrm{d}V^\beta$ is the amount of energy-momentum that flows across the surface spanned by **u** and **v** in the time represented by **w**.

Exercise (5):

Show that in n-dimensional space a totally antisymmetric tensor $A_{\alpha_1,\alpha_2,\ldots,\alpha_{n-1}} = -A_{\alpha_2,\alpha_1,\ldots,\alpha_{n-1}}$ etc, has n independent components and can therefore be represented by a vector. Explain the relevance of this remark to equation (2.28).

6-tuples We have seen that 6-tuples of the form $(u^\alpha v^\beta - u^\beta v^\alpha)$ correspond to elements of area. However, not all six-tuples have such a simple geometrical interpretation. To see show this we use our friend $\epsilon^{\alpha\beta\gamma\delta}$ to form the **dual** $\overline{\mathbf{F}}$ of the Maxwell field tensor **F**

$$\begin{aligned}\overline{F}^{\alpha\beta} &\equiv \tfrac{1}{2}\epsilon^{\alpha\beta\gamma\delta}F_{\gamma\delta} \\ &= \begin{pmatrix} 0 & -B_x & -B_y & -B_z \\ B_x & 0 & E_z/c & -E_y/c \\ B_y & -E_z/c & 0 & E_x/c \\ B_z & E_y/c & -E_x/c & 0 \end{pmatrix},\end{aligned} \tag{2.29}$$

and then contract **F** with its dual:

$$\begin{aligned} f &\equiv \overline{F}^{\alpha\beta}F_{\alpha\beta} \\ &= (\text{each element of } \overline{F}^{\alpha\beta}) \times (\text{corresponding element of } F_{\alpha\beta}) \\ &= \frac{4}{c}\boldsymbol{E}\cdot\boldsymbol{B}. \end{aligned} \tag{2.30}$$

Thus in general $f \neq 0$. On the other hand, when we contract a 6-tuple of the form $S_{\alpha\beta} = u_\alpha v_\beta - u_\beta v_\alpha$ with its dual we obtain

$$\begin{aligned} s &= \tfrac{1}{2}\epsilon^{\alpha\beta\gamma\delta}(u_\alpha v_\beta - u_\beta v_\alpha)(u_\gamma v_\delta - u_\delta v_\gamma) \\ &= 2\det(\mathbf{u},\mathbf{v},\mathbf{u},\mathbf{v}) = 0. \end{aligned}$$

Hence the Maxwell field tensor can be represented as an area element only in the special case $\boldsymbol{E}\cdot\boldsymbol{B}=0$ (as in a radiation field).

The horrid truth is that a generic Maxwell field tensor describes a geometrical entity that has no close analogue in three dimensions. Its simplest representation is in terms of *two* surface elements; one of these may be taken to be defined by two space-like vectors and to describe the magnetic flux, while the other element involves one space-like and one time-like vector and describes $\boldsymbol{E}$. But this decomposition is no more unique than are the three orthogonal projections of a widget produced in a drawing office; they merely serve to describe to lower-dimensional beings a single higher-dimensional entity.

One useful way of thinking of $\mathbf{F}$ is as a machine which tells us how much "electromagnetic flux" Φ is passing through any 2-surface:

$$\Phi(\mathbf{u},\mathbf{v})\equiv F_{\alpha\beta}u^\alpha v^\beta=\tfrac{1}{2}F_{\alpha\beta}S^{\alpha\beta}\quad\text{where}\quad S^{\alpha\beta}\equiv(u^\alpha v^\beta-u^\beta v^\alpha). \tag{2.31}$$

Exercise (6):

Show that in the case $\mathbf{u}=\delta x\,\hat{\mathbf{e}}_x$ and $\mathbf{v}=\delta y\,\hat{\mathbf{e}}_y$ we have $\Phi=B_z\delta x\delta y$, and that in the case $\mathbf{u}=\delta z\,\hat{\mathbf{e}}_z$ and $\mathbf{v}=\delta ct\,\hat{\mathbf{e}}_{ct}$ we have $\Phi=E_z\delta z\delta t$.

2.5 Derivatives of Tensors

Derivatives with respect to any system of coordinates can be expressed in terms of derivatives w.r.t. any other system by use of the chain rule:

$$\frac{\partial}{\partial x'^\mu}=\frac{\partial x^\nu}{\partial x'^\mu}\frac{\partial}{\partial x^\nu}. \tag{2.32}$$

If the primed and unprimed systems are linked by a Lorentz transformation,

$$x'^\nu=\Lambda^\nu{}_\mu x^\mu, \tag{2.33}$$

we have on multiplying by $\Lambda_\nu{}^\kappa$ and summing over ν,

$$\Lambda_\nu{}^\kappa x'^\nu=\Lambda_\nu{}^\kappa\Lambda^\nu{}_\mu x^\mu=x^\kappa,$$

where the last step follows by (2.10). Differentiating we get

$$\frac{\partial x^\kappa}{\partial x'^\nu}=\Lambda_\nu{}^\kappa. \tag{2.34}$$

Thus

$$\frac{\partial}{\partial x'^\mu}=\Lambda_\mu{}^\nu\frac{\partial}{\partial x^\nu}, \tag{2.35}$$

and we see that $\partial/\partial x^\mu$ transforms like a down vector.

Notes:

(i)

$\dfrac{\partial}{\partial x^\mu}$ operates on scalars to produce vectors: $G_\mu\equiv\dfrac{\partial\phi}{\partial x^\mu}\equiv\phi_{,\mu}$

$\dfrac{\partial}{\partial x^\mu}$ operates on vectors to produce 2$^{\text{nd}}$ rank tensors:

$$G_{\nu\mu}\equiv\frac{\partial A_\nu}{\partial x^\mu}\equiv A_{\nu,\mu}$$

$\dfrac{\partial}{\partial x^\mu}$ operates on tensors to produce higher-rank tensors:

$$G_{\lambda\nu\mu}\equiv\frac{\partial B_{\lambda\nu}}{\partial x^\mu}\equiv B_{\lambda\nu,\mu}$$

The operand's indices can be either up or down: $G^\nu{}_\mu = A^\nu{}_{,\mu}$.

(ii) If we contract the tensor produced by operating on a vector, we get a scalar, the 4-divergence $\psi = A^\mu{}_{,\mu}$.

(iii) We can reduce the number of indices on a higher-rank tensor by contraction: $A^\mu = G^{\mu\nu}{}_{,\nu}$.

(iv) The 4-analogue of taking the curl of a vector is to antisymmetrize the tensor formed by operating on a vector: $F_{\mu\nu} = (A_{\nu,\mu} - A_{\mu,\nu})$. If $A_\nu = \phi_{,\nu}$, then $F_{\mu\nu} = 0$ because partial derivatives commute.

Example:

In e.m. the usual vector potential $\boldsymbol{A}$ and the electrostatic potential ϕ form the four components of an up vector

$$A^\mu = (\phi/c, A_x, A_y, A_z) \quad [\Rightarrow \quad A_\mu = (-\phi/c, A_x, A_y, A_z)]. \tag{2.36}$$

Our old friend the Maxwell field tensor $\mathbf{F}$ is then

$$F_{\mu\nu} = -(A_{\mu,\nu} - A_{\nu,\mu}). \tag{2.37}$$

Thus $F_{12} = \dfrac{\partial A_y}{\partial x} - \dfrac{\partial A_x}{\partial y} = B_z$ and $F_{01} = \dfrac{\dot{A}_x}{c} + \dfrac{1}{c}\dfrac{\partial \phi}{\partial x} = -E_x/c$.

Derivatives with respect to proper time The history of a particle defines a curve in space-time. Let λ be a parameter which labels points on the curve in a continuous way. Then the coordinates x^μ of points on the curve are continuous functions $x^\mu(\lambda)$. For $\delta\lambda \ll 1$ the small up vector

$$\delta x^\mu \equiv \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\,\delta\lambda$$

almost joins two points on the curve. Hence it is time-like and $\delta x^\mu \delta x_\mu < 0$. For any two points A and B on the curve, we define

$$\tau \equiv \frac{1}{c}\int_{\mathrm{A}}^{\mathrm{B}} \sqrt{-\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\frac{\mathrm{d}x_\mu}{\mathrm{d}\lambda}}\,\mathrm{d}\lambda \tag{2.38}$$

to be the **proper time** difference between A and B along the curve. If the curve is a straight line, we may transform to the coordinate system in which $x^\mu = (ct, 0, 0, 0)$ at all points on the curve, and then

$$\tau = \frac{1}{c}\int_{\mathrm{A}}^{\mathrm{B}} \sqrt{-\frac{\mathrm{d}ct}{\mathrm{d}\lambda}\frac{\mathrm{d}(-ct)}{\mathrm{d}\lambda}}\,\mathrm{d}\lambda = [t_{\mathrm{B}} - t_{\mathrm{A}}]. \tag{2.39}$$

Hence the name. We regard the coordinates x^μ of events along the trajectory as functions $x^\mu(\tau)$ of the proper time. Differentiating w.r.t. τ and multiplying through by the rest mass m_0 we obtain a 4-vector, the momentum

$$p^\mu \equiv m_0 \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau}. \tag{2.40}$$

From the zeroth component of this equation we have $\mathrm{d}t = \gamma \mathrm{d}\tau$; the hearts of passengers on a fast train appear to beat slowly to a medic on the station platform.

2.6 Laws of e.m. and Mechanics in Tensor Form

The relativistic generalization of Newton's second law is

$$m_0 \frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\tau^2} = \frac{\mathrm{d}}{\mathrm{d}\tau}\left(m_0 \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau}\right) = \frac{\mathrm{d}p^\mu}{\mathrm{d}\tau} = f^\mu, \tag{2.41}$$

where **f** is the **4-force**. The last three components of **f** are just the Newtonian force components f_i. With $\mu = 0$ equation (2.41) states that the zeroth component of **f** is to $1/c$ times the rate of change of the particle's energy cp^0; hence physically f^0 is $1/c$ times the rate of working of the force w. In summary

$$f^\mu = (w/c, f_x, f_y, f_z). \tag{2.42}$$

Raising the indices in (2.37) and forming the divergence, we get four equations

$$F^{\mu\nu}{}_{,\nu} = \begin{pmatrix} \dfrac{1}{c}\dfrac{\partial E_x}{\partial x} + \dfrac{1}{c}\dfrac{\partial E_y}{\partial y} + \dfrac{1}{c}\dfrac{\partial E_z}{\partial z} \\ \partial B_z/\partial y - \partial B_y/\partial z - \frac{1}{c^2}\partial E_x/\partial t \\ -\partial B_z/\partial x + \partial B_x/\partial z - \frac{1}{c^2}\partial E_y/\partial t \\ \partial B_y/\partial x - \partial B_x/\partial y - \frac{1}{c^2}\partial E_z/\partial t \end{pmatrix} = \begin{pmatrix} \frac{1}{c}\nabla\cdot\boldsymbol{E} \\ \nabla\times\boldsymbol{B} - \dfrac{1}{c^2}\dfrac{\partial \boldsymbol{E}}{\partial t} \end{pmatrix}. \tag{2.43}$$

The zeroth component is by Poisson's equation equal to $\rho/(c\epsilon_0) = c\mu_0\rho$, where ρ is the charge density. By Ampere's law, the last three of these equations are equal to $\mu_0\boldsymbol{j}$, where $\boldsymbol{j}$ is the current density. Hence if we form a 4-vector

$$j^\mu = (c\rho, j_x, j_y, j_z), \tag{2.44}$$

we may write four of Maxwell's equations as

$$F^{\mu\nu}{}_{,\nu} = \mu_0 j^\mu. \tag{2.45}$$

It is straightforward to verify that Maxwell's other four equations can be written

$$F_{\mu\nu,\lambda} + F_{\lambda\mu,\nu} + F_{\nu\lambda,\mu} = 0 \quad (\mu \neq \nu \neq \lambda). \tag{2.46}$$

Exercises (7):

(i) Show that when λ, μ and ν equal 1, 2 and 3 respectively, (2.46) becomes $\nabla\cdot\boldsymbol{B} = 0$.

(ii) Show that with equation (2.29) equation (2.46) may also be written $\overline{F}^{\mu\nu}{}_{,\nu} = 0$.

Charge conservation is expressed as

$$\mu_0 j^\mu{}_{,\mu} = F^{\mu\nu}{}_{,\nu\mu} = 0, \tag{2.47}$$

where the last step follows by the antisymmetry of **F**.

The natural definition of the 4-current associated with a particle of charge q is

$$J^\mu = q\frac{\mathrm{d}x^\mu}{\mathrm{d}\tau}. \tag{2.48}$$

Since the force exerted on a charged particle by an e.m. field has to be linear in q, the fields represented by $F^{\mu\nu}$, and the particle's velocity vector, a suitable 4-vector to try as the force is

$$f^\mu = F^{\mu\nu} J_\nu. \tag{2.49}$$

Tentatively inserting this into (2.41) and multiplying through by $\mathrm{d}\tau/\mathrm{d}t = 1/\gamma$ to obtain the acceleration as measured in the laboratory frame, we get

$$\frac{\mathrm{d}p^\mu}{\mathrm{d}t} = qF^\mu{}_\nu \frac{\mathrm{d}x^\nu}{\mathrm{d}t}. \tag{2.50}$$

It is straightforward to check that the last three components of this vector are

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(m_0\gamma\frac{\mathrm{d}x_i}{\mathrm{d}t}\right) = q(\boldsymbol{v}\times\boldsymbol{B} + \boldsymbol{E})_i,$$

while the zeroth component is

$$\frac{\mathrm{d}(m_0 c\gamma)}{\mathrm{d}t} = \frac{q}{c}\boldsymbol{E}\cdot\boldsymbol{v},$$

or, in words, "the rate of change of the particle's energy mc^2 is equal to the rate of working of the Lorentz force."

Mass-energy conservation Consider the flux of mass-energy into and out of some small region W of spacetime. To the past and future W is bounded by the 3-dimensional sets of events that occur in some physical container (an empty beer can?) at the times t_0 and $t_1 > t_0$ in the can's rest frame. In spacetime these sets are represented by 4-vectors $V_\mu^{(0)}$ and $V_\mu^{(1)}$. We orientate $V_\mu^{(0)}$ so that it points into the past, while $V_\mu^{(1)}$ looks to the future. Since the contents of the can may not be uniform, we decompose both $\mathbf{V}^{(0)}$ and $\mathbf{V}^{(1)}$ into a large number of small pieces $\mathrm{d}\mathbf{V}$, each centred on a different position within the can. The balance of W's boundary comprises the 3-dimensional set of events that occur on the can's surface at times between t_0 and t_1. We represent this part of W's boundary by elements $\mathrm{d}V_\mu^{(s)}(\mathbf{x})$, each of which points out of the can.

Let $\mathbf{T}(\mathbf{x})$ be the energy-mometum density in the can. Then the energy-momentum in the can at t_0 is $p^\mu(0) = -\int_{\mathrm{can},t_0} T^{\mu\nu}\mathrm{d}V_\nu^{(0)}$, while that present at t_1 is $p^\mu(1) = \int_{\mathrm{can},t_1} T^{\mu\nu}\mathrm{d}V_\nu^{(1)}$. If energy-momentum is to be conserved, the difference between these two vectors must represent the energy-mometum that flows into the can between t_0 and t_1. Thus the first law of thermodynamics requires that

$$\iiint_{\mathrm{can},t_1} T^{\mu\nu}\mathrm{d}V_\nu^{(1)} + \iiint_{\mathrm{can},t_0} T^{\mu\nu}\mathrm{d}V_\nu^{(0)} = -\iiint_{\substack{\text{surface}\\ t_0<t<t_1}} T^{\mu\nu}\mathrm{d}V_\nu^{(s)}.$$

Thus in a natural notation we have

$$\oint T^{\mu\nu}\mathrm{d}V_\nu = 0. \tag{2.51}$$

Furthermore, the usual proof of Gauss's theorem generalizes easily to the statement that for any sufficiently differentiable tensor $G^{\mu\nu}$ we have

$$\int_\tau G^{\mu\nu}{}_{,\nu}\,\mathrm{d}^4\mathbf{x} = \oint_{\partial\tau} G^{\mu\nu}\mathrm{d}V_\nu, \tag{2.52}$$

where $\partial\tau$ denotes the three-dimensional boundary of the four-dimensional region τ. Thus an alternative expression of the conservation of energy-momentum is

$$T^{\mu\nu}{}_{,\nu} = 0. \tag{2.53}$$

It is interesting to see how this works out in practice. Consider the case of a dust of particles, for which $\mathbf{T}$ is given by (2.17). Thus

$$\begin{aligned} T^{\mu\nu}{}_{,\nu} &= \frac{\partial}{\partial x^\nu}\left(n_0 m_0 v^\mu v^\nu\right) \\ &= n_0 m_0 \frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}\frac{\partial v^\mu}{\partial x^\nu} + m_0 v^\mu \frac{\partial(n_0 v^\nu)}{\partial x^\nu}. \end{aligned} \tag{2.54}$$

The first term vanishes because $\mathrm{d}\mathbf{p}/\mathrm{d}\tau = 0$ for a free particle and the divergence in the second vanishes because particles are conserved.

Exercises (8):

(i) By calculating the divergence of (2.19) demonstrate the conservation in vacuo of electromagnetic energy-momentum.

(ii) Show that when $\mathrm{d}\mathbf{p}/\mathrm{d}\tau$ is given by (2.50) with $q\mathbf{F} \neq 0$, the *sum* of the energy-momentum tensors of the dust and the field is conserved, although the individual divergences do not vanish.

2.7 Summary

The special theory of relativity requires that any physical quantity must be a number that describes some geometrical entity in spacetime. Such an entity must be an n-tuple, where $n = 1, 4, 6, 10, \ldots$. Physical laws must be expressed as equations connecting the n-tuples associated with different physical quantities. These equations must be constructed in accordance with the rules of tensor calculus, which permit only:

(i) the multiplication of n-tuples to form either higher-rank n-tuples (as in $H^{\mu\nu} = x^\mu p^\nu - x^\nu p^\mu$) or lower-rank n-tuples (as in $f^\mu = F^{\mu\nu} J_\nu$), or

(ii) the addition of n-tuples *of the same rank.*

In particular, both sides of every acceptable equation always form valid n-tuples of the same kind.

Rest-mass, electric charge and total spin are scalars (1-tuples). The most important 4-vectors (4-tuples) include any particle's energy-momentum $\mathbf{p}$, e.m. current $\mathbf{J}$ or acceleration $\mathrm{d}\mathbf{p}/\mathrm{d}\tau$, and the potential $\mathbf{A}$ of the e.m. field. Important 6-tuples include any particle's angular momentum $\mathbf{H}$ and the Maxwell field tensor $\mathbf{F}$. The key 10-tuples are the metric tensor $\boldsymbol{\eta}$ that enables us to distinguish future and past and to assign magnitudes to n-tuples, and density of the energy-momentum $\mathbf{T}$ due to either particles or fields.

In 4-vector notation the key equation of mechanics and e.m. are

$$\mathbf{v} = \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}\tau} \quad ; \quad \mathbf{p} = m_0 \mathbf{v} \quad ; \quad \mathbf{J} = q\mathbf{v}$$

$$\mathbf{f} = \mathbf{F} \cdot \mathbf{J} \quad ; \quad \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}\tau} = \mathbf{f}$$

$$F_{\mu\nu} = -(A_{\mu},_{\nu} - A_{\nu},_{\mu}) \quad ; \quad F^{\mu\nu},_{\nu} = \mu_0 j^\mu \quad ; \quad \overline{F}^{\mu\nu},_{\nu} = 0,$$

where $\overline{F}^{\mu\nu} \equiv \frac{1}{2}\epsilon^{\mu\nu\gamma\delta} F_{\gamma\delta}$. The energy-monentum tensors of a zero-pressure gas and of the e.m. field are

$$\mathbf{T} = n_0 m_0 \mathbf{v}\mathbf{v} \quad \text{(dust)}$$

$$\mathbf{T} = \tfrac{1}{4}\mathrm{trace}(\mathbf{F} \cdot \mathbf{F})\mathbf{I} - \mathbf{F} \cdot \mathbf{F} \quad \text{(e.m. field).}$$

3 Newton's Theory & the Principle of Equivalence

3.1 Newton's Theory

According to Newton, every body attracts every other body with a force that is proportional to the product of the masses of the two bodies and inversely proportional to the square of the distance between them. Hence the force on a unit mass at $\boldsymbol{x}$ that is generated by a distribution of matter of density $\rho(\boldsymbol{x}')$ is

$$\boldsymbol{f}(\boldsymbol{x}) = G\int \frac{\boldsymbol{x}'-\boldsymbol{x}}{|\boldsymbol{x}'-\boldsymbol{x}|^3}\rho(\boldsymbol{x}')\,\mathrm{d}^3\boldsymbol{x}', \tag{3.1}$$

where $G = 6.672(4)\times 10^{-11}\ \mathrm{m}^3\,\mathrm{kg}^{-1}\,\mathrm{sec}^{-2}$ is Newton's constant. If we define the **gravitational potential** $\Phi(\boldsymbol{x})$ by

$$\Phi(\boldsymbol{x}) = -G\int \frac{\rho(\boldsymbol{x}')}{|\boldsymbol{x}'-\boldsymbol{x}|}\,\mathrm{d}^3\boldsymbol{x}',$$

and notice that

$$\nabla_{\boldsymbol{x}}\left(\frac{1}{|\boldsymbol{x}'-\boldsymbol{x}|}\right) = \frac{\boldsymbol{x}'-\boldsymbol{x}}{|\boldsymbol{x}'-\boldsymbol{x}|^3},$$

we find that we may write $\boldsymbol{f}$ as

$$\begin{aligned}\boldsymbol{f}(\boldsymbol{x}) &= \nabla_{\boldsymbol{x}}\int \frac{G\rho(\boldsymbol{x}')}{|\boldsymbol{x}'-\boldsymbol{x}|}\,\mathrm{d}^3\boldsymbol{x}' \\ &= -\nabla\Phi.\end{aligned} \tag{3.2}$$

If we take the divergence of equation (3.1), we find

$$\nabla\cdot\boldsymbol{f}(\boldsymbol{x}) = G\int \nabla_{\boldsymbol{x}}\cdot\left(\frac{\boldsymbol{x}'-\boldsymbol{x}}{|\boldsymbol{x}'-\boldsymbol{x}|^3}\right)\rho(\boldsymbol{x}')\,\mathrm{d}^3\boldsymbol{x}'. \tag{3.3}$$

But

$$\nabla_{\boldsymbol{x}}\cdot\left(\frac{\boldsymbol{x}'-\boldsymbol{x}}{|\boldsymbol{x}'-\boldsymbol{x}|^3}\right) = -4\pi\delta(\boldsymbol{x}'-\boldsymbol{x}) \qquad \text{(where } \delta \text{ is the Dirac } \delta\text{-function)} \tag{3.4}$$

as one may show, on the one hand by evaluating the derivative at $\boldsymbol{x}\neq\boldsymbol{x}'$, and on the other hand by using the divergence theorem to integrate the left side through a small sphere centred on $\boldsymbol{x}=\boldsymbol{x}'$. Combining equations (3.2), (3.3) and (3.4) we obtain **Poisson's equation**

$$4\pi G\rho = \nabla^2\Phi \quad = -\nabla\cdot\boldsymbol{f}. \tag{3.5}$$

Elegant though it is, this equation cannot represent the whole truth about gravitational physics since it is not constructed according to the rules of tensor calculus summarized in §2.7; if the right side of equation (3.5) is to form an n-tuple, it must form a scalar since it has only one component. On the other hand, since mass is just a manifestation of energy, we expect the quantity ρ appearing on the left side of equation (3.5) to represent energy density, and this we know to form the 00-component of the 10-tuple $\mathbf{T}$. So we either have to think of some scalar thing to put on the left in the place of ρ, or we have to augment Φ with a whole bunch of extra potentials, its companions in some new 10-tuple $\mathbf{g}$, and somehow extend the single equation (3.5) to a set of ten equations from which the whole set of potentials can be determined.

Consideration of the predicament of a physicist who knows about relativity and electrostatics but not about magnetism will clarify this point. This person looks at the electrostatic form of Poisson's equation

$$\nabla^2\phi = -q/\epsilon_0, \quad \text{where } q \text{ is charge density,}$$

and thinks

" q isn't a scalar because of the Lorentz-Fitzgerald contraction (in fact, q is the 0^{th} component of the current density $\mathbf{j}$),[5] so ϕ can't be a scalar either. Seems I'll have to augment

[5] See equation (2.44).

ϕ with three other potentials, say A_x, A_y and A_z. Then that ∇^2 won't do either, because it's no kind of n-tuple. I'll replace it with the d'Alembertian, which is a scalar. Then I'll have

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\phi = -\frac{q}{\epsilon_0} \quad \text{and} \quad \left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)A_i = -\frac{j_i}{\epsilon_0}.\text{''}$$

By this point our friend would be well on the way to a Nobel prize.

We shall see that the natural generalization of this argument to the case of gravity yields

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\mathbf{g} = \text{constant} \times \mathbf{T}.$$

However, Einstein showed that the way forward is not to tinker thus with Newtonian gravity, but to assign to the gravitational force a unique position as the force generated by the very dynamics of spacetime itself. The stimulus for this remarkable intellectual leap was the modern form of Galileo's famous observation that all bodies fall at the same speed.

3.2 The Principle of Equivalence

Inertial & gravitational mass As conventionally stated Newton's laws of motion are part definition and part empirical law. The purely empirical content can be summed up by the statements:

(i) the more carefully one isolates a body from external influences, the more nearly does its velocity $\boldsymbol{v}$ remain constant;

(ii) when several otherwise isolated bodies $\alpha = 1, \ldots, N$ interact with one another, it is possible to assign a number m_α to each body such that the quantity $\boldsymbol{p} \equiv \sum_\alpha m_\alpha \boldsymbol{v}_\alpha$ remains constant.

We call m_α the **inertial mass** of body α. When bodies are interacting, and therefore have changing individual momenta $\boldsymbol{p}_\alpha \equiv m_\alpha \boldsymbol{v}_\alpha$, it is convenient to imagine that they are acting on one another with a quantity "force", $\boldsymbol{f}_\alpha \equiv \mathrm{d}p_\alpha/\mathrm{d}t$. By statement (ii), $\sum_\alpha \boldsymbol{f}_\alpha = 0$.

Again according to Newton, the gravitational force between bodies α and β is

$$\boldsymbol{f}_{\alpha\beta} = F\frac{\boldsymbol{x}_\alpha - \boldsymbol{x}_\beta}{|\boldsymbol{x}_\alpha - \boldsymbol{x}_\beta|^3},$$

where the constant $F = GM_\alpha M_\beta$ is proportional to the product of two numbers M_α and M_β characteristic of the bodies—we call these masses **gravitational masses** of the bodies. If we place two bodies β and γ at the same distance from α, their accelerations will be in the ratio

$$\frac{|\mathrm{d}\boldsymbol{v}_\beta/\mathrm{d}t|}{|\mathrm{d}\boldsymbol{v}_\gamma/\mathrm{d}t|} = \frac{M_\beta m_\gamma}{m_\beta M_\gamma} = \frac{\Gamma_\beta}{\Gamma_\gamma}, \quad \text{where} \quad \Gamma_\nu \equiv \frac{M_\nu}{m_\nu}.$$

Thus β and γ will fall towards α at the same rate only if $\Gamma_\beta = \Gamma_\gamma$. Newton followed Galileo in thinking that all bodies fall at the same rate, and therefore assumed (with a suitable choice of G) that $\Gamma = 1$ for all particles. But in the 17^{th} century the experimental basis of this step was not strong.

3.3 Dicke–Eötvös Experiments

The most straightforward way to check whether Γ is the same for all masses is to compare the periods of pendulums made of different materials but having the same lengths. However, the impossibility of eliminating frictional resistance to the motion of a pendulum severely restricts the accuracy that can be attained in experiments of this kind.

In 1890 a Hungarian, Baron Roland v. Eötvös carried out a much more sensitive test of the proportionality of inertial and gravitational mass. A modified form of this experiment was repeated with greater accuracy by Robert Dicke and his students in Princeton in the 1960's.

Fig. 3 shows a schematic apparatus for the Dicke experiment. Two balls of approximately equal weight are attached to the ends of a short rod. This is attached to a wire so that it can execute torsional oscillations about a vertical axis. For simplicity we assume that that a new moon is nearly eclipsing the Sun at the time of the experiment, which begins at dusk. Then the acceleration of the balls on account of the Earth's spin lies in the plane of the paper, while that due to the Earth's rotation about the Sun and Moon is perpendicular to the paper. Hence we may forget about the spin of the Earth as we balance the books as regards forces perpendicular to the paper. The bar is aligned North-South and released. If Γ is identical for both balls and equal to Γ for the Earth as a whole, the gravitational force towards the Sun and Moon exactly equals the acceleration due to their instantaneous motion transverse to the Earth-Sun line, and there is no tendency for the wire to twist. But if Γ is abnormally large for one of the balls, say that to the South, this ball will start to fall towards the Sun faster than the other ball, and the rod will start to twist in the direction indicated. Consequently, the bar (which has a period of about one hour) will oscillate about an equilibrium position that is skewed with respect to the N-S line.

Figure 9truecm. 3

Schematic of the Dicke experiment to determine Γ.

During the evening, the torque on the wire due to the extra gravitational force on the southern ball diminishes. After midnight the torque starts to grow again, but with reversed sign. By dawn its displacement of the centre of oscillation is exactly opposite to that operating at dusk. By looking for a component in the motion of the bar with period 24 hrs and the expected phase with respect to solar time, Dicke and his collaborators were able to establish the limit $|\Gamma - 1| < 10^{-11}$.

What material should be used for the balls? Various things were tried but it is most interesting to compare heavy with light atoms, for example aluminium with gold, because:

(i) the nuclei of such atoms have very different proton/neutron numbers (Al = 13/14, Au = 79/118).

(ii) such atoms have very different contributions to their mass from:

  (a) electrostatic energy $[\frac{3}{5}(Ze)^2 r^{-1}/mc^2 \simeq 0.003$ (Al) or 0.008 (Au)];

  (b) overall binding energy [Mass defect$/mc^2 = 0.0089$ (Al) or 0.0084 (Au)];

  (c) virtual positrons $[m_{e+}/mc^2 \simeq 3\times10^{-7}$ (Al) or 2×10^{-6} (Au); see p. 33 of *Gravitation & Relativity* by M. G. Bowler for details].

Hence from these experiments we may conclude that $|\Gamma - 1| \ll 1$ for all forms of mass-energy, with the exceptions of energy associated with weak and gravitational interactions.[6]

Extrapolating wildly from these experiments we hypothesize:

Strong Principle of Equivalence: *No experiment could distinguish between a homogeneous gravitational field and an accelerating frame of reference. In particular, in any frame which falls freely through such a field all the laws of physics are the same as if no field were present.*

Real gravitational fields are never homogeneous, so they *can* be distinguished from an accelerating frame of reference. For example, consider a star-warrior who regains consciousness in a closed cabin some time after being taken prisoner. He reaches for his watch and knocks

[6] These contribute negligibly to the masses of atoms. However, since weak interactions are known to be intimately connected with electromagnetism, it is extremely unlikely that the value of Γ associated with weak-interaction energy differs from that associated with e.m. energy.

it to the floor. Fortunately it falls only slowly, so it continues to tick. Is he in a (possibly elastic) accelerating spaceship, or is he on an asteroid? By now fully alert he determines that plumb bobs on either side of the cabin point towards a spot some ten miles away. He instantly concludes that he is either on an asteroid or that opposite sides of his cabin are accelerating away from one another. Moments later he verifies that his bobs have *not* moved apart. Hence he must be in the gravitational field of an asteroid.

Exercise (9):

What would he have concluded if he had found that his bobs pointed *away* from a spot thirty yards distant?

This example shows that a gravitational field is generally *not* equivalent to an accelerating frame of reference. From the Principle of Equivalence we merely conclude that physics in an accelerating frame of reference must look like physics in a particular type of gravitational field. However, this observation suggests a strategy for discovering how things behave in a strong gravitational field: we first work out the equations governing motion in the absence of a gravitational field (which we understand) when referred to a non-inertial frame of reference. This is a purely mathematical exercise. The equations we derive will contain terms associated with pseudo-forces generated by our accelerating frame of reference. Since there is really no gravitational field present, these pseudo-force terms will be restricted in form. The plan is to obtain equations for physics in the presence of a true gravitational field by lifting these restrictions.

4 Tensors in General Relativity

We start by discovering what the laws of e.m. and mechanics look like in a non-inertial frame. Let x'^μ be such a non-inertial frame and x^μ an inertial frame. Then each primed coordinate is a smooth function $x'^\mu(x^\nu)$ of the four inertial coordinates. Let $x^\mu(\tau)$ be an arbitrary trajectory through space-time and $\psi(x^\mu)$ an arbitrary scalar function of the inertial coordinates x^μ. Then the rate of change of ψ as perceived by an observer who moves along the trajectory $x^\mu(\tau)$ is

$$\frac{\mathrm{d}\psi}{\mathrm{d}\tau} = \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau}\frac{\partial\psi}{\partial x^\mu} \equiv v^\mu\frac{\partial\psi}{\partial x^\mu},$$

where we have defined the observer's 4-velocity $v^\mu \equiv \mathrm{d}x^\mu/\mathrm{d}\tau$. Since by the chain rule

$$\frac{\partial}{\partial x^\mu} = \frac{\partial x'^\nu}{\partial x^\mu}\frac{\partial}{\partial x'^\nu} \tag{4.1}$$

we have

$$\frac{\mathrm{d}\psi}{\mathrm{d}\tau} = v^\mu\frac{\partial x'^\nu}{\partial x^\mu}\frac{\partial\psi}{\partial x'^\nu}.$$

If we define the observer's 4-velocity in the non-inertial primed frame to be

$$v'^\nu \equiv \frac{\partial x'^\nu}{\partial x^\mu}v^\mu, \tag{4.2}$$

then we may write

$$\frac{\mathrm{d}\psi}{\mathrm{d}\tau} = v'^\nu\frac{\partial\psi}{\partial x'^\nu}.$$

A natural extension of this argument leads us to define the primed components of any up vector A^μ as given in terms of the un-primed components by

$$A'^\nu \equiv \frac{\partial x'^\nu}{\partial x^\mu}A^\mu. \tag{4.3}$$

Note that if the primed frame were inertial, we would have $x'^\nu = x_0^\nu + \Lambda^\nu{}_\mu x^\mu$ (x_0^ν a constant 4-vector), so that $\partial x'^\nu/\partial x^\mu = \Lambda^\nu{}_\mu$ and the transformation (4.3) would reduce to a standard Lorentz transformation of an up vector.

If v^μ and u^μ are two up vectors, all inertial observers will agree on the value of the scalar

$$s \equiv \eta_{\mu\nu} u^\mu v^\nu. \tag{4.4}$$

How can we recover this number from the primed components v'^μ and u'^μ? First we express v^μ in terms of v'^μ. We use the chain rule to express $\mathrm{d}x'^\mu$ as

$$\mathrm{d}x'^\mu = \frac{\partial x'^\mu}{\partial x^\nu}\,\mathrm{d}x^\nu. \tag{4.5}$$

Dividing by $\mathrm{d}x'^\kappa$ and proceeding to the limit $\mathrm{d}x'^\kappa \to 0$ at fixed values of all the other coordinates, we get

$$\delta^\mu_\kappa = \frac{\partial x'^\mu}{\partial x'^\kappa} = \frac{\partial x'^\mu}{\partial x^\nu}\frac{\partial x^\nu}{\partial x'^\kappa}. \tag{4.6}$$

Thus the matrix $\partial x^\nu/\partial x'^\kappa$ is the inverse of the matrix $\partial x'^\mu/\partial x^\nu$. Premultiplying equation (4.2) by this matrix we solve for v^μ:

$$v^\mu = \frac{\partial x^\mu}{\partial x'^\nu} v'^\nu. \tag{4.7}$$

Using this relation to eliminate the unprimed components from (4.4) we get

$$s = \left(\eta_{\mu\nu}\frac{\partial x^\mu}{\partial x'^\kappa}\frac{\partial x^\nu}{\partial x'^\lambda}\right) u'^\kappa v'^\lambda.$$

If we define

$$g'_{\kappa\lambda} \equiv \eta_{\mu\nu}\frac{\partial x^\mu}{\partial x'^\kappa}\frac{\partial x^\nu}{\partial x'^\lambda}, \tag{4.8}$$

we have

$$s = g'_{\kappa\lambda} u'^\kappa v'^\lambda. \tag{4.9}$$

Like $\eta_{\kappa\lambda}$ the general **metric tensor** $g'_{\kappa\lambda}$ is symmetric; $g'_{\kappa\lambda} = g'_{\lambda\kappa}$. However, it is not necessarily diagonal. It is called the metric tensor because it allows us to calculate the lengths of vectors such as v'^λ.

We may use $g'_{\kappa\lambda}$ to lower indices;

$$v'_\kappa \equiv g'_{\kappa\lambda} v'^\lambda. \tag{4.10}$$

Let $g'^{\mu\nu}$ be the tensor which raises indices. Then in order that the operations of raising and lowering be mutual inverses we require that for all v'^μ

$$\delta^\mu_\lambda v'^\lambda = v'^\mu = g'^{\mu\kappa} g'{}_{\kappa\lambda} v'^\lambda.$$

i.e. that $g'^{\mu\kappa} g'_{\kappa\lambda} = \delta^\mu_\lambda$ and hence that $g'^{\mu\kappa}$ is the inverse of $g'{}_{\kappa\lambda}$.

Exercise (10):

Show that this definition of $g'^{\mu\kappa}$ is equivalent to the definition

$$g'^{\kappa\lambda} = \frac{\partial x'^\kappa}{\partial x^\mu}\frac{\partial x'^\lambda}{\partial x^\nu}\eta^{\mu\nu}. \tag{4.11}$$

Similarly, if for any tensors **F** and **G** we define

$$F'^{\kappa\lambda} \equiv \frac{\partial x'^\kappa}{\partial x^\mu}\frac{\partial x'^\lambda}{\partial x^\nu} F^{\mu\nu} \qquad \text{and} \qquad G'_{\kappa\lambda} \equiv \frac{\partial x^\mu}{\partial x'^\kappa}\frac{\partial x^\nu}{\partial x'^\lambda} G_{\mu\nu}, \tag{4.12}$$

we ensure that the primed observer will be able to calculate the scalar quantities $F^{\mu\nu} v_\mu u_\nu$ and $G_{\mu\nu} v^\mu u^\nu$ from primed quantities. The generalization to tensors of arbitrary rank is obvious.

Exercise (11):

Show that if x'^{μ} and x''^{μ} are two non-inertial frames, the transformation rules

$$v''^{\mu} = \frac{\partial x''^{\mu}}{\partial x'^{\nu}} v'^{\nu} \quad ; \quad v''_{\mu} = \frac{\partial x'^{\nu}}{\partial x''^{\mu}} v'_{\nu} \tag{4.13a}$$

$$F''^{\mu\nu} = \frac{\partial x''^{\mu}}{\partial x'^{\kappa}} \frac{\partial x''^{\nu}}{\partial x'^{\lambda}} F'^{\kappa\lambda} \quad \text{etc} \tag{4.13b}$$

apply.

[Hint: divide (4.5) by $\mathrm{d}x''^{\kappa}$ to obtain a relation equivalent to $\dfrac{\partial x''^{\mu}}{\partial x^{\kappa}} \dfrac{\partial x^{\kappa}}{\partial x'^{\nu}} = \dfrac{\partial x''^{\mu}}{\partial x'^{\nu}}$].

Notice that there is an easy way to figure out whether to multiply by $\partial x^{\mu}/\partial x'^{\nu}$ or by $\partial x'^{\mu}/\partial x^{\nu}$ when transforming an object $G^{\mu\cdots}$ or $G_{\mu\ldots}$: If the prime are up on the left, put them up on the right by using $\partial x'^{\mu}/\partial x^{\nu}$; if the unprimes are up on the left put them on top on the right with $\partial x^{\mu}/\partial x'^{\nu}$. The other kind of index in the equation will "cancel out" just as in ordinary multiplication of fractions. These rules extend in the obvious way to down vectors.

4.2 Equation of Motion in a Non-Inertial Frame

We have given a prescription for transforming the 4-velocity $v^{\mu} \equiv \mathrm{d}x^{\mu}/\mathrm{d}\tau$ along a trajectory into a non-inertial frame. We may similarly transform the acceleration $a^{\mu} \equiv \mathrm{d}^2 x^{\mu}/\mathrm{d}\tau^2$ into a non-inertial frame:

$$a'^{\mu} = \frac{\partial x'^{\mu}}{\partial x^{\nu}} \frac{\mathrm{d}^2 x^{\nu}}{\mathrm{d}\tau^2}. \tag{4.14}$$

How is v'^{μ} related to a'^{μ}? Rearranging (4.14) we have

$$\begin{aligned} a'^{\mu} &= \frac{\mathrm{d}}{\mathrm{d}\tau}\Big(\frac{\partial x'^{\mu}}{\partial x^{\nu}} \frac{\mathrm{d}x^{\nu}}{\mathrm{d}\tau}\Big) - \frac{\mathrm{d}}{\mathrm{d}\tau}\Big(\frac{\partial x'^{\mu}}{\partial x^{\nu}}\Big) \frac{\mathrm{d}x^{\nu}}{\mathrm{d}\tau} \\ &= \frac{\mathrm{d}v'^{\mu}}{\mathrm{d}\tau} - \frac{\partial^2 x'^{\mu}}{\partial x^{\kappa} \partial x^{\lambda}} \frac{\mathrm{d}x^{\kappa}}{\mathrm{d}\tau} \frac{\mathrm{d}x^{\lambda}}{\mathrm{d}\tau}, \end{aligned}$$

or using (4.7) to express $\mathrm{d}x^{\mu}/\mathrm{d}\tau$ in terms of v'^{μ}

$$\frac{\mathrm{d}v'^{\mu}}{\mathrm{d}\tau} = a'^{\mu} + \Big(\frac{\partial^2 x'^{\mu}}{\partial x^{\kappa} \partial x^{\lambda}} \frac{\partial x^{\kappa}}{\partial x'^{\alpha}} \frac{\partial x^{\lambda}}{\partial x'^{\beta}}\Big) v'^{\alpha} v'^{\beta}. \tag{4.15}$$

If we define the **Christoffel symbol** by

$$\Gamma'^{\mu}_{\alpha\beta} \equiv -\frac{\partial^2 x'^{\mu}}{\partial x^{\kappa} \partial x^{\lambda}} \frac{\partial x^{\kappa}}{\partial x'^{\alpha}} \frac{\partial x^{\lambda}}{\partial x'^{\beta}}, \tag{4.16}$$

then (4.15) may be rewritten

$$\frac{\mathrm{d}v'^{\mu}}{\mathrm{d}\tau} = a'^{\mu} - \Gamma'^{\mu}_{\alpha\beta} v'^{\alpha} v'^{\beta}. \tag{4.17}$$

By the relativistic law of motion (2.41), a'^{μ} is equal to the 4-force per unit mass generated by, for example, an e.m. field. Hence in a non-inertial frame the rate of change of the 4-velocity v'^{μ} is equal to the regular 4-force per unit mass plus a pseudo-force given by the second term on the right of (4.17). This term is made up by contracting a quantity with three loose indices with the square of the 4-velocity. From (4.16) we see that $\Gamma'^{\mu}_{\alpha\beta} = \Gamma'^{\mu}_{\beta\alpha}$ is symmetric in its lower indices. $\mathbf{\Gamma}'$ cannot be a tensor since all its components are zero in an inertial frame, so if it transformed like a third-rank tensor, all its components would be zero in any coordinate system.

The principle of equivalence suggests that gravitational forces will take the same form as pseudo-forces. Thus $\mathbf{\Gamma}$ should play the same role for the gravitational field as $\mathbf{F}$ does for the e.m. field. To make this analogy clearer we equate in equation (4.17) the acceleration $\mathbf{a}'$ with the e.m. force per unit mass $\mathbf{f}'/m_0$ on a particle of charge q and mass m_0 [equations (2.48) and (2.49)]:

$$\frac{\mathrm{d}v'^\mu}{\mathrm{d}\tau} = \frac{q}{m_0} F'^\mu{}_\nu v'^\nu - \Gamma'^\mu_{\alpha\beta} v'^\alpha v'^\beta .$$

This equations shows that $\mathbf{F}'$ and $\mathbf{\Gamma}'$ make contributions to the rate of change of the 4-velocity which differ mainly in the number of copies of $\mathbf{v}'$ that attach to them; $\mathbf{F}'$ uses one copy of $\mathbf{v}'$ because photons are spin-one particles, while $\mathbf{\Gamma}'$ employs two copies of $\mathbf{v}'$ because gravitons are spin-two objects. (Or would be if anyone could figure out how to quantise g.r.) The other significant difference between the e.m. and pseudo-force terms in (4.17) is that the latter contains no factor q/m_0 – from the Dicke–Eötvös experiments we know that all bodies have unit gravitational "charge-to-mass ratio".

Just as $\mathbf{F}$ is a kind of gradient of $\mathbf{A}$, we should be able to express $\mathbf{\Gamma}'$ in terms of derivatives of the gravitational potential. Indeed, we next show that $\mathbf{g}'$ is the relativistic generalization of Newton's Φ by writing $\mathbf{\Gamma}'$ in terms of $\mathbf{g}$'s derivatives.

An application of the chain rule enables us to simplify the definition (4.16) of $\Gamma'^\mu_{\alpha\beta}$ slightly:

$$\begin{aligned} \Gamma'^\mu_{\alpha\beta} &= -\frac{\partial}{\partial x'^\alpha}\Big(\frac{\partial x'^\mu}{\partial x^\lambda}\Big)\frac{\partial x^\lambda}{\partial x'^\beta} \\ &= -\Big[\frac{\partial}{\partial x'^\alpha}\Big(\frac{\partial x'^\mu}{\partial x'^\beta}\Big) - \frac{\partial x'^\mu}{\partial x^\lambda}\frac{\partial^2 x^\lambda}{\partial x'^\alpha \partial x'^\beta}\Big] \\ &= \frac{\partial^2 x^\lambda}{\partial x'^\alpha \partial x'^\beta}\frac{\partial x'^\mu}{\partial x^\lambda}. \end{aligned} \tag{4.18}$$

Even though $\mathbf{\Gamma}'$ is not a tensor, we may define[7]

$$\Gamma'_{\mu,\alpha\beta} \equiv g'_{\mu\nu}\Gamma'^\nu_{\alpha\beta}. \tag{4.19}$$

By (4.18), (4.6) and (4.8), this is

$$\Gamma'_{\mu,\alpha\beta} = \frac{\partial^2 x^\gamma}{\partial x'^\alpha \partial x'^\beta}\frac{\partial x^\delta}{\partial x'^\mu}\eta_{\gamma\delta}. \tag{4.20}$$

This may be expressed in terms of derivatives of the metric tensor as follows. Differentiating (4.8) we get

$$\begin{aligned} \frac{\partial g'_{\mu\alpha}}{\partial x'^\beta} &= \frac{\partial}{\partial x'^\beta}\Big(\frac{\partial x^\gamma}{\partial x'^\mu}\frac{\partial x^\delta}{\partial x'^\alpha}\eta_{\gamma\delta}\Big) \\ &= \Big(\frac{\partial^2 x^\gamma}{\partial x'^\beta \partial x'^\mu}\frac{\partial x^\delta}{\partial x'^\alpha} + \frac{\partial x^\gamma}{\partial x'^\mu}\frac{\partial^2 x^\delta}{\partial x'^\beta \partial x'^\alpha}\Big)\eta_{\gamma\delta}. \end{aligned} \tag{4.21a}$$

The cyclic interchange of indices $\beta \to \alpha \to \mu \to \beta$ yields successively

$$\frac{\partial g'_{\alpha\beta}}{\partial x'^\mu} = \Big(\frac{\partial^2 x^\gamma}{\partial x'^\mu \partial x'^\alpha}\frac{\partial x^\delta}{\partial x'^\beta} + \frac{\partial x^\gamma}{\partial x'^\alpha}\frac{\partial^2 x^\delta}{\partial x'^\mu \partial x'^\beta}\Big)\eta_{\gamma\delta}. \tag{4.21b}$$

$$\frac{\partial g'_{\beta\mu}}{\partial x'^\alpha} = \Big(\frac{\partial^2 x^\gamma}{\partial x'^\alpha \partial x'^\beta}\frac{\partial x^\delta}{\partial x'^\mu} + \frac{\partial x^\gamma}{\partial x'^\beta}\frac{\partial^2 x^\delta}{\partial x'^\alpha \partial x'^\mu}\Big)\eta_{\gamma\delta} \tag{4.21c}$$

If we now add (4.21a,c), subtract (4.21b) and exploit the symmetry of $\eta_{\gamma\delta}$, we obtain the desired expression for $\mathbf{\Gamma}'$ in terms of $\partial\mathbf{x}/\partial\mathbf{g}$:

$$\Gamma'_{\mu,\alpha\beta} = \tfrac{1}{2}\Big(\frac{\partial g'_{\mu\alpha}}{\partial x'^\beta} + \frac{\partial g'_{\beta\mu}}{\partial x'^\alpha} - \frac{\partial g'_{\alpha\beta}}{\partial x'^\mu}\Big). \tag{4.22}$$

[7] The comma amid $\mathbf{\Gamma}$'s subscripts should not be confused with the commas which indicate derivation; the latter are rather larger.

Clearly,

$$\Gamma'^{\mu}_{\alpha\beta} = \tfrac{1}{2} g'^{\mu\nu} \Big(\frac{\partial g'_{\nu\alpha}}{\partial x'^{\beta}} + \frac{\partial g'_{\beta\nu}}{\partial x'^{\alpha}} - \frac{\partial g'_{\alpha\beta}}{\partial x'^{\nu}} \Big). \tag{4.23}$$

Notice the pattern of these important formulae: the three terms in (...) are just the first derivative of g with the indices cyclically permuted. The minus assign attaches to the term which groups the indices in the same way as $\boldsymbol{\Gamma}$.

4.3 Covariant Differentiation

Before we can use (4.18) to calculate the trajectory of a charged particle in a crazy coordinate system we need to know how to calculate $F_{\mu\nu}$ by taking derivatives of the e.m. potential A_ν. So our next task is to find the rules for taking derivatives of vector fields in non-inertial coordinates.

The metric tensor $g'_{\mu\nu}$ enables us to calculate the length s of any curve $x'^{\mu}(\lambda)$ in space-time:

$$s \equiv \int_a^b \sqrt{\Big| g_{\mu\nu} \frac{\mathrm{d}x'^{\mu}}{\mathrm{d}\lambda} \frac{\mathrm{d}x'^{\nu}}{\mathrm{d}\lambda} \Big|}\, \mathrm{d}\lambda. \tag{4.24}$$

If the curve is time-like, s is just c times the elapse $\Delta\tau$ of time on the watch of the observer whose trajectory $x'^{\mu}(\lambda)$ is. If there is an inertial frame in which all the points on the curve have the same value of x^0, s coincides with the length of the curve as measured with metre rules etc by an observer who is stationary in that priviledged frame. We shall call s the **affine parameter** along the curve and use it to characterize points on the curve; hence we write $x'^{\mu}(s)$.

We shall sometimes need to compare vectors at different points on the curve. In an inertial frame this is easy: two vectors are the same iff all their components are the same. But in passing from an inertial to a non-inertial frame by equations (4.3), we change the components of vectors in a position-dependent way. So two vectors that are equal in the sense that in an inertial frame all their components are equal, can have different components in a non-inertial frame. We need a way of diagnosing this condition of hidden equality.

Suppose that in an inertial frame we have a vector field $\mathbf{A}(\mathbf{x})$. By (4.3) this gives rise to a vector field $\mathbf{A}'(\mathbf{x}')$ in a non-inertial frame. As we go along a curve $\mathbf{x}(s)$ the rate of change in the vector of the field is

$$\dot{\mathbf{A}} \equiv \frac{\mathrm{d}}{\mathrm{d}s}\mathbf{A} = \frac{\mathrm{d}x'^{\kappa}}{\mathrm{d}s} \frac{\partial \mathbf{A}}{\partial x'^{\kappa}}. \tag{4.25}$$

Using (4.7) move the $\mathbf{A}$ on the right into the primed system, we get

$$\begin{aligned} \dot{A}^{\mu} &= \frac{\mathrm{d}x'^{\kappa}}{\mathrm{d}s} \frac{\partial}{\partial x'^{\kappa}} \Big(\frac{\partial x^{\mu}}{\partial x'^{\alpha}} A'^{\alpha} \Big) \\ &= \frac{\mathrm{d}x'^{\kappa}}{\mathrm{d}s} \Big(\frac{\partial x^{\mu}}{\partial x'^{\alpha}} \frac{\partial A'^{\alpha}}{\partial x'^{\kappa}} + \frac{\partial^2 x^{\mu}}{\partial x'^{\kappa} \partial x'^{\alpha}} A'^{\alpha} \Big). \end{aligned}$$

Finally, premultiplying by $\partial x'^{\nu}/\partial x^{\mu}$ and using (4.18) we get

$$\dot{A}'^{\nu} \equiv \frac{\partial x'^{\nu}}{\partial x^{\mu}} \dot{A}^{\mu} = \frac{\mathrm{d}x'^{\kappa}}{\mathrm{d}s} \Big(\frac{\partial A'^{\nu}}{\partial x'^{\kappa}} + \Gamma'^{\nu}_{\kappa\alpha} A'^{\alpha} \Big). \tag{4.26}$$

(Notice that $\dot{A}'^{\nu}$, the $\nu^{\rm th}$ component of the vector $\dot{\mathbf{A}}$, is *defined* by this equation. It must not be confused with the rate of change with s of the $\nu^{\rm th}$ component of $\mathbf{A}'$.) If we define a new type of derivative, the **covariant derivative** by

$$A'^{\nu}_{;\kappa} \equiv \nabla_{\kappa} A'^{\nu} \equiv \frac{\partial A'^{\nu}}{\partial x'^{\kappa}} + \Gamma'^{\nu}_{\kappa\alpha} A'^{\alpha}, \tag{4.27}$$

then equation (4.26) can be written

$$\dot{A}'^{\nu} = \frac{\mathrm{d}x'^{\kappa}}{\mathrm{d}s}\nabla_\kappa A'^{\nu}. \tag{4.28}$$

The second term in the definition (4.27) of the covariant derivative has the following physical interpretation. For each value of κ, say $\kappa = 1$, we have a matrix $\Gamma'^{\nu}_{1\alpha}$. When we multiply this matrix by δx^1 we obtain the Lorentz transformation matrix $\mathbf{\Lambda}$ which tells us by how much the speed and orientation of the frame used at $\mathbf{x}$ differs from that used at $(x^0, x^1 + \delta x^1, x^2, x^3)$.[8]

If $\mathbf{A}$ is really the same all along the curve, and only seems to change because we are using a non-inertial coordinate system, we have $\dot{A}'^{\nu} = 0$, and thus that the "gradient" $\nabla_\kappa A'^{\nu}$ of A'^{ν} either vanishes or is "perpendicular" to the direction $\mathrm{d}x'^{\kappa}/\mathrm{d}s$ in which we are moving.

How does ∇ operate on down vectors? Consider

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}s}(A'^{\mu}B'_{\mu}) &= \frac{\mathrm{d}x'^{\kappa}}{\mathrm{d}s}\frac{\partial}{\partial x'^{\kappa}}(A'^{\mu}B'_{\mu}) \\
&= \frac{\mathrm{d}x'^{\kappa}}{\mathrm{d}s}\Big[\Big(\frac{\partial A'^{\mu}}{\partial x'^{\kappa}}\Big)B'_{\mu} + A'^{\mu}\Big(\frac{\partial B'^{\mu}}{\partial x'^{\kappa}}\Big)\Big] \\
&= \frac{\mathrm{d}x'^{\kappa}}{\mathrm{d}s}\Big[(\nabla_\kappa A'^{\mu})B'_{\mu} - \Gamma'^{\mu}_{\kappa\alpha}A'^{\alpha}B'_{\mu} + A'^{\mu}\frac{\partial B'_{\mu}}{\partial x'^{\kappa}}\Big] \\
&= \frac{\mathrm{d}x'^{\kappa}}{\mathrm{d}s}\Big[(\nabla_\kappa A'^{\mu})B'_{\mu} + A'^{\mu}\frac{\partial B'_{\mu}}{\partial x'^{\kappa}} - \Gamma'^{\alpha}_{\kappa\mu}A'^{\mu}B'_{\alpha}\Big].
\end{aligned}$$

This suggests that we define

$$\nabla_\kappa B'_{\mu} \equiv \frac{\partial B'_{\mu}}{\partial x'^{\kappa}} - \Gamma'^{\alpha}_{\kappa\mu}B'_{\alpha} \tag{4.29}$$

for then we will have $\nabla_\kappa(A'_{\mu}B'^{\mu}) = B'^{\mu}\nabla_\kappa A'_{\mu} + A'_{\mu}\nabla_\kappa B'^{\mu}$ and ∇ will operate on such products like any other derivative operator.

The same argument applied to quantities like $G'_{\mu\nu}A'^{\mu}B'^{\nu}$ leads to the rules

$$G'_{\mu\nu;\kappa} \equiv \nabla_\kappa G'_{\mu\nu} \equiv \frac{\partial G'_{\mu\nu}}{\partial x'^{\kappa}} - \Gamma'^{\alpha}_{\kappa\mu}G_{\alpha\nu} - \Gamma'^{\alpha}_{\kappa\nu}G_{\mu\alpha} \tag{4.30a}$$

$$G'^{\mu\nu}_{\ ;\kappa} \equiv \nabla_\kappa G'^{\mu\nu} \equiv \frac{\partial G'^{\mu\nu}}{\partial x'^{\kappa}} + \Gamma'^{\mu}_{\kappa\alpha}G^{\alpha\nu} + \Gamma'^{\nu}_{\kappa\alpha}G^{\mu\alpha}. \tag{4.30b}$$

Notice that each index requires a Γ-symbol, with a plus or a minus sign according as the index is up or down.

In the same spirit we define the operation of ∇ on scalars to be identical with partial differentiation:

$$\nabla_\kappa\psi = \frac{\partial\psi}{\partial x'^{\kappa}}$$

What action does ∇ have on the metric tensor? Suppose that $\mathbf{A}$ and $\mathbf{B}$ are two vector fields that everywhere have the same components in an inertial frame. Then $\nabla_\kappa A'^{\mu} = \nabla_\kappa B'^{\mu} = 0$. Also $A^{\mu}B_{\mu} = g'_{\mu\nu}A'^{\mu}B'^{\nu}$ is everywhere the same. Hence for all curves $x'(s)$

$$0 = \frac{\mathrm{d}}{\mathrm{d}s}\big(g'_{\mu\nu}A'^{\mu}B'^{\nu}\big).$$

[8] In "gauge field theories" this idea is generalized to define covariant derivatives for objects ψ that live in spaces other than space-time. In the simplest case ψ lives in the two-dimensional space of complex numbers, for which the analogue of a Lorentz transformation is mutiplication by another complex number, say iqA_1. The covariant derivative is now $\mathcal{D}_\mu \equiv \partial_\mu + iqA_\mu$. If ψ is the wavefunction of a spin-zero particle of charge q, A_μ proves to be the regular e.m. potential.

Replacing $\mathrm{d}/\mathrm{d}s$ with $(\mathrm{d}x'^{\kappa}/\mathrm{d}s)\nabla_\kappa$ and differentiating each item in the bracket, we get

$$
\begin{aligned}
0 &= \frac{\mathrm{d}x'^{\kappa}}{\mathrm{d}s}\Big\{(\nabla_\kappa g'_{\mu\nu})A'^{\mu}B'^{\nu} + g'_{\mu\nu}\big[(\nabla_\kappa A'^{\mu})B'^{\nu} + A'^{\mu}(\nabla_\kappa B'^{\nu})\big]\Big\} \\
&= \frac{\mathrm{d}x'^{\kappa}}{\mathrm{d}s}A'^{\mu}B'^{\nu}\nabla_\kappa g'_{\mu\nu}.
\end{aligned}
$$

Since $\mathrm{d}x'^{\kappa}/\mathrm{d}s$, A'^{μ} and B'^{ν} are all arbitrary, it follows that

$$\nabla_\kappa g'_{\mu\nu} = 0. \tag{4.31}$$

In words, the *covariant derivative of the metric tensor is always zero.*

If $x^{\mu}(s)$ is a straight line, all components of the "tangent vector" $\mathrm{d}x^{\mu}/\mathrm{d}s$ are constant in an inertial frame. Hence in any coordinate system the tangent vector $x'^{\mu}(s)$ of a straight line satisfies the o.d.e.

$$0 = \frac{\mathrm{d}x'^{\kappa}}{\mathrm{d}s}\nabla_\kappa \frac{\mathrm{d}x'^{\mu}}{\mathrm{d}s}. \tag{4.32}$$

Substituting for ∇_κ this becomes

$$
\begin{aligned}
0 &= \frac{\mathrm{d}x'^{\kappa}}{\mathrm{d}s}\Big(\frac{\partial}{\partial x'^{\kappa}}\frac{\mathrm{d}x'^{\mu}}{\mathrm{d}s} + \Gamma'^{\mu}_{\kappa\alpha}\frac{\mathrm{d}x'^{\alpha}}{\mathrm{d}s}\Big) \\
&= \frac{\mathrm{d}^2x'^{\mu}}{\mathrm{d}s^2} + \Gamma'^{\mu}_{\kappa\alpha}\frac{\mathrm{d}x'^{\kappa}}{\mathrm{d}s}\frac{\mathrm{d}x'^{\alpha}}{\mathrm{d}s} \qquad (x'^{\mu}(s) \text{ a straight line.})
\end{aligned}
\tag{4.33}
$$

Exercise (12):

Obtain (4.33) by extremizing the integral (4.24) with respect to variations of the path $x'^{\mu}(s)$; a straight line is the least distance between two points.

In terms of covariant derivatives, Newton's law of motion (2.41) and the Maxwell equations (2.46) become

$$m_0 v'^{\kappa}\nabla_\kappa v'^{\mu} = f'^{\mu}, \tag{4.34a}$$

$$F'^{\mu\nu}{}_{;\nu} = \mu_0 j'^{\mu}. \tag{4.34b}$$

The other laws of e.m. (2.37) (2.46) remain unchanged because the Christoffel symbols introduced in going over from partial to covariant derivatives magically cancel.

Exercise (13):

Prove that $A'_{\mu;\nu} - A'_{\nu;\mu} = A'_{\mu},_{\nu} - A'_{\nu},_{\mu}$.

4.4 Summary

The rules for transforming between non-inertial frames are the same as those for making regular Lorentz transformation with the substitutions

$$\Lambda_\mu{}^{\nu} \to \frac{\partial x'^{\nu}}{\partial x''^{\mu}} \quad ; \quad \Lambda^{\mu}{}_{\nu} \to \frac{\partial x''^{\mu}}{\partial x'^{\nu}}. \quad \text{Thus} \quad A''_{\mu} = \frac{\partial x'^{\nu}}{\partial x''^{\mu}}A'_{\nu}.$$

The Minkowski metric $\boldsymbol{\eta}$ is replaced by the metric tensor $\mathbf{g}$, which remains symmetric but is no longer its own inverse; consequently the up-up and down-down forms of $\mathbf{g}$ are in general distinct.

In a non-inertial frame $\mathbf{x}$ the partial derivative operator $\partial_\mu \equiv \partial/\partial x^\mu$ should be replaced with the covariant derivative operator ∇_μ:

$$\begin{aligned} \nabla_\mu \psi &= \partial_\mu \psi \\ A^\nu_{;\mu} \equiv \nabla_\mu A^\nu &= \partial_\mu A^\nu + \Gamma^\nu_{\mu\alpha} A^\alpha \quad ; \quad \nabla_\mu B^{\nu\lambda} = \partial_\mu B^{\nu\lambda} + \Gamma^\nu_{\mu\alpha} B^{\alpha\lambda} + \Gamma^\lambda_{\mu\alpha} B^{\nu\alpha} \\ \nabla_\mu A_\nu &= \partial_\mu A_\nu - \Gamma^\alpha_{\mu\nu} A_\alpha \quad ; \quad \nabla_\mu B_{\nu\lambda} = \partial_\mu B_{\nu\lambda} - \Gamma^\alpha_{\mu\nu} B_{\alpha\lambda} - \Gamma^\alpha_{\mu\nu} B_{\nu\alpha} \end{aligned}$$

The Christoffel symbol $\mathbf{\Gamma}$ is

$$\Gamma^\mu_{\alpha\beta} = \tfrac{1}{2} g^{\mu\nu} \Big(\frac{\partial g_{\nu\alpha}}{\partial x^\beta} + \frac{\partial g_{\beta\nu}}{\partial x^\alpha} - \frac{\partial g_{\alpha\beta}}{\partial x^\nu} \Big).$$

The covariant derivative of $\mathbf{g}$ always vanishes: $\nabla \mathbf{g} = 0$

5 Gravity, Geometry & the Einstein Field Equations

Now that we have completed our programme for discovering what physics looks like in a non-inertial frame, it is a good idea to take a rest from all these acres of indices and summarise the physical content of our formulae.

We have defined quantities $g'_{\mu\nu}$, p'_μ, $F'_{\mu\nu}$, j'_μ, $\Gamma'^\mu_{\alpha\beta}$ etc which enable us to use a non-inertial coordinate system $\mathbf{x}'$ to find the space-time trajectory of a charged particle in an e.m. field. We defined these quantities in terms of the momenta, e.m. field tensor etc in an underlying inertial coordinate system $\mathbf{x}$ and the coordinate transformation $\mathbf{x}'(\mathbf{x})$ that couples the inertial and non-inertial systems. But we have found formulae (4.13) and (4.23) which enable us to calculate the values $g''_{\mu\nu}$ etc of all needful quantities in a second non-inertial coordinate system without reference back to the inertial system $\mathbf{x}$.

Since we shall no longer need to refer constantly to an inertial system, we now drop the convention that the unprimed system $\mathbf{x}$ is inertial; from here on *all systems are to be assumed to be non-inertial unless explicitly specified as inertial.*

The principle of equivalence suggests that a gravitational field will look very much like a pseudo-force in an accelerating frame of reference. The Christoffel symbol $\mathbf{\Gamma}$ generates the pseudo-force associated with an accelerating frame, so when a gravitational field is present $\mathbf{\Gamma}$ will play the role of the Newtonian force $\boldsymbol{f}$. We have identified the metric $\mathbf{g}$ as the relativistic generalization of the Newtonian potential Φ on the ground that $\mathbf{\Gamma}$ can be written in terms of derivatives of $\mathbf{g}$ just as $\boldsymbol{f} = -\nabla\Phi$.

In Newton's theory $\boldsymbol{f}$ and Φ are related to the density ρ of gravitating matter via Poisson's equation (3.5). The relativistic generalization of (3.5) should be a second-order p.d.e. in $\mathbf{g}$, or equivalently, a first-order p.d.e. in $\mathbf{\Gamma}$. What is this equation?

Since we can make $\mathbf{\Gamma}$ as large as we like simply by choosing a perverse coordinate system, it is clear that the trick in finding suitable field equations is to find a differential operator on $\mathbf{\Gamma}$ which differentiates away all the contribution to $\mathbf{\Gamma}$ that is caused by mere perversity of the coordinate system. The key to finding this operator proves to be an examination of the geometrical relationships between the lengths of lines and the magnitudes of angles between lines.

We have seen that the metric tensor enables us to define the length of any curve in space-time, and in particular to determine through (4.33) which curves $\mathbf{x}'(s)$ are straight. Now suppose we draw a straight line in a portion of space in which there is no gravitational field and then draw a unit circle around some point on this line. Then no matter what coordinate system we use for the calculations, we shall find that the length s of the circumference is exactly $\pi = 3.14159\ldots$ times the length of the circle's diameter. How come? By changing the coordinate system we can change $\mathbf{g}$ at any given point to almost any value [see (4.8)]. So how come that when we evaluate the integral (4.24) over two completely different sets of points, we always get answers in the same ratio? It must be that $\mathbf{g}$ at one point *is not independent of* $\mathbf{g}$ *at neighbouring points*: $\mathbf{g}$ must satisfy some differential equation. Einstein's idea, and it was pure magic, was that it is *this* differential equation which tells us that there is no gravitational field present, only a perverse coordinate system. Let us find this differential equation.

There are many geometrical relationships in addition to the one just discussed which $\mathbf{g}$ must furnish if there is no gravitational field present. For example, there are 180° in a triangle. But the key to the equation we are seeking turns out to be something slightly odd. It is to consider what happens when we slide a vector around a closed curve while being careful not to rotate the vector. If we do this on a table, the vector (a pencil, say) will be back in its old configuration at the end of the experiment:

But on a sphere things go differently:

In fact, on a sphere of radius r, the angle through which a pencil rotates on being "parallel-transported" around a curve is equal to the area enclosed by this curve divided by r^2.

5.1 The curvature tensor

If we parallel-transport a vector $\mathbf{A}$ around a closed curve $\mathbf{x}(s)$ in space-time, we have that at each point on the curve $\dot{\mathbf{x}} \cdot \nabla \mathbf{A} = 0$ (this is just a statement of the invariance along the curve of $\mathbf{A}$'s components in an inertial frame)

$$0 = \frac{\mathrm{d}x^\alpha}{\mathrm{d}s} \nabla_\alpha A^\mu = \frac{\mathrm{d}x^\alpha}{\mathrm{d}s}\Big(\frac{\partial A^\mu}{\partial x^\alpha} + \Gamma^\mu_{\alpha\beta} A^\beta\Big). \tag{5.1}$$

Consequently, the total change in each component A^μ on going around is

$$\Delta A^\mu = \oint \frac{\partial A^\mu}{\partial x^\alpha}\frac{\mathrm{d}x^\alpha}{\mathrm{d}s}\,\mathrm{d}s = -\oint \Gamma^\mu_{\alpha\beta} A^\beta \frac{\mathrm{d}x^\alpha}{\mathrm{d}s}\,\mathrm{d}s. \tag{5.2}$$

In this integral both $\Gamma^\mu_{\alpha\beta}$ and A^β are functions of s through $\mathbf{x}(s)$. However, if we consider only infinitesimal loops we may expand each component of $\mathbf{\Gamma}$ and $\mathbf{A}$ in power series about some point, say $\mathbf{X}$, on the loop:

$$\begin{aligned} \Gamma^\mu_{\alpha\beta}(\mathbf{x}) &= \Gamma^\mu_{\alpha\beta}(\mathbf{X}) + (x^\nu - X^\nu)\frac{\partial \Gamma^\mu_{\alpha\beta}}{\partial x^\nu} + \cdots \\ A^\mu(\mathbf{x}) &= A^\mu(\mathbf{X}) + (x^\nu - X^\nu)\frac{\partial A^\mu}{\partial x^\nu} + \cdots \end{aligned} \tag{5.3}$$

Multiplying these two expansions together and substituting the result into (5.2), we get

$$\Delta A^\mu = -\oint \Big\{ \big[\Gamma^\mu_{\alpha\beta} A^\beta\big]_{\mathbf{X}} + \Big[\Gamma^\mu_{\alpha\beta}\frac{\partial A^\beta}{\partial x^\nu} + A^\beta \frac{\partial \Gamma^\mu_{\alpha\beta}}{\partial x^\nu}\Big]_{\mathbf{X}} (x^\nu - X^\nu) + \cdots \Big\} \frac{\mathrm{d}x^\alpha}{\mathrm{d}s}\,\mathrm{d}s.$$

Since the first square bracket is constant, it can be taken outside the integral sign. Integrating its coefficient $\mathrm{d}x^\alpha/\mathrm{d}s$ around our closed contour we then obtain zero. The second square bracket may also be taken outside the integral sign. Then eliminating $\partial A^\beta/\partial x^\nu$ with (5.1) we obtain

$$\begin{aligned} \Delta A^\mu &= \Big[\Gamma^\mu_{\alpha\beta}\Gamma^\beta_{\nu\gamma} A^\gamma - \frac{\partial \Gamma^\mu_{\alpha\beta}}{\partial x^\nu} A^\beta\Big]_{\mathbf{X}} \oint (x^\nu - X^\nu)\mathrm{d}x^\alpha + \cdots \\ &= \Big[\Gamma^\mu_{\alpha\beta}\Gamma^\beta_{\nu\gamma} - \frac{\partial \Gamma^\mu_{\alpha\gamma}}{\partial x^\nu}\Big]_{\mathbf{X}} A^\gamma \oint x^\nu\,\mathrm{d}x^\alpha + \cdots \end{aligned} \tag{5.4}$$

The integrals in (5.4) for which $\nu = \alpha$ vanish because each such integral is simply the change in $\frac{1}{2}(x^\alpha)^2$ on going around the loop. Furthermore, when $\alpha \neq \nu$, the integral $\oint x^\nu \mathrm{d}x^\alpha$ is equal in magnitude and opposite in sign to the integral $\oint x^\alpha \mathrm{d}x^\nu$ as this picture of the (x^α, x^ν) plane shows:

We define the directed area enclosed by the loop to be the antisymmetric tensor

$$\Delta S^{\nu\alpha} \equiv \oint x^\nu\,\mathrm{d}x^\alpha. \tag{5.5}$$

This done we may write

$$\Delta A^\mu = \left[\Gamma^\mu_{\alpha\beta}\Gamma^\beta_{\nu\gamma} - \frac{\partial \Gamma^\mu_{\alpha\gamma}}{\partial x^\nu}\right]_{\mathbf{X}} A^\gamma \Delta S^{\nu\alpha} + \ldots \tag{5.6}$$

In the absence of a gravitational field, $\Delta A^\mu = 0$ for any A^μ. Furthermore, by an appropriate choice of loop $\Delta S^{\nu\alpha}$ can be set equal to any given antisymmetric tensor.[9] So it is tempting to conclude that the square bracket in the last equation vanishes. However, when we contract an antisymmetric tensor with a tensor of mixed symmetry, only the antisymmetric portion of the mixed tensor contributes to the sums. Hence from the vanishing of ΔA^μ for arbitrary A^μ and $\Delta S^{\nu\alpha}$ we can infer only the vanishing of the portion of the square bracket that is antisymmetric on exchange of ν and α. We therefore define the **curvature tensor** as minus twice this part of the square bracket in (5.6)

$$R^\mu_{\gamma\alpha\nu} \equiv \frac{\partial \Gamma^\mu_{\alpha\gamma}}{\partial x^\nu} - \frac{\partial \Gamma^\mu_{\nu\gamma}}{\partial x^\alpha} + \Gamma^\mu_{\nu\beta}\Gamma^\beta_{\alpha\gamma} - \Gamma^\mu_{\alpha\beta}\Gamma^\beta_{\nu\gamma}, \tag{5.7}$$

and rewrite (5.6) as

$$\Delta A^\mu = -\tfrac{1}{2} R^\mu_{\gamma\alpha\nu} A^\gamma \Delta S^{\nu\alpha} + \cdots \tag{5.8}$$

Since ΔA^μ is the difference between two vectors that are based at the same point, it is itself a vector. Furthermore, both A^γ and $\Delta S_{\nu\alpha}$ are tensors. Hence $R^\mu_{\gamma\alpha\nu}$ must also be a tensor as its name implies. In the absence of a gravitational field we have

$$R^\mu_{\gamma\alpha\nu} = 0. \tag{5.9}$$

This is the relativistic generalization of Laplace's equation $\nabla^2\Phi = 0$. As promised, it is first-order in $\mathbf{\Gamma}$ and second-order in $\mathbf{g}$. Notice that it is non-linear in both these quantities; this is highly significant (and very inconvenient!).

5.2 Derivation of the Field Equations

If we are to upgrade (5.9) into the relativistic genelarization of Poisson's equation (3.5), we must replace the zero on the right with something that involves the density of mass-energy. We have seen [equations (2.22) and (2.17)] that the mass-energy density forms one component of a symmetric second-rank tensor $\mathbf{T}$. If we want a covariant theory of gravity we are going to have to allow the mass-energy density to bring along all its friends in $\mathbf{T}$ into the field equations. So consider replacing the zero in (5.9) with

$$\text{constant} \times T_{\alpha\beta}.$$

This has only two indices, whereas the left of (5.9) has four indices. Hence we must either use $\mathbf{g}$ (which is the only generally available tensor) to add two more indices on the right, or we must contract away two indices on the left. It is not hard to see that these two courses are equivalent. We do it the second way.

Which two indices should we contract? Well, from the defining expression (5.7) one may show that $R_{\mu\nu\alpha\beta}$ has the following symmetries:

$$R_{\mu\nu\alpha\beta} = R_{\alpha\beta\mu\nu} \quad ; \quad R_{\mu\nu\beta\alpha} = -R_{\mu\nu\alpha\beta} = R_{\nu\mu\alpha\beta}. \tag{5.10}$$

In words; $\mathbf{R}$ is symmetric on interchange of the first pair of indices with the second pair, and antisymmetric under interchange of the indices within each of these pairs. Thus we get zero if we contract within any pair, and the same answer (to within a sign) if we contract between pairs. It is conventional to define the **Ricci tensor** by

$$R_{\alpha\beta} \equiv R^\mu_{\alpha\mu\beta}. \tag{5.11}$$

Exercise (14):

Show that $R^\mu_{\nu\alpha\beta}$ has 18 independent indices.

[9] This is a lie, as the discussion of 6-tuples in §2.5 shows. However, the argument can be fixed up by adding the changes $\Delta\mathbf{A}$ around two non-coplanar paths.

Note:

In terms of $\mathbf{\Gamma}$, $R_{\alpha\beta}$ is by (5.7)

$$R_{\alpha\beta} = \frac{\partial \Gamma^{\mu}_{\mu\alpha}}{\partial x^{\beta}} - \frac{\partial \Gamma^{\mu}_{\alpha\beta}}{\partial x^{\mu}} + \Gamma^{\lambda}_{\alpha\mu}\Gamma^{\mu}_{\beta\lambda} - \Gamma^{\mu}_{\lambda\mu}\Gamma^{\lambda}_{\alpha\beta}. \tag{5.12a}$$

Furthermore, by (4.23)

$$\Gamma^{\mu}_{\alpha\mu} = \Gamma^{\mu}_{\mu\alpha} = \tfrac{1}{2} g^{\mu\nu} \frac{\partial g_{\mu\nu}}{\partial x^{\alpha}}. \tag{5.12b}$$

While $R_{\alpha\beta}$ has the right number of indices to go on the left of our field equations, the law we seek is not $R_{\alpha\beta} = T_{\alpha\beta}$ because mass-energy conservation is expressed by the vanishing of the covariant divergence of $\mathbf{T}$ [generalization to non-inertial coordinates of eq. (2.53)]. Hence whatever goes on the left of our field equations must have zero divergence. Unfortunately, the divergence of $R_{\alpha\beta}$ is not always zero. However, it turns out that (see Appendix B)

$$R_{\alpha}{}^{\beta}{}_{;\beta} = \tfrac{1}{2} R_{;\alpha}, \tag{5.13}$$

where the **Ricci scalar** R is defined by

$$R \equiv R_{\beta}{}^{\beta}. \tag{5.14}$$

From (5.13) it follows that a tensor made out of $R^{\mu}_{\nu\alpha\beta}$ which *has* zero divergence is

$$G^{\alpha\beta} \equiv (R^{\alpha\beta} - \tfrac{1}{2} g^{\alpha\beta} R). \tag{5.15}$$

$\mathbf{G}$ is called the **Einstein tensor** because the p.d.e.'s which describe the generation of a gravitational field by matter are

$$G^{\alpha\beta} = -\frac{8\pi G}{c^4} T^{\alpha\beta}. \tag{5.16}$$

Here G is Newton's gravitational constant, as we shall shortly show. An alternative, and often handier, form of (5.16) is obtained by contracting both sides of the equation to obtain

$$G_{\alpha}{}^{\alpha} = (R_{\alpha}{}^{\alpha} - \tfrac{1}{2}\delta^{\alpha}_{\alpha} R) = -R = -\frac{8\pi G}{c^4} T_{\alpha}{}^{\alpha}.$$

Substituting this value of R into (5.16) we get

$$R^{\alpha\beta} = -\frac{8\pi G}{c^4}\big(T^{\alpha\beta} - \tfrac{1}{2} g^{\alpha\beta} T_{\gamma}{}^{\gamma}\big). \tag{5.17}$$

Equations (5.16) and (5.17) are the relativistic equivalents of Poisson's equation $\nabla^2\Phi = 4\pi G\rho$. As expected, these equations are second-order in the ten potentials $g^{\mu\nu}$ and involve all the energy-density's friends in $\mathbf{T}$.

There is a close analogy between (5.17) and its e.m. counterpart $F^{\mu\nu}{}_{;\nu} = \mu_0 j^{\mu}$ as may be seen by substituting for $\mathbf{R}$ from (5.12)

$$\frac{\partial \Gamma^{\mu}_{\mu\alpha}}{\partial x^{\beta}} - \frac{\partial \Gamma^{\mu}_{\alpha\beta}}{\partial x^{\mu}} + \Gamma^{\lambda}_{\alpha\mu}\Gamma^{\mu}_{\beta\lambda} - \Gamma^{\mu}_{\lambda\mu}\Gamma^{\lambda}_{\alpha\beta} = -\frac{8\pi G}{c^4}\big(T_{\alpha\beta} - \tfrac{1}{2} g_{\alpha\beta} T_{\gamma}{}^{\gamma}\big). \tag{5.18}$$

Worse still, the relationship (4.23) between $\mathbf{\Gamma}$ and the tensor potential $\mathbf{g}$ is a good deal more complex than the corresponding e.m. relation $F_{\mu\nu} = A_{\nu},_{\mu} - A_{\mu},_{\nu}$. So it is hardly surprising that not many exact solutions of the Einstein equations are known! But we shall be able to deduce some extremely interesting solutions nevertheless.

5.3 The Newtonian Limit

If the gravitational field is very weak, we can find a nearly inertial coordinate system. In this system

$$g_{\alpha\beta} = \eta_{\alpha\beta} + h_{\alpha\beta} \quad \text{where} \quad |h_{\alpha\beta}| \ll 1. \tag{5.19}$$

We neglect squares and higher powers of $\mathbf{h}$. By (4.23) we then have

$$\Gamma^\mu_{\alpha\beta} = \tfrac{1}{2}\eta^{\mu\nu}\Big(\frac{\partial h_{\nu\alpha}}{\partial x^\beta} + \frac{\partial h_{\nu\beta}}{\partial x^\alpha} - \frac{\partial h_{\alpha\beta}}{\partial x^\nu}\Big). \tag{5.20}$$

Consider the equation of motion (4.17) to which this gives rise for a non-relativistic free particle ($a'^\mu = 0$). The motion is governed by a gravitational force

$$f^\mu = -\Gamma^\mu_{\alpha\beta} v^\alpha v^\beta, \tag{5.21}$$

where $\mathbf{v}$ is the particle's 4-velocity. Since the zeroth component $v^0 = \gamma c$ of the 4-velocity of a non-relativistic particle is very much larger than any of $\mathbf{v}$'s spatial components, we expect the dominant term in the implied sum of (5.21) to be that for which $\alpha = \beta = 0$. Thus we expect

$$f^\mu \simeq -\gamma^2 c^2 \Gamma^\mu_{00}. \tag{5.22}$$

A typical spatial component of the equations of motion is then

$$\frac{\mathrm{d}v^j}{\mathrm{d}t} = -\gamma^2 c^2 \Gamma^j_{00} \simeq -c^2 \tfrac{1}{2}\Big(2\frac{\partial h_{j0}}{\partial x^0} - \frac{\partial h_{00}}{\partial x^j}\Big).$$

If the field is stationary in our chosen coordinate system (and we are free to boost until it is), then $\partial h_{j0}/\partial x^0 = 0$ and to leading order in v/c

$$\frac{\mathrm{d}v^j}{\mathrm{d}t} = \frac{\partial}{\partial x^j}\big(\tfrac{1}{2}c^2 h_{00}\big). \tag{5.23}$$

If this is to agree with Newton's theory, we require

$$\Phi = -\tfrac{1}{2}c^2 h_{00}, \tag{5.24}$$

where Φ is the Newtonian gravitational potential.

We now check whether Einstein's field equations (5.17) reduce in the same weak-field limit to Poisson's equation for Φ. We expect the source of Φ to be the energy density $\rho = T^{00}/c^2$, where $\mathbf{T}$ is the energy-momentum tensor, so we concentrate on the 00-component of (5.17).

From (5.12a,b), (5.19) and (5.20), $R_{\alpha\beta}$ is to first order in $\mathbf{h}$

$$R_{\alpha\beta} = \tfrac{1}{2}\eta^{\mu\nu}\Big(\frac{\partial^2 h_{\mu\nu}}{\partial x^\alpha \partial x^\beta} - \frac{\partial^2 h_{\alpha\nu}}{\partial x^\mu \partial x^\beta} - \frac{\partial^2 h_{\nu\beta}}{\partial x^\mu \partial x^\alpha} + \frac{\partial^2 h_{\alpha\beta}}{\partial x^\nu \partial x^\mu}\Big). \tag{5.25}$$

In particular, for a time-independent field

$$R^{00} = R_{00} = \tfrac{1}{2}\nabla^2 h_{00} = -\frac{1}{c^2}\nabla^2\Phi.$$

If the only contributor to the energy-momentum tensor is a dust of stationary particles, $\mathbf{T}$ is given by (2.17) with $\mathbf{v} = (c, 0, 0, 0)$. Hence $T^\gamma_\gamma = -n_0 m_0 c^2$, where m_0 is the rest mass of each particle and n_0 is the number-density of particles. Thus the 00-component of (5.17) is

$$R^{00} = -\frac{1}{c^2}\nabla^2\Phi = -\frac{4\pi G}{c^2} n_0 m_0$$

as expected.

5.4 Gravitational Redshift

The analysis at the end of the last section shows that in a weak gravitational field $g_{00} \simeq \eta_{00} - 2\Phi/c^2$ is closely related to the Newtonian gravitational potential. This conclusion has interesting physical consequences. Consider an observer at rest in a weak gravitational field. We choose spatial coordinates so that the field and the observer are stationary. Let the observer be at potential Φ_o and observe a stationary atom at potential Φ_a. Setting $\lambda = x^0$ in (4.24) and differentiating both sides of this equation we find that the observer's proper time elapses at a rate

$$\begin{aligned} \frac{\mathrm{d}\tau_o}{\mathrm{d}t} &= \sqrt{-g_{\mu\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}x^0}\frac{\mathrm{d}x^\nu}{\mathrm{d}x^0}} \\ &= \sqrt{-g_{00}} \simeq \sqrt{1+2\frac{\Phi_o}{c^2}} \\ &\simeq 1 - \frac{|\Phi_o|}{c^2} \qquad \text{(because } \Phi_o < 0). \end{aligned} \tag{5.26}$$

Similarly, the atom's proper time elapses at a rate

$$\frac{\mathrm{d}\tau_a}{\mathrm{d}t} = 1 - \frac{|\Phi_a|}{c^2}. \tag{5.27}$$

If the atom is emitting e.m. radiation of frequency ν, then during an interval $\Delta\tau_o$ on the observer's clock it will emit $(\nu\Delta\tau_o) \times (\mathrm{d}\tau_a/\mathrm{d}\tau_o)$ wave fronts. Of course, these wavefronts will take some time (as measured by either clock) to reach the observer, but because our situation is static the delay before each front reaches the observer is always the same. Hence the fronts will be received in time $\Delta\tau_o$ on the observer's clock and the observer measures frequency

$$\left(\frac{\mathrm{d}\tau_a}{\mathrm{d}\tau_o}\right)\nu = \frac{1-|\Phi_a|/c^2}{1-|\Phi_o|/c^2}\nu \simeq \left(1 - \frac{|\Phi_a - \Phi_o|}{c^2}\right)\nu. \tag{5.28}$$

In words: radiation that comes up out of a gravitational well is redshifted.

Exercise (15):

Consider a machine which lowers boxes full of excited atoms on ropes down a well, deexcites the atoms at the bottom, pulls the atoms back up and then reexcites the atoms with the photons released at the bottom and beamed up to the top. Show that this machine will violate energy conservation unless the photons' frequencies at top and bottom of the well satisfy (5.28).

5.5 Summary

The curvature tensor $R^\mu_{\nu\alpha\beta}$ tells us by how much a vector changes on being parallel transported around a small circuit. Hence we detect the use of a crazy coordinate system for flat space-time by seeing if the curvature tensor $\mathbf{R} = 0$. If $\mathbf{R} \neq 0$ there is a true gravitational field.

The presence of matter at $\mathbf{x}$ is signalled by $R_{\alpha\beta}(\mathbf{x}) \equiv R^\mu_{\alpha\mu\beta}(\mathbf{x}) \neq 0$.

Formally, there is a far-reaching analogy between g.r. and e.m.:

Parallelism of e.m. and g.r.

A_μ	$\leftrightarrow$	$g_{\mu\nu}$
$F_{\mu\nu} = -(A_{\mu,\nu} - A_{\nu,\mu})$	$\leftrightarrow$	$\Gamma_{\mu,\alpha\beta} = \frac{1}{2}(g_{\mu\alpha,\beta} + g_{\mu\beta,\alpha} - g_{\alpha\beta,\mu})$
$f^\mu = \dfrac{q}{m_0} F^\mu{}_\alpha v^\alpha$	$\leftrightarrow$	$f^\mu = -\Gamma^\mu_{\alpha\beta} v^\alpha v^\beta$
$F^{\mu\nu}{}_{,\nu} = \mu_0 j^\mu$	$\leftrightarrow$	eq. (5.18)
$F^{\mu\nu}{}_{,\rho} + F^{\nu\rho}{}_{,\mu} + F^{\rho\mu}{}_{,\nu} = 0$	$\leftrightarrow$	$R^\kappa_{\lambda\mu\nu;\rho} + R^\kappa_{\lambda\nu\rho;\mu} + R^\kappa_{\lambda\rho\mu;\nu} = 0$ (Bianchi identity)

The parallel between Newton's theory and g.r. is less tight: $\Phi \leftrightarrow \mathbf{g}$, $\boldsymbol{f} \leftrightarrow \boldsymbol{\Gamma}$, $\nabla^2\Phi \leftrightarrow R_{\alpha\beta}$.

In a weak gravitational field we can have $\mathbf{g} \simeq \boldsymbol{\eta}$ with $-2\Phi/c^2$ as an estimate of $(g_{00} - \eta_{00})$. The gravitational redshift follows immediately from this estimate.

6 The Schwarzschild Solution

Now that we have the field equations (5.17) it is natural to seek the solution $\mathbf{g}$ that describes the gravitational field in the solar system. A useful step in this direction would be to find the metric associated with a point mass in an otherwise empty universe.

The way we derive most solutions to Einstein's equations is at root the same as that by which we are accustomed to solve other partial differential equations, for example Maxwell's equations. If we want to find the electrostatic potential inside a charged spherical surface, we start by looking for potentials of the special form $\Phi(r, \theta, \phi) = R(r)\Theta(\theta)e^{im\phi}$. We are not initially certain that such solutions exist, but we try the idea out anyway in the knowledge that if there are no such solutions we shall derive inconsistent conditions on R and Θ and thus discover our mistake, but if no inconsistencies arise, we shall get a valid solution and it will not matter that we found it by leaping into the dark.

Proceeding in this spirit towards the metric outside a point mass, we first argue that we should be able to find coordinates in which the metric is diagonal. To see why this is so, suppose we are given a metric tensor $\mathbf{g}$ for some two-dimensional space. Then from simple matrix algebra we know that at any point in the space we can find two mutually perpendicular directions, the eigenvectors $\mathbf{u}$ and $\mathbf{v}$ of $\mathbf{g}$, such that $\mathbf{g}$ would be a diagonal matrix if our coordinate directions coincided with $\mathbf{u}$ and $\mathbf{v}$. Now imagine marking the directions $\mathbf{u}$, $\mathbf{v}$ as small crosses on a grid of points in the space. Since $\mathbf{g}$ is a smoothly varying function of position, the orientation of neighbouring crosses will be similar. Hence we may draw smooth curves through neighbouring crosses, thus covering the space with a curvilinear grid. Finally, if we are able to label each curve of this doubly infinite family of curves with numbers (a, b), these numbers will constitute a valid coordinate system for the space and $\mathbf{g}$ will be diagonal in this coordinate system.

If we start from the metric tensor of a 4-space, the situation is fundamentally the same as in our two-dimensional example; the only difference is that there are now four special directions at each point. So it is reasonable to conjecture that we can find coordinates in which the metric of any simple spacetime is everywhere diagonal.

Furthermore, since the gravitational field we seek to describe is time-independent, we should be able to choose coordinates in such a way that none of the metric coefficients depends on time. Also the gravitational field will be spherically symmetric, so there must be closed

2-surfaces on which the geometry is that of a sphere. If we label these surfaces with the coordinates (r, t) and indicate position on each surface with the angle variables (θ, ϕ), we have

$$\begin{aligned} \mathrm{d}s^2 &\equiv g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu \\ &= -D(r)c^2\mathrm{d}t^2 + A(r)(\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2) + B(r)\mathrm{d}r^2. \end{aligned} \tag{6.1}$$

We next fix the meaning of r by determining that the sphere with labels (r, t) should have area $4\pi r^2$. This yields

$$\mathrm{d}s^2 = -D(r)c^2\mathrm{d}t^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2) + B(r)\mathrm{d}r^2. \tag{6.2}$$

The metric now takes the form

$$g_{\mu\nu} = \begin{matrix} t \\ r \\ \theta \\ \phi \end{matrix} \begin{pmatrix} -c^2 D & 0 & 0 & 0 \\ 0 & B & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2\sin^2\theta \end{pmatrix}. \tag{6.3}$$

$g^{\mu\nu}$ is simply

$$g^{\mu\nu} = \begin{matrix} t \\ r \\ \theta \\ \phi \end{matrix} \begin{pmatrix} -(c^2 D)^{-1} & 0 & 0 & 0 \\ 0 & B^{-1} & 0 & 0 \\ 0 & 0 & r^{-2} & 0 \\ 0 & 0 & 0 & r^{-2}\sin^{-2}\theta \end{pmatrix}. \tag{6.4}$$

Exercise (16):

By making an appropriate coordinate transformation $\mathbf{x}'(\mathbf{x})$ show that when, as here, one uses t rather than ct for the 0^{th} coordinate, the 4-vector of a photon becomes $k^\mu = (\omega/c^2, \boldsymbol{k})$.

Since the metric is diagonal with coefficients which depend only on r and θ, we have from (4.22) that the only non-zero Christoffel symbols are those with at least two indices the same and at least one equal to either r or θ. Specifically,

$$\begin{aligned} &\Gamma_{r,rr} = \tfrac{1}{2}B' \qquad \Gamma_{r,\theta\theta} = -r \qquad \Gamma_{r,\phi\phi} = -r\sin^2\theta \qquad \Gamma_{r,tt} = \tfrac{1}{2}c^2 D' \\ &\Gamma_{\theta,\phi\phi} = -r^2\sin\theta\cos\theta \\ &\Gamma_{\theta,\theta r} = \Gamma_{\theta,r\theta} = r \qquad \Gamma_{t,tr} = \Gamma_{t,rt} = -\tfrac{1}{2}c^2 D' \\ &\Gamma_{\phi,\phi r} = \Gamma_{\phi,r\phi} = r\sin^2\theta \qquad \Gamma_{\phi,\phi\theta} = \Gamma_{\phi,\theta\phi} = r^2\sin\theta\cos\theta. \end{aligned}$$

We raise the first index of each symbol by multiplying by the appropriate element of $g^{\mu\nu}$ to yield

$$\begin{aligned} &\Gamma^r_{rr} = \frac{B'}{2B} \qquad \Gamma^r_{\theta\theta} = -\frac{r}{B} \qquad \Gamma^r_{\phi\phi} = -\frac{r\sin^2\theta}{B} \qquad \Gamma^r_{tt} = \frac{c^2 D'}{2B} \\ &\Gamma^\theta_{\phi\phi} = -\sin\theta\cos\theta \qquad \Gamma^\theta_{\theta r} = \Gamma^\theta_{r\theta} = \frac{1}{r} \\ &\Gamma^\phi_{\phi r} = \Gamma^\phi_{r\phi} = \frac{1}{r} \qquad \Gamma^\phi_{\phi\theta} = \Gamma^\phi_{\theta\phi} = \cot\theta \\ &\Gamma^t_{tr} = \Gamma^t_{rt} = \frac{D'}{2D} \end{aligned} \tag{6.5}$$

Hence

$$\Gamma^\mu_{r\mu} = \frac{B'}{2B} + \frac{2}{r} + \frac{D'}{2D}, \qquad \Gamma^\mu_{\theta\mu} = \cot\theta, \qquad \Gamma^\mu_{\phi\mu} = 0, \qquad \Gamma^\mu_{t\mu} = 0. \tag{6.6}$$

By hard slog and (5.12) one can now obtain

$$R_{tt} = -\frac{c^2 D''}{2B} + \frac{c^2 D'}{4B}\Big(\frac{B'}{B} + \frac{D'}{D}\Big) - \frac{c^2 D'}{rB} \tag{6.7a}$$

$$R_{rr} = \frac{D''}{2D} - \frac{D'}{4D}\Big(\frac{B'}{B} + \frac{D'}{D}\Big) - \frac{B'}{rB} \tag{6.7b}$$

$$R_{\theta\theta} = -1 + \frac{r}{2B}\Big(-\frac{B'}{B} + \frac{D'}{D}\Big) + \frac{1}{B} \tag{6.7c}$$

$$R_{\phi\phi} = \sin^2\theta R_{\theta\theta} \tag{6.7d}$$

$$R_{\mu\nu} = 0 \qquad \mu \neq \nu. \tag{6.7e}$$

We require $R_{\mu\nu} = 0$ everywhere except at $r = 0$ (where these expressions fail anyway). Multiplying (6.7a) by B/c^2D and adding the result to (6.7b) yields

$$\frac{B'}{B} = -\frac{D'}{D} \qquad \Rightarrow \qquad BD = \text{constant}. \tag{6.8}$$

As $r \to \infty$ the metric should become that of flat spacetime for which $B = D = 1$ Thus

$$B(r) = \frac{1}{D(r)} \quad \forall \quad r > 0. \tag{6.9}$$

By (6.7c) the equation $R_{\theta\theta} = 0$ now becomes

$$0 = R_{\theta\theta} = -1 + rD' + D \qquad \Rightarrow \qquad D = 1 + \text{constant}/r. \tag{6.10}$$

By (5.24) we know that as $r \to \infty$ and the field becomes weak, $D \to 1 + 2\Phi/c^2 = 1 - r_s/r$, where M is the mass at the centre and the **Schwarzschild radius** r_s is defined by

$$r_s \equiv \frac{2GM}{c^2}. \tag{6.11}$$

Hence we may identify the constant in (6.11) as $-r_s$, giving

$$D = 1 - \frac{r_s}{r}. \tag{6.12}$$

Collecting everything together we have the **Schwarzschild metric**

$$g_{\mu\nu} = \begin{matrix} t \\ r \\ \theta \\ \phi \end{matrix} \begin{pmatrix} -c^2(1 - r_s/r) & & & \\ & (1 - r_s/r)^{-1} & & \\ & & r^2 & \\ & & & r^2 \sin^2\theta \end{pmatrix}. \tag{6.13}$$

The metric (6.13) deviates markedly from the metric associated with spherical polar coordinates (which has $g_{tt} = -c^2$ and $g_{rr} = 1$) for values of r up to a few times larger than r_s. If M has the same mass as the Sun, $M_\odot = 1.99 \times 10^{30}$ kg, we find $r_s = 2.95$ km.

6.1 Equations of Motion

Now that we know what B and D are we can immediately write down the equations of motion (4.17) of a particle moving in the field of a point mass. It is straightforward to check that a possible solution to the θ-equation of motion is $\theta = \frac{\pi}{2}$; that is, that the particle may move always in the equatorial plane of the coordinate system. We shall assume that our coordinate system has been oriented to ensure $\theta = \frac{\pi}{2}$. We now set $a'^{\mu} = 0$ and $v'^{\mu} = \mathrm{d}x^{\mu}/\mathrm{d}\tau$ in (4.17) to find with (6.5) and (6.9) for the t equation of motion

$$0 = \frac{\mathrm{d}^2 t}{\mathrm{d}\tau^2} + \frac{D'}{D}\frac{\mathrm{d}r}{\mathrm{d}\tau}\frac{\mathrm{d}t}{\mathrm{d}\tau} = \frac{\mathrm{d}^2 t}{\mathrm{d}\tau^2} + \frac{\mathrm{d}\ln D}{\mathrm{d}\tau}\frac{\mathrm{d}t}{\mathrm{d}\tau}.$$

This is a first-order linear differential equation for $y \equiv \mathrm{d}t/\mathrm{d}\tau$. The integrating factor is D, so $\mathrm{d}\tau/\mathrm{d}t = \text{constant} \times D$. The equation simply says that a freely-falling particle suffers a gravitational time dilation in that its proper time elapses at a fraction $D = 1 - r_s/r$ of the rate of elapse of the proper time t of an observer at infinity. We evaluate the constant by observing that for a stationary particle at infinity $\tau = t$, so at any radius r

$$\frac{\mathrm{d}\tau}{\mathrm{d}t} = D. \tag{6.14}$$

Similarly, (4.17) and (6.5) yield the ϕ equation of motion as

$$0 = \frac{\mathrm{d}^2\phi}{\mathrm{d}\tau^2} + \frac{2}{r}\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\frac{\mathrm{d}r}{\mathrm{d}\tau} \qquad \Rightarrow \qquad \frac{\mathrm{d}}{\mathrm{d}\tau}\left(r^2\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\right) = 0.$$

Hence the angular momentum

$$L \equiv r^2\frac{\mathrm{d}\phi}{\mathrm{d}\tau} \tag{6.15}$$

is conserved.

With $\theta = \frac{\pi}{2}$ the r-equation of motion is

$$0 = \frac{\mathrm{d}^2 r}{\mathrm{d}\tau^2} + \frac{1}{2}\frac{c^2 D'}{B}\left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^2 + \frac{1}{2}\frac{B'}{B}\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2 - \frac{r}{B}\left(\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\right)^2.$$

With (6.9), (6.14) and (6.15) this becomes

$$0 = \frac{\mathrm{d}^2 r}{\mathrm{d}\tau^2} - \frac{DL^2}{r^3} + \frac{c^2}{2}\frac{D'}{D} - \frac{1}{2}\frac{D'}{D}\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2. \tag{6.16}$$

We shall see that in Newton's theory slightly modified forms of the first, second and third terms occur. The third last represents a new, speed dependent force.

Exercise (17):

From (6.15) show that the angular frequency of a circular orbit as seen by an observer at infinity is

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = \sqrt{\frac{GM}{r^3}}$$

exactly as in Newton's theory.

If we multiply (6.16) by $\dfrac{1}{D}\dfrac{\mathrm{d}r}{\mathrm{d}\tau}$, we may integrate it to an energy equation

$$\frac{1}{2D}\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2 - \frac{c^2}{2D} + \frac{L^2}{2r^2} = \text{constant} \equiv E - \tfrac{1}{2}c^2. \tag{6.17}$$

The first and third terms on the left of equation (6.17) represent kinetic energy of radial and tangential motion respectively. To first order in r_s/r the second term on the left is $\frac{1}{2}c^2$ plus the Newtonian potential energy $-GM/r$.

6.2 The Perihelion of Mercury

When Einstein introduced g.r. in 1916, the only significant discrepancy between Newtonian dynamics and solar system observations was the rate of advance of the perihelion of Mercury. One of g.r.'s early triumphs was to account for this discrepancy. We start by reviewing Newton's results for motion in the gravitational field of a point mass.

Newtonian motion around a point mass The equation of motion of a particle in the Newtonian field of a mass M located at the origin is $\ddot{\boldsymbol{r}} = -GM\boldsymbol{r}/r^3 = -\frac{1}{2}c^2 r_s \boldsymbol{r}/r^3$. On crossing this equation through by $\boldsymbol{r}$ we obtain $\dot{\boldsymbol{L}} = 0$ where $\boldsymbol{L}$ is the angular momentum vector $\boldsymbol{L} \equiv \boldsymbol{r}\times\dot{\boldsymbol{r}}$. From the constancy of $\boldsymbol{L}$ we deduce that the motion is confined to the plane $\boldsymbol{L}\cdot\boldsymbol{r} = 0$ perpendicular to the angular momentum vector $\boldsymbol{L}$. Let r and ϕ be polar coordinates for this plane. Conservation of angular momentum requires $r^2\dot{\phi} = L$, while the equation of motion of r is $\ddot{r} - r\dot{\phi}^2 = -\frac{1}{2}c^2 r_s/r^2$. Eliminating $\dot{\phi}$ in favour of L the latter reads

$$0 = \frac{\mathrm{d}^2 r}{\mathrm{d}t^2} + \frac{c^2 r_s}{2r^2} - \frac{L^2}{r^3}. \tag{6.18}$$

This is the Newtonian analogue of (6.16): to see this recall that $D = 1 - r_s/r$ and $D'/D \simeq r_s/r^2$.

We obtain the shape of Newtonian orbits by eliminating t from (6.18) through the substitution $\mathrm{d}t = (r^2/L)\mathrm{d}\phi$, and eliminating r in favour of a new variable $u \equiv 1/r$. We then find

$$\frac{\mathrm{d}^2 u}{\mathrm{d}\phi^2} + u = \frac{c^2 r_s}{2L^2}. \tag{6.19}$$

This is just the equation of motion of a simple harmonic oscillator. So the orbit is given by

$$r(\phi) = \frac{1}{u} = \frac{1}{A\cos(\phi - \phi_0) + \frac{1}{2}c^2 r_s/L^2}, \tag{6.20}$$

where A and ϕ_0 are suitable constants of integration. This is actually the equation of an ellipse with one focus at the origin. But the most important point is that since the right side of (6.20) is periodic in ϕ with period 2π, $r(\phi + 2\pi) = r(\phi)$ for any ϕ and thus (6.20) defines a *closed* curve. Consequently, a planet in undisturbed orbit around the Sun would always come closest to the Sun (in the jargon, "move through perihelion") at the same value of ϕ. Actually the perihelia of all the planets precess, that is, they move very slowly around the plane of the planet's orbit.

The planet with the most rapidly precessing perihelion is Mercury because it is the planet with the shortest year. Its perihelion precesses by 576 seconds of arc ($576''$) per century. Most of this precession is caused by the gravitational field of Jupiter.[10] In the late 19$^{\rm th}$ century Bessel showed that disturbance of Mercury's orbit by all the planets gives rise to a net precession of $532''$ per century. Thus Bessel was able to account for all but $44''$ per century of Mercury's precession. Since Mercury's year is 0.24 siderial years long, $44''$ per century corresponds to $0.106''$ per Mercury year.

Relativistic precession Working from (6.16) in close analogy with the our Newtonian calculation, we eliminate τ between (6.15) and (6.16) to obtain

$$0 = \frac{L}{r^2}\frac{\mathrm{d}}{\mathrm{d}\phi}\left(\frac{L}{r^2}\frac{\mathrm{d}r}{\mathrm{d}\phi}\right) + \tfrac{1}{2}c^2\frac{D'}{D} - \tfrac{1}{2}\frac{D'}{D}\frac{L^2}{r^4}\left(\frac{\mathrm{d}r}{\mathrm{d}\phi}\right)^2 - \frac{DL^2}{r^3}.$$

[10] One may understand how Jupiter causes Mercury's perihelion to precess by imagining Jupiter's mass to be uniformly distributed in an annulus centred on Jupiter's orbit. This material pulls Mercury outwards. Hence Mercury's net acceleration towards the Sun falls off with r more steeply than as r^{-2}. This in turn slightly depresses the frequency at which Mercury's radius oscillates around its mean value, and these radial oscillations gradually get out of phase with the overall rotation about the Sun.

We define $u \equiv 1/r$, substitute for D and divide through by $-L^2u^2$ to obtain

$$\frac{\mathrm{d}^2u}{\mathrm{d}\phi^2} + u(1 - r_s u) = \tfrac{1}{2}\frac{c^2 r_s}{(1 - r_s u)L^2}\Big[1 - \Big(\frac{L}{c}\frac{\mathrm{d}u}{\mathrm{d}\phi}\Big)^2\Big]. \tag{6.21}$$

The Newtonian equivalent of (6.21) is equation (6.19). Clearly the former is much harder to solve than (6.19) since on the left where we had 1 we now have $(1 - r_s u)$, and on the right the constant $\frac{1}{2}c^2 r_s/L^2$ has been replaced by a complicated function of u. But it is immediately apparent that solutions to (6.21) are unlikely to be periodic with period 2π and thus we do not expect relativistic orbits around a point mass to be closed. Let us calculated the angle between successive perihelia and compare it with Bessel's discrepancy of $0.106''$.

We first obtain the "energy equation" associated with (6.21) by multiplying through by $\dfrac{2}{(1 - r_s u)}\dfrac{\mathrm{d}u}{\mathrm{d}\phi}$ and integrating:

$$\frac{1}{(1 - r_s u)}\Big(\frac{\mathrm{d}u}{\mathrm{d}\phi}\Big)^2 + u^2 = \frac{c^2}{L^2(1 - r_s u)} + K, \tag{6.22}$$

where K is a constant. The angle $\Delta\phi$ between apo- and perihelion is therefore

$$\Delta\phi = \int_{u_1}^{u_2} \frac{\mathrm{d}u}{\sqrt{c^2/L^2 + K(1 - r_s u) - u^2(1 - r_s u)}}, \tag{6.23}$$

where u_1 and u_2 are the smallest and largest values of u along the orbit. The denominator in (6.23) involves a cubic in u. Two roots of the cubic are u_1 and u_2, so if the third root is u_3 the cubic may be written

$$H(u - u_1)(u_2 - u)(1 - u/u_3), \tag{6.24}$$

where H is a constant to be determined. Comparing coefficients of u^2 and u^3 in (6.24) and the denominator of (6.23) we find

$$u^2: \qquad -H\left(1 + \frac{u_1 + u_2}{u_3}\right) = -1 \qquad\qquad u^3: \qquad \frac{H}{u_3} = r_s,$$

so

$$u_3 = \frac{1}{r_s} - (u_1 + u_2) \simeq \frac{1}{r_s} \qquad \text{and} \qquad H = 1 - r_s(u_1 + u_2). \tag{6.25}$$

Thus $u_3 \gg \max(u_1, u_2)$ and with equations (6.24) and (6.25) we can rewrite equation (6.23) as

$$\begin{aligned}\Delta\phi &= \frac{1}{\sqrt{H}}\int_{u_1}^{u_2} \frac{\mathrm{d}u}{\sqrt{(u - u_1)(u_2 - u)}}\Big(1 + \tfrac{1}{2}\frac{u}{u_3} + \cdots\Big) \\ &\simeq [1 + \tfrac{1}{2}r_s(u_1 + u_2)]\int_{u_1}^{u_2} \frac{\mathrm{d}u}{\sqrt{(u - u_1)(u_2 - u)}}(1 + \tfrac{1}{2}ur_s) \\ &\simeq \pi[1 + \tfrac{3}{2}r_s\tfrac{1}{2}(u_2 + u_1)].\end{aligned} \tag{6.26}$$

For Mercury $\frac{1}{2}(u_1 + u_2) \simeq 1/r_{\mathrm{Merc}} = 1/(5.83 \times 10^7\,\mathrm{km})$, so the perihelion of Mercury should advance in one Mercury year by

$$3\pi\frac{r_s}{r_{\mathrm{Merc}}} \simeq 0.0983''$$

in excellent agreement with Bessel's discrepancy.

6.3 Deflection of Light by the Sun

Naive treatment A simple back-of-the-envelope argument based on the Strong Principle of Equivalence shows that light must be deflected by the Sun and allows us to obtain a quick order-of-magnitude estimate of the magnitude of this effect: the S. P. of E. implies that the path of a photon beam must be approached by a particle beam in the limit as the particles' speed $v \to c$. So let's calculate the deflection of fast (but non-relativistic) particles by the Sun.

Since the beam is fast, its deflection will be small, and we can estimate the net gravitational impulse delivered to each particle by integrating the gravitational force along a straight line. We neglect variations in the particle's speed parallel to this line, so $z \simeq vt$. Hence after a fly-by to within distance b of the Sun, the particle has a component of velocity perpendicular to the original line of magnitude

$$v_\perp \simeq \frac{1}{m}\int_{-\infty}^{\infty} F_\perp \, \mathrm{d}t = 2\int_0^\infty \frac{G\mathrm{M}_\odot}{r^2}\frac{b}{r}\frac{\mathrm{d}z}{v} = \frac{c^2 r_s(\odot)}{bv}\int_0^\infty \frac{\mathrm{d}\zeta}{(1+\zeta^2)^{3/2}},$$

where Pythagoras' useful result has been pressed into service. The substitution $\zeta = \sinh\theta$ enables one to show that the integral equals 1. So the beam is deflected through the small angle

$$\theta_{\rm defl} \simeq \frac{v_\perp}{v} \simeq \frac{r_s c^2}{v^2 b}.$$

In the limit $v \to c$, this tends to $r_s/b \simeq 0.875''$ for $b = R_\odot$. This is just a little too small to be measured with confidence through the haze of the Earth's atmosphere. Fortunately a proper calculation shows that our neglect of relativity has cost us a factor of 2, and Murphy's law notwithstanding, the true deflection is larger than our naive estimate predicts.

Relativistic treatment The definition (4.24) of the affine parameter s fails when applied to a trajectory $x^\mu(\lambda)$ of a photon. Instead we define s by requiring that

$$\frac{\mathrm{d}x^\mu}{\mathrm{d}s} = k^\mu(s), \tag{6.27}$$

where k^μ is the wavevector $(\omega/c^2, \boldsymbol{k})$ of the photon.[11] The equation of motion of the photon is $k^\mu \nabla_\mu k^\nu = 0$, so

$$\frac{\mathrm{d}x^\mu}{\mathrm{d}s}\nabla_\mu \frac{\mathrm{d}x^\nu}{\mathrm{d}s} = 0.$$

More generally, if

$$\tau \equiv \alpha s, \tag{6.28}$$

where α is any constant, we have

$$\frac{\mathrm{d}x^\mu}{\mathrm{d}\tau}\nabla_\mu \frac{\mathrm{d}x^\nu}{\mathrm{d}\tau} = 0,$$

which is identical with the equation (4.33) that governs the motion of material particles in a gravitational field. We may make the analogy with particles of finite rest mass complete and use the expressions we already have to hand to calculate the deflection of light by the Sun, by choosing the constant α such that (6.14) is satisfied. Since by (6.27) $\mathrm{d}t/\mathrm{d}s = \omega/c^2$, we must set

$$\alpha = \frac{\mathrm{d}t}{\mathrm{d}s}\frac{\mathrm{d}\tau}{\mathrm{d}t} = \frac{\omega(1-r_s u)}{c^2}. \tag{6.29}$$

[11] See Exercise (17).

Exercise (18):

Use the photon's zeroth equation of motion $\frac{\mathrm{d}x^\mu}{\mathrm{d}s}\left(\frac{\partial\omega/c^2}{\partial x^\mu}+\Gamma^t_{\mu\nu}k^\nu\right)=0$ to show that $\mathrm{d}(\omega D)/\mathrm{d}s=0$ and that α *is* a constant.

With α thus determined, we may take over the formulae of the last subsection wholesale. In particular, by (6.23) the increment in ϕ as a photon passes by the Sun is

$$\Delta\phi = 2\int_0^{u_0}\frac{\mathrm{d}u}{\sqrt{c^2/L^2+K(1-r_su)-u^2(1-r_su)}}, \tag{6.30}$$

where u_0 is the value of u at closest approach and by (6.15) and (6.27) we have

$$L=\left(r^2\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\right)_{r\to\infty}=\left(\frac{r^2}{\alpha}\frac{\mathrm{d}\phi}{\mathrm{d}r}\frac{\mathrm{d}r}{\mathrm{d}s}\right)_{r\to\infty}. \tag{6.31}$$

With ϕ and b as in the figure, at early times $r=b\csc\phi$ and $k^\mu=(\omega/c^2,k,0,0)$. Hence by (6.27)

$$\left.\frac{\mathrm{d}r}{\mathrm{d}s}\right|_{r\to\infty}=-k \qquad \text{and} \qquad \left(r^2\frac{\mathrm{d}\phi}{\mathrm{d}r}\right)_{r\to\infty}=-b. \tag{6.32}$$

Equation (6.31) with (6.29) and (6.32) now gives

$$L=\frac{bkc^2}{\omega}=bc. \tag{6.33}$$

The constant K in (6.30) is determined by evaluating (6.22) as $r\to\infty$. We have

$$\left(\frac{\mathrm{d}u}{\mathrm{d}\phi}\right)^2_{u\to0}=\frac{c^2}{L^2}+K.$$

Using on the left the fact that as $u\to0$, $u\to\sin\phi/b$, we find

$$K=\frac{1}{b^2}-\frac{c^2}{L^2}=0 \tag{6.34}$$

by (6.33). Thus (6.30) may be written

$$\Delta\phi=2\int_0^{u_0}\frac{\mathrm{d}u}{\sqrt{b^{-2}-u^2(1-r_su)}}. \tag{6.35}$$

Since u_0 is a root of the denominator, this becomes

$$\begin{aligned}\Delta\phi&=2\int_0^{u_0}\frac{\mathrm{d}u}{\sqrt{u_0^2-r_su_0^3-u^2+r_su^3}}\\&=2\int_0^{u_0}\frac{\mathrm{d}u}{\sqrt{(u_0^2-u^2)[1-r_s(u_0^3-u^3)/(u_0^2-u^2)]}}\\&=2\int_0^{u_0}\frac{\mathrm{d}u}{\sqrt{u_0^2-u^2}}\left(1+\tfrac12 r_s\frac{u_0^2+uu_0+u^2}{u_0+u}+\cdots\right)\\&\simeq2\int_0^{u_0}\frac{\mathrm{d}u}{\sqrt{u_0^2-u^2}}\left[1+\tfrac12 r_s\left(u+\frac{u_0^2}{u_0+u}\right)\right]\\&=\pi-r_s\left[\sqrt{u_0^2-u^2}+\sqrt{\frac{u_0-u}{u_0+u}}\right]_0^{u_0}\\&=\pi+2r_su_0.\end{aligned} \tag{6.36}$$

The increment $\Delta\phi$ over the Newtonian value π is $\frac{4}{6\pi}u_0r_{\mathrm{Merc}}$ times Mercury's perihelion advance per Mercury year. If $u_0\simeq1/R_\odot$, $\Delta\phi-\pi=1.75''$. In 1919 Eddington led an expedition to S. America to photograph bright stars around the Sun at a total eclipse. The expedition obtained $\Delta\phi-\pi=1.98\pm0.16''$.

6.4 Gravitational Lenses

Consider lines of sight past a star with the mass of the Sun but at such a distance that the disk of the star subtends an angle of less than an arcsecond. Then as lines of sight that graze the "top" and "bottom" of the star are bent by $\approx 1.75''$ we should be able to see objects that lie behind the star by looking either above or below the star:

The gravitational field is clearly acting as a kind of lens. It can make lensed objects much bigger and brighter than they really are.

While we have not seen lensing by a single star, there are now a handful of known cases of lensing by galaxies. The images are typically separated by a few arcseconds and have almost identical spectra.[12] Because galaxies have extended mass distributions one can often see more than two images. In fact, arcs of brightness are seen in some clusters of galaxies which are thought to be vast smeared-out images of background objects; in these cases it is as if one were observing a street light through a bubbly bathroom window.

6.5 Modern Solar-System Tests

In the last 30 years two developments have led to a big improvement in the exactitude with which g.r. can be tested in the solar system. These are (i) radar ranging to planets and (ii) ranging with radar and lasers to space probes.

Ranging to planets The earliest work involved bouncing radar signals off the inner planets. One measures the delay before the first signals return. This gives $\Delta\tau$

If one claims to know the orbits of Mercury & the Earth and $\mathbf{g}$ in the intervening space, one can calculate $\Delta\tau(t)$. This is a complex function of the parameters ("orbital elements") defining the planetary orbits. There are two important difficulties:

(i) The reflecting planetary surface is not a smooth mirror. Hence the returning pulse has a complex shape. One looks for the leading edge of the pulse and tries to use frequency information:

[12] We don't expect the spectra to be absolutely identical because the underlying object is always variable and one image typically shows the object at an earlier time than the other.

(ii) The most interesting lines of sight pass close to the Sun. Free electrons near the Sun cause the refractive index to differ from unity.

Ranging to planetary probes Since a satellite is too small to give a detectable radar reflection, one programmes the satellite to respond to a pulse from Earth by emitting a similar pulse after a known small delay. With this technique one does not have to worry about planetary topography. By sending signals at several frequencies one can eliminate the effect of dispersion by free electrons along the line of sight.

Analysis of these data has to proceed via a computer program which adjusts orbital elements, the masses of the planets and asteriods, the oblateness of the Sun, the orientation of an inertial coordinate system etc. until the fit of the predicted $\Delta\tau$'s to the observed $\Delta\tau$'s is optimized. One finds that the agreement with g.r. is excellent.

The quality of the fit is normally judged by calculating predictions from the metric[13]

$$\mathrm{d}s^2 = -\Big[1 - \alpha\frac{r_s}{\rho} + \tfrac{1}{2}\beta\Big(\frac{r_s}{\rho}\Big)^2\Big]c^2\,\mathrm{d}t^2 + \Big(1 + \gamma\frac{r_s}{\rho}\Big)[\mathrm{d}\rho^2 + \rho^2(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2)], \qquad (6.37)$$

where α, β and γ are dimensionless parameters to be determined by fitting the calculated to the observed $\Delta\tau$'s. If we identify ρ with

$$\rho \equiv \tfrac{1}{2}\big[r - \tfrac{1}{2}r_s + \sqrt{r(r - r_s)}\,\big], \qquad (6.38)$$

this metric agrees with the Schwarzschild metric (6.13) up to order r_s/r in space and $(r_s/r)^2$ in time when $\alpha = \beta = \gamma = 1$. (In the equations of motion the tt-component of $g_{\mu\nu}$ is multiplied by the largest components of v^μ.) Hence if Einstein was right, the observations should lead to $\alpha \simeq 1$ etc. Data from missions to Mercury & Mars give

$$\begin{aligned} \alpha - 1 &= (2.1 \pm 1.9) \times 10^{-4} \\ \beta - 1 &= (-2.9 \pm 3.1) \times 10^{-3} \\ \gamma - 1 &= (-0.7 \pm 1.7) \times 10^{-3} \\ J_2 &= (-1.4 \pm 1.5) \times 10^{-6} \end{aligned}$$

where J_2 is a parameter describing the oblateness of the Sun.

It is interesting that the precision of these measurements is such that

(i) they determine the interial frame of reference as accurately as can be done by looking right across the Universe at quasars with redshift $z = 2$ (see below);

(ii) they furnish the best estimates of the mass of the asteroid Ceres (the old value proved to be in error by 15%);

(iii) Dirac speculated that Newton's "constant" might decrease as the Universe expands. These measurements yield $\dot{G}/G = (0.2 \pm 0.4) \times 10^{-11}\,\mathrm{yr}^{-1}$.

13 This may be thought of as generated by expanding the functions B and D of (6.2) in powers of r_s/r.

6.6 The Schwarzschild Singularity

For $r = r_s \equiv 2GM/c^2$, the component g_{tt} of the Schwarzschild metric (6.13) vanishes. Hence the trajectory $r = r_s$ is null rather than time-like. Furthermore, since g_{tt} changes sign at $r = r_s$, the trajectory $r =$ constant $< r_s$ is space-like. Consequently an explorer who penetrates to $r < r_s$ is doomed: no matter how hard he fires his rockets, his trajectory must remain time-like. Hence he cannot pass from the condition $\mathrm{d}r/\mathrm{d}\tau < 0$ through the condition $\mathrm{d}r/\mathrm{d}\tau = 0$ as he must if he is to escape. He is carried down to $r = 0$ as surely as you and I are carried into next year.

It is interesting to investigate this predicament more closely. Suppose for simplicity that our explorer's angular momentum L is zero and that at $t = \tau = 0$ he is falling towards the centre at radius r_0 with the speed he would have picked up had he fallen all the way from rest at infinity. Then evaluating (6.17) at infinity we find that the constant E is zero. Hence by (6.17) the elapse of time on his watch as he falls to r_s is

$$\begin{aligned}\Delta\tau &= \int_{r_0}^{r_s} \frac{\mathrm{d}\tau}{\mathrm{d}r}\,\mathrm{d}r = \frac{1}{c}\int_{r_s}^{r_0} \frac{\mathrm{d}r}{\sqrt{1-D}} \\ &= \frac{1}{c\sqrt{r_s}}\int_{r_s}^{r_0} \sqrt{r}\,\mathrm{d}r = \frac{2}{3c\sqrt{r_s}}\big(r_0^{3/2} - r_s^{3/2}\big),\end{aligned} \tag{6.39}$$

which is perfectly finite. Furthermore, he clearly reaches $r = r_s$ with $\mathrm{d}r/\mathrm{d}\tau < 0$. Hence he would be well advised to fire his rockets before he reaches r_s.

Why does g_{rr} diverge at $r = r_s$? Is this divergence caused by gravity or our choice of coordinates? It is straightforward, if tedious, to check that no components of the curvature tensor $R^{\mu}{}_{\nu\alpha\beta}$ diverge at r_s. So our explorer can endure the tidal forces he experiences if he is stocky enough. The reason g_{rr} diverges at r_s turns out to be that Schwarzschild's coordinate system assigns to all events that occur at r_s the time coordinate $t = \infty$. As a specific example, let us calculate the time coordinate at which our explorer crosses $r = r_s$:

$$t = \int_0^{\tau} \frac{\mathrm{d}t}{\mathrm{d}\tau}\,\mathrm{d}\tau = \int_{r_0}^{r_s} \frac{\mathrm{d}t}{\mathrm{d}\tau}\frac{\mathrm{d}\tau}{\mathrm{d}r}\,\mathrm{d}r.$$

With (6.14) and (6.39) this becomes

$$t = \int_{r_0}^{r_s} \frac{\mathrm{d}r}{D\sqrt{1-D}} = \frac{1}{\sqrt{r_s}}\int_{r_0}^{r_s} \frac{r^{3/2}\mathrm{d}r}{r - r_s} = \infty. \tag{6.40}$$

Thus no matter when our explorer sets off, an observer who uses Schwarzschild's coordinates always assigns $t = \infty$ to the event at which the explorer crosses $r = r_s$. We should not be surprised that such a foolish convention leads to a singular metric; if we choose coordinates q_i in ordinary space in such a way that all points on the edge of a ruler are assigned the same three numbers q_i, an expression for the length of the ruler in terms of the coordinates of the ruler's ends is going to involve multiplication by some awfully big numbers!

To bring this problem under control we need to choose a new coordinate system. In 1960 M. Kruskal showed that when new coordinates (r', t') are defined through

$$\begin{aligned} r'^2 - t'^2 &= r_s^2\Big(\frac{r}{r_s} - 1\Big)e^{r/r_s} \\ t' &= r'\frac{\cosh(ct/r_s) - 1}{\sinh(ct/r_s)} = r'\tanh\Big(\frac{ct}{2r_s}\Big)\end{aligned} \tag{6.41a}$$

the metric takes the non-singular form

$$\mathrm{d}s^2 = r^2(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2) + 4(\mathrm{d}r'^2 - \mathrm{d}t'^2)\frac{r_s}{r}e^{-r/r_s}. \tag{6.41b}$$

The lines $r' =$ constant are always timelike. Radially directed photons move along the 45° lines $\mathrm{d}r' = \pm\mathrm{d}t'$ in the (r', t') plane. In particular, the null line $r = r_s$ becomes $r' = t'$. If we plot curves of constant r and t in the (r', t') plane, we get a picture like this

It is now obvious that Schwarzschild's coordinates (r, t) break down as $r' = t'$ is approached. To first order in ct/r_s (6.41a) becomes $t' \simeq \frac{1}{2}ctr'/r_s$, so t' may be considered a stretched form of t at $r = \infty$. Near $r = r_s$, $t' \simeq r'$ and by (6.41a) all events correspond to large t as expected. The region $t' > r'$ corresponds to $r < r_s$. At $r = 0$, corresponding to $t'^2 - r'^2 = r_s^2$, there is a bona-fide singularity in the gravitational field.

The Schwarzschild radius r_s corresponding to the mass of the Sun is 2.96 km. The black holes that probably power quasars and other very active galactic nuclei are likely to have Schwarzschild radii between the radius of the Sun and that of the Earth's orbit.

Exercise (19):

Show that a cubic light-year of water (supposed incompressible) would be contained within its Schwarzschild radius.

6.7 Summary

The metric outside a point mass can be written to look like that of ordinary spherical polar coordinates with $1 \to (1-r_s/r)$ in the tt slot and $1 \to 1/(1-r_s/r)$ in the rr slot. The singularity of these corrction factors when $r = r_s = 2GM/c^2$ is not physically interesting. However the geometry of spacetime *is* singular at $r = 0$ and $r = r_s$ *is* special in that an "outward" running photon on this sphere would actually not move away from the centre.

The Schwarzschild metric accounts for the last 10% of the precession of Mercury's perihelion and for the measured bending of light by the Sun. The magnitude of both these effects is of order $n \times r_s/r$, where $n \sim 4$ and r is the smallest distance of the the test body from the Sun. Detailed studies of the Solar System's dynamics show that any errors in the g.r.'s corrections to Newtonian dynamics are smaller that $\sim 0.1\%$. There is evidence that many distant galaxies are multiply imaged by gravitational lenses just as g.r. predicts.

7 Cosmology

7.1 Empirical Basis

Between 1920 and 1928 it became clear that the Universe is populated by countless galaxies like the Milky Way, and that these are receeding from one another with velocities that are proportional to separation. If we follow the trajectories of these galaxies back in time, we find that some 10^{10} yr ago the mean density of the Universe must have been extremely high. Indeed, a naive extrapolation leads to the conclusion that a finite time in the past any density was reached, no matter how great.

In 1946 G. Gamow at Cornell, and 20 years later R. Dicke in Princeton, argued that the large abundance (about 25% by weight) of He in the present Universe could have been generated some minutes after the formation of the Universe if a black-body radiation field fills the present Universe. The first estimate of the current temperature of this radiation field was 25 K, but this later fell to ≈ 3 K. In 1964 A. Penzias & R. Wilson at Bell Labs discovered this cosmic background serendipitously. This triumph of the big-bang theory quickly killed all interest in attempts to construct a steady-state cosmology.

It is now known that the spectrum of the cosmic background is accurately Planckian with $T = 2.7 \pm 0.1$ K. An observer who moves with respect to the centre of our Galaxy at $\approx 400\,\mathrm{km\,s^{-1}}$ in a certain direction would see the same spectrum in all directions, to within at least a few parts in 10,000. At any point in the Universe a natural standard of rest is defined as that of an observer whose cosmic background is isotropic. Such observers are called **fundamental observers**. Any two fundamental observers receed from one another with a speed $v \approx D \times (25 \pm 8)\,\mathrm{km\,s^{-1}}$, where D is their separation in millions of light years. A more suggestive way of expressing this result is $v = D/\tau$, where $20\,\mathrm{Gyr} \gtrsim \tau \gtrsim 10\,\mathrm{Gyr}$.

As the Universe expands, the photons of the cosmic background are doppler shifted to lower frequencies and the temperature characterizing their distribution falls.

7.2 Friedmann Metrics

The first step towards finding a solution of Einstein's equations to describe the expanding Universe is to choose a good coordinate system. The cosmic radiation background is a great help in this: we may say that two events occur at the same place if they occur on the world-line of a single fundamental observer. Similarly, two events that occur at different places may be said to occur simultaneously if the background temperature measured by fundamental observers local to those events are the same. With this natural division into space and time we would expect $\mathrm{d}s^2$ to be of the form

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + g_{ij}\mathrm{d}x^i\mathrm{d}x^j, \tag{7.1}$$

$\boldsymbol{g}$ is the metric of a 3-space of simultaneous events.

The structure of $\boldsymbol{g}$ is strongly restricted by the fact that fundamental observers observe the cosmic background to be highly isotropic: the photons they receive were last scattered at a point several thousands of millions of light years away, at a time when the mean density of the Universe was about 10^9 times its present value. In fact, until these photons collide with an observer's telescope they have been flying freely through space since the Universe was a mere 10^{-4} of its present age. Consequently, when a fundamental observer compares the temperature he sees in the forward and backward directions, he is comparing physical conditions in the early Universe at points that are now separated by thousands of millions of light years. Since these conditions are found to be identical to within a few parts in 10,000 we conclude that the Universe is extremely homogeneous on any time-slice $t =$ constant. Hence the geometry of such a space, which is described by $\boldsymbol{g}$, should be extremely homogeneous too.

A theorem in differential geometry states that any homogeneous and isotropic 3-space must be a scaled version of one of three basic models:

(i) Flat space Obviously this admits spherical polar coordinates in which the line element can be written

$$\mathrm{d}s^2 = \mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2). \tag{7.2}$$

Constructing the unit n-sphere

1-sphere:	$(x_1, x_2) = (\sin\phi, \cos\phi)$
2-sphere:	$(x_1, x_2, x_3) = (\sin\phi\sin\theta, \cos\phi\sin\theta, \cos\theta)$
3-sphere:	$(x_1, x_2, x_3, x_4) = (\sin\phi\sin\theta\sin\eta, \cos\phi\sin\theta\sin\eta, \cos\theta\sin\eta, \cos\eta)$
n-sphere:	$(x_1, \ldots, x_{n+1}) = (\sin\theta_1\sin\theta_2\ldots\sin\theta_n, \ldots, \cos\theta_{n-1}\sin\theta_n, \cos\theta_n)$

(ii) The 3-sphere Suppose we parametrize the coordinates of points $\mathbf{x}$ in a 4-dimensional Euclidean space (nothing to do with spacetime) by

$$(x_1, x_2, x_3, x_4) = a(\sin\psi\sin\theta\cos\phi, \sin\psi\sin\theta\sin\phi, \sin\psi\cos\theta, \cos\psi).$$

Then it is easy to show that $\sum_\mu x_\mu^2 = a^2$. Hence as we vary the three angles (ψ, θ, ϕ) the point $\mathbf{x}$ moves over a 3-sphere. The small vector $\mathbf{\Delta}^{(\phi)}$ that joins two points whose coordinates differ only by a small change $\delta\phi$ in ϕ is

$$\begin{aligned}\mathbf{\Delta}^{(\phi)} &= \frac{\partial\mathbf{x}}{\partial\phi}\,\delta\phi \\ &= a(-\sin\psi\sin\theta\sin\phi, \sin\psi\sin\theta\cos\phi, 0, 0)\,\delta\phi.\end{aligned}$$

Similarly,

$$\begin{aligned}\mathbf{\Delta}^{(\theta)} &= a(\sin\psi\cos\theta\cos\phi, \sin\psi\cos\theta\sin\phi, -\sin\psi\sin\theta, 0)\,\delta\theta \\ \mathbf{\Delta}^{(\psi)} &= a(\cos\psi\sin\theta\cos\phi, \cos\psi\sin\theta\sin\phi, \cos\psi\cos\theta, -\sin\psi)\,\delta\psi.\end{aligned}$$

It is straightforward to check that these three small vectors are mutually perpendicular. Hence when we move by an arbitrary small amounts $(\delta\psi, \delta\theta, \delta\phi)$ over the sphere, the distance traversed δs is given by

$$\begin{aligned}\delta s^2 &= |\mathbf{\Delta}^{(\psi)}|^2 + |\mathbf{\Delta}^{(\theta)}|^2 + |\mathbf{\Delta}^{(\phi)}|^2 \\ &= a^2(\delta\psi^2 + \sin^2\psi\delta\theta^2 + \sin^2\psi\sin^2\theta\delta\phi^2).\end{aligned} \tag{7.3}$$

If we introduce a new coordinate in place of ψ

$$r \equiv a\sin\psi \quad \Rightarrow \quad \mathrm{d}r^2 = (a^2 - r^2)\,\mathrm{d}\psi^2, \tag{7.4}$$

and define the **curvature** K of the sphere as

$$K \equiv \frac{1}{a^2}, \tag{7.5}$$

then (7.3) becomes

$$\mathrm{d}s^2 = \frac{\mathrm{d}r^2}{1 - Kr^2} + r^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2). \tag{7.6}$$

Notice that the 2-sphere with area $4\pi r^2$ has radius $a\psi > r$. Thus within the 3-sphere the areas of the members of a nested sequence of 2-spheres increase more slowly than they would in Euclidean space. (Similarly, for concentric small cicles on a two sphere circumference/2π increases more slowly than radius.)

(iii) Hyperbolic space If we set $K = 0$, the line element (7.6) of the 3-sphere becomes the line-element (7.2) of flat Euclidean space. The line element of the only other homogeneous, isotropic 3-space is given by (7.6) with K set equal to a negative number. This space is caled **hyperbolic space**. It is harder to visualize than the 3-sphere because it cannot be embedded in Euclidean 4-space. The characteristic property of hyperbolic space is that in it a 2-sphere with area $4\pi r^2$ has radius

$$R = \int_0^r \frac{\mathrm{d}r}{\sqrt{1+|K|r^2}} = \frac{1}{\sqrt{|K|}} \sinh^{-1}\left(r\sqrt{|K|}\right) < r.$$

That is, in this space the areas of a sequence of nested 2-spheres increase *faster* than in Euclidean space.

In summary, a spatial section of simultaneous events must form either a 3-sphere, flat space or hyperbolic space. In each case the line element may be expressed in the form (7.6) with an appropriate value of K.

We want to use coordinates on these spatial sections such that the coordinates of each fundamental observer are constant. These are called **comoving coordinates**. Since fundamental observers are receeding from one another, it follows that our desired coordinates cannot at all times coincide with those in which the line element takes the form (7.6). However, if at one time, for example now, the comoving coordinates (r, θ, ϕ) are such that the line element is of this form, then at an earlier time, when fundamental observers were closer to one another, the separation δs between neighbouring observers was some fraction $a(t)$ of their current separation. Hence at all times the metric of spacetime can be written

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2\left[\frac{\mathrm{d}r^2}{1-Kr^2} + r^2(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2)\right], \tag{7.7}$$

where K is the curvature of the *current* time-slice $t = t_0$ and $a(t_0) = 1$.

Equations (4.22) yield for the Γ's that involve the time index:

$$\begin{aligned} \Gamma^t_{t\alpha} &= \Gamma_{t,t\alpha} = 0 \qquad & \Gamma^t_{ij} &= -\frac{1}{c^2}\Gamma_{t,ij} = \frac{\dot a}{a}\frac{g_{ij}}{c^2} \\ \Gamma^i_{tt} &= \Gamma_{i,tt} = 0 \qquad & \Gamma^i_{tj} &= g^{ik}\Gamma_{k,tj} = \frac{\dot a}{a}\delta^i_j. \end{aligned} \tag{7.8}$$

7.3 The Cosmological Redshift

We know that the Universe is expanding because we observe the frequencies of spectral lines from distant galaxies to be shifted towards lower frequencies. It turns out that the magnitude of this spectral shift is related in a remarkably simple way to the scale of the Universe when the light by which we see galaxies set out towards us.

The **redshift** z is defined by

$$1 + z \equiv \frac{\omega_{\mathrm{emit}}}{\omega_{\mathrm{observe}}}.$$

If we elevate our status to that of a fundamental observer, and suppose that the atoms that emit the radiation we receive were stationary with respect to a local fundamental observer, then the 4-momenta of photons have zeroth component $\omega_{\mathrm{emit}}/c^2$ on emission and $\omega_{\mathrm{obs}}/c^2$ on observation. By direct analogy with the deflection of light by the Sun we may follow the decrease of ω by integrating the zeroth component of

$$k^\mu \nabla_\mu k^\nu = 0. \tag{7.9}$$

As in §6.3 we write $k^\mu = \mathrm{d}x^\mu/\mathrm{d}s$ [eq. (6.27)], and multiply (7.9) through by $\mathrm{d}s/\mathrm{d}t$ to get

$$\begin{aligned} 0 &= \frac{\mathrm{d}s}{\mathrm{d}t}\frac{\mathrm{d}x^\mu}{\mathrm{d}s}\left[\frac{\partial\omega/c^2}{\partial x^\mu} + \Gamma^t_{\mu\gamma}k^\gamma\right] \\ &= \frac{\mathrm{d}\omega/c^2}{\mathrm{d}t} + \frac{\mathrm{d}s}{\mathrm{d}t}\Gamma^t_{\mu\gamma}k^\mu k^\gamma. \end{aligned}$$

With (6.27) and (7.8) this becomes for a radially propogating photon

$$\frac{\mathrm{d}\omega}{\mathrm{d}t} = -\frac{\dot a}{a}\left(g_{rr}k^r k^r\right)\frac{c^2}{\omega} = -\frac{\dot a}{a}\omega,$$

where we have used the null property of k^μ in the form $g_{rr}k^r k^r + g_{tt}(\omega/c^2)^2 = 0$. Integrating we get

$$1 + z = \frac{\omega_{\rm emit}}{\omega_{\rm obs}} = \frac{a(t_{\rm obs})}{a(t_{\rm emit})}.$$

In words, $1+z$ gives the factor by which the Universe has expanded since the photons we receive were emitted. Notice that this result has been obtained without using Einstein's equations to determine the dynamics of the Universe.

7.4 Field Equations for Friedmann Cosmologies

Let's now go back to equations (7.8) and the job of calculating the Christoffel symbols of the Friedmann metric (7.7). Since each Γ^i_{jk} is unaffected by a position-independent scaling of $\boldsymbol{g}$, these Γ's can be obtained from the expressions (6.5) for the spatial Γ's of the Schwarzschild solution with $B = 1/(1 - Kr^2)$:

$$\begin{aligned} &\Gamma^r_{rr} = \frac{Kr}{1-Kr^2} \quad \Gamma^r_{\theta\theta} = -r(1-Kr^2) \quad \Gamma^r_{\phi\phi} = -r\sin^2\theta(1-Kr^2) \\ &\Gamma^\theta_{\phi\phi} = -\sin\theta\cos\theta \qquad \Gamma^\theta_{\theta r} = \Gamma^\theta_{r\theta} = \frac{1}{r} \\ &\Gamma^\phi_{\phi r} = \Gamma^\phi_{r\phi} = \frac{1}{r} \qquad \Gamma^\phi_{\phi\theta} = \Gamma^\phi_{\theta\phi} = \cot\theta. \end{aligned} \tag{7.10}$$

When using these results in (5.12) to calculate $R_{\alpha\beta}$ it is helpful to isolate all terms that involve a t index. One finds

$$\begin{aligned} R_{it} = R_{ti} &= 0 \qquad R_{tt} = \frac{\partial\Gamma^\mu_{t\mu}}{\partial t} + \Gamma^j_{tk}\Gamma^k_{tj} = 3\frac{\ddot a}{a} \\ R_{ij} &= \widetilde{R}_{ij} - \frac{\partial\Gamma^t_{ij}}{\partial t} + 2\Gamma^t_{ik}\Gamma^k_{jt} - \Gamma^t_{ij}\Gamma^k_{tk} \\ &= \widetilde{R}_{ij} - \left[\frac{\ddot a}{a} + 2\left(\frac{\dot a}{a}\right)^2\right]\frac{g_{ij}}{c^2}, \end{aligned}$$

where $\widetilde{R}_{ij}$ is the Ricci tensor of the 3-space whose metric is g_{ij}. A tedious calculation yields[14]

$$\widetilde{R}_{ij} = -\frac{2K}{a^2}g_{ij}. \tag{7.11}$$

14 Since the 3-space is homogeneous and isotropic, it is obvious that $\widetilde{\boldsymbol{R}} \propto \boldsymbol{g}$. Hence it is only necessary to calculate one non-zero component of $\widetilde{\boldsymbol{R}}$, say $\widetilde{R}_{rr}$.

Hence

$$R_{\alpha\beta} = \begin{pmatrix} \frac{3\ddot{a}}{a} & & & \\ & -\left[\frac{2Kc^2}{a^2} + \frac{\ddot{a}}{a} + 2\left(\frac{\dot{a}}{a}\right)^2\right]\frac{g_{rr}}{c^2} & & \\ & & -\left[\frac{2Kc^2}{a^2} + \frac{\ddot{a}}{a} + 2\left(\frac{\dot{a}}{a}\right)^2\right]\frac{g_{\theta\theta}}{c^2} & \\ & & & -\left[\frac{2Kc^2}{a^2} + \frac{\ddot{a}}{a} + 2\left(\frac{\dot{a}}{a}\right)^2\right]\frac{g_{\phi\phi}}{c^2} \end{pmatrix}. \tag{7.12}$$

We now turn our attention to the right side of the Einstein equations (5.17). At the present epoch the energy density contributed by the cosmic background is $a_s(2.7)^4 \simeq 1.9\times 10^5 e.v.\,\mathrm{m}^{-3}$. The rest mass energy density on the other hand is at least $10^{-27}\,\mathrm{kg\,m}^{-3} = 5.6\times 10^8\, e.v.\,\mathrm{m}^{-3}$, so that the cosmic energy density is currently dominated by rest mass energy. In this case we may adopt for $\mathbf{T}$ the formula (2.17) for dust. In our chosen frame of reference $\mathbf{v} = (0,0,0,c)$, so we now have

$$T_{tt} = \rho c^2 \quad ; \quad T^\alpha_\alpha = -\rho c^2 \qquad \text{(dust)}, \tag{7.13}$$

where ρ is the rest-mass density.

During the first few thousands of years of the Universe's evolution rest-mass energy will not have been dominant. To see this recall that when a box of volume V that contains radiation is slowly expanded, the radiation behaves like an ideal gas with ratio of principal specific heats $\gamma = \frac{4}{3}$. Hence the radiation's energy density, which is three times its pressure, falls faster than the rest-mass denergy density of any dust that is also uniformly distributed through the box: $U_{\rm rad} = 3P_{\rm rad} \propto V^{-4/3} \propto U_{\rm dust}^{4/3}$. Hence $U_{\rm rad}/U_{\rm dust} \propto 1/a$ and although the cosmic radiation density is now at least a thousand times smaller than the rest-mass energy density, back when $a \lesssim 10^{-4}$ the radiation density would have exceeded the matter density. Therefore we take $\mathbf{T}$ to be of the general isotropic form [cf. (2.20)]

$$T^\beta_\alpha = \begin{pmatrix} -\rho c^2 & 0 & 0 & 0 \\ 0 & P & 0 & 0 \\ 0 & 0 & P & 0 \\ 0 & 0 & 0 & P \end{pmatrix} \qquad \Rightarrow \qquad T^\alpha_\alpha = 3P - \rho c^2. \tag{7.14}$$

With $\mathbf{T}$ of the form (7.14) the tt-equation of the set (5.17) reads

$$\frac{3\ddot{a}}{ac^2} = -\frac{8\pi G}{c^4}\left(\tfrac{3}{2}P + \tfrac{1}{2}\rho c^2\right). \tag{7.15a}$$

The rr-equation reads

$$-\left[\frac{2Kc^2}{a^2} + \frac{\ddot{a}}{a} + 2\left(\frac{\dot{a}}{a}\right)^2\right]\frac{g_{rr}}{c^2} = -\frac{8\pi G}{c^4}\tfrac{1}{2}(\rho c^2 - P)g_{rr}. \tag{7.15b}$$

Eliminating $\ddot{a}$ between these equations yields

$$\dot{a}^2 + Kc^2 = \tfrac{8}{3}\pi G\rho a^2. \tag{7.16}$$

We also have the equation of mass-energy conservation $T^\beta_{\alpha\,;\beta} = 0$, which for $\alpha = t$ gives

$$\begin{aligned} 0 &= \frac{\partial T^t_t}{\partial t} - \Gamma^\gamma_{\beta t}T^\beta_\gamma + \Gamma^\beta_{\beta\gamma}T^\gamma_t \\ &= -\dot{\rho}c^2 - 3\frac{\dot{a}}{a}P + 3\frac{\dot{a}}{a}(-\rho c^2) \quad \Rightarrow \quad \frac{\mathrm{d}\rho a^3}{\mathrm{d}a} = -\frac{3a^2P}{c^2}. \end{aligned} \tag{7.17}$$

If $P = 0$ this states that the mass in each comoving volume is conserved, while if $P = \frac{1}{3}\rho c^2$, we have that ρa^4 is conserved. Since $a(t_0) = 1$, we therefore have

$$\rho(t) = \begin{cases} \dfrac{\rho(t_0)}{a^3(t)} & \text{(dust)} \\ \dfrac{\rho(t_0)}{a^4(t)} & \text{(radiation)} \end{cases} \tag{7.18}$$

and (7.16) becomes

$$\dot{a}^2 = \begin{cases} \dfrac{8\pi G}{3a}\rho(t_0) - Kc^2 & \text{(dust)} \\ \dfrac{8\pi G}{3a^2}\rho(t_0) - Kc^2 & \text{(radiation).} \end{cases} \tag{7.19}$$

Currently the Universe is matter-dominated and expanding, so $\dot{a} > 0$. Equation (7.19) states that it will expand for ever if $K \leq 0$. But if $K > 0$ (the case in which spatial sections are 3-spheres), the expansion will cease when

$$a = \frac{8\pi G\rho(t_0)}{3c^2 K} = \frac{1}{(7.5 \times 10^{10}\,\text{light yr})^2 K} \times \frac{\rho(t_0)}{10^{-27}\,\text{kg}\,\text{m}^{-3}}.$$

Thus our longevity hangs ultimately on how the radius of curvature of the Universe compares with some tens of billions of light years.

Exercise (20):

Integrate (7.19) in the case of dust to show

$$\frac{c\sqrt{|K|}}{a_m}t(a) = \begin{cases} \theta - \frac{1}{2}\sin 2\theta & \text{when } K > 0 \quad [\theta \equiv \arcsin(\sqrt{a/a_m})] \\ \frac{1}{2}\sinh 2\theta - \theta & \text{when } K < 0 \quad [\theta \equiv \text{arcsinh}(\sqrt{a/a_m})] \end{cases}$$

Sketch $a(t)$ in the two cases.

The special case $K = 0$ divides a doom-laden future from one of ultimate boredom. In this case the present density is given by

$$\rho_{\text{crit}}(t_0) = \left.\frac{3\dot{a}^2}{8\pi G a^2}\right|_{t_0}. \tag{7.20}$$

The distance between nearby fundamental observers, $\Delta s \simeq a(t)\Delta r$, increases at a rate $\dot{a}\Delta r = (\dot{a}/a)\Delta s$. Thus $(\dot{a}/a)$ is the quantity H in Hubble's relation $v = Hs$. Its current value lies near $75\,\text{km}\,\text{s}^{-1}\,\text{Mpc}^{-1}$ in idiotic astronomical units; this translates to $2.43 \times 10^{-18}\,\text{s}^{-1}$, so

$$\rho_{\text{crit}}(t_0) = 1.06 \times 10^{-26}\,\text{kg}\,\text{m}^{-3}. \tag{7.21}$$

The best observational evidence suggests that the actual density is about a factor ten lower than this: the future is more likely to be boring than otherwise. However, it is widely believed on semi-philosophical grounds that $\rho = \rho_{\text{crit}}$. Note that if $\rho \leq \rho_{\text{crit}}$, the Universe is spatially infinite and contains infinite mass, while if $\rho > \rho_{\text{crit}}$ the total mass is finite.

Exercises (21):

(i) Show for a dust-dominated universe with $K = 0$ that $a = (t/t_0)^{2/3}$. Hence estimate the age of the Universe if $\rho(t_0) = \rho_{\text{crit}}(t_0)$.

(ii) Show for a radiation-dominated universe with $K = 0$ that $a = \sqrt{t/t_0}$.

(iii) Show that in Newton's theory the radial coordinate $a(t)$ of a particle embedded in a homogeneous spherical cloud of mutually gravitating particles which are initially receeding from the origin with speeds proportional to radius, obeys (7.16). Identify the analogue of K in this case.

7.5 Inflation

In 1981 Alan Guth of M.I.T. pointed out[15] that grand unified theories of particle physics, which attempt to unite the strong, electromagnetic and weak forces, suggest that the cosmic scale factor $a(t)$ may for a period have grown exponentially rather than at the leisurely rate $a \propto \sqrt{t}$ expected of a conventional radiation-dominated early Universe. Exponential growth is caused by the vacuum temporarily stumbling into a so-called "false vacuum" state. A false vacuum differs from the usual vacuum in that it has a large energy density even at zero temperature: $\rho(T=0) \approx 10^{77}\,\mathrm{kg\,m^{-3}}$. Obviously the zero-temperature vacuum must be Lorentz invariant, so the energy-momentum tensor of this vacuum must be a multiple of the metric tensor. Thus

$$T_{\mu\nu} = -\lambda g_{\mu\nu} \qquad (\lambda \text{ a constant}). \tag{7.22}$$

In a locally freely-falling frame $g_{\mu\nu} = \eta_{\mu\nu}$, so a positive energy density corresponds to $\lambda > 0$. It follows that the false vacuum exerts a negative pressure; $P = -\lambda$.[16] When we plug $P = -\rho c^2 = \lambda$ into (7.15a) we get

$$\ddot{a} = \frac{8\pi G\lambda}{3c^2} a \qquad \Rightarrow \qquad a(t) = a(0)\exp\left(\sqrt{\frac{8\pi G\lambda}{3c^2}}\, t\right). \tag{7.23}$$

Grand unified theories suggest that the time constant associated with this exponential growth is $\approx 10^{-34}$ s.

Exercise (22):

Let the present age of the Universe be t_H and the distance over the *current* time-slice $t = t_H$ to the most distant fundamental observer it is in principle possible to see be D_H. Show that if the Universe had inflated from $t = 0$ to the present day we would have $D_H = ct_H$, while we would have $D_H = 2ct_H$ if the Universe had been always flat and radiation-dominated. The furthest fundamental observer we can see is said to be on the **particle horizon**. [Hint: use $0 = g_{rr}\mathrm{d}r^2 + g_{tt}\mathrm{d}t^2$.]

Guth's inflationary conjecture has two very seductive properties:

(i) It offers an explanation of why the Universe is so homogeneous on a large scale by suggesting that everything we see may have emerged from the explosive expansion of a single causally-connected fluctuation in the preinflationary Universe.

(ii) It offers an explanation of why $\rho(t_0)/\rho_{\mathrm{crit}}(t_0) \simeq 1$: with the definition (7.20) of ρ_{crit} the cosmic energy equation (7.16) can be written

$$\frac{\rho(t)}{\rho_{\mathrm{crit}}(t)} = 1 + \frac{Kc^2}{\dot{a}^2}. \tag{7.24}$$

Whatever the initial value of K, after a sufficient number of e-folding times $\dot{a}$ becomes enormous and the deviation of each side of (7.24) from unity becomes extremely small.

The Universe's inflationary episode is supposed to have begun when the cosmic temperature dropped below the temperature at which it became thermodynamically favourable for the vacuum to move to a configuration of lower symmetry—an oft-quoted analogy is with a transition to ferromagnetism at the Curie temperature. It is argued that the cosmic vacuum may have been slow to accomplish this transition, just as water-vapour in a cloud chamber is slow to form water droplets. The excess of the vacuum's actual energy-density over its theoretical lowest-energy state is presumed to be physically real (unlike the zero-point energy of the vacuum's normal modes) and to require representation on the right side of the Einstein equations. The inflationary period is supposed to have ended when the vacuum made a phase transition into the lower-energy configuration, releasing its former energy density as normal thermal radiation.

[15] *Phys. Rev.,* **D23**,347.

[16] The physical origin of this negative pressure can be understood by imagining what happens when we increase by $\mathrm{d}V$ the volume of a cylinder containing the false vacuum. The false vacuum's mass increases by $\rho_{\mathrm{vac}}\mathrm{d}V$, so its energy increases by $\rho_{\mathrm{vac}}c^2\mathrm{d}V$. The latter increase must equal the work done on the piston, $-P\mathrm{d}V$. Thus the pressure of the false vacuum is $P = -\rho_{\mathrm{vac}}c^2$.

7.6 Cosmic Strings

It is thought that when the vacuum changed its phase from a symmetric high-temperature form to a less symmetrical low-temperature form, discontinuities may have arisen that would have persisted to the present day. The general idea is illustrated by what happens when a lump of iron cools in zero magnetic field through the Curie temperature T_c (at which iron becomes ferromagnetic). At T_c groups of atoms here and there in the lump decide to align their spins in some common direction. Since the direction is chosen at random, widely separated groups choose different directions. So long as the groups remain isolated they can all grow by convincing adjacent uncommited atoms to align with them. But eventually the swelling groups touch each other – the lump has become a mass of interlocking domains. Between the domains are regions of high B and therefore of large magnetic energy. So it is energetically desirable for each domain boundary to shrink. But usually the boundary around one domain can shrink only if the boundaries of adjacent domains grow. So the domains are effectively locked into place.

When the Universe cools two-dimensional domain boundaries may form, but the most important discontinuities are one-dimensional – strings. The complex field ψ associated with charged particles such as electrons can give rise to a string like this.[17] Imagine that it is decided that the field shall everywhere have amplitude $|\psi| = 1$ and you are told to specify its phase $0 \le \arg(\psi) \le 2\pi$ throughout space. You decide to set $\arg[\psi(\boldsymbol{x})] = \phi(\boldsymbol{x})$, where ϕ is the usual cylindrical-polar coordinate of the point $\boldsymbol{x}$. This assignment works fine everywhere except at your coordinate origin, $r = 0$. Here $\nabla \arg(\psi)$ diverges since any phase can be reached arbitrarily close to $r = 0$. It is not hard to persuade oneself that by adjusting the values of ψ in any finite volume you can move but not eliminate this singularity, which is associated with a line of energy-momentum. This is a cosmic string.

What does the energy momentum tensor $\mathbf{T}$ look like in the narrow tube around $r = 0$ in which $\mathbf{T} \neq 0$? We'd expect $\mathbf{T}$ to be Lorentz invariant with respect to boosts parallel to the string's line. So the in the (t, z) plane $\mathbf{T}$ has to be proportional to the Minkowski metric. Also it's hard to see how the string could be carrying anything in the x or y directions. So

$$T_{\mu\nu} = -\rho c^2 \begin{pmatrix} -c^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{7.25}$$

where ρ is a constant.

Now consider the line element

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + r_0^2(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2) + \mathrm{d}z^2, \tag{7.26}$$

where r_0 is a constant. This is almost the line element $\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + \mathrm{d}r^2 + r^2\mathrm{d}\phi^2 + \mathrm{d}z^2$ of flat spacetime in cylindrical polars; $r_0\theta$ is a kind of radial variable. The only non-zero Christoffel symbols generated by (7.26) are

$$\Gamma^\theta_{\phi\phi} = -\tfrac{1}{2}\sin 2\theta \quad ; \quad \Gamma^\phi_{\phi\theta} = \Gamma^\phi_{\theta\phi} = \cot\theta.$$

The only non-zero components of the Ricci tensor are

$$R^\theta_\theta = R^\phi_\phi = -r_0^{-2}.$$

[17] The treatment here is a little oversimplified inasmuch as it neglects the fact that for electrons ψ is a Dirac spinor rather than a scalar.

Thus $R = -2r_0^{-2}$ and the Einstein equations (5.16) read

$$R^\beta_\alpha - \tfrac{1}{2}\delta^\beta_\alpha R = \begin{pmatrix} r_0^{-2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & r_0^{-2} \end{pmatrix} = -\frac{8\pi G}{c^4} T^\beta_\alpha$$
$$= \frac{8\pi G\rho}{c^2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{7.27}$$

Hence with $\rho > 0$ (which corresponds to a positive energy density and tension in the string) the metric (7.26) solves Einstein's equations inside the string.

What we really need is the metric outside the string, where we live. Let the outer surface of the string be $\theta = \theta_m$. Then the exterior metric is

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + r_0^2\Big(\frac{\cos^2\theta}{\cos^2\theta_m}\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\Big) + \mathrm{d}z^2. \tag{7.28}$$

This metric obviously joins smoothly to the interior metric (7.26) on $\theta = \theta_m$. To show that it is a vacuum solution of Einstein's equations, we transform to a new coordinate set (t, r', ϕ', z), where the t and z coordinates are the old ones and

$$r' \equiv r_0\frac{\sin\theta}{\cos\theta_m} \quad ; \quad \phi' \equiv \cos(\theta_m)\phi. \tag{7.29}$$

The metric (7.28) now becomes

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + \mathrm{d}r'^2 + r'^2\mathrm{d}\phi'^2 + \mathrm{d}z^2, \tag{7.30}$$

which is just the cylindrical-polar metric of flat spacetime. But on a large scale the spacetime outside the string is very odd because the range of ϕ' is $(0, 2\pi\cos\theta_m)$. [This follows from (7.29) and the fact that ϕ is in $(0, 2\pi)$]. Consider for example a large circle $r' = a \gg r_0$. The radius of this circle is

$$R = \int_0^a \sqrt{g_{r'r'}}\,\mathrm{d}r' \simeq a, \tag{7.31a}$$

while its circumference is

$$C = \int \sqrt{g_{\phi'\phi'}}\,\mathrm{d}\phi' = a2\pi\cos(\theta_m). \tag{7.31b}$$

So the usual flat-space relation $C = 2\pi R$ does not apply. Thinking about a cone may help to clarify this strange state of affairs. At each point a cone is flat in the sense that it can be made out of a piece of paper without stretching the paper (you can't make a paper sphere as easily), but circles distance a from the cone's apex have a circumference smaller than $2\pi a$.

How could we detect a cosmic string? Our best bet is to look for lines of gravitationally lensed objects. To understand how a string lenses an object, think of the exterior space as a piece of paper with a wedge of angle

$$\theta_{\mathrm{def}} \equiv 2\pi(1 - \cos\theta_m) \tag{7.32}$$

cut out and corresponding points along the cuts identified. Place the object to be lensed at radius $r' = a_q$ on the cut and yourself directly opposite at $r' = a_o$.

Rays travel over the paper in straight lines, so you can see the object along two lines of sight separated by $2\alpha_s$, where

$$\frac{\sin(\pi - \frac{1}{2}\theta_{\rm def})}{\sqrt{a_o^2 + a_q^2 + 2a_o a_q \cos(\pi - \frac{1}{2}\theta_{\rm def})}} = \frac{\sin\alpha_s}{a_q}.$$

The largest possible value of α_s is clearly $\frac{1}{2}\theta_{\rm def}$. It should be possible to detect a cosmic string by looking for a line in the sky either side of which lie members of pairs of similar objects.

The mass per unit length μ of the string would follow immediately from $\theta_{\rm def}$: from the interior metric (7.26) it follows that the string's cross-sectional area is

$$A = \int_0^{\theta_m} r_0 {\rm d}\theta \int_0^{2\pi} r_0 \sin\theta {\rm d}\phi = 2\pi r_0^2(1 - \cos\theta_m).$$

Hence using (7.27) we have that the string's mass per unit length is $\mu = \rho A = c^2(1 - \cos\theta_m)/(4G) = c^2\theta_{\rm def}/(8\pi G)$ dependently of the string's physical width r_0. There won't be room outside the string for the Universe as we know it unless $\mu < \frac{1}{4}c^2/G = 3.37\times10^{26}\,{\rm kg\,m^{-1}}$. Particle theorists think strings may exist with line densities of order a thousandth of this.

7.7 Summary

The cosmic microwave background defines a natural coordinate system for cosmology. On large scales the Universe appears to be strikingly homogeneous and isotropic. This implies that equal-time hypersurfaces must have the geometry of either (i) the 3-sphere, (ii) flat space, or (iii) hyperbolic space according as the mean cosmic density ρ is greater than, equal to, or less than $\rho_{\rm crit} \simeq 10^{-26}\,{\rm kg\,m^{-3}}$. It is widely believed that $\rho = \rho_{\rm crit}$ although measurements suggest a slightly smaller value.

The cosmic scale when the light we detect from a distant object was emitted can be deduced from the redshift z of the object's spectrum: $1 + z = \omega_{\rm emit}/\omega_{\rm obs} = a(t_{\rm obs})/a(t_{\rm emit})$. The most distant objects are seen at an epoch when a was smaller than now by more than a factor 5.

The expansion of the Universe will cease only if $\rho > \rho_{\rm crit}$. At early times we always have $\rho \simeq \rho_{\rm crit}$ and the cosmic scale grows as $a \propto t^{2/3}$. If the wild speculations of high-energy physicists are to be believed, very early on there may have been an inflationary phase in which $a \propto e^{\gamma t}$ and the entire observable Universe grew out of a single quantum fluctuation.