1. Introduction

Physicists consider Relativity to be a theory about how the world works. Mathematicians consider it as a particularly beautiful application of geometry and differential equations. As a physicist in a maths school, I must remain agnostic.

Be that as it may, it is still a good idea to begin this course with a non-technical discussion of the physical and mathematical ingredients of relativity. Figure 1.1 summarizes what we are going to cover as a flowchart.

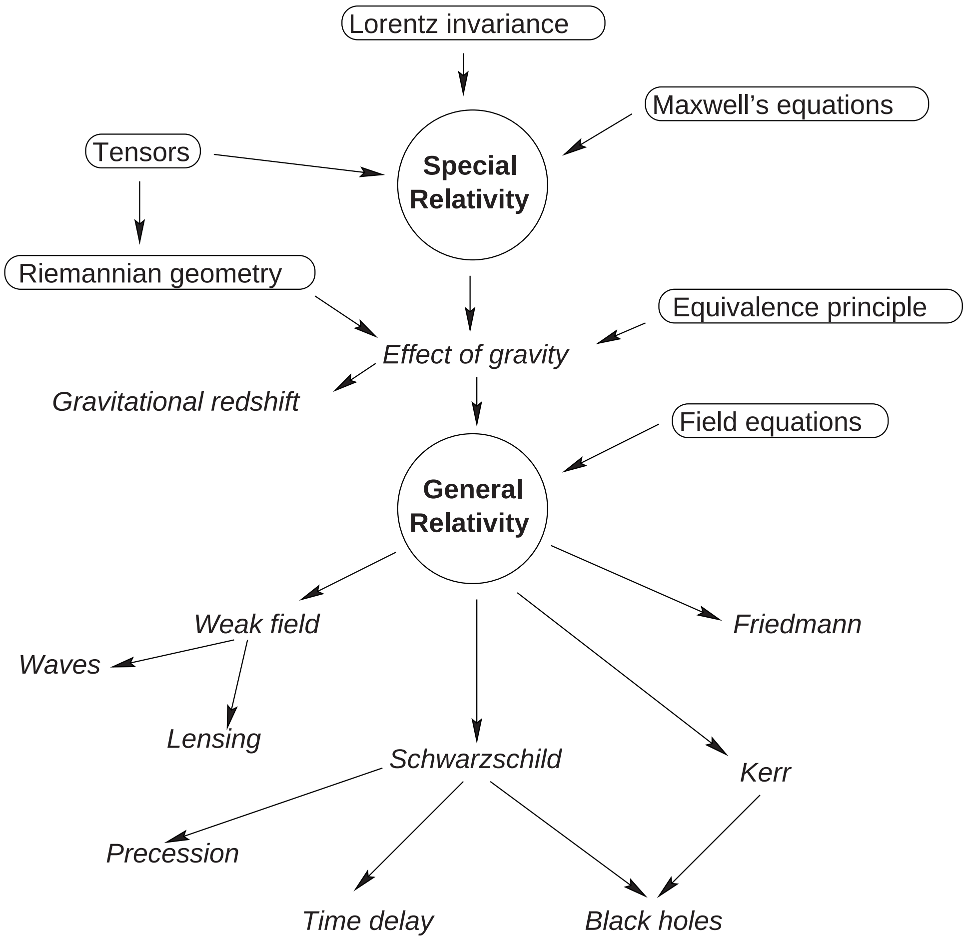

Figure 1.1: A flowchart for relativity. The circles denote the parts of the Theory of Relativity. The boxed items are the ingredients that go in to make up the theory. The unboxed items are applications of the theory.

1.1 Lorentz Invariance

'Relativity' in everyday language means that what someone observes depends on which way they are facing, how fast they are moving, and so on. This is not very useful or profound. The technical meaning of Relativity, however, is something much more specific and useful. It is about expressing physics in a way which *doesn't* depend on what the observer is doing.

The first, and probably most significant, aspect of this is the concept Lorentz invariance. It is expressed as:

I. The laws of physics, suitably expressed, are the same for all **inertial** observers. (An inertial observer is someone not in a gravitational fields and not accelerating.)

II. The speed of light in a vacuum is the same for all observers.

These are often called the two postulates of relativity. The first one looks innocuous enough, but the second one seems very strange. Think of two spacecraft elsewhere in the solar system shining light pulses towards the earth. The two spacecraft have a relative speed of (say) 50 km/sec. What about the speeds of the two light pulses? Should they differ by 50 km/sec? Experiments of this type reveal no difference in speed, even in measurements accurate to 1 m/sec. How can we possibly make sense of this?

The way to make sense of it was pointed out by Lorentz, Poincaré and (most fully and importantly) Einstein. We have to revise our notions of space and time.

Consider events, occurring at two different places and times, the separation in space and time between them being $(\Delta x, \Delta y, \Delta z, \Delta t)$. The spatial distance between them is of course

$$\left(\Delta x^2 + \Delta y^2 + \Delta z^2\right)^{\frac{1}{2}} .$$

Let us define time as an extra dimension, and define a squared distance in four-dimensional 'space-time' as

$$\Delta s^2 = \Delta x^2 + \Delta y^2 + \Delta z^2 - c^2 \Delta t^2, \tag{1.1.1}$$

where c is the speed of light. (Δs^2 may not be positive, so we won't take the square root.) As mathematicians, we are free to define anything we like and call it anything we like [as long as we do it consistently]. The question is, it is useful?

Yes it is, because a concise way of stating postulates I and II above is that Δs^2 is the same for all inertial observers. This is the principle of 'Lorentz invariance'. This is a piece of physics, subject to experimental tests. But it is a very sweeping piece of physics, because it claims to apply to all of physics, including physics that is yet to be discovered! When Einstein first wrote about it in 1905 Lorentz invariance was controversial, but when Einstein died in 1955 it was just about the most secure thing in physics. It still is—a lot of physics has been discovered since 1955, and it is all Lorentz invariant.

1.2 Tensors

Although Lorentz invariance is a piece of physics, to follow up its consequences we need to develop some mathematics. In particular, we need to develop a language for coordinate systems and geometry in four-dimensional spacetime. This is tensor calculus. Like any new language, this initially sounds and looks like gibberish. And also like any language one gets used to it, until it becomes effortless.

1.3 Special relativity

The **special theory of relativity**, developed by Einstein in 1905, is about making then-known physics Lorentz invariant. The dynamics developed by Newton and later generations was not Lorentz invariant. Einstein showed how to modify Newtonian dynamics to make it Lorentz invariant, which involved things like $E = mc^2$. This new 'relativistic' dynamics is (as far as we now know) is how nature behaves.

One important thing is lacking in special relativity. In 1905 Einstein could not make Newtonian gravity Lorentz invariant, and had to leave it out of the theory. It is in this sense that special relativity is 'special' or 'not general'. The general theory of relativity, including gravity, would take another decade.

1.4 Electromagnetism

Strangely enough, long before anyone was thinking about Lorentz invariance, there was one physical theory that was *already* Lorentz invariant. This was electromagnetism (the theory of electric charges and electric and magnetic fields) as given by Maxwell's equations of 1869. Around 1900, it led to a crisis in physics. People realized that electromagnetism was Lorentz-invariant, Newtonian dynamics was not. They couldn't *both* be right! Einstein eventually solved the problem in 1905 by modifying Newtonian dynamics to make it consistent with electromagnetism. In fact his paper is called not *Theory of relativity* but *On the electrodynamics of moving bodies.*

It turns out that Maxwell's equations are a bit of a dress rehearsal for the more complicated theory of gravity.

1.5 Principle of equivalence

The inertia of a body (i.e., how much force you have to apply to accelerate something) is proportional to its mass. The gravitational force produced by a body is also proportional to its mass. So are gravity and inertia of a body *always* in the same ratio? This is an experimentally testable thing, and it appears to be true. This led Einstein to formulate the 'principle of equivalence of gravitational and inertial mass' or **principle of equivalence** for short. It states that you can make the effect of gravity go away locally by going into free fall. So if you want to work out what physics looks like in the presence of gravity, just use coordinates moving with a bungee-jumper and use the zero-gravity equations in those coordinates.

Again, the principle of equivalence is a piece of physics, but to follow up its consequences it helps to invent some more mathematics.

Now, things that stay fixed in space or move with constant speed (in fact, inertial observers) move from event to event in space time in straight lines. In other words, they connect events through paths of minimum distance, where distance is defined as in (1.1.1). Bungee-jumper coordinates, on the other hand, are accelerating, and do *not* minimize the distance (1.1.1). The principle of equivalence implies (and will see this in gory detail later) that bungee-jumper coordinates will make the distance between events minimal provided we redefine distance in a certain new way. This new kind of distance—called the **metric**—agrees with (1.1.1) over infinitesimal regions, but not over finite regions.

Mathematicians can and do define distances in whatever way they feel like. But great mathematicians also have a sense for which definitions will turn out to be important long after they are dead. Thus it happened that in the 19th century first Gauss and then in more generality Riemann had already studied metrics that look like (1.1.1) over infinitesimal regions. The subject is called Riemannian geometry. The principle of equivalence tells us that gravity in effect puts a Riemannian metric on spacetime, and bungee-jumpers follow the shortest-distance paths between events.

Actually Riemannian metrics are not that far from our intuition. Imagine you are taking a plane from London to Tokyo. The shortest route on a map will take you sort of east-south-east. But that's not what any airline will do—they'll take you far *north*, nearly to the north pole, and then south again. Because, of course, the Earth is curved, and distances on a map are okay over short distances but not between London and Tokyo. We usually explain this by saying that the surface of the Earth is embedded in three dimensions and a two-dimensional object like a map can't get it quite right. But Riemann showed that if you keep the two-dimensional map coordinates and cunningly redefine the distance, there's no need to worry about the third dimension.

Thus the principle of equivalence tells us that gravity induces a Riemannian metric, and if we know what that metric is we can work out bungee-jumper coordinates and calculate everything we want. This leads to an observational prediction, the so-called gravitational redshift. But the principle of equivalence doesn't tell us what that metric is, for that we need one last piece of input.

1.6 General relativity

That last piece consists of Einstein's field equations. These are essentially six coupled differential equations. They are nevertheless a piece of physics, because they don't follow from anything simpler. But nobody knows how to express their physical content in a simple way, which may be a sign that we don't understand them properly. Einstein wrote them down in an inspired guess. Others have tried their own guesses since then, but all the evidence is that Einstein guessed correctly.

The result is called the **general theory of relativity** and was published in 1916. It is no longer confined to inertial observers. Accelerating observers are now allowed, as are of course gravitational fields.

1.7 Applications of general relativity

The differential equations (1916) and good evidence that they are the correct ones (c. 1919) was only the beginning. Finding exact and approximate solutions, and working out what they mean, has kept generations of researchers busy. We will have time only for a few highlights.

The first of the exact solutions is Schwarzschild's from 1917, for the metric around a spherical mass. Using it we can work out the general relativistic effect of the sun. Schwarzschild's metric also predicts black holes.

Another exact solution is Friedmann's, which is the metric associated with an expanding universe and the basis of cosmology.

When gravitational fields are weak, approximate metrics are comparatively easy to compute, and very useful. We'll discuss one of their uses, gravitational lensing.

1.8 Books

There are many many books on Relativity, at many different levels. The following are some examples.

- For physical insight with a minimum of mathematics, try

  Principles of Cosmology and Gravitation by M.V. Berry

  and even the comic book

  Einstein for Beginners

  is surprisingly good for understanding special relativity (it doesn't discuss general relativity) and has plenty of historical material. If you're curious to read some of Einstein's own writings, try

  The Principle of Relativity

- Two well-known textbooks at about the level of this course are

  A first course in general relativity by Bernard F. Schutz

  and

  Essential Relativity: Special, General and Cosmological, by W. Rindler.

  Also similar to this course is

  Classical Fields by James Binney,

  not yet published, but on the web at

  `http://www-thphys.physics.ox.ac.uk/users/JamesBinney/`

- The main source material for these lecture notes is

  Gravitational and Cosmology by Steven Weinberg

  but this is an advanced text which I don't recommend trying to work through yourself.

Notation is notoriously variable between different books in this subject. We will follow Weinberg.

1.9 The attraction of General Relativity

People often speak of the beauty of General Relativity. But different people are drawn to relativity for different reasons. The great contributors to relativity in recent times have had very different styles. Thus Penrose and Hawking seem most attracted to the mathematical elegance of the theory. Chandrasekhar loved the richness of the field equations and their solutions. Feynmann was less interested in the mathematics as such, he wanted to understand what all the physical concepts were about. Taylor is even less interested in mathematics per se, he has devoted his life to observable applications. Among authors of books, you can also sense their personalities. Thus Schutz seems sympathetic to Penrose and Hawking, Weinberg and Berry are more in the mould of Feynman, while Binney's loyalties lie somewhere between Chandrasekhar and Feynman.

Einstein's own views evolved during his lifetime. When he first developed the theory, he put great emphasis on physical ideas and thought experiments. Later on, he was drawn to the geometrical ideas, and spend the second half of his life trying out more geometrically-motivated modifications and extensions of his theory.

You will have to decide for yourself where you belong. But I hope you will enjoy this course.

2. Lorentz invariance

We develop notation for coordinates in four dimensions, and an idea which generalizes rotations to spacetime.

2.1 Spacetime index notation

In order to write about 4D coordinate systems concisely, we use **index notation**. First we say

$$(x^0, x^1, x^2, x^3) \quad \text{means} \quad (ct, x, y, z), \tag{2.1.1}$$

i.e., we make $c \times$ time one of the coordinates. Note that the superscripts here are just labels, not powers! Then, we agree to use Greek indices to denote any component. Thus x^α or x^μ stands for any coordinate—so it's really a way of referring to *all* the coordinates. And we agree to use Roman indices for the spatial components. Thus x^i stands for any of x^1, x^2, x^3, but not x^0. We'll also use boldface notation: $\mathbf{x}$ for the spatial components and $\boldsymbol{x}$ for all four components.

Let us also define a 4D creature with two indices, $\boldsymbol{\eta}$ or $\eta_{\alpha\beta}$, which we may conveniently write as a matrix

$$\boldsymbol{\eta} = \begin{pmatrix} -1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix} \tag{2.1.2}$$

In index notation, the squared distance between infinitesimally close events is

$$ds^2 = \sum_{\alpha\beta} \eta_{\alpha\beta}\, dx^\alpha dx^\beta. \tag{2.1.3}$$

This ds^2 is known as the **interval**. If $ds^2 < 0$ it is called a **timelike** interval, if > 0 a **spacelike** interval, and if $= 0$ **null** or **lightlike**. Related to the interval is the important concept of **proper time**, which equals $d\tau$ where $d\tau^2 = -ds^2$.

Beware differences in convention between books! Some authors use the opposite sign for $\boldsymbol{\eta}$ and ds^2, and sometimes time becomes x^4 rather than x^0.

2.2 Summation convention

The **summation convention** is another device to save writing.

If the index appears twice in a term, once as a superscript and once as a subscript, it is implied to be summed over. No $\sum$ is written. Using the summation convention, our interval is just

$$ds^2 = -d\tau^2 = \eta_{\alpha\beta}\, dx^\alpha dx^\beta. \tag{2.2.1}$$

The summation convention will apply throughout this course, unless we explicitly write 'no sum'. A repeated index in a term is called a **dummy index**, because it can be replaced by a different letter in that term of an equation. A non-dummy or **free index** can also be changed of course, but we have to make the change in the whole equation. We will often need to use index-changing tricks.

If the same index appears twice in a term, it *must* be once as a superscript and once as a subscript, never twice as superscript or subscript; and it may never appear more than twice in a term. Without these prohibitions, expressions could become ambiguous. If you ever really need to write such prohibited things, you can suspend the summation convention with 'no sum'. But you may find you never need to.

A small supplement to the summation convention: if a superscript (subscript) appears in the denominator of a derivative it counts as a subscript (superscript).

Exercise 2.1

Write the following in index notation:

$$d\phi = \frac{\partial\Phi}{\partial x^0}\, dx^0 + \frac{\partial\Phi}{\partial x^1}\, dx^1 + \frac{\partial\Phi}{\partial x^2}\, dx^2 + \frac{\partial\Phi}{\partial x^3}\, dx^3$$

$$a_1 x^1 x^0 + a_2 x^2 x^0 + a_3 x^3 x^0$$

Evaluate

$$\frac{\partial}{\partial x^\mu}\, A_{\alpha\beta} x^\alpha x^\beta$$

where $A_{\alpha\beta}$ is constant.

2.3 Relativistic units

It is a common practice among relativists to measure time in equivalent lengths. It's like measuring distances in light years but the other way round. Thus 29.9792457 cm becomes a way of saying 1 nanosecond.

In such units, the speed of light is just 1, so we can dispense with writing the c's. This saves a lot of writing, but if you want to convert something to (say) SI units, you have to put all the c's back. To do this, you need to replace every time t by ct, and every speed v by v/c.

Some authors use similar tricks to save write the gravitational constant G or Planck's constant $\hbar$, or both. But we won't do that.

2.4 Lorentz transformations

Lorentz invariance means that the interval (2.2.1) when measured in different inertial coordinate systems remains the same. This implies that different inertial coordinate systems must be somehow related.

The general relation between cartesian coordinate systems in spacetime is

$$x'^{\alpha} = \Lambda^{\alpha}{}_{\beta}\, x^{\beta} + a^{\alpha}, \qquad \text{subject to} \qquad \Lambda^{\alpha}{}_{\gamma}\, \eta_{\alpha\beta}\, \Lambda^{\beta}{}_{\delta} = \eta_{\gamma\delta}. \tag{2.4.1}$$

It is easy to verify that (2.4.1) does preserve the interval.

Derivation Just substitute.

$$\eta_{\alpha\beta}\, dx'^{\alpha} dx'^{\beta} = \eta_{\alpha\beta}\, \Lambda^{\alpha}{}_{\gamma}\, dx^{\gamma}\, \Lambda^{\beta}{}_{\delta}\, dx^{\delta} = \eta_{\gamma\delta}\, dx^{\gamma} dx^{\delta}.$$

□

The transformation (2.4.1) is also the most general transformation that preserves the interval. (See e.g., p 27 of Weinberg.) It is called a **Lorentz transformation**.

The expression (2.4.1) is very opaque–what does it mean? Specifically, what does $\Lambda^{\alpha}{}_{\beta}$ mean (since a^{α} is simply a translation in spacetime)? In fact, $\Lambda^{\alpha}{}_{\beta}$ is a kind of rotation in spacetime.

Spatial rotations

$$\begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \cos\varphi & -\sin\varphi \\ & & \sin\varphi & \cos\varphi \end{pmatrix}, \quad \begin{pmatrix} 1 & & & \\ & \cos\varphi & & \sin\varphi \\ & & 1 & \\ & -\sin\varphi & & \cos\varphi \end{pmatrix}, \quad \begin{pmatrix} 1 & & & \\ & \cos\varphi & -\sin\varphi & \\ & \sin\varphi & \cos\varphi & \\ & & & 1 \end{pmatrix}. \tag{2.4.2}$$

are already familiar. The three matrices in (2.4.2) represent rotations by φ about the x, y, z axes respectively.

A rotation-like thing where the time axis is involved is called a **boost** and is composed of

$$\begin{pmatrix} \cosh\varphi & \sinh\varphi & & \\ \sinh\varphi & \cosh\varphi & & \\ & & 1 & \\ & & & 1 \end{pmatrix}, \quad \begin{pmatrix} \cosh\varphi & & \sinh\varphi & \\ & 1 & & \\ \sinh\varphi & & \cosh\varphi & \\ & & & 1 \end{pmatrix}, \quad \begin{pmatrix} \cosh\varphi & & & \sinh\varphi \\ & 1 & & \\ & & 1 & \\ \sinh\varphi & & & \cosh\varphi \end{pmatrix}. \tag{2.4.3}$$

To see what a boost actually does, let's rewrite the last line using

$$v = \tanh\varphi, \qquad \gamma = 1/\sqrt{1-v^2}, \tag{2.4.4}$$

to get

$$\begin{pmatrix} \gamma & \gamma v & & \\ \gamma v & \gamma & & \\ & & 1 & \\ & & & 1 \end{pmatrix}, \quad \begin{pmatrix} \gamma & & \gamma v & \\ & 1 & & \\ \gamma v & & \gamma & \\ & & & 1 \end{pmatrix}, \quad \begin{pmatrix} \gamma & & & \gamma v \\ & 1 & & \\ & & 1 & \\ \gamma v & & & \gamma \end{pmatrix}. \tag{2.4.5}$$

This amounts to moving the coordinate system with velocity v, while keeping the origin the same.

To summarize, a Lorentz transformation is an arbitrary combination of translations, spatial rotations, and boosts.

2.5 Transformation conventions

Different books use different conventions for interpreting coordinate transformations. The following is the convention these notes will follow.

We imagine two sets of spacetime coordinates, "room" coordinates and "trolley" coordinates which may be rotated or moving with respect to room coordinates. The three matrices (2.4.2), which we may denote as

$$\mathbf{R}_x(\varphi), \quad \mathbf{R}_y(\varphi), \quad \mathbf{R}_z(\varphi) \tag{2.5.1}$$

transform from trolley coordinates to room coordinates when the trolley is rotated. The three matrices (2.4.5), which we may denote as

$$\mathbf{B}_x(v), \quad \mathbf{B}_y(v), \quad \mathbf{B}_z(v) \tag{2.5.2}$$

transform from trolley coordinates to room coordinates when the trolley is given a velocity. For the inverse transformations (i.e., going from room to trolley coordinates) we use R and B matrices with minus the argument.

Transformations can be composed: thus $\mathbf{R}_z(\varphi)\,\mathbf{B}_x(v)\,\mathbf{R}_z(-\varphi)$ amounts to rotating the trolley about z by φ, then giving a boost of v along its new x axis, and finally rotating about z by $-\varphi$.

The only new formula you really need to remember is that a boost along x is

$$\boxed{\big(t,x,y,z\big) \xrightarrow[\text{along } x]{\text{trolley moves}} \big(\gamma[t+vx], \gamma[x+vt], y, z\big), \qquad \gamma = 1/\sqrt{1-v^2} \geq 1} \tag{2.5.3}$$

and boosts along y and z have analogous formulas.

Many books have a boost formula that looks like (2.5.3) but with $-v$ instead of v. Such a formula represents transforming from room coordinates to trolley coordinates, the reverse convention to ours. Either convention is fine, but getting them mixed up is lethal.

Exercise 2.2

A laser beam is at angle θ to the x-axis in room coordinates. Write down possible spacetime coordinates for the emission and detection of a photon from this laser beam. Transform these to a trolley boosted by v along x.

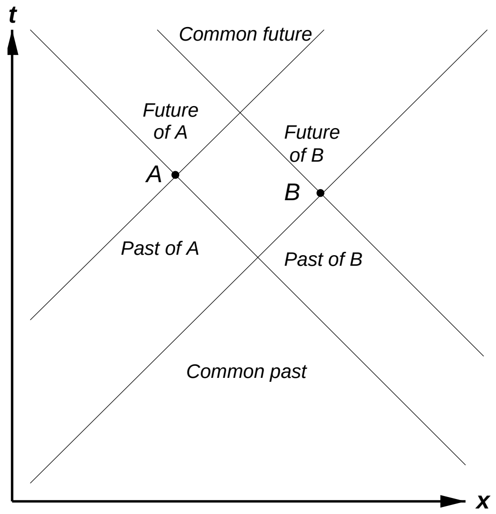

Figure 2.1: Spacetime diagram showing two events and some light paths.

2.6 Spacetime diagrams

Spacetime diagrams are plots of sections of spacetime, and often used for illustrations. They show x^0 as if it were a spatial axis, so one or two of x^i have to be suppressed.

Points on a spacetime diagram are of course events. The spacetime trajectories of particles are called **world lines**; they are parallel to the t axis for stationary objects and inclined for moving objects. Light paths are always at 45°. The region between lights paths from an event is called a **light cone**, and it marks the past and future of an event. Figure 2.1 illustrates.

A boost has the effect of making the coordinate axes oblique on a spacetime diagram. The rotation angle is $\arctan v$.

Figure 2.2 uses a boost to illustrate *why* past and future depend on space as well as time. Here we imagine a little demonstration happening on a moving trolley. Two light pulses set out from the same point on the trolley, they are reflected at the same time from different points on the trolley and return to the initial point. In room coordinates, the emitting and detecting points are not the same, and the reflections are not simultaneous.

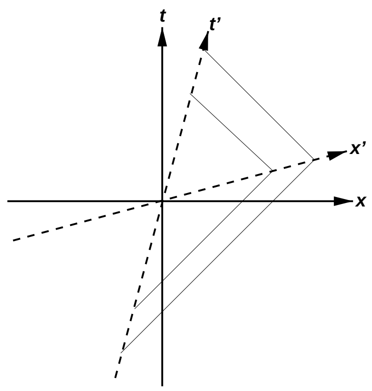

Figure 2.2: Here the solid axes refer to the room. The dashed inclined axes refer to a trolley moving forwards along x. (The inclination of the axes shown here would amount to $v \simeq 0.25$.) Events which are co-spatial or simultaneous in the trolley frame are not so in the room frame.

2.7 Some Lorentzian effects

An immediate consequence of Lorentz transformations is time dilation, or 'moving clocks run slower' or

$$\boxed{dt = \gamma\, d\tau} \tag{2.7.1}$$

Derivation In the fixed-clock frame $d\tau$ is just the time differential, but in the moving-clock frame $d\tau = \sqrt{dt^2 - d\mathbf{x}^2}$. Since $d\tau$ is Lorentz invariant, dt must get bigger, by a factor of $\left(1 - (d\mathbf{x}/dt)^2\right)^{-1/2}$, or γ. □

This presupposes that we know where the clock is and correct for its motion. If we just measure the arrival times of light pulses from a moving clock the answer is different. If a clock is emitting light pluses at intervals of $\Delta\tau$ while moving at speed v at angle θ to our line of sight, we will receive the ticks at intervals of

$$\Delta t = \gamma(1 + v\cos\theta)\Delta\tau. \tag{2.7.2}$$

Derivation In our frame, the light ticks are emitted $\gamma\Delta\tau$ apart. But the distance the light has to travel to get to us has changed by $v\cos\theta \times \gamma\Delta\tau$. □

This is the relativistic Doppler effect.

Another consequence is relativistic length contraction: moving objects appear shortened by γ in the direction of motion.

There is a subtlety involved in the meaning of length. Imagine a fish at rest in room coordinates, lying along the x direction. To measure its length we measure the x coordinates of its head and tail, and it doesn't matter if we make these two measurements at different times. However, to measure the fish's length in the coordinates of a trolley with $\mathbf{B}_x(v)$, we must measure head and tail coordinates at the same trolley-t.

Derivation Say the trolley coordinate (t, x) for tail and head are $(0, 0)$ and $(0, l)$. In room coordinates, the same events would be $(0, 0)$ and $(\gamma vl, \gamma l)$. Thus, a fish-length of γl in the room has become a length of l in the trolley. □

The observation requires two friends at different places on the trolley with synchronized clocks. Their measurements won't be simultaneous in room coordinates.

Finally, velocities don't add in the usual way. If we apply two successive boosts in the same direction, the boost velocities v_1 and v_2 don't add linearly. The boost 'angles' φ_1 and φ_2 *do* add linearly, however. The net velocity is

$$\frac{v_1 + v_2}{1 + v_1 v_2} \tag{2.7.3}$$

Derivation We use the identities

$$\cosh(\varphi_1 + \varphi_2) = \cosh\varphi_1 \cosh\varphi_2 + \sinh\varphi_1 \sinh\varphi_2,$$
$$\sinh(\varphi_1 + \varphi_2) = \sinh\varphi_1 \cosh\varphi_2 + \cosh\varphi_1 \sinh\varphi_2.$$

Multiplying the boost matrices shows that the result is a boost with $\varphi_+\varphi_2$. Working out $\tanh(\varphi_1 + \varphi_2)$ gives the result.

An alternative derivation would be to compose two boosts: $\mathbf{B}_x(v_1)\mathbf{B}_x(v_2)$. □

2.8 The ultimate speed?

The velocity addition formula brings up the well-known consequence of relativity that one can't accelerate things to speeds faster than light.

It is important to be precise about the ultimateness of the speed of light. The speed of light is the fastest *signalling* speed. This means that no particles or anything else that carries information can go faster than light. There is nothing to prevent formally superluminal speeds which don't carry information, e.g., a spotlight moving across a stage.

There is also some literature on 'tachyons', particles which *always* move faster than light. These do not seem forbidden by relativity. But we have no idea about how they might interact with light or ordinary particles and hence carry information.

3. Special Relativity

We introduce the mathematical ideas of vectors and tensors in spacetime, and the physical ideas of relativistic dynamics.

3.1 Contravariant and covariant vectors

A vector is something with magnitude and direction. We can express these properties as components in a coordinate system. On the other hand, we don't want the magnitude and direction of a vector to change whenever we change coordinate system; for that not to happen, the components must change in a suitable way under coordinate transformations.[1]

In fact, vectors are *defined* in terms of their transformation properties. Vectors in relativity are 4-tuples of numbers which transform in one of two ways. **Contravariant vectors** transform like the coordinate differentials,

$$\boxed{v'^{\alpha} = \Lambda^{\alpha}{}_{\beta}\, v^{\beta}} \tag{3.1.1}$$

while **covariant vectors** transform inversely to the coordinate differentials

$$\boxed{w'_{\alpha} = \Lambda_{\alpha}{}^{\beta}\, w_{\beta}} \tag{3.1.2}$$

Here $\Lambda_{\alpha}{}^{\beta}$ is the inverse of $\Lambda^{\alpha}{}_{\beta}$:

$$\boxed{\Lambda^{\alpha}{}_{\beta}\, \Lambda_{\alpha}{}^{\gamma} = \delta^{\gamma}_{\beta}} \tag{3.1.3}$$

where δ^{α}_{γ} is the Kronecker delta.

The archetypical contravariant vector is displacement. The archetypical covariant vector is the gradient of a scalar.

Derivation For a scalar ϕ

$$\frac{\partial \phi}{\partial x'^{\alpha}} = \frac{\partial x^{\beta}}{\partial x'^{\alpha}} \frac{\partial \phi}{\partial x^{\beta}} = \Lambda_{\alpha}{}^{\beta}\, \frac{\partial \phi}{\partial x'^{\beta}} \tag{3.1.4}$$

□

Some books call contravariant vectors just 'vectors' and covariant vectors 'one forms'. For this course, **up vectors** and **down vectors** will do.

[1] They may also change as a function of spacetime location, but that is a separate issue.

3.2 Raising and lowering indices

In relativity, contravariant and covariant vectors are really two different ways of writing the same thing. We can freely change from one to the other using metric. (In non-metric spaces, one can't do that, and contravariant and covariant vectors are different species of things.)

To see how to do this, let us first define $\eta^{\alpha\beta}$ as the inverse of $\eta_{\alpha\beta}$:

$$\eta^{\alpha\gamma}\eta_{\beta\gamma} = \delta^{\alpha}_{\beta} \tag{3.2.1}$$

(As it happens, $\eta^{\alpha\beta}$ is numerically equal to $\eta_{\alpha\beta}$, but writing it with indices up lets us use the summation convention, and also helps prepare for a more general situation later.) We have

$$\Lambda_{\alpha}{}^{\gamma}\,\eta^{\alpha\beta}\,\Lambda_{\beta}{}^{\delta} = \eta^{\gamma\delta} \tag{3.2.2}$$

corresponding to the second part of (2.4.1).

Derivation Easiest to see by considering

$$\Lambda^{\alpha}{}_{\gamma}\,\eta_{\alpha\beta}\,\Lambda^{\beta}{}_{\delta} = \eta_{\gamma\delta}$$

as a matrix equation and inverting. □

We now have

$$\Lambda_{\alpha}{}^{\beta} = \eta_{\alpha\gamma}\eta^{\beta\delta}\,\Lambda^{\gamma}{}_{\delta}\,. \tag{3.2.3}$$

Derivation From the definitions

$$\Lambda_{\alpha}{}^{\gamma}\;\Lambda^{\alpha}{}_{\beta} = \delta^{\gamma}_{\beta} = \eta_{\epsilon\beta}\eta^{\gamma\epsilon} = \eta_{\alpha\delta}\eta^{\gamma\epsilon}\,\Lambda^{\delta}{}_{\epsilon}\,\Lambda^{\alpha}{}_{\beta}\,. \tag{3.2.4}$$

□

We can now use $\eta_{\alpha\beta}$ ($\eta^{\alpha\beta}$) to define a covariant (contravariant) dual for any contravariant (covariant) vector, thus:

$$v_{\alpha} = \eta_{\alpha\beta}v^{\beta}, \qquad w^{\alpha} = \eta^{\alpha\beta}w_{\beta}, \tag{3.2.5}$$

and the dual vectors will satisfy the appropriate transformation properties.

Derivation

$$v'_{\alpha} = \eta_{\alpha\beta}v'^{\beta} = \eta_{\alpha\beta}\,\Lambda^{\beta}{}_{\gamma}\,v^{\gamma} = \eta_{\alpha\beta}\eta^{\gamma\delta}\,\Lambda^{\beta}{}_{\gamma}\,v_{\delta} = \Lambda_{\alpha}{}^{\delta}\,v_{\delta}. \tag{3.2.6}$$

and similarly w_{α} in (3.2.5) is covariant. □

This is known as lowering and raising indices.

3.3 Tensors

A tensor is defined as a quantity which transforms like the product of vectors. Thus $f^{\alpha\beta}$ is a tensor with two up indices if

$$f'^{\gamma\delta} = \Lambda^{\gamma}{}_{\alpha}\, \Lambda^{\delta}{}_{\beta}\, f^{\alpha\beta} \tag{3.3.1}$$

holds.

The **rank** of a tensor is the number of indices. Thus a scalar is a tensor of rank 0, a vector is a tensor of rank 1, and so on.

We have already come across two curious second-rank tensors, $\boldsymbol{\eta}$ and the Kronecker delta, whose components don't change at all. The unchanging components do nevertheless satisfy the appropriate transformation laws.

Derivation From the definitions

$$\begin{aligned}
&\Lambda_{\alpha}{}^{\gamma}\, \Lambda^{\beta}{}_{\delta}\, \delta^{\alpha}_{\beta} = \Lambda_{\alpha}{}^{\gamma}\, \Lambda^{\alpha}{}_{\delta} = \delta'^{\gamma}_{\delta}\\
&\Lambda^{\alpha}{}_{\gamma}\, \Lambda^{\beta}{}_{\delta}\, \eta_{\alpha\beta} = \eta'_{\gamma\delta}\\
&\Lambda_{\alpha}{}^{\gamma}\, \Lambda_{\beta}{}^{\delta}\, \eta^{\alpha\beta} = \eta'^{\gamma\delta}
\end{aligned} \tag{3.3.2}$$

□

For this reason $\boldsymbol{\eta}$ is known as the **metric tensor**—metric because it defines a distance i.e., the interval.

Tensors can be produced from other tensors in several ways.

(a) A linear combination of tensors with the same indices is a tensor. This is obvious from the transformation laws.

(b) A **direct product** of two tensors (meaning expressions of the type $v_{\alpha} f^{\mu\nu}$) is a tensor, as is again obvious from the transformation laws.

(c) **Contraction**, or setting an upper and a lower index equal and then summing, produces a tensor with rank reduced by two.

Derivation Consider

$$f'^{\alpha}_{\ \beta} = \Lambda^{\alpha}{}_{\gamma}\, \Lambda_{\beta}{}^{\delta}\, f^{\gamma}{}_{\delta}. \tag{3.3.3}$$

Contracting gives

$$f'^{\alpha}_{\ \alpha} = \Lambda^{\alpha}{}_{\gamma}\, \Lambda_{\alpha}{}^{\delta}\, f^{\gamma}{}_{\delta} = \delta^{\delta}_{\gamma} f^{\gamma}{}_{\delta} = f^{\gamma}{}_{\gamma}.$$

The argument immediately generalizes to arbitrary rank. □

A direct product combined with a contraction is called an **inner product**. If the inner product of something and a tensor is a tensor, then that something is also a tensor.

Derivation Suppose we are given that

$$f'^{\alpha\beta} v'_\beta = \Lambda^\alpha{}_\beta \, f^{\beta\gamma} v_\gamma \tag{3.3.4}$$

and that v_γ is a tensor, but we aren't told about $f'^{\alpha\beta}$.

Define $\tilde{f}^{\alpha\beta}$ to equal $f^{\alpha\beta}$ in *one* coordinate system and transform it like

$$\tilde{f}'^{\alpha\beta} = \Lambda^\alpha{}_\gamma \, \Lambda^\beta{}_\delta \, f^{\gamma\delta}. \tag{3.3.5}$$

Then

$$\left(\tilde{f}'^{\alpha\beta} - f'^{\alpha\beta} \right) v'_\beta = \Lambda^\alpha{}_\gamma \, \Lambda^\beta{}_\delta \, f^{\gamma\delta} v'_\beta - \Lambda^\alpha{}_\gamma \, f^{\gamma\delta} v_\delta. \tag{3.3.6}$$

Substituting $v'_\beta = \Lambda_\beta{}^\epsilon \, v_\epsilon$ makes the right hand side zero. And since the components of v'_β can be varied arbitrarily by changing coordinate system, the term in brackets must also be zero. Hence $\tilde{f}'^{\alpha\beta} = f'^{\alpha\beta}$.

Again, this argument generalizes to arbitrary rank. □

This is called the **quotient rule**.

(d) Differentiating a tensor with respect to the coordinates increases the rank by one covariant index. The derivation (3.1.4) for the gradient of a scalar generalizes.

But not everything with indices on it is a tensor. In particular, $\Lambda^\alpha{}_\beta$ and $\Lambda_\alpha{}^\beta$ cannot possibly be tensors; they don't even have components in any one coordinate system, since they relate different coordinate systems.

3.4 Gradient, divergence, curl, Laplacian

We adopt the shorthand

$$\begin{aligned} \phi_{,\alpha} &\equiv \frac{\partial \phi}{\partial x^\alpha} \\ \phi^{,\alpha} &\equiv \eta^{\alpha\beta} \, \frac{\partial \phi}{\partial x^\beta} \end{aligned} \tag{3.4.1}$$

and so on with any tensor in place of ϕ.

In this notation $\phi_{,\alpha}$ denotes the gradient of ϕ. We can also take the gradient of a vector or any tensor.

Divergence are of the type $v^\alpha{}_{,\alpha}$. We can take the divergence of any vector or higher rank tensor.

Curl has a non-obvious generalization. The curl of v_α is defined as

$$v_{\alpha,\beta} - v_{\beta,\alpha} \tag{3.4.2}$$

i.e., as not a vector but a second-rank antisymmetric tensor. In three dimensions (but only in three dimensions) it is possible to define a vector which carries exactly the same information as a second-rank antisymmetric tensor, and thus defining curl of a vector as a vector works. In general we can take curls of higher rank tensors too.

The Laplacian is a gradient contracted with a divergence. Thus

$$\nabla^2 \phi \equiv \phi^{,\alpha}{}_{,\alpha} \tag{3.4.3}$$

Exercise 3.1

Evaluate the expressions

$$x^{\alpha}{}_{,\beta}\left(=\frac{\partial x^{\alpha}}{\partial x^{\beta}}\right) \quad \text{and} \quad x^{\alpha}{}_{,\alpha}$$

and show that

$$\begin{aligned} &\operatorname{curl}(\phi x^{\alpha}) = x^{\alpha}\phi_{,\beta} - \eta_{\beta\gamma}x^{\gamma}\phi^{,\alpha} \\ &f^{\alpha\beta}{}_{,\beta} = f^{\alpha}{}_{\beta}{}^{,\beta} \end{aligned}$$

3.5 Four-momentum

Consider a particle of mass m and define

$$\boxed{p^{\alpha} = m\,\frac{dx^{\alpha}}{d\tau}} \tag{3.5.1}$$

which clearly *is* a vector because τ is invariant.

Consider now

$$f^{\alpha} = \frac{dp^{\alpha}}{d\tau} \tag{3.5.2}$$

which is also a vector. Now, in the instantaneous rest frame of the particle $dt = d\tau$; hence $f^0 = 0$ and f^i is the Newtonian force. Thus (3.5.2) is the Lorentz-invariant generalization of Newtonian dynamics, and p^{α} is a vector generalizing 3D momentum to spacetime: it is called the **four-momentum**. In the absence of external forces, four-momentum is conserved.

The four-momentum is fundamental in relativistic dynamics. In four-dimensional notation it looks simple, but when we try to interpret in 3D and time, weird and wonderful things happen. Let us examine the space and time parts of p^{α} separately. We have

$$\begin{aligned} p^0 &= m\gamma = m + \tfrac{1}{2}mv^2 + O(v^4) \\ p^i &= m\gamma v^i = mv^i + O(v^2) \end{aligned} \tag{3.5.3}$$

Compared with Newtonian dynamics the mass seems to be enhanced by a factor of γ. Also, since in Newtonian dynamics, $\frac{1}{2}mv^2$ is the kinetic energy, mass seems to increase in amount by the kinetic energy. This is nothing but the statement of $E = mc^2$. It suggests that mass and energy are mutually convertible, but does not prove it; convertibility is an extra physical postulate. We will call p^0 the **mass-energy**.

Since p^{α} is a vector, $\sqrt{-\eta_{\alpha\beta}p^{\alpha}p^{\beta}}$ is a scalar. For a single particle it equals the mass m.[2] The p^{α} for a system of particles is the sum of their individual four momenta; the corresponding scalar is called the total mass which may be different from the sum of individual masses.

[2] Some books refer to $m\gamma$ as 'relativistic mass' or even just 'mass'. We won't do this because we'd then need a new name, such as 'rest mass' for m. We'll reserve 'mass' for the scalar m.

3.6 Force in 4 and 3 dimensions

Although relativistic dynamics calculations are usually simplest in 4D notation, it is also useful to have transformation formulas for forces in 3D notation.

The relativistic form of Newton's second law, equation (3.5.2), in full, is

$$(f^0, f^x, f^y, f^z) = m\,\frac{d}{d\tau}\,(\gamma, \gamma v_x, \gamma v_y, \gamma v_z). \tag{3.6.1}$$

In the particle's instantaneous rest frame, we have

$$(0, F_x, F_y, F_z) = m\,\frac{d}{d\tau}\,(1, v_x, v_y, v_z), \tag{3.6.2}$$

where **F** is the Newtonian force. Now **F** is a 3D vector and does *not* transform like part of a 4D vector under boosts; however f^i, which happens to equal F_i in the particle's instantaneous frame, is part of a vector. So we can transform (3.6.2) out of its special frame and into an arbitrary inertial frame; applying a Lorentz transformation and using $dt = \gamma d\tau$, we get

$$(vF_x, F_x, F_y/\gamma, F_z/\gamma) = m\,\frac{d}{dt}\,(\gamma, \gamma v_x, \gamma v_y, \gamma v_z). \tag{3.6.3}$$

The space part of the right hand side here is a rate of change of momentum, so we interpret the space part of the left hand side as the 3D force. Thus we have derived how ordinary 3D force changes under a Lorentz transformation:

$$\boxed{(F_x, F_y, F_z) \xrightarrow[\text{along } x]{\text{particle moves}} (F_x, F_y/\gamma, F_z/\gamma)} \tag{3.6.4}$$

We see from (3.6.4) that a moving particle feels less force transverse to its direction of motion. One consequence is that acceleration need not be parallel to force, a very non-Newtonian feature of relativistiv dynamics. Meanwhile the time component of (3.6.3) expresses the rate of change of the particle's mass-energy, another completely non-Newtonian concept.

An odd feature of equation (3.6.3) is that it gives derivatives of v^i, and also the derivative of γ which of course isn't independent. Fortunately, the derivative of γ is the same value as implied by the derivatives of v^i.

Derivation

$$\begin{aligned} \frac{f_i}{m} &= \frac{d}{d\tau}\,(\gamma v_i) = \gamma\,\frac{dv_i}{d\tau} + v_i\,\frac{d\gamma}{d\tau} = \gamma\,\frac{dv_i}{d\tau} + \gamma^3 v_j\,\frac{dv_j}{d\tau}\,v_i; \\ \frac{v_i f_i}{m} &= \gamma v_i\,\frac{dv_i}{d\tau}\left(1 + \gamma^2 v_j v_j\right) = \gamma^3 v_i\,\frac{dv_i}{d\tau} = \frac{d\gamma}{d\tau}\,, \end{aligned} \tag{3.6.5}$$

where we have used the identity

$$\frac{d\gamma}{d\tau} = \gamma^3 v_i\,\frac{dv_i}{d\tau}\,. \tag{3.6.6}$$

(We have used the summation convention for both-down Roman indices here.) □

3.7 Energy-momentum tensor

The four-momentum lets us describe the dynamics of particles or systems of particles. But if the system we are interested in has a very large number of particles, a description in terms of individual particles becomes too complicated to be useful. We then think of the system as a fluid, and describe the dynamics of infinitesimal elements of fluid. The relevant dynamical quantity is the fluid density $\rho(\boldsymbol{x})$, and the dynamics is described through an **energy-momentum tensor**.

The simplest kind of fluid consists of particles with no random motions; relativists call it dust. In this case the energy momentum tensor is defined as

$$T^{\alpha\beta} = \rho(\boldsymbol{x}) \, \frac{dx^\alpha}{d\tau} \, \frac{dx^\beta}{d\tau} \,. \tag{3.7.1}$$

With all components displayed, (3.7.1) becomes

$$T^{\alpha\beta} = \gamma^2 \rho(\boldsymbol{x}) \begin{pmatrix} 1 & v_x & v_y & v_z \\ v_x & v_x^2 & v_x v_y & v_x v_z \\ v_y & v_y v_x & v_y^2 & v_y v_z \\ v_z & v_z v_x & v_z v_y & v_z^2 \end{pmatrix} . \tag{3.7.2}$$

In the instantaneous rest frame of the fluid element the only nonzero component is $T^{00} = \rho$. In general $T^{\alpha\beta}$ is the flux of the α-component of momentum through a surface of constant x^β. Hence the name.

The divergence of the energy-momentum tensor is the density of external (relativistic) force. To see this, we expand $T^{\alpha\beta}{}_{,\beta}$ as

$$\begin{aligned} T^{0\beta}{}_{,\beta} &= \gamma^2 \left(\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x^j} \left(\rho v^j \right) \right) , \\ T^{i\beta}{}_{,\beta} &= v^i T^{0\beta}{}_{,\beta} + \gamma^2 \rho \left(\frac{\partial v^i}{\partial t} + v^j \, \frac{\partial v^i}{\partial x^j} \right) . \end{aligned} \tag{3.7.3}$$

Derivation To get the second line, we write

$$T^{i\beta}{}_{,\beta} = \gamma^2 \left(\frac{\partial}{\partial t} \left(\rho v^i \right) + \frac{\partial}{\partial x^j} \left(\rho v^i v^j \right) \right)$$

and then expand and regroup. □

The upper line in (3.7.3) is zero by mass conservation (i.e., the continuity equation). To interpret the second line in (3.7.3), we can go to the instantaneous rest frame of the fluid element, where the expression is simply $\rho \, dv^i/dt$, i.e., the force density; in the absence of external forces it is zero.

Thus the equations of motion for a fluid are

$$\boxed{T^{\alpha\beta}{}_{,\beta} = 0} \tag{3.7.4}$$

In the presence of external forces, there will be additional terms. We could put new force terms in (3.7.4). But it turns out that new forces are always divergences. For these reason, it is the convention to incorporate new forces by modifying the definition (3.7.1) of the energy-momentum tensor, leaving (3.7.4) unchanged.

The next simplest case is called the perfect fluid. In this the particles in any fluid element have random motions, but inside a fluid element there is no preferred direction for the random velocities. To see how $T^{\alpha\beta}$ will change, we go to the instantaneous rest frame of the fluid element, and consider it as made up of many sub-elements with random velocities with no preferred direction but averaging to zero. Averaging over the random velocites, we will get

$$T^{\alpha\beta}{}_{\text{inst rest fr}} = \rho(\boldsymbol{x}) \begin{pmatrix} 1 & & & \\ & 0 & & \\ & & 0 & \\ & & & 0 \end{pmatrix} + p(\boldsymbol{x}) \begin{pmatrix} 0 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix} \tag{3.7.5}$$

where $p(\boldsymbol{x})$ is the mean square velocity in any direction. The Lorentz invariant generalization of (3.7.5) is

$$T^{\alpha\beta} = \rho(\boldsymbol{x}) \frac{dx^\alpha}{d\tau} \frac{dx^\beta}{d\tau} + p(\boldsymbol{x}) \left(\frac{dx^\alpha}{d\tau} \frac{dx^\beta}{d\tau} + \eta^{\alpha\beta} \right). \tag{3.7.6}$$

and the p-dependent terms in full are

$$p(\boldsymbol{x}) \begin{pmatrix} 0 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix} + \gamma^2 p(\boldsymbol{x}) \begin{pmatrix} v^2 & v_x & v_y & v_z \\ v_x & v_x^2 & v_x v_y & v_x v_z \\ v_y & v_y v_x & v_y^2 & v_y v_z \\ v_z & v_z v_x & v_z v_y & v_z^2 \end{pmatrix} \tag{3.7.7}$$

as expected. The spatial part of the divergence of (3.7.7) is $\nabla p + O(v)$. The interpretation of $p(\boldsymbol{x})$ is fluid pressure.

4. Cartesian Tensors in 3D

In which[1] we express vector and tensor analysis for ordinary three dimensions in index notation, assuming cartesian coordinates. Since the metric is identity and there is no distinction between contravariant and covariant, we will write all indices down. The summation convention applies with down indices.

4.1 Cartesian coordinate transformations

The general form for cartesian coordinate transformations is the spatial part of the Lorentz transformation (2.4.1), thus

$$x'_i = \Lambda_{ij} x_j + a_i, \qquad \text{subject to} \qquad \Lambda_{ik}\delta_{kl}\Lambda_{lj} = \delta_{ij}. \tag{4.1.1}$$

In this case Λ_{ij} comprises rotations and inversions.

Vectors and tensors are defined as before, only now there is no difference between the definitions of contravariant and covariant.

The dot product, gradient, and divergence from ordinary vector analysis are easy to write down in index notation. A dot product looks like $u_i v_i$, a gradient looks like $\partial_i \phi$, and a divergence like $\partial_i u_i$. Here we are using the symbol ∂_i as the index form for ∇, and the convention that in a product it applies only to the next expression. (The comma notation can also be used.)

The cross product and curl need some more notation.

4.2 The permutation tensor

Consider ϵ_{ijk} defined as

$$\begin{cases} \epsilon_{123} = \epsilon_{231} = \epsilon_{312} = 1 \\ \epsilon_{321} = \epsilon_{132} = \epsilon_{213} = -1 \\ \text{all others} = 0. \end{cases} \tag{4.2.1}$$

In other words, $\epsilon_{123} = 1$ and ϵ_{ijk} is antisymmetric in any pair of indices. It is called the **permutation tensor** or the Levi-Civita tensor. As the name suggests, it is a tensor of rank 3.

Derivation Consider

$$\tilde{\epsilon}_{ijk} = \Lambda_{ip}\Lambda_{jq}\Lambda_{kr}\,\epsilon_{pqr}\,. \tag{4.2.2}$$

From the form of (4.2.2), $\tilde{\epsilon}_{ijk}$ is antisymmetric in any pair of indices. Furthermore $\tilde{\epsilon}_{123} = \det|\Lambda|$. Since (4.1.1) implies that $\det|\Lambda| = \pm 1$, we have $\tilde{\epsilon}_{ijk} = \pm\,\epsilon_{ijk}$. Thus ϵ_{ijk} as defined in (4.2.1) transforms like a 3rd rank tensor, except for the factor of ± 1. □

[1] This chapter is for background and/or revision only, and is not examinable.

Because it changes sign under inversion, ϵ_{ijk} is sometimes called a **pseudotensor**.

The permutation tensor satisfies a very important identity:

$$\boxed{\epsilon_{rmn}\,\epsilon_{rpq} = \delta_{mp}\delta_{nq} - \delta_{mq}\delta_{np}} \tag{4.2.3}$$

Derivation We prove (4.2.3) by verifying the possible cases.

(i) If all of $1, 2, 3$ are included in m, n, p, q then both sides give 0. The LHS gives 0 because r cannot be different from all of m, n, p, q. The RHS gives 0 because we cannot have $m = p, n = q$ and we cannot have $n = q, m = p$.

(ii) If $m = n$ or $p = q$ then both sides give 0. If only one of these is true, both terms on the RHS are 0; if both are true, the terms on the RHS cancel.

(iii) If $m = p \neq n = q$ then both sides give $+1$. The ϵ terms have the same sign.

(iv) If $m = q \neq n = p$ then both sides give -1. □

4.3 The cross product, and curl

We can now write the cross product $\mathbf{w} = \mathbf{u}\times\mathbf{v}$ and the curl $\nabla\times\mathbf{u}$ respectively as

$$w_i = \epsilon_{ijk}\, u_j v_k, \qquad \epsilon_{ijk}\,\partial_j u_k. \tag{4.3.1}$$

These two definitions make use of the fact that in 3D we can associate a vector v_i with any second rank antisymmetric tensor a_{ij}:

$$v_i = \tfrac{1}{2}\,\epsilon_{ijk}\, a_{jk}, \qquad a_{ij} = \epsilon_{ijk}\, v_k. \tag{4.3.2}$$

Derivation We verify that the second part of (4.3.2) follows from the first.

$$a_{ij} = \tfrac{1}{2}\,\epsilon_{ijk}\,\epsilon_{klm}\, a_{lm} = \tfrac{1}{2}(\delta_{il}\delta_{jm} - \delta_{jl}\delta_{im})a_{lm} = \tfrac{1}{2}(a_{ij} - a_{ji}) = a_{ij}.$$

□

A permutation tensor can be defined in any number of dimensions; the rank equals the number of dimensions. But it is most useful in three dimensions, for the above reason.

Exercise 4.1

Express and derive the following identities in index notation.

$$\nabla\cdot(\phi\mathbf{u}) = \phi\nabla\cdot\mathbf{u} + \nabla\phi\cdot\mathbf{u}$$
$$\nabla\times(\phi\mathbf{u}) = \phi\nabla\times\mathbf{u} + \nabla\phi\times\mathbf{u}$$
$$\nabla\cdot(\mathbf{u}\times\mathbf{v}) = \mathbf{v}\cdot\nabla\times\mathbf{u} - \mathbf{u}\cdot\nabla\times\mathbf{v}$$
$$\nabla\times(\mathbf{u}\times\mathbf{v}) = \mathbf{u}\nabla\cdot\mathbf{v} - \mathbf{v}\nabla\cdot\mathbf{u} + \mathbf{v}\cdot\nabla\mathbf{u} - \mathbf{u}\cdot\nabla\mathbf{v}$$
$$\nabla\times\nabla\times\mathbf{u} = \nabla\nabla\cdot\mathbf{u} - \nabla\cdot\nabla\mathbf{u}$$
$$\nabla(\mathbf{u}\cdot\mathbf{v}) = \mathbf{u}\times\nabla\times\mathbf{v} + \mathbf{v}\times\nabla\times\mathbf{u} + \mathbf{u}\cdot\nabla\mathbf{v} + \mathbf{v}\cdot\nabla\mathbf{u}$$

5. Electromagnetism

In this part we will study electromagnetism as an example of a relativistic theory. We won't really go into applications of the theory at all (that could easily fill a degree in electrical engineering), but we will develop the basic equations of electromagnetism in relativistic notation, and introduce some important ideas that we will meet again when we come to gravity in general relativity.[1]

5.1 Fields

Electromagnetism is all about how electrically charged particles interact. This interaction does not happen directly—that would imply instantaneous communication between charges—it is mediated by a **field**, which carries information at the speed of light. In one sense fields are an abstraction invented to describe the interaction of particles. On the other hand, fields can do such complex things that they can seem even more important than the particles. People working with them speak of "the world where the electric field is not a symbol merely, but something that crackles".

The equations of electromagnetism come in two sets. First, there are **Maxwell's equations** giving the electromagnetic field in terms of the charges. In four-dimensional notation, the electromagnetic field is a second-rank antisymmetric tensor, usually written $F^{\alpha\beta}$; in three dimensional notation, this can be split into two parts: an electric field $\mathbf{E}$ and a magnetic field $\mathbf{B}$. Then there is the equation for the **Lorentz force**, which gives the effect of the field on charges.

5.2 A thought experiment

Suppose we have some charges, at rest, producing an electric field. In particular, a charge q_1 at $(0,0,0)$ produces a field[2]

$$\mathbf{E} = q_1 \frac{\hat{\mathbf{r}}}{4\pi r^2}. \tag{5.2.1}$$

Now we introduce an extra particle ("the particle") with charge q and mass so small it has negligible effect on the field or the other charges. The particle will feel a force

$$q\mathbf{E}. \tag{5.2.2}$$

So far so good, and we could go on and develop electrostatics, but we won't do that. Instead, we ask: what is the force if the particle is moving? This question can only be answered through experiment, and the answer is that (5.2.2) *always* gives the force on a charge from an electric field—in the "lab" frame where the field is expressed, not the particle's rest frame—regardless of how the particle is moving, and regardless of what charges caused the electric field.

That (5.2.2) is general tells us that (5.2.1) is *not* general; if the charge is moving, the field is different. If the particle is moving along x with some γ, the force in the

[1] The two sections on Thought Experiments and the section on Dipole Radiation are not examinable.

[2] Equation (5.2.1) also serves to define the units we will use.

particle's rest frame (which we can get by reversing equation 3.6.4) must be the same as the force if the particle is stationary and the charge is moving along $-x$. Thus we get

$$(E_x, E_y, E_z) \xrightarrow[\text{along } x]{\text{charge moves}} (E_x, \gamma E_y, \gamma E_z). \tag{5.2.3}$$

Now suppose both the particle and charge are moving along x with the same velocity. Using (5.2.3) and then (3.6.4) in reverse, we get $(F_x, \gamma^2 F_y, \gamma^2 F_z)$ for the force in the particle frame—but this can't be right because in this frame both the charge and the particle are at rest! Have we shown that (5.2.2) is inconsistent with Lorentz invariance?

Not necessarily: it is possible that a new effect appears that counteracts the electric field when both charge and particle are moving. In fact, given the experimental generality of (5.2.2) we are forced to predict a new field, and you can guess what that field is.

Exercise 5.1

Suppose that

- (5.2.2) applied in the particle's instantaneous rest frame, but not generally; and
- the electric field transformed like

$$(E_x, E_y, E_z) \xrightarrow[\text{along } x]{\text{charge moves}} (aE_x, bE_y, bE_z).$$

Now do the following:

(i) move the charge along $+x$ with some γ, and write down how the force changes on a stationary charge; then

(ii) move both charge and particle along $-x$ with the same γ, thus deriving the force on a moving particle from an electric field; and then

(iii) let both charge and particle move along x with the same γ and work out the force in the particle frame.

Are there any choices of a, b which don't need an extra field for consistency?

5.3 Another thought experiment

Consider a charged wire along the x axis, with charge λ per unit length. At a point $(0, r, 0)$ the electric field will be

$$E_y = \frac{\lambda}{2\pi r}. \tag{5.3.1}$$

Derivation

$$E_y = \frac{\lambda}{4\pi} \int_{-\infty}^{\infty} \frac{r}{(r^2 + x^2)^{3/2}} \, dx = \frac{\lambda}{2\pi r} \tag{5.3.2}$$

□

Next let the wire move along x with some γ. In accordance with (5.2.3), the new field will be

$$E_y = \frac{\gamma\lambda}{2\pi r}. \tag{5.3.3}$$

Another interpretation is that length contraction has increased the effective λ by a factor of γ.

Now we make things interesting. Take two wires, coaxial with x; one has charge λ per unit length and moves along $+x$ with (v_0, γ_0), and the other has opposite charge and moves with the same speed in the opposite direction. So there is charge transport but no net charge.

But suppose the particle is moving along $+x$ with (v, γ). In the particle frame, the wires will have velocities of (using the velocity addition formula 2.7.3)

$$v_+ = \frac{v_0 - v}{1 - v_0 v}, \quad v_- = \frac{v_0 + v}{1 + v_0 v}, \tag{5.3.4}$$

and corresponding γ-factors γ_+ and γ_-, and charge densities of

$$\lambda_+ = \gamma_+(\lambda/\gamma_0), \quad \lambda_- = \gamma_-(\lambda/\gamma_0). \tag{5.3.5}$$

We have

$$\gamma_+ - \gamma_- = -2v\gamma v_0\gamma_0. \tag{5.3.6}$$

Derivation Just insert (5.3.4) and simplify. □

Hence the total charge per unit length in the particle frame is $-2v\gamma\lambda v_0$ and the force $-qv\gamma\lambda v_0/(\pi r)$. In the no-field frame, the force becomes $-qv\lambda v_0/(\pi r)$. And defining the current I as the rate of charge transport or $2\lambda v_0$, we have for the force:

$$F_y = -q\frac{vI}{2\pi r}. \tag{5.3.7}$$

Notice that (5.3.7) does not depend on the separate velocities of the wires, only on the total rate of charge transport.

5.4 The equations of electromagnetism

Electromagnetism can be written very concisely in relativistic notation. The ingredients are (i) the source vector J^α, i.e., electric charges and their motions, (ii) A ghostly vector A^α called the **potential** which arises from the source, (iii) the field tensor, which is the curl of the potential, and (iv) the relativistic force density f^α of the field on other sources. The equations are

$$\boxed{\begin{aligned} &F^{\alpha\beta}{}_{,\beta} = J^\alpha \quad \text{where } F_{\alpha\beta} = -(A_{\alpha,\beta} - A_{\beta,\alpha}) \quad \text{(Maxwell's equations)} \\ &F^{\alpha\beta} J_\beta = f^\alpha \quad \text{(Lorentz force)} \end{aligned}} \tag{5.4.1}$$

and everything in electromagnetism follows.

Now let us work out what (5.4.1) means!

There are four ingredients.

First the source: J^α, which is the four-dimensional velocity vector associated with the electric charge density ρ_e

$$\boxed{J^\alpha = \rho_e \frac{dx^\alpha}{d\tau}} = (\rho_e, J_x, J_y, J_z). \tag{5.4.2}$$

The spatial part of J^α is the 3D current density $\mathbf{J}$.

Next we have the potential: from (5.4.1) we see that J^α gives A^α via a linear second-order differential equation. We will return to that equation later. The common names for A^0 and A^i are the **electric potential** Φ and the **magnetic vector potential** $\mathbf{A}$:

$$A^\mu = (\Phi, A_x, A_y, A_z), \quad A_\mu = (-\Phi, A_x, A_y, A_z). \tag{5.4.3}$$

Third is the field tensor $F^{\alpha\beta}$. Since $F^{\alpha\beta}$ is antisymmetric, we can associate it with two 3D vectors, say $\mathbf{E}$ and $\mathbf{B}$, thus:

$$F^{\alpha\beta} = \begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & B_z & -B_y \\ -E_y & -B_z & 0 & B_x \\ -E_z & B_y & -B_x & 0 \end{pmatrix}, \quad F_{\alpha\beta} = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & B_z & -B_y \\ E_y & -B_z & 0 & B_x \\ E_z & B_y & -B_x & 0 \end{pmatrix}. \tag{5.4.4}$$

Then (5.4.1) implies

$$\mathbf{E} = -\nabla\Phi - \frac{\partial \mathbf{A}}{\partial t}, \quad \mathbf{B} = \nabla\times\mathbf{A}, \tag{5.4.5}$$

and

$$\nabla\cdot\mathbf{E} = \rho_e, \quad \nabla\times\mathbf{B} - \frac{\partial \mathbf{E}}{\partial t} = \mathbf{J}. \tag{5.4.6}$$

Derivation Inserting (5.4.3) and (5.4.4) into the the upper right equation in (5.4.1) gives

$$F_{\alpha\beta} = \begin{pmatrix} 0 & A_{x,t} + \Phi_{,x} & A_{y,t} + \Phi_{,y} & A_{z,t} + \Phi_{,z} \\ & 0 & A_{y,x} - A_{x,y} & A_{z,x} - A_{x,z} \\ & & 0 & A_{z,y} - A_{y,z} \\ & & & 0 \end{pmatrix} \tag{5.4.7}$$

with the lower triangle implied by antisymmetry, and this is the same as (5.4.5).

Inserting (5.4.4) into the upper left equation in (5.4.1) gives

$$\begin{pmatrix} E_{x,x} + E_{y,y} + E_{z,z} \\ B_{z,y} - B_{y,z} - E_{x,t} \\ B_{x,z} - B_{z,x} - E_{y,t} \\ B_{y,x} - B_{x,y} - E_{z,t} \end{pmatrix} = \begin{pmatrix} \rho_e \\ J_x \\ J_y \\ J_z \end{pmatrix}$$

which is the same as (5.4.6). □

And fourth, we have the Lorentz force

$$\boxed{\mathbf{f} = \rho_e \mathbf{E} + \mathbf{J} \times \mathbf{B}} \qquad f^0 = \mathbf{J} \cdot \mathbf{E}. \tag{5.4.8}$$

As we have already seen, the magnetic field acts in a rather odd way. To feel it, (i) a charge has to be moving, and (ii) the force is always *perpendicular* to the motion of the charge; (ii) means—and the fact that f^0 doesn't depend on $\mathbf{B}$ means the same thing—that a magnetic field never does any work!

5.5 Gauge invariance

I implied in the previous section that the potential A^α is strange even by the standards of relativity. That's because it can't be measured! Only its curl can. We see from $F_{\alpha\beta} = -(A_{\alpha,\beta} - A_{\beta,\alpha})$ that adding an arbitrary gradient to the potential thus,

$$A_\alpha \longrightarrow A_\alpha + \phi_{,\alpha} \tag{5.5.1}$$

This type of non-uniqueness is known as **gauge invariance**, and transformations of the type (5.5.1) are called **gauge transformations**. We will meet gauges again when we consider the field equations for gravity.

5.6 The retarded potential

We can use the gauge freedom (5.5.1) to our advantage. In particular we can choose a ϕ so that $A^\alpha{}_{,\alpha} = 0$. This is called the **Lorentz gauge**. In this case, Maxwell's equations simplify to

$$F_{\alpha\beta} = -(A_{\alpha,\beta} - A_{\beta,\alpha}) \quad \text{where } \nabla^2 A^\alpha = -J^\alpha. \tag{5.6.1}$$

The physically admissible solution to the wave equation (5.6.1) is

$$A^\alpha(t, \mathbf{x}) = \frac{1}{4\pi} \int \frac{J^\alpha(t - r, \mathbf{x}')}{r} \, d^3\mathbf{x}', \quad r \equiv |\mathbf{x} - \mathbf{x}'|. \tag{5.6.2}$$

It is called the **retarded potential** and it is has the same form as the potential in electrostatics, only retarded in time. Despite its appearance (5.6.2) is Lorentz-invariant. We won't derive it here. (For derivation, see p. 19 of Binney, or any advanced electromagnetism text such as *Classical Electrodynamics* by J.D. Jackson.).

From the form (5.6.2) we can easily deduce several simple facts about electromagnetism. For example: a current-carrying wire will have the **A** parallel to it and **B** circulating around it. A solenoid (made up by stacking loops of wire) has **B** parallel to it. So does a natural magnet like a compass needle. Two parallel wires will attract if they carry parallel currents and repel if they carry anti-parallel currents. (We have already derived this last fact from completely different reasoning!)

5.7 Dipole radiation

We might conclude from (5.6.2) that since at large distances the potentials fall off like $1/r$, the fields would fall off $1/r^2$. This is true for static charges and currents, but for time varying charges and currents, there is an important new effect.

For a localized J^α, viewed from large r, r is almost independent of $\mathbf{x}'$ in (5.6.2). So we may write for the magnetic potential

$$\mathbf{A} = \frac{1}{4\pi r} \int \mathbf{J}(t - r, \mathbf{x}')\, d^3\mathbf{x}'. \tag{5.7.1}$$

But the integral in (5.7.1) equals the rate of change of the **dipole moment**, i.e.,

$$\int \mathbf{J}(\mathbf{x})\, d^3\mathbf{x} = \frac{d\mathbf{p}}{dt}, \qquad \text{where } \mathbf{p} = \int \rho_e(\mathbf{x})\, d^3\mathbf{x}. \tag{5.7.2}$$

Derivation

$$\frac{d\mathbf{p}}{dt} = \int \frac{\partial \rho_e}{\partial t}\, \mathbf{x}\, d^3\mathbf{x} = -\int (\nabla\cdot\mathbf{J})\, \mathbf{x}\, d^3\mathbf{x} = \int (\mathbf{J}\cdot\nabla)\, \mathbf{x}\, d^3\mathbf{x} = \int \mathbf{J}\, d^3\mathbf{x}. \tag{5.7.3}$$

□

Thus

$$\mathbf{A}(t, \mathbf{x}) = \frac{1}{4\pi r} \left(\frac{d\mathbf{p}}{dt} \right)_{(t-r)} \tag{5.7.4}$$

and so if the dipole moment is time varying the fields fall off as $1/r$.

5.8 The field energy-momentum tensor

The tensor

$$T^{\alpha\beta} = F^{\alpha}{}_{\gamma} F^{\beta\gamma} - \tfrac{1}{4} \eta^{\alpha\beta} F_{\gamma\delta} F^{\gamma\delta} \tag{5.8.1}$$

has divergence equal to minus the force density:

$$T^{\alpha\beta}{}_{,\beta} = -f^{\alpha} \tag{5.8.2}$$

Derivation

$$T^{\alpha\beta}{}_{,\beta} = F^{\alpha}{}_{\gamma} F^{\beta\gamma}{}_{,\beta} + F^{\beta\gamma} F^{\alpha}{}_{\gamma,\beta} - \tfrac{1}{2} F_{\gamma\delta} F^{\gamma\delta,\alpha} \tag{5.8.3}$$

We can rearrange the last two terms a little:

$$\begin{aligned} F^{\beta\gamma} F^{\alpha}{}_{\gamma,\beta} &= F_{\beta\gamma} F^{\alpha\gamma,\beta} = -F_{\beta\gamma} F^{\gamma\alpha,\beta} = -F_{\beta\gamma} F^{\alpha\beta,\gamma} \\ F_{\gamma\delta} F^{\gamma\delta,\alpha} &= F_{\beta\gamma} F^{\beta\gamma,\alpha} \end{aligned} \tag{5.8.4}$$

and thus rewrite (5.8.3) as

$$T^{\alpha\beta}{}_{,\beta} = F^{\alpha}{}_{\gamma} F^{\beta\gamma}{}_{,\beta} - \tfrac{1}{2} F_{\beta\gamma} \left(F^{\alpha\beta,\gamma} + F^{\gamma\alpha,\beta} + F^{\beta\gamma,\alpha} \right) \tag{5.8.5}$$

Using Maxwell's equations, the bracketed part is zero, and the remaining term on the right is $-F^{\alpha}{}_{\gamma} J^{\gamma} = -f^{\alpha}$. □

Thus $T^{\alpha\beta}$ in (5.8.1) can be interpreted as an energy-momentum tensor.

In full $T^{\alpha\beta}$ looks like

$$\begin{pmatrix} \frac{1}{2}(E^2+B^2) & E_yB_z - E_zB_y & E_zB_x - E_xB_z & E_xB_y - E_yB_x \\ E_yB_z - E_zB_y & \frac{1}{2}(E^2+B^2) - E_x^2 - B_x^2 & -E_xE_y - B_xB_y & -E_xE_z - B_xB_z \\ E_zB_x - E_xB_z & -E_yE_x - B_yB_z & \frac{1}{2}(E^2+B^2) - E_y^2 - B_y^2 & -E_yE_z - B_yB_z \\ E_xB_y - E_yB_x & -E_zE_x - B_zB_x & -E_zE_y - B_zB_y & \frac{1}{2}(E^2+B^2) - E_z^2 - B_z^2 \end{pmatrix} \tag{5.8.6}$$

Despite its ghastly appearance (5.8.6) can tell us several interesting things.

First, the electromagnetic field carries energy: the energy density is $T^{00} = \frac{1}{2}(E^2 + B^2)$. However $T^{\alpha}{}_{\alpha}$, which for a fluid was the rest mass density, is zero. Which suggests interpreting the electromagnetic field as a fluid consisting of massless particles.

Second, the field carries momentum: T^{0i} is $\mathbf{E} \times \mathbf{B}$, usually called the Poynting vector.

Third, the field exerts pressure. For example, if the only nonzero field component is E_x then $T^{\alpha\beta}$ becomes

$$\frac{1}{2}E_x^2 \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix} \tag{5.8.7}$$

Comparing with (3.7.5) we can interpret this as a pressure perpendicular to the field and tension (negative pressure) along the field.

 5.8 The field energy-momentum tensor

6. Calculus of Variations

The calculus of variations is about finding paths such that some integral along the path is extremized.

6.1 Statement of the problem

We want to find a path $x^\mu(\omega)$ connecting two given points $\boldsymbol{x}_{\rm ini}$ and $\boldsymbol{x}_{\rm fin}$ [ω being a parameter for the path] such that the integral of some given function $L(\boldsymbol{x},\dot{\boldsymbol{x}},\omega)$ along the path is extremized. Or rather, the integral is stationary:

$$\int_{\boldsymbol{x}_{\rm ini}}^{\boldsymbol{x}_{\rm fin}} L(\boldsymbol{x},\dot{\boldsymbol{x}},\omega)\,d\omega \quad \text{stationary} \tag{6.1.1}$$

By this we mean that the value of the integral along the desired path $x^\mu(\omega)$ equals the value along an infinitesimally close path $x^\mu(\omega)+\delta x^\mu(\omega)$. Another way of writing is

$$\delta \int_{\boldsymbol{x}_{\rm ini}}^{\boldsymbol{x}_{\rm fin}} L(\boldsymbol{x},\dot{\boldsymbol{x}},\omega)\,d\omega = 0 \tag{6.1.2}$$

By the way, $\delta x^\mu(\omega)=0$ at the endpoints, since the latter are given.

This is kind of a zero-derivative property, only for functionals instead of functions.

6.2 The Euler-Lagrange equations

Remarkably, the integral condition (6.2.0) can be reduced to a set of differential equations. We have

$$\int_{\boldsymbol{x}_{\rm ini}}^{\boldsymbol{x}_{\rm fin}} \left(\frac{\partial L}{\partial x^\mu}\,\delta x^\mu + \frac{\partial L}{\partial \dot{x}^\mu}\,\delta \dot{x}^\mu\right) d\omega = 0 \tag{6.2.1}$$

Integrating by parts and using the fact that $\delta x^\mu = 0$ at the ends gives

$$\int_{\boldsymbol{x}_{\rm ini}}^{\boldsymbol{x}_{\rm fin}} \left(\frac{\partial L}{\partial x^\mu} - \frac{d}{d\omega}\left(\frac{\partial L}{\partial \dot{x}^\mu}\right)\right) \delta x^\mu\,d\omega = 0 \tag{6.2.2}$$

Since this must be true under arbitrary path-variations δx^μ we have

$$\boxed{\frac{\partial L}{\partial x^\mu} = \frac{d}{d\omega}\left(\frac{\partial L}{\partial \dot{x}^\mu}\right)} \tag{6.2.3}$$

known as the **Euler-Lagrange equations**.

6.3 A special case

There are many variants and special cases of the Euler-Lagrange equations. One interesting special case is when L has no *explicit* dependence on ω:

$$L - \dot{x}^{\mu} \frac{\partial L}{\partial \dot{x}^{\mu}} = \text{const} \tag{6.3.1}$$

Derivation Equation (6.2.3) is equivalent to

$$\frac{\partial L}{\partial \omega} - \frac{d}{d\omega}\left(L - \dot{x}^{\mu} \frac{\partial L}{\partial \dot{x}^{\mu}}\right) = 0 \tag{6.3.2}$$

(Just use the chain rule to verify). If the first term is zero, then (6.3.1) follows. □

6.4 The brachistochrone problem

This is the classic calculus of variations problem, first solved by Newton, no less. We have two points, and we want to connect them by a curved track such that a body sliding frictionlessly down the track under gravity takes the minimum time.

Let x be the horizontal distance (and independent variable) and y the vertical distance, measured downwards. Say the sliding body starts from rest at $(0, 0)$. The velocity of the body at a point (x, y) on the track is $\propto \sqrt{y}$. The differential arclength along the track is $\sqrt{1 + y'^2}\, dx$. Thus the travel time is $\propto$

$$\int y^{-\frac{1}{2}} (1 + y'^2)^{\frac{1}{2}}\, dx \tag{6.4.1}$$

Applying equation (6.3.1) and simplifying, we get

$$y(1 + y'^2) = \text{const} \tag{6.4.2}$$

This is nonlinear a differential equation which we need to solve. There is no general method for doing this, so we basically have to guess.

First note a useful property of (6.4.2): if $y = f(x)$ is a solution for const $= 1$ then

$$y = c_1 f\left(\frac{x - c_2}{c_1}\right) \tag{6.4.3}$$

will be a solution for const $= c_1$. So we really need to write down a solution of (6.4.3) for one value of the constant. Let us put const $= 2$. Then, as we can verify, a solution is

$$y = 1 - \cos\theta, \qquad x = \theta - \sin\theta \tag{6.4.4}$$

It is a cycloid.

Exercise 6.1

What is the shape adopted by electric transmission wires?

To solve this, consider a path problem similar to the above; but instead of minimizing $y^{-\frac{1}{2}}$ times the arclength, we have to minimize y times the arclength, because it's the potential energy and will get minimized by the wire.

Derive the equation analogous to (6.4.2) for this case and try to guess the answer. The answer is called a catenary.

7. Principle of Equivalence

The principle of equivalence of gravitational and inertial mass is a physical assertion about the nature of gravity. It implies the mathematical statement that spacetime has a Riemannian metric.

7.1 Freely falling frames

In the Newtonian theory, gravity causes a force on a body that is proportional to that body's mass. The theory is inconsistent with relativity and must be modified. But the basic principle that gravitational force is proportional to mass (which, we recall, measures inertia)—and therefore that gravitational mass is the same as inertial mass—does not disagree with special relativity provided we interpret mass as relativistic mass-energy. This is the principle of equivalence.

The principle of equivalence is a new piece of physics, it is not implied by what we have done so far. But it is experimentally well tested, most importantly by a class of experiments known as Eötvös experiments. (See pp. 27–28 of Binney for an account of these.)

Now, in an accelerating frame, a body experiences 'fictitious' forces proportional to the inertial mass. (This is a tautology really, it's just another way of saying that the frame is accelerated.) If the principle of equivalence applies, then fictitious forces can be used to cancel the gravitational forces. In other words, in any gravitational field, there are (accelerated) frames in which the gravitational forces vanish. This prompts the following formal statement of the principle of equivalence (quoting from Weinberg): *At every space-time point in an arbitrary gravitational field it is possible to choose a "locally inertial coordinate system" such that, within a sufficiently small region of the point in question, the laws of nature take the same form as in unaccelerated Cartesian coordinate systems in the absence of gravitation.* Such coordinates are also known as **freely falling frames**.

In other words, we can locally nullify the effect of a gravitational field, thus making special relativity applicable, by choosing a freely falling frame. "Locally" is important—outside a small neighbourhood, the gravitational field may have changed, and the same cartesian coordinates will not be freely-falling any more.

7.2 Metrics and Geodesics

Even if we know that freely-falling coordinates exist, it would be cumbersome to have to transform into them every time we wanted to calculate anything. Fortunately, we can use the principle of equivalence to derive the effect of gravity in arbitrary coordinates.

Consider a particle moving under gravitational (and no other) forces. In freely falling coordinates (call them ξ^α) the equation of motion of the particle is

$$\frac{d^2\xi^\alpha}{d\tau^2} = 0, \qquad d\tau^2 = -\eta_{\alpha\beta}\, d\xi^\alpha\, d\xi^\beta. \tag{7.2.1}$$

In an arbitrary coordinate system x^μ (which may be curvilinear, rotating, accelerating), the equations of motion are

$$\frac{d^2x^\lambda}{d\tau^2} + \Gamma^\lambda_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} = 0, \qquad d\tau^2 = -g_{\mu\nu}\, dx^\mu\, dx^\nu \tag{7.2.2}$$

where

$$\Gamma^{\lambda}_{\mu\nu} = \frac{\partial x^{\lambda}}{\partial \xi^{\alpha}} \frac{\partial^2 \xi^{\alpha}}{\partial x^{\mu} \partial x^{\nu}} \tag{7.2.3}$$

is called a **Christoffel symbol** or **affine connection**, and

$$g_{\mu\nu} = \frac{\partial \xi^{\alpha}}{\partial x^{\mu}} \frac{\partial \xi^{\beta}}{\partial x^{\nu}} \eta_{\alpha\beta}, \qquad g^{\lambda\mu} g_{\mu\nu} = \delta^{\lambda}_{\nu} \tag{7.2.4}$$

is the **metric tensor**, about which more presently. The first equation in (7.2.2) is called a **geodesic equation**; and we will interpret it soon.

DERIVATION The expression (7.2.4) for the metric follows immediately from the chain rule.

To get the geodesic equation, we rewrite the first equation in (7.2.1) as

$$0 = \frac{d}{d\tau}\left(\frac{\partial \xi^{\alpha}}{\partial x^{\mu}} \frac{dx^{\mu}}{d\tau}\right) = \frac{\partial \xi^{\alpha}}{\partial x^{\mu}} \frac{d^2 x^{\mu}}{d\tau^2} + \frac{\partial^2 \xi^{\alpha}}{\partial x^{\mu} \partial x^{\nu}} \frac{dx^{\mu}}{d\tau} \frac{dx^{\nu}}{d\tau} \tag{7.2.5}$$

and then multiply by $\left(\partial x^{\lambda}/\partial \xi^{\alpha}\right)$. □

For a massless particle (7.2.2) does not apply, since $d\tau = 0$; instead we have

$$\frac{d^2 x^{\lambda}}{d\sigma^2} + \Gamma^{\lambda}_{\mu\nu} \frac{dx^{\mu}}{d\sigma} \frac{dx^{\nu}}{d\sigma} = 0, \qquad g_{\mu\nu} \frac{dx^{\mu}}{d\sigma} \frac{dx^{\nu}}{d\sigma} = 0 \tag{7.2.6}$$

which is called a **null geodesic**.

DERIVATION Use $\sigma = \xi^0$ instead of τ in the previous derivation. □

7.3 Tensors redefined

We defined tensors in connection with inertial coordinates and Lorentz transformations in chapter 3. But since we are now concerned with general coordinate systems and arbitrary coordinate transformations, we need to generalize those definitions.

Now comes the cunning bit. In chapter 3, we carefully avoided making use of various special properties of $\eta_{\alpha\beta}$, $\eta^{\alpha\beta}$, and $\Lambda^{\alpha}{}_{\beta}$ and $\Lambda_{\alpha}{}^{\beta}$. The properties we did use were:

(i) $\eta_{\alpha\beta}$, $\eta^{\alpha\beta}$ are symmetric and inverses of each other;

(ii) $\Lambda^{\alpha}{}_{\beta}$ and $\Lambda_{\alpha}{}^{\beta}$ are inverses of each other; and

(iii) $\Lambda^{\alpha}{}_{\beta}$ and $\Lambda_{\alpha}{}^{\beta}$ are constants.

Item (iii) we used only to show that the derivative of a vector or higher rank tensor is a tensor.

So if we amend our definitions slightly:

$$\boxed{\eta_{\alpha\beta} \to g_{\alpha\beta}, \quad \eta^{\alpha\beta} \to g^{\alpha\beta}, \quad \Lambda^{\alpha}{}_{\beta} \to \frac{\partial x'^{\alpha}}{\partial x^{\beta}}, \quad \Lambda_{\alpha}{}^{\beta} \to \frac{\partial x^{\beta}}{\partial x'^{\alpha}}} \tag{7.3.1}$$

we can carry over all the properties of tensors we derived in chapter 3, *with one exception.* The exception is that the derivatives vectors or higher rank tensors (of the type $v^{\alpha}{}_{,\beta}$ or $F^{\alpha\beta}{}_{,\lambda}$) are *not* tensors—generalizing the definition of a gradient will take a little more work. But the derivative of a scalar is a still vector.

The old definitions amended in (7.3.1) remain as special cases of the new ones.

If a tensor has all components zero in one coordinate system, that remains true in any coordinate system. This fact is occasionally useful in recognizing tensors and non-tensors.

7.4 Christoffel symbols

The Christoffel symbol was defined in (7.2.3) in terms of freely-falling coordinates. However, it can be expressed in terms of the metric:

$$\Gamma^{\mu}_{\alpha\beta} = \tfrac{1}{2} g^{\mu\nu} \left(g_{\nu\alpha,\beta} + g_{\beta\nu,\alpha} - g_{\alpha\beta,\nu} \right) \tag{7.4.1}$$

Derivation Define

$$\Gamma_{\mu\alpha\beta} = g_{\mu\nu}\Gamma^{\nu}_{\alpha\beta}.$$

From the definition (7.2.3) we get

$$\Gamma_{\mu\alpha\beta} = \frac{\partial^2 \xi^{\gamma}}{\partial x^{\alpha}\partial x^{\beta}} \frac{\partial \xi^{\delta}}{\partial x^{\mu}} \eta_{\gamma\delta}.$$

From the definition (7.2.4) we have

$$g_{\mu\alpha,\beta} = \frac{\partial}{\partial x^{\beta}} \left(\frac{\partial \xi^{\gamma}}{\partial x^{\mu}} \frac{\partial \xi^{\delta}}{\partial x^{\alpha}} \right) \eta_{\gamma\delta} = \left(\frac{\partial^2 \xi^{\gamma}}{\partial x^{\beta}\partial x^{\mu}} \frac{\partial \xi^{\delta}}{\partial x^{\alpha}} + \frac{\partial \xi^{\gamma}}{\partial x^{\mu}} \frac{\partial^2 \xi^{\delta}}{\partial x^{\beta}\partial x^{\alpha}} \right) \eta_{\gamma\delta}. \tag{7.4.2}$$

Permuting indices gives

$$\begin{aligned} g_{\alpha\beta,\mu} &= \left(\frac{\partial^2 \xi^\gamma}{\partial x^\mu \partial x^\alpha} \frac{\partial \xi^\delta}{\partial x^\beta} + \frac{\partial \xi^\gamma}{\partial x^\alpha} \frac{\partial^2 \xi^\delta}{\partial x^\mu \partial x^\beta} \right) \eta_{\gamma\delta} \\ g_{\beta\mu,\alpha} &= \left(\frac{\partial^2 \xi^\gamma}{\partial x^\alpha \partial x^\beta} \frac{\partial \xi^\delta}{\partial x^\mu} + \frac{\partial \xi^\gamma}{\partial x^\beta} \frac{\partial^2 \xi^\delta}{\partial x^\alpha \partial x^\mu} \right) \eta_{\gamma\delta} \end{aligned} \tag{7.4.3}$$

Collating and exploiting the symmetry of $\eta_{\gamma\delta}$ gives

$$\Gamma_{\mu\alpha\beta} = \tfrac{1}{2} \left(g_{\mu\alpha,\beta} + g_{\beta\mu,\alpha} - g_{\alpha\beta,\mu} \right), \tag{7.4.4}$$

from which (7.4.1) follows. □

The purpose of this section is mainly to show that the equations of motion (i.e., 7.2.2 and 7.2.6) can be written without explicit reference to freely falling coordinates. The formula (7.4.1) isn't usually the most efficient way to calculate Christoffel symbols.

7.5 Variational form

Important differential equations in physics often turn out to have an equivalent variational form. This is so for the geodesic equations (7.2.2). In the variational form, the worldline $x^\alpha(\tau)$ of a particle between two fixed events is given by

$$\int \left(g_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} \right) d\tau \quad \text{stationary}, \qquad \text{or} \int d\tau \quad \text{stationary} \tag{7.5.1}$$

The two statements in (7.5.1) are equivalent, because the expression in parenthesis is just $(d\tau/d\tau)^2 = 1$. The second form provides a nice interpretation—the proper time is stationary—but the first form is more useful for calculation.

Applying the Euler-Lagrange equations (6.2.3) to the first statement in (7.5.1) gives

$$\frac{d}{d\tau} \left(g_{\lambda\nu} \frac{dx^\nu}{d\tau} \right) = \tfrac{1}{2} g_{\mu\nu,\lambda} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} \tag{7.5.2}$$

which is just the geodesic equations.

Derivation Using dots to mean $d/d\tau$, (7.5.2) gives

$$g_{\lambda\nu} \ddot{x}^\nu + g_{\lambda\nu,\mu} \dot{x}^\nu \dot{x}^\mu = \tfrac{1}{2} g_{\mu\nu,\lambda} \dot{x}^\mu \dot{x}^\nu \tag{7.5.3}$$

which we can rewrite as

$$g_{\lambda\nu} \ddot{x}^\nu + \tfrac{1}{2} (g_{\lambda\nu,\mu} + g_{\lambda\mu,\nu} - g_{\mu\nu,\lambda}) \dot{x}^\mu \dot{x}^\nu = 0 \tag{7.5.4}$$

which on raising indices gives the geodesic equation (7.2.2). □

7.6 Gravitational time dilation

The principle of equivalence doesn't tell us how to calculate determine $g_{\mu\nu}$ completely, it just tells us about dynamics once we have $g_{\mu\nu}$. To determine the metric in the presence of gravity we need to input more physics.

We can, however, make a start at calculating $g_{\mu\nu}$ by requiring that the Newtonian limit be reproduced.

If the gravitational field is weak then

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \tag{7.6.1}$$

where $h_{\mu\nu}$ is small. Consider such a situation, and further assume (i) the field is static, i.e., $h_{\mu\nu,t} = 0$, (ii) a particle is moving in this field slowly, i.e., $dx^i/d\tau \ll dt/d\tau$. In that case the geodesic equations give

$$\frac{d^2x^i}{dt^2} = \tfrac{1}{2} g_{00,i}. \tag{7.6.2}$$

Derivation We take (7.5.2) to leading order, which gives

$$\begin{aligned} \frac{d^2t}{d\tau^2} &= 0, \\ \frac{d^2x^i}{d\tau^2} &= \tfrac{1}{2} g_{00,i} \left(\frac{dt}{d\tau}\right)^2. \end{aligned} \tag{7.6.3}$$

The first equation in (7.6.3) implies that $dt/d\tau$ is a constant, from which the second equation implies (7.6.2). □

Comparing with the Newtonian equation of motion $d^2x^i/dt^2 = -\Phi_{,i}$ (Φ being the Newtonian gravitational potential) gives

$$g_{00} = -(1 + 2\Phi).$$

Equation (7.6.3) is general for static weak fields; we can now forget the slow-moving particle we used to derive it. Consider the interval between two events at the same point in this metric. We get

$$dt = (1 - \Phi)\, d\tau. \tag{7.6.4}$$

Comparing with the time dilation formula (2.7.1) from special relativity, we see that clocks run slower in a gravitational field.

8. Curvature

We study derivatives of the metric and find that they are related to curvature of space-time.

8.1 Gravity versus coordinate systems

We have seen that $g_{\mu\nu}$ differs from $\eta_{\mu\nu}$ when we have gravity, or even when we don't have gravity but have curvilinear or accelerated coordinates. Can we disentangle the gravitational and coordinate-system contributions?

We know that we can make special relativity hold *locally,* even if there is a gravitational field. Can we make it hold beyond a small neighbourhood? In general we cannot. That is, at any point we can make $g_{\mu\nu} = \eta_{\mu\nu}$ and $g_{\mu\nu,\lambda} = 0$ by choosing a cunning coordinate system, but we cannot make $g_{\mu\nu,\lambda\sigma} = 0$.

Derivation Consider the neighbourhood of an arbitrary event, which we can take to be the origin without loss of generality. We have the metric $g'_{\alpha\beta}$ in some coordinates $\boldsymbol{x}'$ and want to change to some new coordinates $\boldsymbol{x}$ such that $g_{\mu\nu}(\boldsymbol{x})$ looks as much like $\eta_{\mu\nu}$ as possible.

We have

$$g_{\mu\nu} = \Lambda^\alpha{}_\mu \, \Lambda^\beta{}_\nu \, g'_{\alpha\beta}. \tag{8.1.1}$$

Taylor expanding, we have

$$\begin{aligned} g_{\mu\nu}(\boldsymbol{x}) &= g_{\mu\nu}(0) + g_{\mu\nu,\lambda}(0)x^\lambda + \tfrac{1}{2} g_{\mu\nu,\lambda\sigma}(0)x^\lambda x^\sigma + \dots \\ &= \Lambda^\alpha{}_\mu \, \Lambda^\beta{}_\nu \, g'_{\alpha\beta} + \Big[2\,\Lambda^\alpha{}_{\mu,\lambda} \, \Lambda^\beta{}_\nu \, g'_{\alpha\beta} + \Lambda^\alpha{}_\mu \, \Lambda^\beta{}_\nu \, g'_{\alpha\beta,\lambda}\Big] x^\lambda + \\ &\quad \Big[\Big(\Lambda^\alpha{}_{\mu,\lambda\sigma} \, \Lambda^\beta{}_\nu + \Lambda^\alpha{}_{\mu,\lambda} \, \Lambda^\beta{}_{\nu,\sigma}\Big) g'_{\alpha\beta} + \Lambda^\alpha{}_{\mu,\lambda} \, \Lambda^\beta{}_\nu \, g'_{\alpha\beta,\lambda} + \\ &\quad \tfrac{1}{2}\,\Lambda^\alpha{}_\mu \, \Lambda^\beta{}_\nu \, g'_{\alpha\beta,\lambda\sigma}\Big] x^\lambda x^\sigma + \dots \end{aligned} \tag{8.1.2}$$

In (8.1.2) the commas are all derivatives with respect to $\boldsymbol{x}$, and all the Λ and g' terms on the RHS are evaluated at the origin and hence are constants. The Λ constants are for us to choose, because we are defining the transformation $\boldsymbol{x}' \to \boldsymbol{x}$.

Let us compare the upper and lower equations in (8.1.2). In $g_{\mu\nu}$ there are 10 numbers, and we can set them all to the corresponding elements of $\eta_{\mu\nu}$, because we have 16 numbers $\Lambda^\mu{}_\alpha(0)$ to play with. In fact we have six numbers to spare, so we can stick on an extra boost and rotation. In $g_{\mu\nu,\lambda}$ there are 40 numbers, and we can set them to zero by choosing $\Lambda^\mu{}_{\alpha,\lambda}(0)$ suitably; this is because $\Lambda^\mu{}_{\alpha,\lambda}$ is symmetric in μ, λ (refer to the definition 7.3.1) and hence has 40 independent components. In $g_{\mu\nu,\lambda\sigma}$ there are 100 numbers, whereas in $\Lambda^\alpha{}_{\mu,\lambda\sigma}$ has only 80 numbers, because of symmetry with respect to μ, λ, σ.

Thus there are 20 numbers in $g_{\mu\nu,\lambda\sigma}$ at any point that cannot in general be removed by a coordinate transformation. For N dimensions the answer is $N^2(N^2-1)/12$—see page 159 of Schutz. □

We conclude that the effect of gravity is lurking in 20 numbers in the second derivative of the metric.

8.2 Covariant derivatives

So now we want to study derivatives, and eventually second derivatives of the metric. But the derivative of a tensor with respect to the coordinates, such as $g_{\mu\nu,\lambda}$ which we have already used, is not a tensor. It carries information about the coordinate system in a way tensors do not.

Nevertheless, there is an obvious way to differentiate a tensor so as to get another tensor: we can transform to a freely falling frame, work out the derivative there (where it will be a tensor, because special relativity applies) and then transform back using the appropriate tensor transformation rule. This is known as **covariant differentiation** and can be expressed without explicit reference to freely falling frames. We denote it by semicolons instead of commas. The formulas are

$$\boxed{\begin{array}{ll} A^{\mu}{}_{;\lambda} = A^{\mu}{}_{,\lambda} + \Gamma^{\mu}_{\nu\lambda} A^{\nu} & A_{\mu;\lambda} = A_{\mu,\lambda} - \Gamma^{\nu}_{\mu\lambda} A_{\nu} \\ F^{\mu}{}_{\nu;\lambda} = F^{\mu}{}_{\nu,\lambda} + \Gamma^{\mu}_{\sigma\lambda} F^{\sigma}{}_{\nu} - \Gamma^{\sigma}_{\nu\lambda} F^{\mu}{}_{\sigma} & \end{array}} \tag{8.2.1}$$

and so on.

Derivation Suppose we have vector $A^{\alpha}(\boldsymbol{x})$, whose components in a freely-falling frame are $\tilde{A}^{\mu}(\boldsymbol{\xi})$. Consider the rate of change of $\tilde{A}^{\mu}$ along a curve $\boldsymbol{\xi}(\omega)$.

$$\begin{aligned} \frac{d\tilde{A}^{\mu}}{d\omega} &= \frac{dx^{\kappa}}{d\omega} \frac{\partial \tilde{A}^{\mu}}{\partial x^{\kappa}} = \frac{dx^{\kappa}}{d\omega} \frac{\partial}{\partial x^{\kappa}} \left(\frac{\partial \xi^{\mu}}{\partial x^{\alpha}} A^{\alpha} \right) \\ &= \frac{dx^{\kappa}}{d\omega} \left(\frac{\partial \xi^{\mu}}{\partial x^{\alpha}} \frac{\partial A^{\alpha}}{\partial x^{\kappa}} + \frac{\partial^2 \xi^{\mu}}{\partial x^{\kappa} \partial x^{\alpha}} A^{\alpha} \right). \end{aligned} \tag{8.2.2}$$

Multiplying by $\partial x^{\nu}/\partial \xi^{\mu}$ gives

$$\frac{dA^{\nu}}{d\omega} = \frac{dx^{\kappa}}{d\omega} \left(\frac{\partial A^{\nu}}{\partial x^{\kappa}} + \Gamma^{\nu}_{\kappa\alpha} A^{\alpha} \right), \tag{8.2.3}$$

and the covariant derivative for an up index follows.

Now consider

$$\begin{aligned} \frac{d}{d\omega} \left(A^{\mu} B_{\mu} \right) &= \frac{dx^{\kappa}}{d\omega} \left(A^{\mu}{}_{,\kappa} B_{\mu} + A^{\mu} B_{\mu,\kappa} \right) \\ &= \frac{dx^{\kappa}}{d\omega} \left(A^{\mu}{}_{;\kappa} B_{\mu} - \Gamma^{\mu}_{\kappa\alpha} A^{\alpha} B_{\mu} + A^{\mu} B_{\mu,\kappa} \right) \\ &= \frac{dx^{\kappa}}{d\omega} \left[A^{\mu}{}_{;\kappa} B_{\mu} + A^{\mu} \left(B_{\mu,\kappa} - \Gamma^{\alpha}_{\kappa\mu} B_{\alpha} \right) \right] \end{aligned} \tag{8.2.4}$$

which gives the answer for a down index. □

With the definition of covariant derivatives in hand, we can express any expression from special relativity in a general coordinate system. All we have to do is replace commas in the special-relativistic equations with semi-colons. This is worth emphasizing: *any equation involving only tensors and first derivatives of tensors that is valid in a freely falling frame will become valid in general coordinate systems if coordinate derivatives are replaced by covariant derivatives.*

8.3 Parallel displacement

The strange-looking formulas for covariant derivatives actually have a simple geometrical meaning. The derivative of a vector measures how much it changes between the events $\boldsymbol{x}$ and $\boldsymbol{x} + d\boldsymbol{x}$. But in a non-inertial coordinate system the axes also change between the two events. So to differentiate in a way that's not influenced by the writhings of a coordinate system, we need to take the vector at $\boldsymbol{x} + d\boldsymbol{x}$, find its components with respect to the coordinate axes at $\boldsymbol{x}$ and *then* compare with the vector at $\boldsymbol{x}$. The Γ terms in the covariant derivative do this job of transplantation.

More formally

$$dA^\lambda = \Gamma^\lambda_{\mu\nu} A^\mu \, dx^\nu \tag{8.3.1}$$

defines the **parallel displacement** of a vector.

Parallel displacement is easily visualized in two dimensions: one slowly moves the vector along while keeping its orientation relative to the surface fixed. If we parallel-displace through a closed path, we return to the same vector only if the surface is flat, not if it is curved—see the Figures on pp. 164–165 of Schutz, or p. 39 of Binney.

8.4 Riemann-Christoffel tensor

For sufficiently smooth functions (which are all we will be concerned with) partial derivatives commute. Covariant derivatives, on the other hand may *not* commute, because the involve parallel displacement.

An interesting and important thing is that commutator of covariant derivatives can be expressed in terms of a tensor.

$$A^\mu{}_{;\alpha\beta} - A^\mu{}_{;\beta\alpha} = R^\mu{}_{\delta\alpha\beta} A^\delta \tag{8.4.1}$$

where

$$R^\mu{}_{\delta\alpha\beta} = \Gamma^\mu_{\alpha\delta,\beta} - \Gamma^\mu_{\beta\delta,\alpha} + \Gamma^\mu_{\gamma\beta}\Gamma^\gamma_{\alpha\delta} - \Gamma^\mu_{\gamma\alpha}\Gamma^\gamma_{\beta\delta} \tag{8.4.2}$$

Derivation

$$\begin{aligned}
A^\mu{}_{;\alpha\beta} &= A^\mu{}_{;\alpha,\beta} + \Gamma^\mu_{\gamma\beta} A^\gamma{}_{;\alpha} - \Gamma^\gamma_{\alpha\beta} A^\mu{}_{;\gamma} \\
&= (A^\mu{}_{,\alpha} + \Gamma^\mu_{\alpha\gamma} A^\gamma)_{,\beta} + \Gamma^\mu_{\gamma\beta}(A^\gamma{}_{,\alpha} + \Gamma^\gamma_{\alpha\delta} A^\delta) \\
&- \Gamma^\gamma_{\alpha\beta}(A^\mu{}_{,\gamma} + \Gamma^\mu_{\gamma\delta} A^\delta) \\
&= A^\mu{}_{,\alpha\beta} + \Gamma^\mu_{\alpha\delta,\beta} A^\delta + \Gamma^\mu_{\alpha\delta} A^\delta{}_{,\beta} + \Gamma^\mu_{\delta\beta} A^\delta{}_{,\alpha} + \Gamma^\mu_{\gamma\beta}\Gamma^\gamma_{\alpha\delta} A^\delta \\
&- \Gamma^\gamma_{\alpha\beta} A^\mu{}_{,\gamma} - \Gamma^\gamma_{\alpha\beta}\Gamma^\mu_{\gamma\delta} A^\delta
\end{aligned} \tag{8.4.3}$$

On substracting $A^\mu{}_{;\beta\alpha}$ only the second and fifth terms in the last expression survive. □

The expression $R^\mu{}_{\delta\beta\alpha}$ in (8.4.2) is a tensor by the quotient rule. It is known as the **Riemann-Christoffel curvature tensor**.

In a freely falling frame, the Γ are zero but their derivatives are not, and we have

$$R_{\alpha\beta\mu\nu} = \tfrac{1}{2}(g_{\alpha\mu,\beta\nu} - g_{\alpha\nu,\beta\mu} + g_{\beta\nu,\alpha\mu} - g_{\beta\mu,\alpha\nu}) \tag{8.4.4}$$

Derivation In a freely falling frame, expanding out the Γ derivatives gives

$$\begin{aligned}
R^\alpha{}_{\beta\mu\nu} &= \Gamma^\alpha_{\beta\mu,\nu} - \Gamma^\alpha_{\beta\nu,\mu} \\
&= \tfrac{1}{2} g^{\alpha\sigma}(g_{\sigma\beta,\mu\nu} + g_{\sigma\mu,\beta\nu} - g_{\beta\mu,\sigma\nu} - g_{\sigma\beta,\nu\mu} - g_{\sigma\nu,\beta\mu} + g_{\beta\nu,\sigma\mu})
\end{aligned} \tag{8.4.5}$$

Here the first and fourth terms cancel, and lowering the first index gives (8.4.4). □

Although the Riemann-Christoffel tensor has 256 components, only 20 are independent because of its many symmetries: $R_{\alpha\beta\mu\nu}$ is antisymmetric in α, β and in μ, ν, and moreover $R_{\alpha\beta\mu\nu} = R_{\mu\nu\alpha\beta}$.

Derivation The symmetries are evident from the form of (8.4.4). Because of the antisymmetries, there are 6 independent possibilities for the first pair of indices and 6 for the last pair. But since the first and last pair of indices can be swapped, that leaves 21 components. Finally, the identity

$$R_{\alpha\beta\mu\nu} + R_{\alpha\nu\beta\mu} + R_{\alpha\mu\nu\beta} = 0, \tag{8.4.6}$$

also verifiable from (8.4.4) leaves 20 independent components. □

The 20 independent components of the curvature hint that curvature has something to do with gravity.

8.5 Ricci Tensor

Contractions of the Riemann-Christoffel tensor turn out to be important enough to merit their own names. The **Ricci tensor** is defined as

$$\begin{aligned} R_{\mu\nu} &= R^{\alpha}{}_{\mu\alpha\nu} \\ &= \Gamma^{\alpha}_{\alpha\mu,\nu} - \Gamma^{\alpha}_{\mu\nu,\alpha} + \Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu} - \Gamma^{\alpha}_{\alpha\beta}\Gamma^{\beta}_{\mu\nu} \end{aligned} \tag{8.5.1}$$

Because of the symmetries, any other contraction would at most change the sign. The Ricci tensor is symmetric.

A useful identity that slightly simplifies the the Ricci tensor is

$$\Gamma^{\alpha}_{\alpha\mu} = (\ln\sqrt{})_{,\mu} \qquad \sqrt{} \text{ means } \sqrt{\det|g_{\gamma\delta}|} \tag{8.5.2}$$

Derivation From (7.4.1) we have

$$\Gamma^{\alpha}_{\alpha\mu} = \tfrac{1}{2} g^{\alpha\beta} g_{\alpha\beta,\mu} \tag{8.5.3}$$

Since $g^{\alpha\beta}$ is the inverse matrix of $g_{\alpha\beta}$ we have

$$g^{\alpha\beta} = \frac{1}{\det|g_{\gamma\delta}|} \frac{\partial}{\partial g_{\alpha\beta}} \det|g_{\gamma\delta}| \tag{8.5.4}$$

and hence

$$\begin{aligned} \Gamma^{\alpha}_{\alpha\mu} &= \tfrac{1}{2} \frac{\partial g_{\alpha\beta}}{\partial x^{\mu}} \frac{1}{\det|g_{\gamma\delta}|} \frac{\partial}{\partial g_{\alpha\beta}} \det|g_{\gamma\delta}| \\ &= \tfrac{1}{2} \frac{1}{\det|g_{\gamma\delta}|} \frac{\partial}{\partial x^{\mu}} \det|g_{\gamma\delta}| = (\ln\sqrt{})_{,\mu} \end{aligned} \tag{8.5.5}$$

(No sum over γ, δ in this derivation.) □

A further contraction

$$R = g^{\mu\nu} R_{\mu\nu} \tag{8.5.6}$$

is called the **Ricci scalar**.

8.6 The Bianchi identities

The **Bianchi identity** is

$$R_{\alpha\beta\mu\nu;\lambda} + R_{\alpha\beta\lambda\mu;\nu} + R_{\alpha\beta\nu\lambda;\mu} = 0. \tag{8.6.1}$$

Derivation Equation (8.6.1) with commas instead of semicolons (i.e., in a freely falling frame) follows from the antisymmetry of $R_{\alpha\beta\mu\nu}$ with respect with μ, ν and the commutativity of ordinary derivatives. We then replace commas with semicolons. □

Contracting twice leads to

$$\boxed{(R^{\mu\nu} - \tfrac{1}{2} R g^{\mu\nu})_{;\nu} = 0} \tag{8.6.2}$$

The tensor inside the brackets is called the **Einstein tensor**.

Derivation Contracting the Bianchi identity, thus

$$g^{\beta\nu} g^{\alpha\mu} (R_{\alpha\beta\mu\nu;\lambda} + R_{\alpha\beta\lambda\mu;\nu} + R_{\alpha\beta\nu\lambda;\mu}) = 0 \tag{8.6.3}$$

gives

$$g^{\beta\nu} (R_{\beta\nu;\lambda} - R_{\beta\lambda;\nu} + R^{\mu}{}_{\beta\nu\lambda;\mu}) = 0 \tag{8.6.4}$$

because $g^{\alpha\mu}$ can go through a covariant derivative, and further

$$R_{;\lambda} - R^{\nu}{}_{\lambda;\nu} - R^{\mu}{}_{\lambda;\mu} = 0 \tag{8.6.5}$$

and raising indices leads to (8.6.2). □

9. Field Equations

We now introduce Einstein's field equations, the last and most remarkable ingredient of the General Theory of Relativity. The principle of equivalence told us the effect of the gravitational field (i.e., the metric) on matter. The field equations specify that gravitational field is generated.

9.1 Einstein's field equations

In developments so far, we have come across two important second-rank tensors, both symmetric and with zero divergence. One is the Einstein tensor, and it comes from studying the geometry of spacetime, and its zero divergence is a geometrical identity. The other is the energy-momentum tensor, and it comes from studying the dynamics of matter and energy. The contents of the energy-momentum tensor depend on the system. For dust we have (cf. equation 3.7.2)

$$T^{\alpha\beta} = \gamma^2 \rho \begin{pmatrix} 1 & v_j \\ v_i & v_i v_j \end{pmatrix} \tag{9.1.1}$$

for a perfect fluid we have (cf. equation 3.7.5)

$$T^{\alpha\beta} = \gamma^2 \rho \begin{pmatrix} 1 & v_j \\ v_i & v_i v_j \end{pmatrix} + p \begin{pmatrix} 0 & 0 \\ 0 & \delta_{ij} \end{pmatrix} + \gamma^2 p \begin{pmatrix} v^2 & v_j \\ v_i & v_i v_j \end{pmatrix} \tag{9.1.2}$$

and for electromagnetic field we have (cf. equation 5.8.6)

$$T^{\alpha\beta} = \begin{pmatrix} \frac{1}{2}(E^2 + B^2) & (\mathbf{E} \times \mathbf{B})_j \\ (\mathbf{E} \times \mathbf{B})_i & \frac{1}{2}(E^2 + B^2)\delta_{ij} - E_i E_j - B_i B_j \end{pmatrix} \tag{9.1.3}$$

But in each case zero divergence (that is to say $T^{\alpha\beta}{}_{,\beta} = 0$) is a dynamical property.

Einstein's field equations are the assertion that these two very different tensors are in fact equal:

$$\boxed{R_{\mu\nu} - \tfrac{1}{2} R g_{\mu\nu} = -8\pi G T_{\mu\nu}} \tag{9.1.4}$$

Here G is the Newtonian gravitational constant. (We could set $G = 1$ by a suitable choice of units.) The proportionality constant is set so as to give the correct Newtonian limit, which we will verify later.

Recall that $R_{\mu\nu}$ consists of $g_{\mu\nu}$ and its first and second derivatives (quoting 8.5.1 and 7.4.1)

$$\begin{aligned} R_{\mu\nu} &= \Gamma^{\alpha}_{\alpha\mu,\nu} - \Gamma^{\alpha}_{\mu\nu,\alpha} + \Gamma^{\alpha}_{\beta\mu}\Gamma^{\beta}_{\alpha\nu} - \Gamma^{\alpha}_{\alpha\beta}\Gamma^{\beta}_{\mu\nu} \\ \Gamma^{\mu}_{\alpha\beta} &= \tfrac{1}{2} g^{\mu\nu} \left(g_{\nu\alpha,\beta} + g_{\beta\nu,\alpha} - g_{\alpha\beta,\nu} \right) \end{aligned} \tag{9.1.5}$$

Thus it is symmetric and therefore has 10 independent components. Thus, the field equations are 10 second-order partial differential equations for the 10 independent components of $g_{\mu\nu}$. But the zero-divergence property expresses 4 relations between these equations, so there are actually only 6 independent differential equations. This leaves four degrees of freedom in $g_{\mu\nu}$; these are known as **gauge conditions** or **coordinate conditions** and are analogous to the gauge freedom we found in the electromagnetic field.

The field equations are not a consequence of anything that has come before. Einstein just guessed that these equations describe the gravitational field. But we can gain some understanding of why he guessed these particular equations by studying special cases.

10. Weak-field Theory

The field equations are coupled nonlinear partial differential equations, so exact solutions are very very hard to come by. But if the gravitational fields are not too strong, meaning that $g_{\mu\nu}$ is not too different from $\eta_{\mu\nu}$, approximate solutions are fairly easy to find and very useful.

10.1 Weak-field metrics

In the weak field regime, we write

$$\begin{aligned} g_{\mu\nu} &= \eta_{\mu\nu} + h_{\mu\nu} \\ g^{\mu\nu} &= \eta^{\mu\nu} - h^{\mu\nu} \end{aligned} \tag{10.1.1}$$

and neglect all terms of $O(h^2)$. We have

$$h^{\mu\nu} = (\text{numerically}) h_{\mu\nu} \tag{10.1.2}$$

.

Derivation Taking $g_{\alpha\beta} = g_{\alpha\mu} g_{\beta\nu} g^{\mu\nu}$ to $O(h)$ leads to

$$h_{\alpha\beta} = h^{\mu\nu} \eta_{\alpha\mu} \eta_{\beta\nu} \tag{10.1.3}$$

and nonzero terms survive only when $\alpha = \mu$ and $\beta = \nu$. □

We have to be a little careful when using the weak field approximation. In particular we have to restrict ourselves to nearly-inertial coordinates, otherwise (10.1.3) will not apply. Also, not that while $g_{\mu\nu}$ is certainly a tensor, the separate pieces $\eta_{\mu\nu}$ and $h_{\mu\nu}$ may not be tensors. (See pp. 200–201 of Schutz for a longer discussion of this point.)

In the weak field approximation, the Ricci tensor simplifies considerably. Neglecting $O(h^2)$ gives

$$R_{\mu\nu} = \tfrac{1}{2} \left[h^{\lambda}{}_{\lambda,\mu\nu} + h_{\mu\nu}{}^{,\lambda}{}_{,\lambda} - h^{\lambda}{}_{\mu,\lambda\nu} - h^{\lambda}{}_{\nu,\lambda\mu} \right] \tag{10.1.4}$$

Derivation Using

$$\begin{aligned} R_{\mu\nu} &= \Gamma^{\lambda}_{\lambda\mu,\nu} - \Gamma^{\lambda}_{\mu\nu,\lambda} + O(h^2) \\ \Gamma^{\gamma}_{\alpha\beta} &= \eta^{\gamma\delta} \left[h_{\alpha\delta,\beta} + h_{\beta\delta,\alpha} - h_{\alpha\beta,\delta} \right] + O(h^2) \end{aligned} \tag{10.1.5}$$

gives

$$\begin{aligned} R_{\mu\nu} = \tfrac{1}{2} \eta^{\lambda\sigma} \big[& h_{\lambda\sigma,\mu\nu} + h_{\mu\sigma,\lambda\nu} - h_{\lambda\mu,\sigma\nu} \\ & - h_{\nu\sigma,\mu\lambda} - h_{\mu\sigma,\nu\lambda} + h_{\mu\nu,\sigma\lambda} \big] + O(h^2) \end{aligned} \tag{10.1.6}$$

and hence (10.1.4). □

10.2 The harmonic gauge

We can simplify further by applying a gauge transformation. In the weak field regime a gauge transformation is a coordinate transformation of the type

$$x'^{\mu} = x^{\mu} + \varepsilon^{\mu}(\boldsymbol{x}) \tag{10.2.1}$$

where ε^{μ} is of $O(h)$. It has the effect of changing the metric thus

$$h'_{\mu\nu} = h_{\mu\nu} - \varepsilon_{\mu,\nu} - \varepsilon_{\nu,\mu} \tag{10.2.2}$$

while leaving the Ricci tensor unchanged.

Derivation To derive (10.2.2), consider

$$\begin{aligned} g'^{\mu\nu} &= \frac{\partial x'^{\mu}}{\partial x^{\alpha}} \frac{\partial x'^{\nu}}{\partial x^{\beta}} g^{\alpha\beta} \\ &= \left(\delta^{\mu}_{\alpha} + \varepsilon^{\mu}{}_{,\alpha}\right)\left(\delta^{\nu}_{\beta} + \varepsilon^{\nu}{}_{,\beta}\right)\left(\eta^{\alpha\beta} - h^{\alpha\beta}\right) \\ &= \eta^{\mu\nu} - h^{\mu\nu} + \varepsilon^{\mu,\nu} + \varepsilon^{\nu,\mu} \end{aligned} \tag{10.2.3}$$

Now inserting (10.2.2) into the Ricci tensor (10.1.4) we get

$$\begin{aligned} R'_{\mu\nu} = R_{\mu\nu} + \tfrac{1}{2}\big[&- 2\varepsilon^{\lambda}{}_{,\lambda\mu\nu} - \varepsilon_{\mu,\nu}{}^{,\lambda}{}_{,\lambda} - \varepsilon_{\nu,\mu}{}^{,\lambda}{}_{,\lambda} \\ &+ \varepsilon^{\lambda}{}_{,\mu\lambda\nu} + \varepsilon_{\mu}{}^{,\lambda}{}_{,\lambda\nu} + \varepsilon^{\lambda}{}_{,\nu\lambda\mu} + \varepsilon_{\nu}{}^{,\lambda}{}_{,\lambda\mu}\big] \end{aligned} \tag{10.2.4}$$

and the extra terms all cancel. □

A suitable gauge transformation will give us

$$R_{\mu\nu} = \tfrac{1}{2} h_{\mu\nu}{}^{,\lambda}{}_{,\lambda} \tag{10.2.5}$$

This is known as the **harmonic gauge** and is analogous to the Lorentz gauge in electromagnetism.

Derivation Under a gauge transformation

$$h'^{\lambda}{}_{\mu,\lambda} - \tfrac{1}{2} h'^{\lambda}{}_{\lambda,\mu} = h^{\lambda}{}_{\mu,\lambda} - \tfrac{1}{2} h^{\lambda}{}_{,\lambda\mu} - \varepsilon^{\lambda}{}_{,\mu\lambda} - \varepsilon_{\mu}{}^{,\lambda}{}_{,\lambda} + \varepsilon^{\lambda}{}_{,\lambda\mu} \tag{10.2.6}$$

Here the third and fifth terms on the right cancel. Hence by choosing ε_{μ} such that

$$\varepsilon_{\mu}{}^{,\lambda}{}_{,\lambda} = h^{\lambda}{}_{\mu,\lambda} - \tfrac{1}{2} h^{\lambda}{}_{\lambda,\mu} \tag{10.2.7}$$

we can make

$$h'^{\lambda}{}_{\mu,\lambda} - \tfrac{1}{2} h'^{\lambda}{}_{\lambda,\mu} = 0 \tag{10.2.8}$$

in which case three of the terms in (10.1.4) will cancel, and

$$R'_{\mu\nu} = \tfrac{1}{2} h'_{\mu\nu}{}^{,\lambda}{}_{,\lambda}$$

from which (10.2.5) without loss of generality. □

Writing $\nabla^2 h_{\mu\nu}$ for $h_{\mu\nu}{}^{,\lambda}{}_{,\lambda}$ we get

$$\nabla^2 h_{\mu\nu} = -16\pi G \left(T_{\mu\nu} - \tfrac{1}{2} T \eta_{\mu\nu}\right) \tag{10.2.9}$$

This is a wave equation, and has exactly the same form as the Maxwell's equations in (5.6.1). In particular it predicts the existence of gravitational waves, analogous to electromagnetic waves. (But note that the electromagnetic equation 5.6.1 is exact, whereas 10.2.9 is a weak-field approximation.)

10.3 The Post-Newtonian metric

By this we mean the case of slow-moving matter. The leading term in the energy-momentum tensor is simply

$$T_{\mu\nu} = \rho \begin{pmatrix} 1 & & & \\ & 0 & & \\ & & 0 & \\ & & & 0 \end{pmatrix} \tag{10.3.1}$$

and (10.2.9) has the solution

$$h_{\mu\nu} = -2\Phi \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix} \tag{10.3.2}$$

where Φ is the Newtonian potential, i.e., the solution of $\nabla^2\Phi = 4\pi G\rho$.

Derivation We have $T = -\rho$, leading to

$$\nabla^2 h_{\mu\nu} = -8\pi G\rho \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix} \tag{10.3.3}$$

□

The interval is

$$ds^2 = -(1+2\Phi)dt^2 + (1-2\Phi)\eta_{ij}\,dx^i\,dx^j \tag{10.3.4}$$

The fact that $g_{00} = -(1+2\Phi)$ is enough to reproduce Newtonian dynamics in the low-velocity limit (cf. equation 7.6.3). The $(1-2\Phi)$ term describes the leading order deviations from Newtonian dynamics, and this approximation is often known as post-Newtonian dynamics.

10.4 Gravitational lensing

When we look for null-geodesics in the post-Newtonian metric, we find that gravity has two interesting effects on light.

The first (though historically later) effect is the Shapiro time delay. The speed of light is always the same, provided the correct metric is used to calculate distances. But if we just use the Euclidean line element $dl = \sqrt{\eta_{ij}\, dx^i\, dx^j}$ then light seems to be slowed down and light travel time increased, by

$$\Delta t = -2 \int \Phi \, dl \tag{10.4.1}$$

Derivation Since for light, ds^2 in (10.3.4) must be zero, we have

$$dt = \left(\frac{1-2\Phi}{1+2\Phi}\right)^{\frac{1}{2}} \sqrt{\eta_{ij}\, dx^i\, dx^j} \simeq (1-2\Phi)\, dl \tag{10.4.2}$$

□

This effect is measurable in radar echos from other planets in the solar system, and is one of the experimental tests of general relativity.

A useful interpretation of the Shapiro time delay is to think of light as travelling through a medium of refractive index $(1 - 2\Phi)$. This makes a nice analogy with glass lenses.

The second important effect is the deflection of light rays. A weak gravitational field can cause only a very small change in light paths, but an effective transverse acceleration of

$$\frac{d^2\mathbf{x}_\perp}{dt^2} = -2\,\frac{\partial \Phi}{\partial \mathbf{x}_\perp} \tag{10.4.3}$$

is measurable.

Derivation To find a null geodesic, we use (7.5.1) but with a different parameter σ in place of τ, and use $ds^2 = 0$ as a constraint. (This is equivalent of 7.2.6.) The parameter σ can later be eliminated in favour of t.

The Euler-Lagrange equations are

$$\begin{aligned} &\frac{d}{d\sigma}\left[(1+2\Phi)\left(\frac{dt}{d\sigma}\right)\right] = 0 \\ &\frac{d}{d\sigma}\left[(1-2\Phi)\left(\frac{dx^k}{d\sigma}\right)\right] = -\Phi_k\left[\left(\frac{dt}{d\sigma}\right)^2 + \eta_{ij}\left(\frac{dx^i}{d\sigma}\right)\left(\frac{dx^j}{d\sigma}\right)\right] \end{aligned} \tag{10.4.4}$$

Using the $ds^2 = 0$ condition, the second equation becomes

$$\frac{d^2x^k}{d\sigma^2} = -2\Phi_k\left(\frac{dt}{d\sigma}\right)^2 + O(h^2) \tag{10.4.5}$$

Now the first equation in (10.4.4) makes $(dt/d\sigma)$ a constant at leading order. Using this fact to change variables in (10.4.5) from σ to t and taking the transverse part gives (10.4.3). □

The deflection formula (10.4.3) is the basis of gravitational lensing, a trendy topic in astrophysics.

11. The Schwarzschild Solution

There are very few known exact solutions of the field equations. Of these, the Schwarzschild solution is the best understood. It is the metric produced by a spherical mass.

The section on Kruskal-Szekeres coordinates is not examinable.

11.1 A static spherical system

To look for the metric produced by a spherical mass, we start by making an ansatz (= a guess) that the metric is static and spherical. (We don't know in advance that any such solution exists, but if look for one we can at least expect to reach either an inconsistency or a valid solution.)

With no gravity, the interval is

$$ds^2 = -dt^2 + dr^2 + r^2\,d\theta^2 + r^2 \sin^2\theta\,d\phi^2. \tag{11.1.1}$$

With gravity, but if the metric is still *static* and *spherical* the metric will change, but not that much. Static implies that replacing t by $-t$ should make no difference, so no terms odd in dt can appear in the interval. The angular part of the interval must be of the form

$$g_{\theta\theta}(d\theta^2 + \sin^2\theta\,d\phi^2) \tag{11.1.2}$$

with $g_{\theta\theta}$ depending only on r, since every event is assumed to lie on the surface of a sphere, and on a sphere the interval is of the form (11.1.2).

The above argument applies to any symmetric second-rank tensor derived from $g_{\mu\nu}$. In particular, the Ricci tensor must be

$$R_{\mu\nu} = \begin{pmatrix} R_{tt}(r) & & & \\ & R_{rr}(r) & & \\ & & R_{\theta\theta}(r) & \\ & & & \sin^2\theta R_{\theta\theta}(r) \end{pmatrix} \tag{11.1.3}$$

because the scalar $R_{\mu\nu}\,dx^\mu\,dx^\nu$ must have the same static and spherical-symmetry properties as the interval.

For the metric tensor itself, we can simplify further: by redefining r we can make $g_{\theta\theta}$ anything we like, and we choose to make it r^2. Thus the interval

$$ds^2 = -e^{\nu(r)}\,dt^2 + e^{\lambda(r)}\,dr^2 + r^2\,d\theta^2 + r^2\sin^2\theta\,d\phi^2 \tag{11.1.4}$$

actually expresses the most general static spherical metric.

A word of caution about the metric (11.1.4). The 'radial' coordinate r is effectively defined not as the distance from the centre but as circumference by 2π of a circle at constant r. A circle at constant r may not even *have* a centre—see the cover of Schutz.

11.2 Geodesics and Christoffel symbols

The condition

$$\int \left(-e^{\nu}\dot{t}^2 + e^{\lambda}\dot{r}^2 + r^2\dot{\theta}^2 + r^2\sin^2\theta\,\dot{\phi}^2\right) d\tau \quad \text{stationary} \tag{11.2.1}$$

where dots denote $d/d\tau$, gives the geodesic equations:

$$\begin{aligned}
&\frac{d}{d\tau}\left(e^{\nu}\dot{t}\right) = 0\\
&\frac{d}{d\tau}\left(e^{\lambda}\dot{r}\right) + \tfrac{1}{2}\nu' e^{\nu}\dot{t}^2 - \tfrac{1}{2}\lambda' e^{\lambda}\dot{r}^2 - r\dot{\theta}^2 - r\sin^2\theta\,\dot{\phi}^2 = 0\\
&\frac{d}{d\tau}\left(r^2\dot{\theta}\right) - r^2\sin\theta\cos\theta\,\dot{\phi}^2 = 0\\
&\frac{d}{d\tau}\left(r^2\sin^2\theta\,\dot{\phi}\right) = 0
\end{aligned} \tag{11.2.2}$$

Simplifying, we can read off the nonzero Christoffel symbols:

$$\begin{aligned}
&\Gamma^t_{tr} = \Gamma^t_{rt} = \tfrac{1}{2}\nu' \\
&\Gamma^r_{tt} = \tfrac{1}{2}\nu' e^{\nu-\lambda} && \Gamma^r_{rr} = \tfrac{1}{2}\lambda' \\
&\Gamma^r_{\theta\theta} = -re^{-\lambda} && \Gamma^r_{\phi\phi} = -r\sin^2\theta\, e^{-\lambda} \\
&\Gamma^\theta_{r\theta} = \Gamma^\theta_{\theta r} = \tfrac{1}{r} && \Gamma^\theta_{\phi\phi} = -\sin\theta\cos\theta \\
&\Gamma^\phi_{r\phi} = \Gamma^\phi_{\phi r} = \tfrac{1}{r} && \Gamma^\phi_{\theta\phi} = \Gamma^\phi_{\phi\theta} = \cot\theta
\end{aligned} \tag{11.2.3}$$

11.3 Ricci Tensor

We now have to put all the pieces from (11.2.3) into the Ricci tensor

$$R_{\mu\nu} = \Gamma^\alpha_{\alpha\mu,\nu} - \Gamma^\alpha_{\mu\nu,\alpha} + \Gamma^\alpha_{\beta\mu}\Gamma^\beta_{\alpha\nu} - \Gamma^\alpha_{\alpha\beta}\Gamma^\beta_{\mu\nu} \tag{11.3.1}$$

Further, on inserting (11.1.4) in (8.5.2) we have

$$(\ln\sqrt{\;}) = \tfrac{1}{2}(\nu+\lambda) + 2\ln r + \ln|\sin\theta| \tag{11.3.2}$$

With all these substitutions, the Ricci tensor is

$$\begin{aligned}
R_{tt} &= -\tfrac{1}{2}e^{\nu-\lambda}\left(\nu'' + \tfrac{1}{2}\nu'^2 - \tfrac{1}{2}\lambda'\nu' + 2\frac{\nu'}{r}\right)\\
R_{rr} &= \tfrac{1}{2}\left(\nu'' + \tfrac{1}{2}\nu'^2 - \tfrac{1}{2}\lambda'\nu' - 2\frac{\lambda'}{r}\right)\\
R_{\theta\theta} &= -1 + \tfrac{1}{2}e^{-\lambda}r\left(\nu' - \lambda'\right) + e^{-\lambda}\\
R_{\phi\phi} &= \sin^2\theta\, R_{\theta\theta}
\end{aligned} \tag{11.3.3}$$

Derivation

$$\begin{aligned}
R_{tt} &= -\Gamma^r_{tt,r} + 2\Gamma^r_{tt}\Gamma^t_{tr} - \Gamma^r_{tt}\,(\ln\sqrt{\;})_{,r}\\
R_{rr} &= (\ln\sqrt{\;})_{,rr} - \Gamma^r_{rr,r} + \Gamma^{t\,2}_{rt} + \Gamma^{r\,2}_{rr} + \Gamma^{\theta\,2}_{\theta r} + \Gamma^{\phi\,2}_{\phi r} - \Gamma^r_{rr}\,(\ln\sqrt{\;})_{,r}\\
R_{\theta\theta} &= (\ln\sqrt{\;})_{,\theta\theta} - \Gamma^r_{\theta\theta,r} + \Gamma^r_{\theta\theta}\Gamma^\theta_{\theta r} + \Gamma^\theta_{\theta r}\Gamma^r_{\theta\theta} + \Gamma^{\phi\,2}_{\theta\phi} - \Gamma^r_{\theta\theta}\,(\ln\sqrt{\;})_{,r}
\end{aligned} \tag{11.3.4}$$

□

11.4 Solution for the metric

Outside our spherical mass, we have $T_{\mu\nu} = 0$, and hence the components in (11.3.3) must be equated to zero.

Taking $R_{tt} + R_{rr} = 0$ we get $\lambda' + \nu' = 0$, and thus $\lambda + \nu$ constant. Inspecting (11.1.4) we see that we can set that constant to zero by rescaling t. Thus we have

$$\lambda = -\nu \tag{11.4.1}$$

Inserting (11.4.1) into (11.3.3) we get

$$\begin{aligned} R_{tt} &= -\tfrac{1}{2}\frac{e^{\nu}}{r}\,(re^{\nu})'' \\ R_{rr} &= \tfrac{1}{2}\frac{e^{-\nu}}{r}\,(re^{\nu})'' \\ R_{\theta\theta} &= (re^{\nu})' - 1 \\ R_{\phi\phi} &= \sin^2\theta\, R_{\theta\theta} \end{aligned} \tag{11.4.2}$$

Now it is easy to see that the solution to $R_{\mu\nu} = 0$ is

$$e^{\nu} = e^{-\lambda} = 1 - r_{\mathrm{s}}/r \tag{11.4.3}$$

r_{s} being a constant of integration. In other words, the metric is

$$\boxed{ds^2 = -(1 - r_{\mathrm{s}}/r)dt^2 + \frac{dr^2}{1 - r_{\mathrm{s}}/r} + r^2\big(d\theta^2 + \sin^2\theta\, d\phi^2\big)} \tag{11.4.4}$$

The constant r_{s} is called the **Schwarzschild radius**. It must equal Gm where m is the spherical mass, because we need $g_{tt} = 1 - 2Gm/r$ at large r to reproduce Newtonian dynamics.

11.5 Perihelion precession

We now consider the dynamics of a planet in the Schwarzschild metric, and the first major success of general relativity.

The equations of motion of a planet (assumed too light to alter the metric) are the geodesic equations (11.2.2).

We can integrate three of the four geodesic equations immediately. The third equation says that $\dot{\theta} = 0$ if $\theta = \pi/2$. So we choose our axes such that

$$\theta = \tfrac{1}{2}\pi \tag{11.5.1}$$

initially, and it will stay that way. The third equations says that

$$r^2\dot{\phi} = \text{const} = h \text{ (say)}. \tag{11.5.2}$$

This is Kepler's second law (a planet sweeps equal area in equal time) from classical celestial mechanics, and as we see here it continues to hold in general relativity. The first equation in (11.2.2) says that

$$(1 - r_\text{s}/r)\dot{t} = \text{const} = \gamma \text{ (say)}. \tag{11.5.3}$$

Equations (11.5.1) and (11.5.2) are statements that angular momentum is conserved, while (11.5.3) is a relativistic generalization of energy conservation.

We could continue now with the second equation in (11.2.2), but we get an easier differential equation if we take $(d\tau/d\tau)^2 = 1$ and substitute the metric and the above constants of motion into it:

$$\frac{\gamma^2 - \dot{r}^2}{1 - r_\text{s}/r} - \frac{h^2}{r^2} = 1 \tag{11.5.4}$$

We could integrate (11.5.4), but it is rather awkward. It gets simpler if we (i) change the dependent variable from r to $u = 1/r$, and (ii) change the independent variable from τ to ϕ. Then $\dot{r} = -hu'$ where a prime denotes $d/d\phi$.

Derivation

$$\dot{r} = -u^{-2}\dot{u} = -u^{-2}\dot{\phi}u' = -hu' \tag{11.5.5}$$

□

We get

$$u'^2 + u^2 = \frac{\gamma^2 - 1}{h^2} + \frac{r_\text{s}u}{h^2} + r_\text{s}u^3 \tag{11.5.6}$$

Differentiating, we get two equations: $u' = 0$ (a circular orbit) and the more interesting

$$u'' + u = \tfrac{1}{2}r_\text{s}/h^2 + \tfrac{3}{2}r_\text{s}u^2 \tag{11.5.7}$$

In classical celestial mechanics, the last term does not appear. (Which is still a very decent approximation, because the constant $\frac{3}{2}r_\text{s}$ is typically much smaller than the constant $\frac{1}{2}r_\text{s}/h^2$: $\sim 10^{-6}$ for Mercury) Also, in classical celestial mechanics, the constant $\frac{1}{2}r_{\mathrm{s}}/h^2$ is conventionally written as $1/(a(1-e^2))$. In other words we have the equation

$$u'' + u = a^{-1}(1-e^2)^{-1} \tag{11.5.8}$$

which has the solution

$$r = \frac{1}{u} = \frac{a(1-e^2)}{1-e\cos\phi} \tag{11.5.9}$$

i.e., an ellipse with semi-major axis a, eccentricity e, and the sun at one focus.

We can't solve exactly for the effect of the $\frac{3}{2}r_{\mathrm{s}}u^2$ in (11.5.7) but we can get an approximate solution using perturbation theory.

DIGRESSION: PERTURBATION THEORY There is general technique for approximately solving differential equations like

$$\frac{d^2x}{dt^2} + x = A + \epsilon x^2 \tag{11.5.10}$$

where ϵ is small. (In this digression, x, t, τ have no spacetime significance, they are just variables.) The trick is to expand our variables as power series in ϵ.

$$\begin{aligned} x &= x_0 + \epsilon x_1 \\ \tau &= t(1+\epsilon k_1) \Rightarrow \frac{d}{dt} = (1+\epsilon k_1)\,\frac{d}{d\tau} \end{aligned} \tag{11.5.11}$$

where k_1 is a constant we are free to choose. We can go to any order in ϵ.

We substitute (11.5.11) in (11.5.10). Collecting terms without ϵ gives

$$\frac{d^2x_0}{d\tau^2} + x_0 = A \tag{11.5.12}$$

for which a solution is

$$x_0 = A + B\cos\tau \tag{11.5.13}$$

where B is an integration constant; the other integrations constant has been chosen to eliminate a $\sin\tau$ term.

Next we substitute (11.5.13) and (11.5.11) in (11.5.10), and collect terms of $O(\epsilon)$:

$$\frac{d^2x_1}{d\tau^2} + x_1 - 2k_1B\cos\tau = A^2 + 2AB\cos\tau + B^2\cos^2\tau \tag{11.5.14}$$

Now, terms involving $\cos\tau$ in the equation are dangerous, because they lead to terms like $\frac{1}{2}\tau\sin\tau$ in the solution for x_1, which will make x_1 blow up; terms like $\cos 2\tau$ in the equation only produce periodic terms in the solution. However, if we choose $k_1 = -A$ (as we are free to do) then the dangerous terms cancel.

The solution for x to order ϵ is

$$x = A + B\cos(t - \epsilon At) + \epsilon \times \langle\text{constant and periodic terms}\rangle \tag{11.5.15}$$

We could find the periodic terms easily, and we can go to higher order too. But the main effect of the ϵx^2 term is an apparent slowing down of the unperturbed periodic solution, by a factor $(1-\epsilon A)$. □

Applying perturbation theory we get

$$r = \frac{a(1-e^2)}{1 - e\cos\tilde{\phi}}, \qquad \tilde{\phi} = \phi\left(1 - \frac{\frac{3}{2}r_s}{a(1-e^2)}\right) \tag{11.5.16}$$

which amounts to a forward-precessing ellipse.

Einstein did a version of the above calculation circa 1916 (using an approximate metric—the Schwarzschild solution was still in the future) and found he could account for the observed perihelion shift in Mercury. This was the first major success of his theory.[1] More recently, perihelion precession has been tested much more precisely in binary pulsars, where the general relativistic effect is larger ($\sim 10^{-4}$); Taylor and Hulse received the Physics Nobel prize in 1993 for their work on general relativity in binary pulsars.

11.6 The horizon

The surface $r = r_s$ is a strange place, and the simplest example of something relativists call a **horizon**. First of all, $g_{tt} = 0$, so the trajectory $r = r_s$ is null rather than spacelike. Then, inside $r = r_s$ g_{tt} and g_{rr} change sign, so $r = \text{const}$ is spacelike and all timelike trajectories have $\dot{r} < 0$. An observer inside $r = r_s$ is "carried down to $r = 0$ as surely as you and I are carried into next year."

Although g_{rr} becomes singular at $r = r_s$, there is nothing singular about the horizon itself. Inspecting the Christoffel symbols in (11.2.3), we can see that none of them contain the dangerous $e^\lambda = 1/(1 - r_s/r)$, so none of them or their derivatives becomes singular at $r = r_s$. Thus neither $R_{\alpha\beta\mu\nu}$ nor any contraction of it will be singular. The singularity of g_{rr} is really only a coordinate singularity, and is just a consequence of the fact that Schwarzschild coordinates happen to assign $t = \infty$ to any event on the horizon. To see this, consider the geodesic for a particle falling radially to $r = r_s$. The coordinate time for reaching r_s

$$t(r_s) = \int_{r_0}^{r_s} \frac{dt}{d\tau}\,\frac{d\tau}{dr}\,dr \tag{11.6.1}$$

is infinite, but the proper time

$$\tau(r_s) = \int_{r_0}^{r_s} \frac{d\tau}{dr}\,dr \tag{11.6.2}$$

remains is finite.

> Derivation Taking equations (11.5.3) and (11.5.4), and putting $h = 0$ for radial geodesics, we have
>
> $$\frac{dt}{d\tau} = \frac{\gamma}{1 - r_s/r}, \qquad \frac{dr}{d\tau} = \sqrt{\gamma^2 + r_s/r - 1} \tag{11.6.3}$$
>
> whence clearly the integral (11.6.1) will blow up but the integral (11.6.2) will not. □

[1] Incidentally, you may find the observed precession of Mercury's perihelion variously quoted as 43 arcsec/century, 532 arcsec/century and 5600 arcsec/century. The middle number is the correct one. The large number arises because observations are made from the Earth, whose spin axis is precessing; after correcting for that one gets the middle number. This precession is dominated by perturbation from Jupiter (i.e., classical dynamics), and after subtracting that out one gets 43 arcsec/century, which is the general relativistic effect.

Thus, an observer can pass into a horizon in a finite time (provided they can withstand the tidal forces). But not out again of course.

11.7 Kruskal-Szekeres coordinates

The cure for the coordinate singularity is a new coordinate system, the Kruskal-Szekeres coordinates

$$\begin{aligned} r' &= \sqrt{r/r_\mathrm{s} - 1}\, e^{r/2r_\mathrm{s}} \cosh(t/2r_\mathrm{s}) \\ t' &= \sqrt{r/r_\mathrm{s} - 1}\, e^{r/2r_\mathrm{s}} \sinh(t/2r_\mathrm{s}) \\ &\text{or} \\ r' &= \sqrt{1 - r/r_\mathrm{s}}\, e^{r/2r_\mathrm{s}} \sinh(t/2r_\mathrm{s}) \\ t' &= \sqrt{1 - r/r_\mathrm{s}}\, e^{r/2r_\mathrm{s}} \cosh(t/2r_\mathrm{s}) \end{aligned} \tag{11.7.1}$$

where the upper or lower pair is adopted according to the sign of $r/r_\mathrm{s} - 1$. The interval in these coordinates is

$$ds^2 = \frac{4{r_\mathrm{s}}^3}{r}\, e^{r/2r_\mathrm{s}} \left(- dt'^2 + dr'^2 \right) + r^2 \left(d\theta^2 + \sin^2\theta\, d\phi^2 \right) \tag{11.7.2}$$

Note that in (11.7.2) r is not a coordinate but a function of t', r'.

Derivation Differentiating (11.7.1) gives

$$\begin{aligned} dr' &= \frac{e^{r/2r_\mathrm{s}}}{2r_\mathrm{s}\sqrt{r/r_\mathrm{s} - 1}} \Big((r/r_\mathrm{s} - 1)\sinh(t/2r_\mathrm{s})dt + r/r_\mathrm{s}\cosh(t/2r_\mathrm{s})dr \Big) \\ dt' &= \frac{e^{r/2r_\mathrm{s}}}{2r_\mathrm{s}\sqrt{r/r_\mathrm{s} - 1}} \Big((r/r_\mathrm{s} - 1)\cosh(t/2r_\mathrm{s})dt + r/r_\mathrm{s}\sinh(t/2r_\mathrm{s})dr \Big) \\ &\text{or} \\ dr' &= \frac{e^{r/2r_\mathrm{s}}}{2r_\mathrm{s}\sqrt{1 - r/r_\mathrm{s}}} \Big((1 - r/r_\mathrm{s})\cosh(t/2r_\mathrm{s})dt - r/r_\mathrm{s}\sinh(t/2r_\mathrm{s})dr \Big) \\ dt' &= \frac{e^{r/2r_\mathrm{s}}}{2r_\mathrm{s}\sqrt{1 - r/r_\mathrm{s}}} \Big((1 - r/r_\mathrm{s})\sinh(t/2r_\mathrm{s})dt - r/r_\mathrm{s}\cosh(t/2r_\mathrm{s})dr \Big) \end{aligned} \tag{11.7.3}$$

and then taking $-dt'^2 + dr'^2$ we recover the first two terms in the Schwarzschild interval. □

There is no singularity at $r = r_\mathrm{s}$, but there *is* a singularity at $r = 0$.

Figure 11.1 shows a spacetime diagram in Kruskal-Szekeres coordinates and it nicely summarizes Schwarzschild geometry.

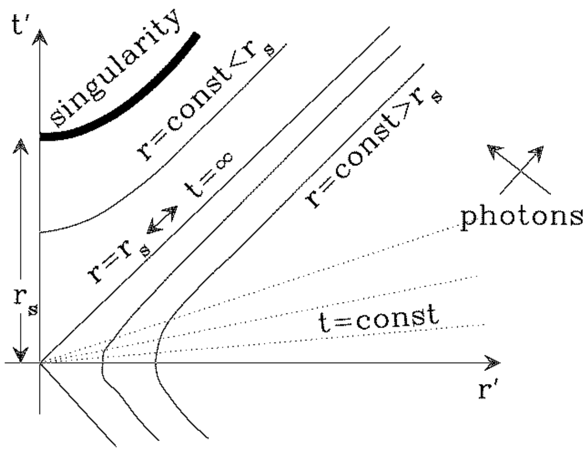

Figure 11.1: Spacetime diagram with Kruskal-Szekeres coordinates. Light rays are always at 45° to the axes on this plot. (This is true in all special-relativity spacetime diagrams but not always in general relativity.) Lines of constant r are hyperbolae; these are timelike if $r > r_\mathrm{S}$, spacelike if $r < r_\mathrm{S}$. The hyperbola $r = 0$ is a singularity and is 'the end of spacetime'. Lines of constant t are straight lines through the origin. The limit of the constant-r and constant-t curves is the horizon.

11.8 Black holes

The concept of a horizon—a surface on which $d\tau^2 = 0$ and which is one-way permeable by timelike geodesics—is quite general and does not require spherical symmetry or time-independence. The presence of a horizon is sort of a formal definition of a black hole.

The most general *stationary* black hole—at least if there is no charge—has the so-called Kerr metric, which is a rotating generalization of the Schwarzschild metric. A Kerr black hole has several features not present in a Schwarzschild bkac hole, notably an **ergosphere**, a region outside the horizon where all timelike geodesics are rotating.

There is a remarkable connection between horizons and thermodynamics. The two most important aspects of this were discovered by Hawking.

The first is Hawking's area theorem, which says that within general relativity the area of a horizon can only increase, never decrease. So two black holes can merge and produce a new horizon larger than either, but a black hole cannot split. This leads to the interpretation of the area of the horizon as a sort of entropy.

The second aspect is about how the area of a black hole *can* decrease, which is through quantum field theory. The effect originates in vacuum fluctuations. A vacuum in quantum field theory is a very active place, where virtual particles are continually created and destroyed. These particles cannot be directly observed because their energy times their lifetime is below the bounds given by the uncertainty principle

$$\Delta E\,\Delta t \le h/(4\pi) \tag{11.8.1}$$

where h is Planck's constant. If a pair of virtual particles is created just outside the horizon, one of them may fall in before they annihilate each other again; the other particle then has a chance to escape and become a real particle. The effect is of particles quantum tunnelling out through the horizon. For a black hole of mass M, the characteristic energy of particles escaping in this way is

$$\left(\frac{m_\mathrm{Pl}}{4\pi}\right)^2 \frac{1}{M} \tag{11.8.2}$$

where $m_{\rm Pl} = \sqrt{hc/G} = 5.46 \times 10^{-5}$g is known as the Planck mass.

> HEURISTIC DERIVATION We take the maximum distance a virtual pair can separate in the lifetime allowed by (11.8.1) as half the circumference
>
> $$\tfrac{1}{2}\Delta t = 2\pi GM \tag{11.8.3}$$
>
> and take ΔE as the typical energy. □

Hawking actually found that black holes should radiate with a temperature corresponding to the characteristic energy (11.8.2).

12. Cosmology

In which General Relativity predicts an expanding universe.

12.1 The Cosmological Principle

The cosmological principle is that on large-enough scales and at a given time, the universe looks the same everywhere and there are no distinguished directions, i.e., the universe is **homogeneous** and **isotropic**. But the universe evolves with time.

This is a physical assumption. It could be wrong, but there is a fair body of observational evidence that it is true. Hence it is a working hypothesis for most work in cosmology at present.

What exactly 'large-enough' means is still uncertain, but it appears to be $\gtrsim 10^8$ light-years. On smaller scales, there are clear structures: clusters of galaxies, galaxies, stars and so on. But the average properties (density, temperature, curvature, and so on) over large volumes appear to be the same.

12.2 The Robertson-Walker Metric

The cosmological principle implies that the metric can be put in a rather simple form. First we can set $g_{00} = 1$; it cannot depend on space, and any time-dependence we can eliminate by redefining the time coordinate. Then we can set $g_{0i} = 0$; this quantity specifies a spatial direction and if it were impossible to set it to zero that would imply a distinguished direction, contradicting isotropy. Also, homogeneity implies that the time dependence is the same everywhere, so we can factor out the spatial and time dependencies of the rest of the metric, thus

$$ds^2 = -dt^2 + a^2(t)\,\tilde{g}_{ij}(\mathbf{x})\,dx^i\,dx^j \tag{12.2.1}$$

where $a(t)$ is called the **scale factor** and $\tilde{g}_{ij}$ (and other symbols with tildes later in this chapter) refer to 3D space. Further invoking isotropy at the origin, we can use spherical polar coordinates and write

$$ds^2 = -dt^2 + a^2(t)\left(e^{\lambda(r)}\,dr^2 + r^2\,d\theta^2 + r^2\sin^2\theta\,d\phi^2\right) \tag{12.2.2}$$

Requiring the curvature to be constant further restricts the metric to the **Robertson-Walker** form

$$\boxed{ds^2 = -dt^2 + a^2(t)\left(\frac{dr^2}{1-kr^2} + r^2\,d\theta^2 + r^2\sin^2\theta\,d\phi^2\right)} \tag{12.2.3}$$

Here k is a constant; we could make it one of $-1, 0, 1$ if we want to, by rescaling r.

Derivation Consider the case of $a(t) = 1$. We then have just the spherically symmetric metric (11.1.4) with $\nu = 0$. Putting $\nu = 0$ in the corresponding Ricci tensor (11.3.3) and raising one index, we have

$$\begin{aligned} R^t{}_t &= 0 \\ R^r{}_r &= -e^{-\lambda}\,\frac{\lambda'}{r} \\ R^\theta{}_\theta &= R^\phi{}_\phi = e^{-\lambda}\left(\frac{1}{r^2} - \frac{1}{2}\frac{\lambda'}{r}\right) - \frac{1}{r^2} \end{aligned} \tag{12.2.4}$$

From this, the Ricci scalar is

$$R = \frac{2}{r^2}\left(\left(re^{-\lambda}\right)' - 1\right) \tag{12.2.5}$$

This must be constant in space, and we set it to $-6k$. Integrating (12.2.5) gives

$$e^{\lambda} = \frac{1}{1 - kr^2 + C/r} \tag{12.2.6}$$

The integration constant C must be zero to avoid a singularity at the origin. Substituting back in (12.2.4) gives

$$R^i{}_j = -2k\,\delta^i_j \tag{12.2.7}$$

Below we will call this $\tilde{R}^i{}_j$, as it refers to a subspace of constant t. □

12.3 Cosmological redshift

In our universe the scale factor $a(t)$ is growing with time. We infer this because of an observational consequence: cosmological redshift.

Light emitted at time t_{emit} and observed elsewhere at time t_{obs} has its wavelength increased by a factor

$$\frac{\lambda_{\text{obs}}}{\lambda_{\text{emit}}} = \frac{a(t_{\text{obs}})}{a(t_{\text{emit}})} \tag{12.3.1}$$

The factor in (12.3.1) is conventionally written as $1 + z$, and z is called the redshift.

Derivation Without loss of generality we can put the observer at the origin. This means that observed light travels at constant θ, ϕ. Say the emitter is at $r = r_{\text{emit}}$.

Consider one wave crest in the light ray. Since $ds^2 = 0$ for light

$$\int_{t_{\text{emit}}}^{t_{\text{obs}}} \frac{dt}{a(t)} = \int_0^{r_{\text{emit}}} \frac{dr}{\sqrt{1 - kr^2}} \tag{12.3.2}$$

For the next wave crest

$$\int_{t_{\text{emit}}+\delta t_{\text{emit}}}^{t_{\text{obs}}+\delta t_{\text{obs}}} \frac{dt}{a(t)} = \int_0^{r_{\text{emit}}} \frac{dr}{\sqrt{1 - kr^2}} \tag{12.3.3}$$

Since the right hand sides of (12.3.3) and (12.3.2) are the same, the left hand sides must be equal, hence

$$\int_{t_{\text{emit}}+\delta t_{\text{emit}}}^{t_{\text{emit}}} \frac{dt}{a(t)} = \int_{t_{\text{obs}}+\delta t_{\text{obs}}}^{t_{\text{obs}}} \frac{dt}{a(t)} \tag{12.3.4}$$

Assuming $a(t)$ does not change much in δt_{emit} or δt_{obs} we have

$$\frac{\delta t_{\text{emit}}}{a(t_{\text{emit}})} = \frac{\delta t_{\text{obs}}}{a(t_{\text{obs}})} \tag{12.3.5}$$

whence (12.3.1) follows because the speed of light remains unity. □

12.4 Ricci tensor

Under the cosmological principle, the Ricci tensor has only two independent components.

$$\begin{aligned} R_{tt} &= 3\frac{\ddot{a}}{a} \\ R_{ij} &= -\left(\frac{\ddot{a}}{a} + 2\frac{\dot{a}^2}{a^2} + 2\frac{k}{a^2}\right) g_{ij} \end{aligned} \tag{12.4.1}$$

Derivation From the condition

$$\int \left(-\dot{t}^2 + a^2\tilde{g}_{ij}\dot{x}^i\dot{x}^j\right) d\tau \quad \text{stationary} \tag{12.4.2}$$

we write down the geodesic equations

$$\begin{aligned} &\ddot{t} + \dot{a}a\dot{x}^i\dot{x}^j = 0 \\ &\tilde{g}_{kj}\ddot{x}^j + 2(\dot{a}/a)\tilde{g}_{kj}\dot{t}\dot{x}^j + \left(\tilde{g}_{kj,l} - \tfrac{1}{2}\tilde{g}_{lj,k}\right)\dot{x}^l\dot{x}^j = 0 \end{aligned} \tag{12.4.3}$$

and (after multiplying the second of these by $\tilde{g}^{ik}$) we read off the nonzero Christoffel symbols

$$\Gamma^t_{ij} = \dot{a}a\tilde{g}_{ij} \qquad \Gamma^i_{tj} = \frac{\dot{a}}{a}\delta^i_j \qquad \Gamma^i_{lj} = \tilde{\Gamma}^i_{lj} \tag{12.4.4}$$

Substituting in the Ricci tensor

$$R_{\mu\nu} = \Gamma^\alpha_{\alpha\mu,\nu} - \Gamma^\alpha_{\mu\nu,\alpha} + \Gamma^\alpha_{\beta\mu}\Gamma^\beta_{\alpha\nu} - \Gamma^\alpha_{\mu\nu}\Gamma^\beta_{\beta\alpha} \tag{12.4.5}$$

we get

$$\begin{aligned} R_{tt} &= \Gamma^\alpha_{\alpha t,t} - \Gamma^\alpha_{tt}{}_{,\alpha} + \Gamma^\alpha_{\beta t}\Gamma^\beta_{\alpha t} - \Gamma^\alpha_{tt}\Gamma^\beta_{\beta\alpha} \\ &= \Gamma^i_{it,t} + \Gamma^i_{jt}\Gamma^j_{it} = 3(\dot{a}/a)_{,t} + 3(\dot{a}/a)^2 = 3\ddot{a}/a \\ R_{ti} &= \Gamma^\alpha_{\alpha t,i} - \Gamma^\alpha_{ti}{}_{,\alpha} + \Gamma^\alpha_{\beta i}\Gamma^\beta_{\alpha t} - \Gamma^\alpha_{ti}\Gamma^\beta_{\beta\alpha} \\ &= \Gamma^j_{li}\Gamma^l_{jt} - \Gamma^l_{ti}\Gamma^j_{jl} = \left(\Gamma^j_{li}\delta^l_j - \delta^l_i\Gamma^j_{jl}\right)(\dot{a}/a) = 0 \\ R_{ij} &= \tilde{R}_{ij} - \Gamma^t_{ij,t} + \Gamma^t_{li}\Gamma^l_{tj} + \Gamma^l_{ti}\Gamma^t_{lj} - \Gamma^t_{ij}\Gamma^l_{lt} \\ &= \tilde{R}_{ij} - (\dot{a}a\tilde{g}_{ij})_{,t} + \dot{a}a\tilde{g}_{li}\frac{\dot{a}}{a}\delta^l_j + \frac{\dot{a}}{a}\delta^l_i\dot{a}a\tilde{g}_{lj} - 3\dot{a}a\tilde{g}_{ij}\frac{\dot{a}}{a} \\ &= \tilde{R}_{ij} - (\ddot{a}a + 2\dot{a}^2)\tilde{g}_{ij} \end{aligned} \tag{12.4.6}$$

Also, from (12.2.7) $\tilde{R}_{ij} = -2k\tilde{g}_{ij}$. □

12.5 Energy-momentum tensor

Under the cosmological principle, the energy-momentum tensor must be of the form

$$T^{\mu}{}_{\nu} = \begin{pmatrix} -\rho & & & \\ & p & & \\ & & p & \\ & & & p \end{pmatrix} \tag{12.5.1}$$

This is because spatial isotropy implies that the spatial part must be a multiple of δ^i_j, and that $T^i{}_t = 0$. Moreover ρ and p can depend only on t. Of course $\rho(t)$ and $p(t)$ are just functions, but the form of (12.5.1) reminds us of the perfect fluid in (3.7.5). We can expect that ρ and p will turn out to be density and pressure.

Because of homogeneity, there is only one nontrivial component to the equation $T^{\beta}{}_{\alpha,\beta} = 0$, and that is for $\alpha = 0$. This equation simplifies to

$$\frac{d}{da}\left(a^3\rho\right) = -3a^2 p \tag{12.5.2}$$

Derivation

$$\begin{aligned} T^{\beta}{}_{t;\beta} &= T^{\beta}{}_{t,\beta} - \Gamma^{\gamma}_{\beta t} T^{\beta}{}_{\gamma} + \Gamma^{\beta}_{\beta\gamma} T^{\gamma}{}_{t} \\ &= T^{t}{}_{t,t} - \frac{\dot a}{a}\delta^i_j T^j{}_i + \frac{\dot a}{a}\delta^i_i T^t{}_t \\ &= -\dot\rho - 3\frac{\dot a}{a}p - 3\frac{\dot a}{a}\rho \end{aligned} \tag{12.5.3}$$

where in the second line we have used the Christoffel symbols (12.4.4) and the fact that $T^{\beta}{}_{\alpha}$ is diagonal. Equating to zero gives (12.5.2). □

The relation between ρ and p depends on what is in the universe.

(i) If it is dust (matter) dominated then $p = 0$ and (12.5.2) gives

$$\rho \propto a^{-3} \tag{12.5.4}$$

(ii) If it is radiation dominated, then $p = \frac{1}{3}\rho$ (see the energy-momentum tensor for electromagnetism) and

$$\rho \propto a^{-4} \tag{12.5.5}$$

(iv) A third possibility is that the universe is dominated by a weird thing called 'vacuum energy', which has $p = -\rho$. Note that pressure is negative in this case. It gives

$$\rho = \text{const} \tag{12.5.6}$$

For historical reasons, this third case is sometimes known as the cosmological constant.

12.6 The Friedmann equation

The energy momentum tensor (12.5.1) and the Ricci tensor (12.4.1) both have two independent components. We can now combine them in the field equations. (If they had unequal numbers of independent components, it would mean that the cosmological principle was inconsistent with the field equations.)

The field equations reduce to the single equation

$$\dot{a}^2 + k = \frac{8\pi G}{3}\rho a^2 \tag{12.6.1}$$

This is known as the Friedmann equation. Friedmann developed the idea of an evolving universe circa 1922.

Derivation From (12.5.1) we can express $T_{\mu\nu} - \frac{1}{2}Tg_{\mu\nu}$ as

$$\begin{aligned} T_{00} - \tfrac{1}{2}Tg_{00} &= \tfrac{1}{2}(\rho + 3p) \\ T_{ij} - \tfrac{1}{2}Tg_{ij} &= \tfrac{1}{2}(\rho - p)g_{ij} \end{aligned} \tag{12.6.2}$$

Combining (12.6.2) and (12.4.1) gives

$$\begin{aligned} 3\frac{\ddot{a}}{a} &= -4\pi G(\rho + 3p) \\ \frac{\ddot{a}}{a} + 2\frac{\dot{a}^2}{a^2} + 2\frac{k}{a^2} &= 4\pi G(\rho - p) \end{aligned} \tag{12.6.3}$$

and we can eliminate $\ddot{a}$ between these two equations. □

It is usual to define a ‘critical density’

$$\rho_{\text{crit}} = \frac{3}{8\pi G}\left(\frac{\dot{a}}{a}\right)^2 \tag{12.6.4}$$

in which case (12.6.1) becomes

$$\dot{a}^2 + k = \frac{\rho}{\rho_{\text{crit}}}\dot{a}^2 \tag{12.6.5}$$

This shows that k is negative, zero, or positive (and recall that we can make one of $-1, 0, 1$ by choosing units of length suitably) according to whether ρ is less than, equal to, or greater than ρ_{crit}. Thus k, which is a geometrical constant appearing in the metric, is connected to the density ρ.

At present $\dot{a} > 0$, i.e., we live in an expanding universe. The sign of k is not known, but the latest indications are that it is zero.

12.7 Conformal time

Sometimes it is convenient to introduce a variable η [unrelated to the $\eta_{\alpha\beta}$ of special relativity] such that

$$d\eta = a\,dt \tag{12.7.1}$$

and η is called the **conformal time**. The Robertson-Walker metric then takes the form

$$ds^2 = a^2(\eta)\left(-d\eta^2 + \frac{dr^2}{1-kr^2} + r^2\,d\theta^2 + r^2\sin^2\theta\,d\phi^2\right) \tag{12.7.2}$$