# Lecture notes for Spacetime Geometry and General Relativity

2015-2016

## Section 1: Introduction to Gravity and Special Relativity

# 1 Special Relativity and Newtonian Gravity – Introduction.

In 1905 a new theory of gravitation was needed to solve the following two problems:

1. A conceptual and mathematical incompatibility between Einstein's theory of Special Relativity and Newton's theory of gravity.
2. The failure of Newtonian gravity to correctly describe the observed behaviour of the universe, in particular the observed precession of the perihelion of Mercury in its orbit around the Sun.

While the eventual solution, Einstein's theory of General Relativity, was formed by an attempt to reconcile special relativity and Newtonian gravity, the reason it was accepted was its success in describing the orbit of Mercury, for which Newton's theory had failed.

If Mercury were the only planet orbiting the Sun, then its orbit as predicted by Newton's theory of gravity would be a perfect ellipse, with the Sun at one of the two foci of the ellipse, as in figure 1a.

However, Mercury is not the only planet, and the other planets push and pull on Mercury in such a way that its orbit is not an ellipse. The point in its orbit where Mercury is closest to the Sun is called its *perihelion*, and one can draw a line through Mercury and the Sun when Mercury is at perihelion. Then, one sees that after Mercury goes around the Sun once, the point at which it is closest to the Sun has changed, and this line does not point in the same direction; one can measure by how much this line has changed direction by finding the angle between the lines on successive orbits (as in fig. 1b), and when all the observations of Mercury in it its orbit over 100 years were compared, it was seen that this line had in total over those 100 years moved through an angle of $1.555483^\circ$.

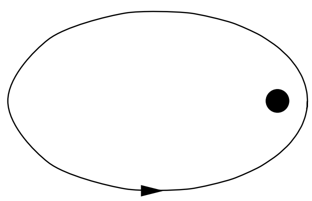


Fig 1a: Unperturbed Newtonian planetary orbit – an ellipse

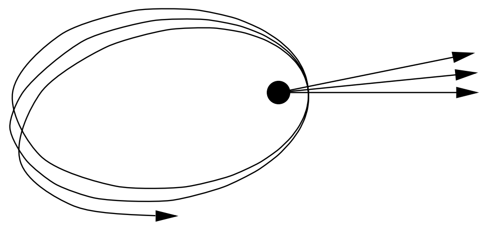


Fig 1b: Actual planetary orbit – close to an ellipse but the perihelion precesses

Since people wanted to check Newton's theory of gravity, they then proceeded to add up all the effects that the other planets would have had on Mercury's orbit during this time, and found that most of this angle was accounted for by the pushing and pulling of the other planets – however they were short by an amount of $0.0119^\circ$, and this was the discrepancy which any new theory of gravity had to account for.

However, simply noticing that a theory disagrees slightly with observation doesn't help in finding a new theory. Much more profitable was trying to reconcile special relativity with Newtonian gravity, and so we shall first review very briefly the main elements of these two theories.

## 1.1 Newton's laws and Newtonian gravity.

Long before Einstein came up with the theory of Special Relativity, Newton had given us his three laws of motion:

**N1** *Free* particles move with constant velocity, or zero acceleration, i.e. with constant speed along straight lines.[1]

**N2** The acceleration of a particle is proportional to the forces on it, with constant of proportionality the inverse of its mass, i.e. $\mathbf{F} = m\,\boldsymbol{a}$, or equivalently the force is equal to the rate of change of the momentum $\mathbf{F} = \dot{\mathbf{P}}$.

**N3** The forces of action and reaction are equal, i.e. if $A$ exerts force $\mathbf{F}$ on $B$, then $B$ exerts force $-\mathbf{F}$ on $A$.

He also gave the formula to find the gravitational force exerted by one body on another, i.e. Newton's law of gravity:

**NG** If a body $A$ has mass $m_A$, and body $B$ of mass $m_B$, and they are separated by a distance $r_{AB}$, then the gravitational force exerted by $B$ on $A$ is of magnitude $G\,m_A\,m_B\,/\,r_{AB}^2$, directed along the line from $A$ to $B$. Or in other words,

$$\mathbf{F}_{AB}^{(\text{gravitational})} = \frac{G\,m_A\,m_B}{r_{AB}^2}\,\hat{\boldsymbol{r}}_{AB}\ . \tag{1}$$

This definition will be sufficient for the discussions in this section, but it will be useful for later to have a slightly more sophisticated treatment.

From the expression (1), the force exerted on a point particle $A$ of mass $m_A$ at position $\boldsymbol{x}$ by a point particle $B$ of mass $m_B$ at $\boldsymbol{y}$ is

$$\mathbf{F}_{AB}^{(\text{gravitational})} = -\,G\,m_A\,m_B\,\frac{\widehat{(\boldsymbol{x}-\boldsymbol{y})}}{|\boldsymbol{x}-\boldsymbol{y}|^2} = -\,G\,m_A\,m_B\,\frac{(\boldsymbol{x}-\boldsymbol{y})}{|\boldsymbol{x}-\boldsymbol{y}|^3}\ .$$

Using (N2), $\mathbf{F} = m\,\ddot{\boldsymbol{x}}$, the acceleration of the particle at position $\boldsymbol{x}$ is due to the gravitational force of the particle at $\boldsymbol{y}$ is

$$\ddot{\boldsymbol{x}} = -\,G\,m_B\frac{(\boldsymbol{x}-\boldsymbol{y})}{|\boldsymbol{x}-\boldsymbol{y}|^3}\ .$$

If instead of a single point particle $B$ of mass $m_B$, we consider a mass distribution of density $\rho(\boldsymbol{y})$, the acceleration of the particle at $\boldsymbol{x}$ as a result of the gravitational attraction of this distribution is the integral

$$\ddot{\boldsymbol{x}} = -\int \mathrm{d}^3y\;\,G\,\rho(\boldsymbol{y})\frac{(\boldsymbol{x}-\boldsymbol{y})}{|\boldsymbol{x}-\boldsymbol{y}|^3}\ . \tag{2}$$

Using the result that

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

[1] Note that this implies the existence of a preferred class of particles, *free particles*, and of a preferred class of coordinate systems or observers, called *inertial coordinate systems*

we see that this can be written as

$$\ddot{\boldsymbol{x}} \;=\; -\nabla\Phi(\boldsymbol{x}) \;=\; \mathbf{g}\,, \quad \text{the gravitational field}\,, \tag{3}$$

where

$$\Phi(\boldsymbol{x}) \;=\; -\int \mathrm{d}^3y \;\frac{G\,\rho(\boldsymbol{y})}{|\boldsymbol{x}-\boldsymbol{y}|}\;. \tag{4}$$

$\Phi(\boldsymbol{x})$ is called the gravitational potential, and we shall henceforth regard equation (3) as the equation defining the acceleration of a particle in the gravitational field of the mass distribution $\rho(\boldsymbol{y})$. The potential is given by equation (4), but using the fact that

$$\nabla^2\left(\frac{1}{|\boldsymbol{x}-\boldsymbol{y}|}\right) = -4\pi\delta^3(\boldsymbol{x}-\boldsymbol{y})\;,$$

where the delta-function is defined by its property that for any region $V$ and function $f(\boldsymbol{x})$,

$$\int_V \mathrm{d}^3y\;\; f(\boldsymbol{y})\,\delta^3(\boldsymbol{x}-\boldsymbol{y}) \;=\; f(\boldsymbol{x}) \;=\; \begin{cases} f(\boldsymbol{x})\,, & \boldsymbol{x}\in V\,, \\ 0\,, & \boldsymbol{x}\notin V\,. \end{cases}\;,$$

so long as the region of integration includes the point $\boldsymbol{x}$, it is equally well given as the solution of the second order differential equation

$$\nabla^2\Phi(x) = 4\pi G\rho(\boldsymbol{x})\;. \tag{5}$$

Later (in section 3) we take these two equations, (3) and (5), as defining Newton's theory of gravity,

$$\begin{aligned} \ddot{\boldsymbol{x}} \;&=\; -\nabla\Phi(\boldsymbol{x})\;, \\ \nabla^2\Phi(\boldsymbol{x}) \;&=\; 4\pi G\rho(\boldsymbol{x})\;. \end{aligned}$$

## 1.2 Special Relativity.

The history of special relativity is not really necessary for us, but it grew out of a need to understand some very strange consequences of Maxwell's equations for the electric and magnetic fields.

They describe, for example, the electric and magnetic fields of charged objects, and the forces felt by charged objects in these fields. So, let us consider a very simplified model of an atom with a charged nucleus creating an electric field and an electron moving in this field. If the nucleus us at rest, the electric field will be spherically symmetric, of size $d$ say, and the electron will orbit in time $T$ say[2].

[2]In fact, an electron will not move in a closed orbit in such an electric field: Maxwell's equations predict that it will eventually spiral into the centre of the atom - it was this problem amongst others which was solved by quantum mechanics. However, if the electron is far from the atom then it will *almost* follow an elliptical orbit and one can make sense of the time $T$ as the time taken to move through an angle of $2\pi$.

However, Maxwell's equations also describe the electric field of a nucleus moving at speed $v$. When you solve for the electric field of the atom you find that it is squashed by a factor $1/\gamma$ in the direction of motion, where

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}} \,. \tag{6}$$

and $c$ is the speed of light. (This is shown schematically in figure 2) Similarly, when you solve for the time it takes an electron to orbit, you find it takes time $\gamma T$, i.e. a factor $\gamma$ longer in the field of the moving atom then in the field of the stationary atom. People spent a lot of

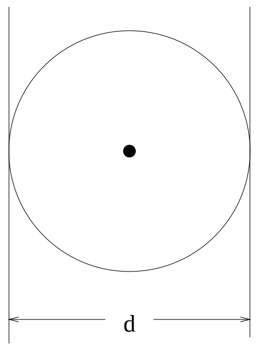



Fig 2a: An atom at rest, of size $d$.

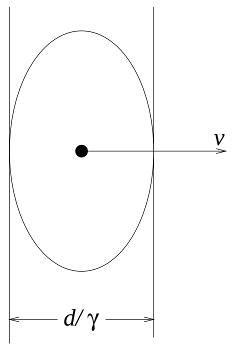



Fig 2b: The same atom moving - the electric field is squashed by a factor $1/\gamma$ in the direction of motion

effort trying to reconcile these facts with the usual Galilean transformation of Newton's laws, by saying that somehow this was only the way things 'appeared'.

Einstein's great breakthrough was to state this was not just a funny quirk of Maxwell's equations but was real – that all moving bodies are actually squashed by a factor $1/\gamma$ in the direction of their motion and that all physical processes run slower by a factor $\gamma$ when an object is moving. If you consider that the objects in the lecture room are essentially governed by (the quantum versions of) Maxwell's equations, then if all the atoms in a table shrink when the table is moving, then the table will also shrink. If the motions of all the electrons and all the atoms in a watch go slower by a factor of $\gamma$, then the watch will also go slower by the same factor.

If you've studied Maxwell's equations before, you would know that you could have derived these basic properties (the relation between stationary and moving systems) using the fact that the equations are invariant under the Lorentz group, or alternatively that the Lorentz transformations are a symmetry of Maxwell's equations. This fact means that the solutions of Maxwell's equations for moving systems can be found from those of static systems by applying these Lorentz transformations, and the general properties of length contraction and time dilation will always arise since they follow directly from the Lorentz transformations.

It will be useful to recall the now exact form of the Lorentz transformations and some other facts about Special Relativity which will be necessary later.

## 1.3 A short review of Lorentz Transformations

Maxwell's equations are invariant under linear transformations of the coordinates $(t, x, y, z) \to (t', x', y', z')$ which leave

$$\Delta^2 = c^2 t^2 - x^2 - y^2 - z^2 \ , \tag{7}$$

invariant. Such transformations are called Lorentz transformations[3].

It is possible to write $\Delta^2$ in a mathematically elegant way. The first observation is that it is much better to consider the combination $x^0 = ct$, as this has dimensions of length, just as the coordinates $x^i$. Then, if we consider

$$x^\mu = (x^0, x^1, x^2, x^3) = (ct, x, y, z) \ ,$$

then we can write $\Delta^2$ as

$$\Delta^2 = \sum_{\mu=0}^{3} \sum_{\nu=0}^{3} \ \eta_{\mu\nu} \, x^\mu \, x^\nu \ , \tag{9}$$

where $\eta_{\mu\nu}$ is the special 4×4 matrix

$$\eta_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}_{\mu\nu} \ . \tag{10}$$

The matrix $\eta_{\mu\nu}$ is called the Lorentz metric and is fundamental to a mathematical formulation of special relativity.

In equation (10), there are two sums over $\mu$ and $\nu$. Such sums occur repeatedly in special and general relativity, and for this reason Einstein developed a convention by means of which we do not have to write such signs explicitly in the majority of cases. This convention is called the *Einstein summation convention* and says that whenever an expression contains an index repeated exactly twice, this means that it is being summed over unless it is stated explicitly that this is not the case. So, using this convention, we can rewrite eqn. (9) as

$$\Delta^2 = \eta_{\mu\nu} \, x^\mu \, x^\nu \ , \tag{11}$$

Using this notation of 4–vectors, Lorentz transformations are linear transformations

$$x'^\mu = \Lambda^\mu{}_\nu \, x^\nu \ , \quad \begin{pmatrix} x'^0 \\ x'^1 \\ x'^2 \\ x'^3 \end{pmatrix} = \begin{pmatrix} \Lambda^0{}_0 & \Lambda^0{}_1 & \Lambda^0{}_2 & \Lambda^0{}_3 \\ \Lambda^1{}_0 & \Lambda^1{}_1 & \Lambda^1{}_2 & \Lambda^1{}_3 \\ \Lambda^2{}_0 & \Lambda^2{}_1 & \Lambda^2{}_2 & \Lambda^2{}_3 \\ \Lambda^3{}_0 & \Lambda^3{}_1 & \Lambda^3{}_2 & \Lambda^3{}_3 \end{pmatrix} \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} \tag{12}$$

of the coordinates which leave

$$\Delta^2 \ = \ \eta_{\mu\nu} \, x^\mu \, x^\nu \ , \tag{13}$$

[3]Maxwell's equations are also invariant under spatial and temporal translations, and the larger group of transformations generated by Lorentz transformations and translations in spacetime are known as *Poincaré transformations*. These take the general form

$$x'^\mu = \Lambda^\mu{}_\nu \, x^\nu + a^\mu \ , \tag{8}$$

where $\Lambda^\mu{}_\nu$ satisfies (15) and $a^\mu$ is a constant.

invariant. This requirement is

$$\begin{aligned} \Delta'^2 &= \Delta^2 \\ \Rightarrow \ \eta_{\mu\nu}\, x'^{\mu}\, x'^{\nu} &= \eta_{\mu\nu}\, \Lambda^{\mu}{}_{\rho}\, x^{\rho}\, \Lambda^{\nu}{}_{\sigma}\, x^{\sigma} = (\eta_{\mu\nu}\, \Lambda^{\mu}{}_{\rho}\, \Lambda^{\nu}{}_{\sigma})\; x^{\rho}\, x^{\sigma} = \eta_{\rho\sigma}\, x^{\rho}\, x^{\sigma}\ , \end{aligned} \tag{14}$$

or

$$(\eta_{\mu\nu}\, \Lambda^{\mu}{}_{\rho}\, \Lambda^{\nu}{}_{\sigma}) = \eta_{\rho\sigma}\ . \tag{15}$$

It is outside the scope of this course, but it can be shown that any transformation satisfying (15) can be decomposed into a series of transformations of two kinds - spatial rotations and boosts[4]. A spatial rotation takes the usual form

$$\begin{aligned} ct' &= ct\ , \\ x'^{i} &= R^{i}{}_{j}\, x^{j}\ , \end{aligned} \tag{16}$$

where $x^i = (x^1, x^2, x^3) = (x, y, z)$ are the spatial coordinates and $R^i{}_j$ is a rotation matrix.

We can think of Lorentz transformations as relating the coordinates of points in space times with respect to the coordinate axes of two different observers. Supposing that we use coordinates $\{ct, x, y, z\}$, and that other observer uses coordinates $\{ct', x', y', z'\}$. We say that we are in a *Lorentz frame* $\mathcal{S}$, and the other observer in a frame $\mathcal{S}'$.

A boost corresponds to changing coordinates to a frame moving at a constant velocity. If the moving observer is moving past us at constant speed $v$ in the positive $x$-direction, so that we both agree that we pass each other at time $t = t' = 0$, and at position $x = x' = 0, y = y' = 0, z = z' = 0$, then this means that our coordinate systems are related by the standard Lorentz boost:

$$\begin{aligned} ct' &= \gamma\,(ct - \frac{v}{c}x) & \qquad ct &= \gamma\,(ct' + \frac{v}{c}x') \\ x' &= \gamma\,(x - \frac{v}{c}\,ct) & x &= \gamma\,(x' + \frac{v}{c}\,ct') \\ y' &= y & y &= y' \\ z' &= z & z &= z' \end{aligned} \tag{17}$$

where $\gamma$ is the constant $\gamma = (1 - v^2/c^2)^{-1/2}$, which is greater than 1 for $v$ non-zero.

### 1.3.1 Inertial Frames

Although the laws of physics in Special Relativity are unchanged by Lorentz transformations and so the idea of an absoulte time is lost, this does not mean that all coordinate systems are equivalent: instead, there is a preferred set of coordinate systems which are related by Lorentz transformations. These are called Inertial Frames.

[4]Technically, to get the full Lorentz group we should also include time reflections $\{t \to -t, x^i \to x^i\}$ and spatial reflections $\{t \to t, x^i \to -x^i\}$. These reflections are unimportant for us, and so we shall not consider them any more. If we consider only what we get by combining by rotations and boosts, we get only a *part* of the Lorentz group, called the *proper orthochronous Lorentz group.* The full Lorentz group is often denoted by $O(3,1)$ and the part we are interested in, the proper orthochronous Lorentz group by $SO(3,1)^{\uparrow}$

The implications of special relativity can be visualised well by considering a physical example of the idea of a frame (or coordinate system) and what the Lorentz transformations imply for the relation between relatively boosted frames.

We can imagine a frame, or coordinate system, as a collection of clocks (which give the time coordinate at each point) distributed along a set of rules (which give the spatial coordinates of the points):

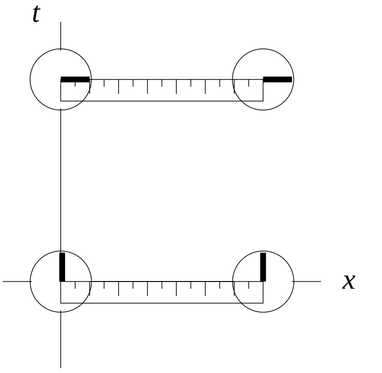



Fig 3: Our coordinate system or frame visualised as a system of clocks and rulers

Equally, we can consider another observer and who has constructed their own frame with their own set of rules and clocks. Let's suppose that we are at rest, and that another observer is at a constant speed $v$ relative to us. The time coordinate $t'$ that the observer assigns to a point in spacetime is the value that their clock at that point shows; the space coordinate $x'$ they give to points in spacetime is the distance along their ruler of that point in spacetime, as shown in figure 4.

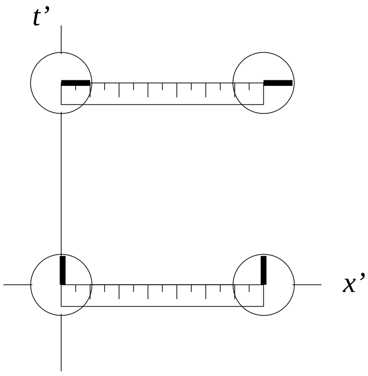



Fig 4: The coordinates of the other observer visualised as a system of clocks and rulers

Supposing that this observer is moving at speed $v$ past us. What do we see? The relation between these two coordinate systems is given by the Lorentz transformations, (17), and using these it is easy to find the coordinates in our frame of particular points in the moving observer's frame, as is shown in figure 5.

In particular, the coordinates of the end of the rod with $x' = 0$ satisfy $x = \gamma v t', t = \gamma t'$ and the end follows the path $x = vt$, while the coordinates of the other end of the rod with $x' = 1$ satisfy $x = \gamma(vt' + 1), t = \gamma(t' + v^2/c^2)$ and the end follows the path $x = vt + 1/\gamma$.

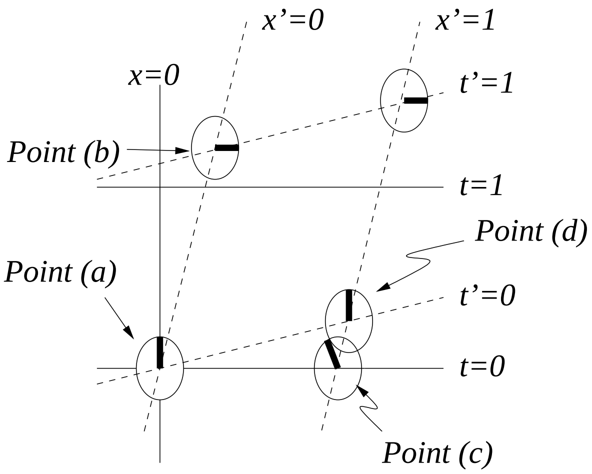



Figure 5

| Coordinates | $(t, x)$ | $(t', x')$ |
|---|---|---|
| Point $(a)$ | $(0, 0)$ | $(0, 0)$ |
| Point $(b)$ | $(\gamma, \gamma\, v)$ | $(1, 0)$ |
| Point $(c)$ | $(0, 1/\gamma)$ | $(-\frac{v}{c^2}, 1)$ |
| Point $(d)$ | $(\gamma\frac{v}{c^2}, \gamma)$ | $(0, 1)$ |

This shows the three main implications of Special Relativity. If two observers are at non-zero speed with respect to each other then:

**Time-dilation:**

Moving clocks go slow by a factor of $\gamma$.

The first clock of the observer in frame $S'$ at coordinate $x' = 0$ shows time 0 at the point $(a)$ with coordinates $\{t' = 0, x' = 0\}$ and time 1 at point $(b)$ with coordinates $\{t' = 1, x' = 0\}$. In frame $S$, these points have coordinates $\{t = 0, x = 0\}$ and $\{t = \gamma, x = \gamma\, v\}$. Thus the moving clock appears in frame $S$ to take $\gamma$ seconds to show a change of 1 second, i.e. it is going slow by a factor $\gamma$.

**Length-contraction:**

Moving rulers oriented along the direction of motion are shortened by a factor $1/\gamma$

The two clocks of the observer in frame $S'$ are unit distance apart in frame $S'$. To find out how far apart they are in our coordinates, we need to find their $x$ coordinates at a given time, say $t = 0$. The first clock has coordinate $x' = 0$ and hence is at coordinates $\{t = \gamma t', x = \gamma v t'\}$. The second clock is at coordinate $x' = 1$, and hence is at coordinates $\{t = \gamma(t' + v/c^2), x = \gamma(1 + vt')\}$. Taking $t = 0$, we see that the first clock is at point $(a)$ and has coordinate $x = 0$ and the second is at point $(c)$ and has $x = 1/\gamma$. Hence the distance between the two clocks at our time $t = 0$ has shrunk from 1 to $1/\gamma$.

**Lack of simultaneity:**

Events which are separated in space in the direction of relative motion which appear simultaneous to one observer do not appear simultaneous to the other — i.e. the idea of objective simultaneity which existed in Newton's theories is lost.

At $t = 0$ the first clock shows time $t' = 0$ whereas the second shows $t' = -v/c^2$. Hence the two points in spacetime $\{t = 0, x = 0\}$ and $\{t = 0, x = 1\}$ which we think are simultaneous are given different time coordinates by the observer in frame $S'$. Equally, the points $\{t' = 0, x' = 0\}$ and $\{t' = 0, x' = 1\}$ which the moving observer thinks happen at the same time happen at times $t = 0$ and $t = v/c^2$ as far as we are concerned.

Although simultaneity and absolute distance are each separately lost, they are replaced by a new idea, that of *proper time*, or the *invariant interval*.

### 1.3.2 Proper time

Given two points $\mathcal{P}$ and $\mathcal{Q}$ at coordinates $(t, x, y, z)$ and $(t', x', y', z')$ respectively, we can consider the combination

$$\begin{aligned}\Delta s^2 &= c^2(t'-t)^2 - (x'-x)^2 - (y'-y)^2 - (z'-z)^2 \\ &= c^s \Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2 \\ &= (\Delta x^0)^2 - (\Delta x^1)^2 - (\Delta x^2)^2 - (\Delta x^3)^2 \\ &= \eta_{\mu\nu}\Delta x^\mu \Delta x^\nu\end{aligned} \quad . \tag{18}$$

This combination has the property (by definition) that it is unchanged under a Lorentz transformation. As a result, all observers will assign the same value of this combination to a pair of points, independent of their velocity.

In particular, we can work this quantity out in the coordinate system of an observer who moves at constant velocity from the point $\mathcal{O}$ with coordinates $(0, 0, 0, 0)$ to a nearby point $\mathcal{P}$ with coordinates $(c\Delta t, \Delta x, \Delta y, \Delta z)$. In their coordinates, the origin still has coordinates $(0, 0, 0, 0)$, but since they regard themselves as stationary, the point $\mathcal{P}$ has coordinates $(\Delta\tau, 0, 0, 0)$. Hence they find

$$\Delta s^2 = c^2 \Delta\tau^2 \ . \tag{19}$$

In this way we see that $\Delta\tau^2 = \Delta s^2/c^2$ is the square of the time taken to travel between two points from the point of view of a person travelling between them at constant velocity. $\Delta\tau$ is called the *proper time* between these two points.

This enables us to divide spacetime into those points $\mathcal{P}$ which are separated from the point $\mathcal{O}$ with coordinates $(0, 0, 0, 0)$ by the 4–vector $x^\mu$ according to whether $\Delta s^2 > 0$, $\Delta s^2 < 0$ and $\Delta s^2 = 0$.

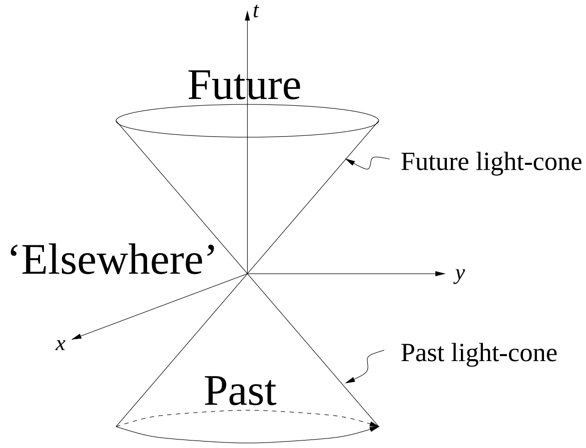
If $\Delta s^2 > 0$ then we say that the point $\mathcal{P}$ is *timelike–separated* from $\mathcal{O}$. The time difference $t$ between the two points can be positive or negative - if it is positive then it is possible to travel forward from the origin to that point, i.e. that point is in the future of the origin. Conversely, if $t < 0$, then that point is in the past of the origin.

If, however, $\Delta s^2 < 0$ then we say that $\mathcal{O}$ and $\mathcal{P}$ are *spacelike–separated*. In this case it is not possible to travel either from the origin to that point, or from that point to the origin. Such points are viewed by some observers to be earlier than the origin, and by some to be later; the truth is that these comparisons are not sensible. Such points are simply neither in the future nor the past of the origin - they are somewhere we can call *elsewhere*. In figure 6 we show these regions.

The surface dividing the future and past from elsewhere has a mathematical and a physical interpretation. Mathematically it is defined by

$$\Delta s^2 = 0 \ .$$

Physically it is the set of points which can be connected to the origin by a particle moving at the speed of light, $c$. For this reason, the set of points with $\Delta s^2 = 0$ and $t > 0$ is known as the *forward* or *future* lightcone; equally those points with $\Delta s^2 = 0$ and $t < 0$ are known as the *backward* or *past* lightcone.




| | |
|---|---|
| Future | $\Delta s^2 > 0, \quad \Delta\tau > 0$ |
| Past | $\Delta s^2 > 0, \quad \Delta\tau < 0$ |
| Elsewhere | $\Delta s^2 < 0$ |

There are a few more bits of nomenclature which it will be good to get straight now. In figure 6, we labelled the various regions as the future, the past, etc. The points in the future of the origin are also said to be *inside the future lightcone* of the origin, and those in the past of the origin are said to be *inside the past lightcone of the origin.*

All points $x^\mu$ such that $\Delta s^2 > 0$ are said to be *timelike separated* from the origin since to some observer, each of these points will lie on the time-axis passing through the origin.

Equally, the points with $\Delta s^2 < 0$ are called *spacelike separated* from the origin and the points with $\Delta s^2 = 0$ are lightlike separated.

These names can also be applied to curves or paths $x^\mu(\lambda)$, which are parametrised by some arbitrary parameter $\lambda$ in the following way. If the path through each point lies *inside* the light cone at that point, this path is said to be *timelike*; if the path through each point lies outside the lightcone, it is said to be *spacelike*; if the path through each point lies actually in the lightcone, it is said to be *lightlike.* An example of each of these types of path is shown in figure 7.

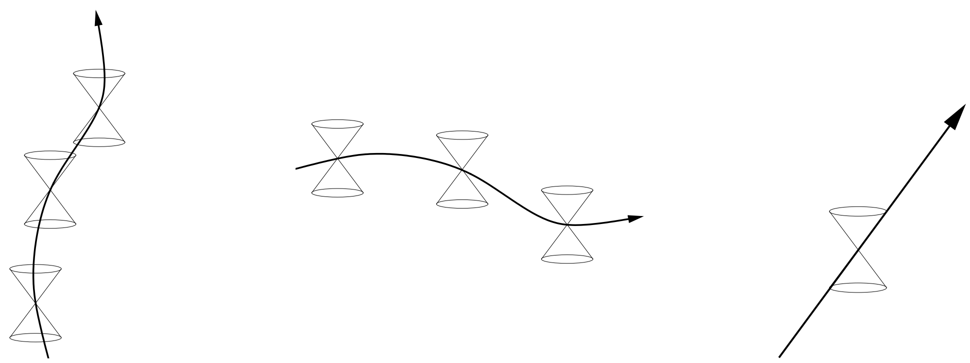


Figure 7a: A timelike path: the path always lies inside the lightcone

Figure 7b: A spacelike path: the path always lies outside the lightcone

Figure 7c: A lightlike path: the path always lies on the lightcone

If a path is the trajectory through spacetime of a massive-particle, then that path must be timelike. The reason for this is that at each point it is possible to transform coordinates to the rest-frame of the particle, that is to the coordinates in which the particle is momentarily at rest. Since in that coordinate system it's instantaneously at rest, its path will lie close to the time-axis at that point and hence is timelike at that point – but since the condition of being timelike or spacelike is invariant under Lorentz transformations, the path will be timelike at that point in all coordinate systems. This will apply equally to all points along the trajectory and so the whole trajectory of a massive-particle is timelike.

Applying similar considerations to massless particles which move at the speed of light, we see that massless particles such as photons always travel along lightlike paths.

We can calculate the proper time along a timelike path (i.e. the time that somebody following this trajectory would measure to pass) by using the infinitesimal version of equations (18) and (19). If along an infinitesimal part of the path the four coordinates change by $\mathrm{d}x^\mu$, then the proper time take $\mathrm{d}\tau$ is given by

$$c^2\mathrm{d}\tau^2 = \mathrm{d}s^2 = c^2\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2 = \eta_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu \ . \tag{20}$$

If we can parametrise this path by some parameter $\lambda$, so that as $\lambda$ varies, the coordinates of the path are $x^\mu(\lambda)$, then

$$\mathrm{d}x^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\,\mathrm{d}\lambda = \dot{x}^\mu\mathrm{d}\lambda \ , \tag{21}$$

and so

$$\mathrm{d}s^2 = \eta_{\mu\nu}\dot{x}^\mu\dot{x}^\nu\mathrm{d}\lambda^2 \ , \tag{22}$$

or

$$c\,\mathrm{d}\tau = \mathrm{d}s = \sqrt{\,\eta_{\mu\nu}\dot{x}^\mu\dot{x}^\nu\,}\;\mathrm{d}\lambda \ . \tag{23}$$

Hence the proper time along a particle trajectory is given by the integral

$$c\,\Delta\tau = \Delta s = \int \sqrt{\,\eta_{\mu\nu}\dot{x}^\mu\dot{x}^\nu\,}\;\mathrm{d}\lambda \ . \tag{24}$$

There is a similar expression for the total proper length $\Delta l$ of a spacelike path, although for this we must remember that $\Delta s^2$ is negative along a spacelike path, and so the formula is

$$\Delta l = \int \mid \eta_{\mu\nu}\dot{x}^\mu\dot{x}^\nu \mid^{1/2}\,\mathrm{d}\lambda \ . \tag{25}$$

### 1.3.3 End-note

One might wonder why it is that we are paying such close attention to ideas from special relativity – the answer is that the Strong Equivalence Principle (which we meet in section 1.5) says that at each point in spacetime we can find coordinates in which locally, i.e. in a small region, the effects of gravity can be made as small as one likes, and that physical properties such as the paths particles can follow are approximately given by the laws of special relativity. So, in general relativity, it is still possible to label paths as timelike, spacelike and lightlike, and massive particles always follow timelike paths and photons always follow lightlike paths.

## 1.4 Trying to compare Newton's laws and Special Relativity.

It is interesting to note that Newton's laws of motion are not all incompatible with special relativity. In particular, N1 holds in SR, so that it is still true in SR that free bodies move in straight lines at constant speeds.

**N2** also holds, provided one takes the second definition $\mathbf{F} = \dot{\mathbf{P}}$ and replaces this with the 4-vector equation $F^\mu = d/d\tau\, P^\mu$. Of course there are some changes – the spatial components of $P^\mu$ are not quite the same as $\mathbf{P}$, and $\dot{} \equiv d/dt$ has been replaced by $d/d\tau$ – but the idea of momentum changing under the action of force remains intact.

The main difference is in N3. To formulate N3, Newton had to assume that forces act instantaneously across space (he seems not to have liked this idea, but that was the best equation he could find). This is only possible for the reason that Newton assumed that there is an objective idea of simultaneity – in this way the forces at two different points at the same time can be equal and opposite. This clearly cannot hold in general in SR, since as we have seen different observers have different ideas of what is simultaneous and what is not.

This shows that the most obvious inconsistency between Newtonian Gravity and SR is that **NG** also assumes the existence of an objective notion of simultaneity. The gravitational force in **NG** is assumed to act instantaneously across space.

However, if different observers have different ideas about what is simultaneous, then they may come up with different ideas about the accelerations of particles under gravity. It may of course happen by some miracle that the motions predicted by the two observers are the same, but that is not the case: Newton's laws are indeed incompatible with special relativity.

This very same problem would have occurred if we had been trying to reconcile the laws of electrostatics with special relativity[5].

The force between static charged particles is given by Coulomb's law: the force between two static bodies $A$ and $B$ with electric charges $q_A$ and $q_B$ takes the same form as the gravitational force, i.e. the force of $B$ on $A$ is given by

$$\mathbf{F}_{AB}^{\text{(electrostatic)}} = -\frac{q_A\, q_B}{4\,\pi\,\epsilon_0\, r_{AB}^2}\, \hat{\boldsymbol{r}}_{AB} \ .$$

We can write both Coulomb's law and **NG** in the same form as

$$\mathbf{F}_{AB} = \text{constant} \times \frac{\hat{\boldsymbol{r}}_{AB}}{r_{AB}^2} \ ,$$

so it appears that it might be possible that **NG** can be reconciled with special relativity in the same manner as Coulomb's law. So, how was Coulomb's law reconciled with special relativity?

In Maxwell's theory, electro-magnetic interactions cope with the lack of an objective idea of simultaneity by saying that the force between two charged particles is carried by the electric and magnetic fields. If the particle $A$ is moved so that the distance between it and $B$ changes, $B$ will only feel this effect after enough time has passed for the electro-magnetic waves (which naturally move at $c$, the speed of light) have had time to travel from $A$ to $B$, as in figure 8.

[5] Of course this was never a problem since special relativity was discovered as a result of the proper unification of electrostatics and other phenomena into Maxwell's equations

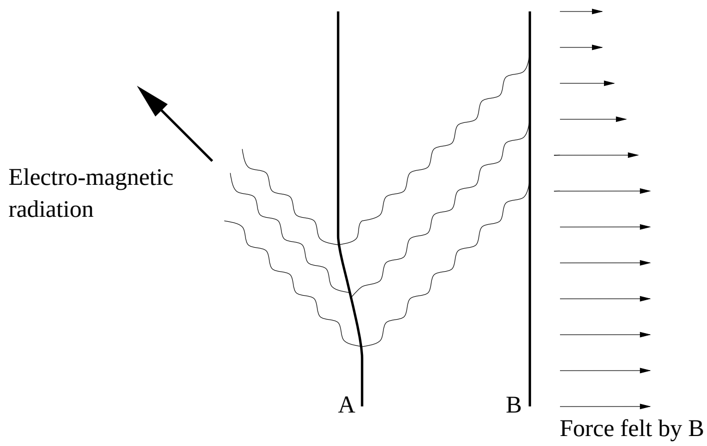



Fig 8: Diagram showing how the force felt by charge $B$ due to charge $A$ is mediated by electro-magnetic radiation, and hence it takes a finite time for the movement of $A$ to be reflected in a change of the force on $B$.

Since Einstein knew about Maxwell's equations and how they solved the problem of transmission of force in electrodynamics, he naturally thought that it might be possible to find a similar solution to the problem in gravitation. He spent several years working in this way, with different numbers and types of *fields* (analogues of the electric and magnetic fields) and was able to find theories of this sort, with gravitational force being carried by some fields on spacetime, and which at least were compatible with Special Relativity. However, they did not predict the observed precession of the perihelion of Mercury correctly! It needed another new idea to be able to make the new sort of mathematical theory, and this was the *principle of equivalence.*

## 1.5 The Equivalence Principle.

The equivalence principle comes in two forms, known as the *weak* (or WEP) and *strong* (or SEP). The weak form was known for some time before Einstein formulated the strong version, as we will see know it really is rather weak.

In Newtonian gravity, there is a coincidence that is often unnoticed, and that is that mass is defined in both **N2** and in **NG**, and it is not obvious that the two definitions should be the same.

We can make it more apparent by introducing different notations, so that mass as used in **N2** will be denoted $m^{(I)}$ for inertial mass, and mass as used in **NG** will be denoted $m^{(G)}$ for gravitational mass. We can now find the acceleration of particle $A$ caused by its gravitational attraction to $B$ in this new notation,

$$m_A^{(I)}\,\ddot{\boldsymbol{x}}_A \;=\; \frac{G\,m_A^{(G)}\,m_B^{(G)}}{r_{AB}^2}\,\hat{\boldsymbol{r}}_{AB}\;. \tag{26}$$

or,

$$\ddot{\boldsymbol{x}}_A \;=\; \left(\frac{m_A^{(G)}}{m_A^{(I)}}\right)\times\frac{G\,m_B^{(G)}}{r_{AB}^2}\,\hat{\boldsymbol{r}}_{AB}\;. \tag{27}$$

So, (if $A$ is so light that we can take $B$ to be static), the acceleration of $A$ in the gravitational field of $B$ depends only upon the ratio $(m_A^{(G)}/m_A^{(I)})$ .

In principle, this factor could differ from one body to another, depending, for example, on the chemical composition, or other factors. Although people firmly believed that this ratio was exactly 1 for all bodies, it still had to be checked. The first person to do this was Eötvös, in Budapest in 1889 (long before SR had been formulated) where he checked that the ratio was 1 to within $10^{-9}$ for such different materials as gold and aluminium. More recent measurements by Dicke et al in 1964 found that the ratio was 1 to within $10^{-11}$. This is the weak equivalence principle:

**Weak Equivalence Principle:** The inertial and gravitational masses of a body, as defined in **N2** and **NG**, are equal for all bodies no matter what their composition.

As a result, the acceleration of a body $A$ in the gravitational field of a body $B$ does not depend on the mass of $A$. However, this independence on the mass of $A$ means that in a uniform gravitational field Newton's equations of motion can be dramatically simplified.

Let's suppose that the gravitational field inside some region (such as the lecture room) is uniform in space and time, so that the acceleration (27) is a constant vector which we denote by $\boldsymbol{g}$. (For the lecture room this is approximately $9{\cdot}8\,ms^{-2}$ vertically down).

Let's also suppose that there are several particles, labelled $i$, with positions $\boldsymbol{x}_i$, which are moving in this region, such that the gravitational forces between them are negligible and that the net force of particle $i$ on $j$ is $\mathbf{F}_{ij}$.

Then we can write the combined result of **N2** and **NG** as

$$m_j^{(I)}\,\ddot{\boldsymbol{x}}_j \;=\; \left(\textstyle\sum_i \mathbf{F}_{ij}\right) \;+\; m_j^{(G)}\,\boldsymbol{g}\;.$$

Now, by the WEP we must take $m^{(I)} \equiv m^{(G)}$, and we find that the formula for the acceleration is

$$\ddot{\boldsymbol{x}}_j \;=\; \frac{1}{m_j}\left(\sum_i \mathbf{F}_{ij}\right) \;+\; \boldsymbol{g}\;.$$

Now a great simplification can be made if we change coordinates to those relative to an observer freely falling in the gravitational field. Such a freely-falling observer would have position

$$\boldsymbol{z} = \boldsymbol{z}_0 + \boldsymbol{v}\,t + \tfrac{1}{2}\boldsymbol{g}\,t^2\;,$$

where $\boldsymbol{z}_0$ and $\boldsymbol{v}$ are constants. Relative to this freely falling observer, the $j$–th particle has position

$$\boldsymbol{y}_j = \boldsymbol{x}_j - \boldsymbol{z} = \boldsymbol{x}_j - \boldsymbol{z}_0 - \boldsymbol{v}\,t - \tfrac{1}{2}\boldsymbol{g}\,t^2\;,$$

so that in terms of $\boldsymbol{y}$, the gravitational field has vanished entirely:

$$\ddot{\boldsymbol{y}}_j \;=\; \ddot{\boldsymbol{x}}_j \;-\; \ddot{\boldsymbol{z}} \;=\; \frac{1}{m_j}\left(\sum_i \mathbf{F}_{ij}\right)\;.$$

This simple result, that changing coordinates can entirely remove the effects of a gravitational field, and the fact that all gravitational fields look constant when restricted to a small enough region, led Einstein to formulate the *strong equivalence principle*:

**Strong Equivalence Principle:** At every spacetime point in an arbitrary gravitational field it is possible to choose a "locally inertial coordinate system" such that, within a sufficiently small region around the point in question, the laws of nature take the same form as in unaccelerated Cartesian coordinate systems in the absence of gravitation.

Note that it is crucially important that the SEP includes the words "within a sufficiently small region around the point in question". The SEP does not mean that gravity can be completely thrown away - clearly it has effects (in the next section we shall see what effect a varying gravitational field has.) What it does mean however is that within small regions one can choose coordinates - locally inertial coordinate systems, also known as local inertial frames or LIFs for short - in which the effects of gravity can be made as small as one likes, as in figure 9

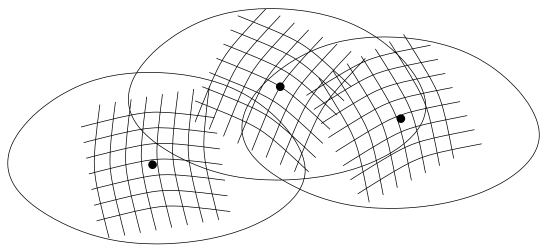


Figure 9: Three points in spacetime with overlapping LIFs (Local Inertial Frames) around each point

## 1.6 How Newtonian gravity shows itself by relative acceleration

We have seen that the effects of a *constant* Newtonian gravitational field can be entirely removed by a change of coordinates - but this is not possible for a varying field. The effects shows up as relative accelerations of neighbouring bodies, as we shall show now.

We first consider a particle moving freely under Newtonian gravity with coordinates

$$X^i(t) = (X^1(t), X^2(t), X^3(t)) \ . \tag{28}$$

If the Newtonian gravitational potential is $\Phi(x)$, then Newton says that the acceleration of the particle is given by:

$$\frac{\mathrm{d}^2 X^i}{\mathrm{d}t^2} = - \left.\frac{\partial \Phi}{\partial x^i}\right|_{x=X} \ . \tag{29}$$

We now want to consider a second particle close to this first particle at coordinates $Y^i(t)$. Again according to Newton's laws its motion satisfies

$$\frac{\mathrm{d}^2 Y^i}{\mathrm{d}t^2} = - \left.\frac{\partial \Phi}{\partial x^i}\right|_{x=Y} \ . \tag{30}$$

The separation of these two particles, $\Delta^i = Y^i - X^i$ then satisfies

$$\ddot{\Delta}^i = -\frac{\partial \Phi}{\partial x^i}(\mathbf{Y}) + \frac{\partial \Phi}{\partial x^i}(\mathbf{X}) \ . \tag{31}$$

If we assume the separation is small, that $Y^i = X^i + \Delta^i$ where $\Delta^i$ is small, then we can use the many variable version of Taylor's theorem to expand the right-hand-side of the equation about the point $X^i$. For a scalar function, this theorem states

$$f(\mathbf{X} + \mathbf{\Delta}) = f(\mathbf{X}) \ + \ \Delta^i \frac{\partial f}{\partial x^i}(\mathbf{X}) \ + \ \frac{1}{2}\Delta^i \Delta^j \frac{\partial^2 f}{\partial x^i \partial x^j}(\mathbf{X}) \ + \ldots \ . \tag{32}$$

The functions that appear in equation 31 are the components of the gradient of $\Phi$, and so Taylor's theorem implies that

$$-\frac{\partial \Phi}{\partial x^i}(\mathbf{X} + \mathbf{\Delta}) = -\frac{\partial \Phi}{\partial x^i}(\mathbf{X}) \ - \ \Delta^j \frac{\partial}{\partial x^j}\frac{\partial \Phi}{\partial x^i}(\mathbf{X}) \ - \ \frac{1}{2}\Delta^j \Delta^k \frac{\partial^2}{\partial x^j \partial x^k}\frac{\partial \Phi}{\partial x^i}(\mathbf{X}) \ + \ldots \ ,$$

where the indices have been changed to avoid repeating the index $i$.

Applying this to the equation for the separation, we get

$$\ddot{\Delta}^i = -\Delta^j \, \frac{\partial^2 \Phi}{\partial x^i \partial x^j}(\mathbf{X}) \ + O(|\mathbf{\Delta}|^2) \ . \tag{33}$$

This shows that a varying gravitational field leads to a relative acceleration of nearby particles: since the field is carying, nearby particles experience different fields and so accelerate differently. There is an exercises on this at the end of the section.

We shall use the formula (33) to motivate the formulation of Einstein's theory of gravity.

## 1.7 Local Inertial frames and the geometric formulation of gravity

The idea and importance of LIFs is the clue to the correct mathematical formalism, which is that of differential geometry. This is a treatment of curved surfaces and spaces formulated by Riemann in which curved surfaces are described in terms of coordinate patches - small regions of the space in which the space is *almost* the same as usual flat space, $\mathbb{R}^n$, but with the property that when these patches are all joined together the result is something interesting.

Einstein's formulation of general relativity is something very similar. At each point we can find LIFs in which the laws of nature are those of an unaccelerated coordinate system and, in particular, in which free particles move approximately in straight lines at constant speed. However, when all these little LIFs are patched together, it turns out that particles have not moved in straight lines, but instead they have been mysteriously affected by gravity.

A close analogy to this is the description of geodesics on curved surfaces: a geodesic is the best approximation to a straight line on a curved surface. For example, on a sphere the geodesics are great circles. If we imagine two particles at the equator of a sphere starting to move North, then initially they are moving parallel. Each particle then tries to follow a straight line as best it can, and as a result follows a great circle, which in this case is a line of constant longitude. However, all these pass through the North pole, so that particles which were trying to go in straight lines end up being 'drawn together' at the North pole by the 'force' of geometry, as in figure 10.

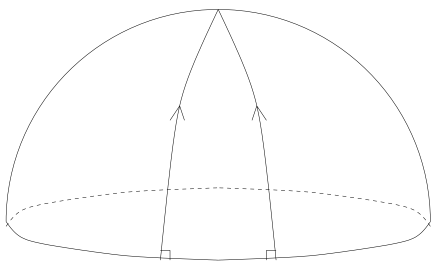


Figure 10a: Two geodesics on a sphere. They each try to follow as straight a path as possible.

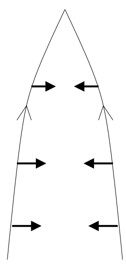


Figure 10b: From the point of view of the paths, they are 'attracted' towards each other by the 'force' of geometry.

This is half of the story of general relativity: particles move dictated by geometry in the simplest possible way. The other half is that the geometry is itself determined by the position of the masses, just as the gravitational potential in Newton's theory is determined by the distribution of masses.

In the rest of the course we shall spend some time learning the mathematics necessary to find the motion of particles that is dictated by the geometry of spacetime. We shall then discuss Einstein's equations which show how the geometry of spacetime is in turn determined by the distribution of the masses, and then in the last third of the course we shall find some simple solutions of Einstein's equations and discuss the motion of particles in the resulting geometries.

It's now time to examine some of the effects that we can discover just by applying the SEP to various situations. These are the bending of light, the gravitational redshift and the gravitational time-dilation effects. To apply the SEP to understand what happens at a point in spacetime, we simply have to find a LIF at that point and use our knowledge of special relativity, hence we first need to have some examples of LIFs to be able to apply this.

## 1.8 Examples of "locally inertial coordinate systems".

One of the reasons why it took so long to formulate the SEP is that we are normally very constrained by gravity, and only in special cases do we adopt such "locally inertial coordinate systems" as natural. These cases are when we are freely falling and not rotating.

Until this century, this experience was limited to people falling off buildings, cliffs, down wells, etc, when they were not usually in a such a condition afterwards as to be able to describe coherently what they experienced.

However, there is now a small group of people for whom this feeling is more normal, and that is astronauts, for whenever they are in a spaceship which is not feeling the resistance of air, or the force of its rockets, or a space station which is not spinning wildly out of control, they are freely falling under the influence of gravity. What do they see? Objects stay where they are put, or continue to move in straight lines at constant speeds with respect to the spaceship.

When dealing with LIFs on the earth, rather than talk about people falling out of planes etc, it is traditional to think of a lift falling freely in a lift shaft. The lift itself provides a visualisation of the spatial coordinates of the LIF.

To see how effective this is, let's see the SEP shows that the path of light must be bent in a gravitational field.

## 1.9 The bending of light from the SEP

Let us consider the following set up: an observer is in a lift which for $t < 0$ is stationary in a lift-shaft. At $t = 0$, simultaneously a photon of enters the lift from the left, and the lift is allowed to fall freely down the lift shaft.

Since the observer in the lift is in free-fall, the photon appears to move in a straight-line across the lift without changing direction or frequency, as in the series of snapshots of the photon's position in figure 11a, and the total trajectory as seen by the observer in the lift in 11b.

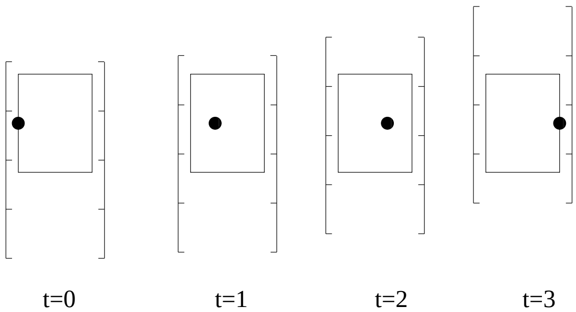



Fig. 11a: The photon moving across the lift as viewed from the lift

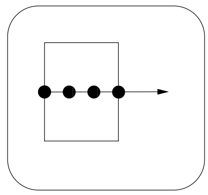


Fig. 11b: The trajectory of the photon as seen in the lift

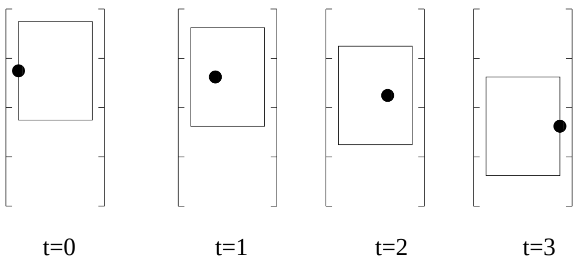



Fig. 11c: The photon moving across the lift as viewed from the lift-shaft

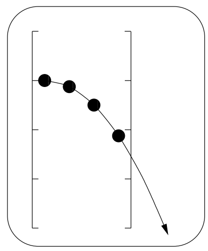


Fig. 11b: The trajectory of the photon as seen from the lift-shaft

However, when viewed by a stationary observer outside the lift, it is clear[6] that the photon appears to be accelerating downwards, as shown in figures 11c and 11d which give the stationary observer's viewpoint of the events in 11a and 11b.

It is important to realised that this result is in fact only *qualitative*. It shows that in each small

[6]There is one important caveat about this argument. We have made a major assumption that a path that is *horizontal* (i.e. at all points perpendicular to the force of gravity) is also *straight* (i.e. a *geodesic*, which we can think of as a straight line in a LIF.) We made this assumption when we drew the lift in the liftshaft. The bottom of the lift is meant to be a straight line in a LIF, but it is an assumption that it will look straight in the coordinate system we have used for the liftshaft, where a line of constant $z$ joining points of equal height in the liftshaft is meant to be horizontal.

It turns out that this assumption is correct in the description of a constant gravitational field in Einstein's theory, but it is not correct in an alternative (and wrong) theory proposed by Nordstrom. Nordstrom's theory also respects the SEP and is geometric, but it is also wrong since in Nordstrom's theory light is not deflected. One way to understand this is that there is a difference between straight lines and horizontal lines. There is a coordinate patch light is deflected, but in order to know by how much starlight is deflected on its way to us as it passes by the sun, we have to know how to add these patches together, as in figure 13.

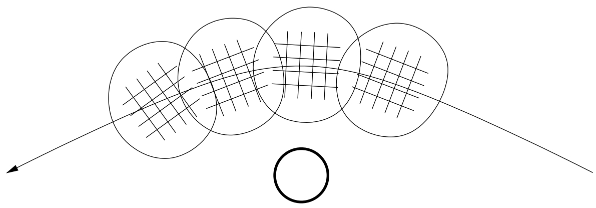


Figure 13: The net effect of the bending of light as it passes by a massive body is given by the result of adding together several small patches. There is more than one way to do this - Einstein gave a prescription which works, as far as we can tell

In Einstein's theory the result is that light is deflected by more than one might guess based on the argument above, the reason being that we have neglected some of the the curvature of spacetime.

One can also view this as an explanation of how light is not deflected in Nordstrom's theory which nevertheless obeys the SEP - the small LIFs are added together in such a way that the bending effects in each little patch all cancel out and after moving through the gravitational field of a planet, the net effect is exactly the same as if the light had gone in a straight line.

So it is very important to measure this effect, and when this was done Einstein's theory was shown to agree with nature. Later in this course we shall calculate exactly by how much light is bent as it passes a massive object.

In this context, the sun is not very heavy - a much more dramatic example of the bending of light by gravitational sources is the effect of *gravitational lensing* – the very large collections of matter in distant galaxies can be so large that the light of sources further away can be bent quite dramatically.

One result is that we can often see more than one image of a distant light source such as a bright galaxy, because a galaxy close to us has bent the light. Figure 14 gives a sketch of this

sketch of how the LIF and liftshaft appear in Nordstrom's theory in figure 12.

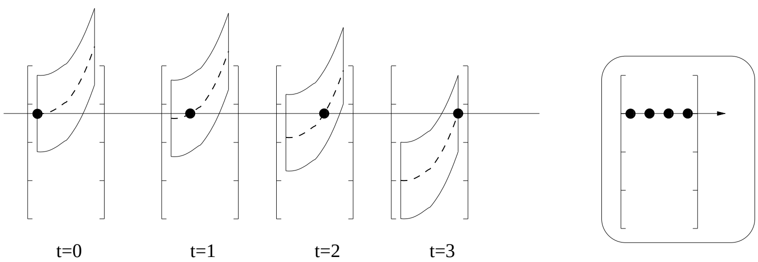
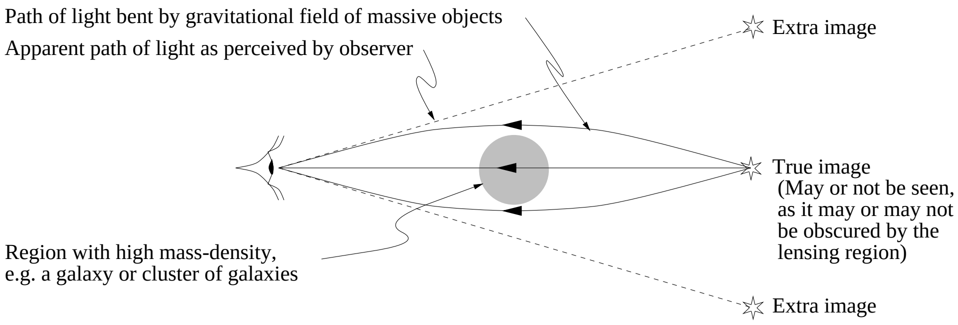



Fig. 12a: In Nordstrom's theory, the LIF is embedded in spacetime in such a way that the photon is not deflected by gravity. The lift falls but in such a way that the photon still moves in a straight line in the liftshaft

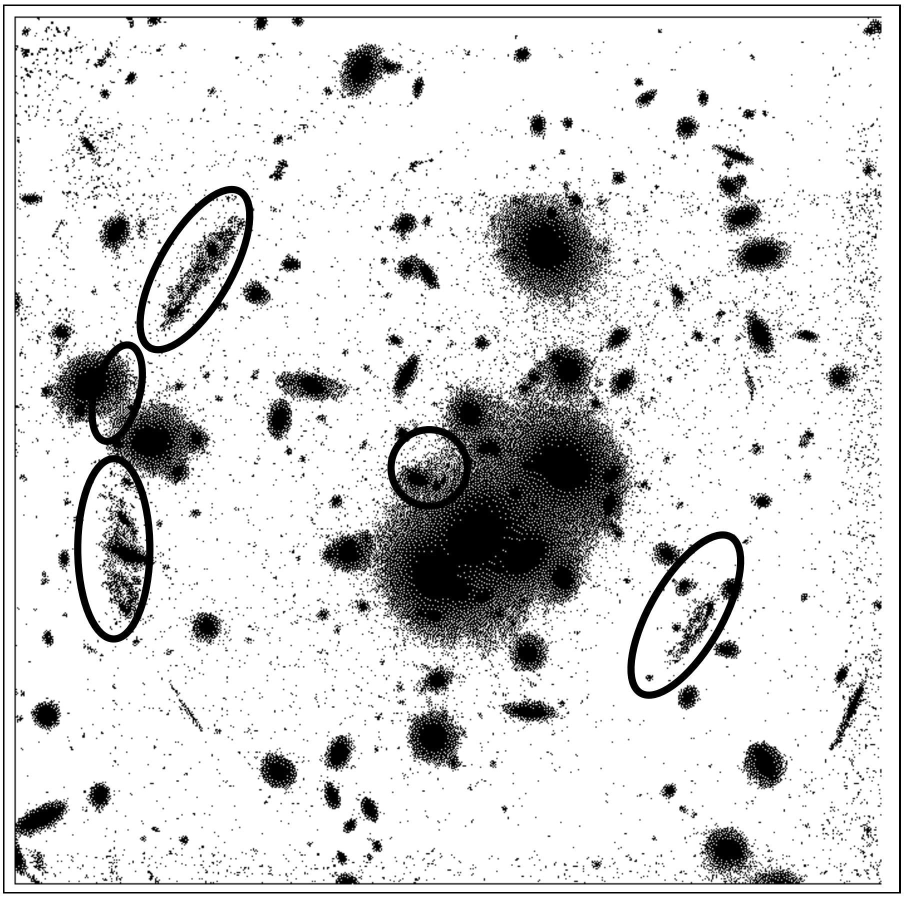
Fig. 12b: The trajectory of the photon as seen in the lift phenomenon.

Figure 14:




Some very dramatic examples of this phenomenon have been found using the Hubble space telescope. The next page has some details of one of these gravitational lenses, although the pictures on the course web page `http://www.mth.kcl.ac.uk/courses/cm334`, and on the sites linked to these pages, look much better.

The next phenomenon we will look at is *gravitational red-shift*



Hubble Space Telescope press release:
This Hubble Space Telescope image shows several loop-shaped objects that actually are multiple images of the same galaxy. They have been duplicated by the gravitational lens of the cluster of elliptical and spiral galaxies - called 0024+1654 - near the photograph's centre. The gravitational lens is produced by the cluster's tremendous gravitational field that bends light to magnify, brighten and distort the image of a more distant object. How distorted the image becomes and how many copies are made depends on the alignment between the foreground cluster and the more distant galaxy, which is behind the cluster.
In this photograph, light from the distant galaxy bends as it passes through the cluster, dividing the galaxy into five separate images (ringed). The light also has distorted the galaxy's image from a normal spiral shape into a more arc-shaped object. Astronomers are certain the loop-shaped objects are copies of the same galaxy because the shapes are similar. The cluster is 5 billion light-years away in the constellation Pisces, and the loop-shaped galaxy is about 2 times farther away. Though the gravitational light-bending process is not new, Hubble's high resolution image reveals structures within the loop-shaped galaxy that astronomers have never seen before. Some of the structures are as small as 300 light-years across. The bits embedded in the loop-shaped galaxy represent young stars; the core inside the ring is dust, the material used to make stars. This information, together with its blue colour and unusual "lumpy" appearance, suggests a young, star-making galaxy.
The picture was taken October 14, 1994 with the Wide Field Planetary Camera-2. Separate exposures in blue and red wavelengths were taken to construct this (colour) picture.
Credit: W.N. Colley and E. Turner (Princeton University), J.A. Tyson (Bell Labs, Lucent Technologies) and NASA

### 1.10 Gravitational Red-shift.

This is known as the gravitational redshift effect as it says how the frequency of light changes as it passes though a gravitational field. Photons can be thought of as small oscillations of the electric and magnetic fields moving at the speed of light and of a certain frequency $\nu$ and wavelength $\lambda$ satisfying $\nu\lambda = c$. The colour of visible light is determined by its wavelength: red light has longer wavelength and lower frequency than blue light. Hence, if the frequency of light decreases as it passes through a gravitational field, we say that it is *redshifted*. Conversely, if the frequency is increased and wavelength decreased, the light is said to be *blueshifted*.

There is a nice line of reasoning leading to the gravitational red-shift through the SEP, but since it is a little lengthy, we shall leave it to a question on the exercise sheets, and instead give a different simple argument based on conservation of energy that was found by Einstein in 1911.

To understand Einstein's argument we have to accept the following:

• The energy of photons is quantised, so that each photon of frequency $\nu$ has energy $h\,\nu$, where $h$ is Planck's constant.

• It is possible, in principle, to convert a photon of energy $E$ to a particle of mass $m = E/c^2$, and vice-versa, with no loss or gain of energy in the process.

• We can define a static gravitational field through a gravitational potential defined at all points. A particle of mass $m$ will gain kinetic energy $m\,g\,H$ if it falls through a distance $H$ in a gravitational field of strength $g$. In general, if it falls freely from a point with gravitational potential $\Phi_0$ to one with potential $\Phi_1$, it will gain energy $m(\Phi_0 - \Phi_1)$.

• We should keep in mind that this is *not* a mathematically completely consistent description — that will only come later — but just an attempt to extract some results from simple assumptions (which may later need to be altered)

Now, we imagine an experiment in which two experimenters $A$ and $B$ are at different gravitational potentials, $\Phi_A$ and $\Phi_B$, and that initially experimenter $A$ has a massive particle of mass $m$. The two experimenters then perform the steps, as illustrated in figure 15.

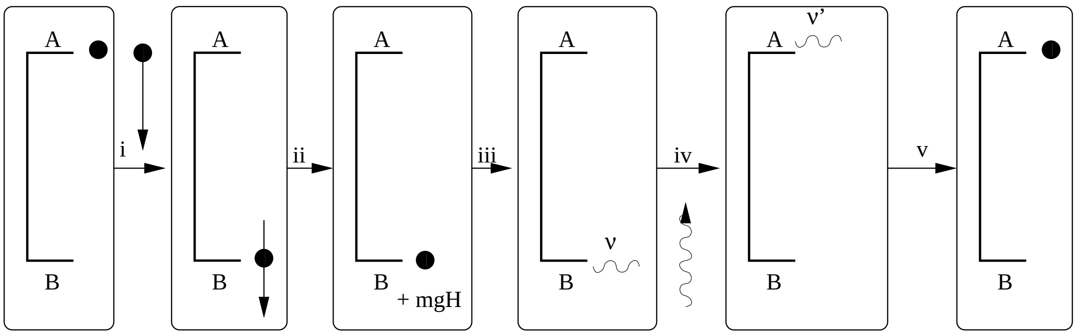



Fig 15: (i) $A$ drops the particle to $B$, so that $B$ receives a particle of mass $m$ with kinetic energy $m(\Phi_A - \Phi_B)$.

(ii) $B$ removes the kinetic energy of the particle somehow and stores this, e.g. in a battery.

(iii) $B$ now converts the particle of mass $m$ together with the stored energy into a photon of total energy $E = h\,\nu = m\,c^2 + m\,g\,(\Phi_A - \Phi_B)$.

(iv) $B$ sends this photon up to $A$.

(v) $A$ converts the photon of frequency $\nu'$ and energy $E' = h\,\nu'$ into a particle of mass $m' = E'/c^2$.

The question is, how is the frequency of the photon altered by its passage through the gravitational field, i.e. how are the frequencies $\nu$ and $\nu'$ related?

We can easily calculate the frequency of the photon as emitted by $B$ as

$$\nu = \frac{E}{h} = \frac{mc^2 + m(\Phi_A - \Phi_B)}{h}$$

By conservation of energy, the particles that $A$ has at the beginning and end should have the same energy, i.e. the same mass, so that $m' = m$. We can now calculate the frequency of the photon as received by $A$

$$\nu' = \frac{E'}{h} = \frac{m'c^2}{h} = \frac{mc^2}{h}$$

Comparing the two, we find

$$\nu = \nu' + \nu' \, \frac{\Phi_A - \Phi_B}{c^2} \; .$$

As a result, we see that the frequency of the photon sent by $B$ to $A$ has changed by $\delta\nu = \nu' - \nu$,

$$\delta\nu = \nu' - \nu = -\nu' \, \frac{\Phi_A - \Phi_B}{c^2} \approx -\nu \, \frac{\Phi_A - \Phi_B}{c^2} \; .$$

Since the photon has changed from potential $\Phi_B$ to potential $\Phi_A$, its change in potential is $\delta\Phi = (\Phi_A - \Phi_B)$, or in other words

$$\frac{\delta\nu}{\nu} \approx -\frac{\delta\Phi}{c^2} \; . \tag{34}$$

Equation (34) is only correct for infinitesimally small changes in potential or frequency, and correctly should be thought of as equation between differentials, or as a differential equation and integrated to find

$$\begin{aligned} \frac{\mathrm{d}\nu}{\nu} &= -\frac{\mathrm{d}\Phi}{c^2} \; , \\ \int_{\nu_0}^{\nu_1} \frac{\mathrm{d}\nu}{\nu} &= -\int_{\Phi_0}^{\Phi_1} \frac{\mathrm{d}\Phi}{c^2} \; , \\ \log\left(\nu_1/\nu_0\right) &= -(\Phi_1 - \Phi_0)/c^2 \; , \\ \nu_1/\nu_0 &= \mathrm{e}^{-(\Phi_1 - \Phi_0)/c^2} \; . \end{aligned} \tag{35}$$

This is the *gravitational red-shift formula*, which shows how the frequency of photons, (e.g. the colour of visible light), changes when they pass through a (weak, static) gravitational field. (Of course we must reformulate the red-shift formula when we turn to General Relativity, but in the limit in which GR can be well approximated by Newtonian Gravity and the Newtonian potential doesn't change with time, the correct GR expression for the red-shift is well-approximated by equation (35). )

### 1.11 Clocks in gravitational fields.

We don't normally think very hard about how we measure time, but the SI definition of 1 second is (as of 1967):

**One second** is defined to be the time taken for 9,192,631,770 periods of the unperturbed microwave transition between the 2 hyperfine levels of the ground state of $Cs^{133}$.

What this means is that our international standard definition of a second is intimately connected with the frequency of the photons (microwaves) emitted by a Caesium atom.

So, suppose we want to tell someone far away how long we think one second is? What we can do is send them one of these microwaves emitted by Caesium 133, and say we think that one second is given by the frequency $\nu$ of this photon by the formula

$$\frac{9,192,631,770}{\nu}\ . \tag{36}$$

Alternatively, if we beam them photons at frequency, then after 9,192,631,770 periods of the photons' oscillations, our clocks will have changed by 1 second.

So, we can tell someone that all they have to do is to count 9,192,631,770 periods of the oscillations of the photons that we have sent them, and that will be one second on our clock. However, by the time the photons have arrived at the distant observer, the frequency may well have changed from $\nu$ to a new frequency $\nu'$. This means that the time they take to count 9,192,631,770 periods of the oscillations would be different - it would take them $\nu/\nu'$. So they would think that one second as shown on our clock actually takes $\nu/\nu'$ seconds (see figure 16)

This is in fact the case – as we have seen, the frequency of a photon can change as it passes through a gravitational field, with the consequence that time really does go slower or faster depending on the gravitational potential.

This effect can actually be measured. There is an atomic clock kept at Boulder in Colorado (altitude 1650 m) and another one at Greenwich (altitude 25 m). This difference is roughly $5\mu$s per year[7]. These clocks are themselves accurate to about $1\,\mu$s per year, so that this difference is quite observable, and is indeed observed.

[7] To see this, use equation (34) with the potential $\Phi = gz$ and with the values $\delta z$=1625 m, $g$=9·8 ms$^{-2}$, $c \sim 3\times10^8$ ms$^{-1}$ and 1 year $\sim 3\times10^7$ s, so that $\Delta$, the number of seconds difference per year, is roughly

$$\Delta = (3\times10^7) \times \left(\frac{1}{1 - g\,\delta z/c^2}\ -\ 1\right)$$

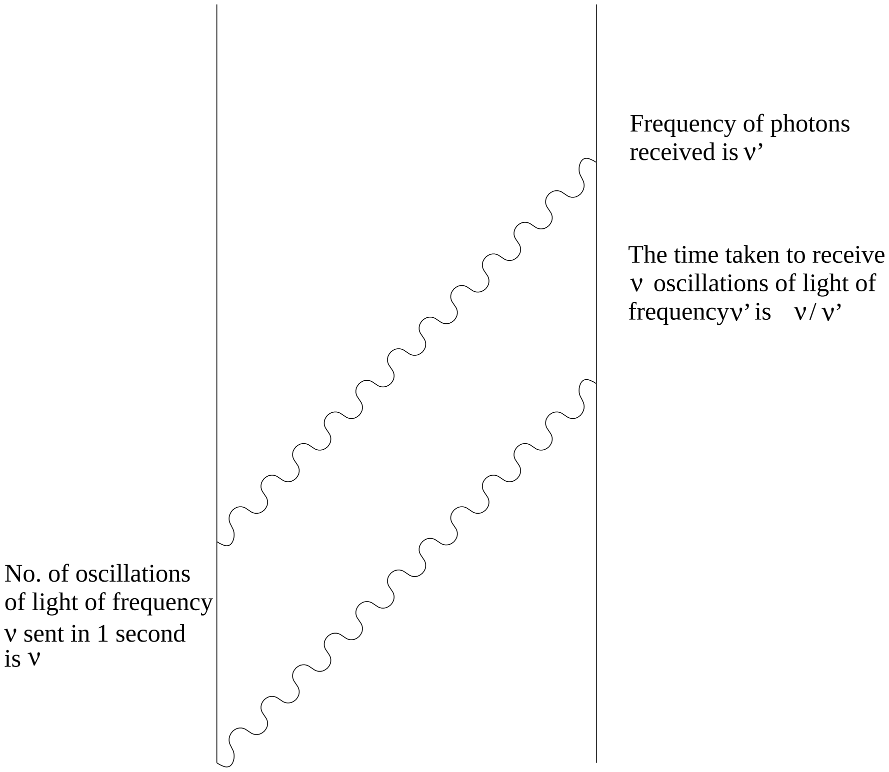

Frequency of photons
received is ν'
The time taken to receive
ν oscillations of light of
frequencyν' is ν/ ν'
No. of oscillations
of light of frequency
ν sent in 1 second
is ν

Figure 16:

## 1.12 Why the results we have so far suggest spacetime is curved.

This last result – gravitational time dilation – provides convincing evidence that the geometry of spacetime is curved.

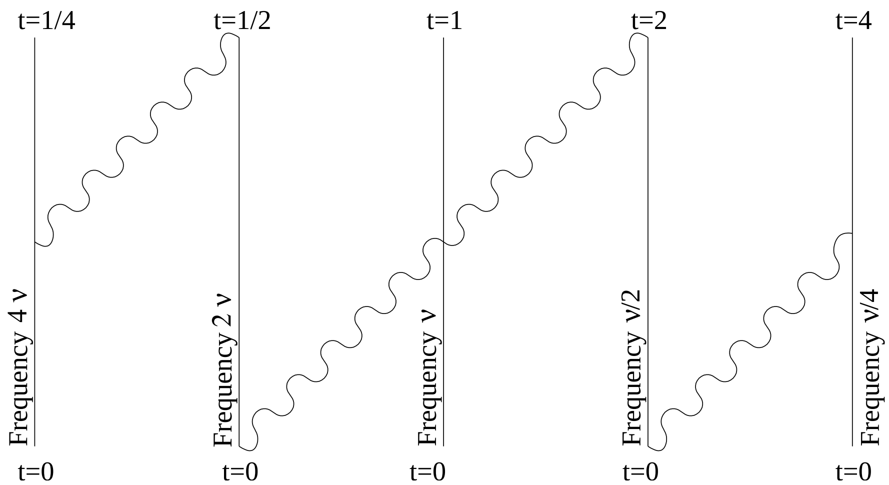
Firstly, we have to agree that we shall measure spatial distance in spacetime using rules, and temporal distance using clocks.

Suppose we now assemble a collection of identical clocks and rulers and use these to try to assemble a concrete relaisation of a frame. We can bolt the rulers together and use these to keep the colcks at fixed a distance from us.

The rulers measure the spatial separation of points in spacetime, and to measure the temporal separation we have to wait and see what elapsed time the clocks show.

As an example, let us suppose that we are in a gravitational field so strong that the frequency of light is halved as it passes from one clock to the next on the right, then each clock will think the clock on the left is going at half speed, the clock two on the left is going at quarter speed, the next clock on the right is going at twice speed, etc, as in figure 17.

Figure 17:




Each clock runs at twice the speed of its left neighbour and half the speed of its right neighbour

To see why this shows that spacetime is curved, we can imagine trying to construct a network of rods with the same arrangement of separations - equally separated in the $x$–direction but of succesively halving separations in the $y$-direction. If we tried to do this we would quickly find that the set of rods wanted to curl up into a curved surface in three dimensions. A set of mutual spatial separations such as these shows that the surface is curved, as is shown in figure 18

Equally, a set of mutual temporal and spatial separations as in figure 17 shows that the *spacetime* is curved.

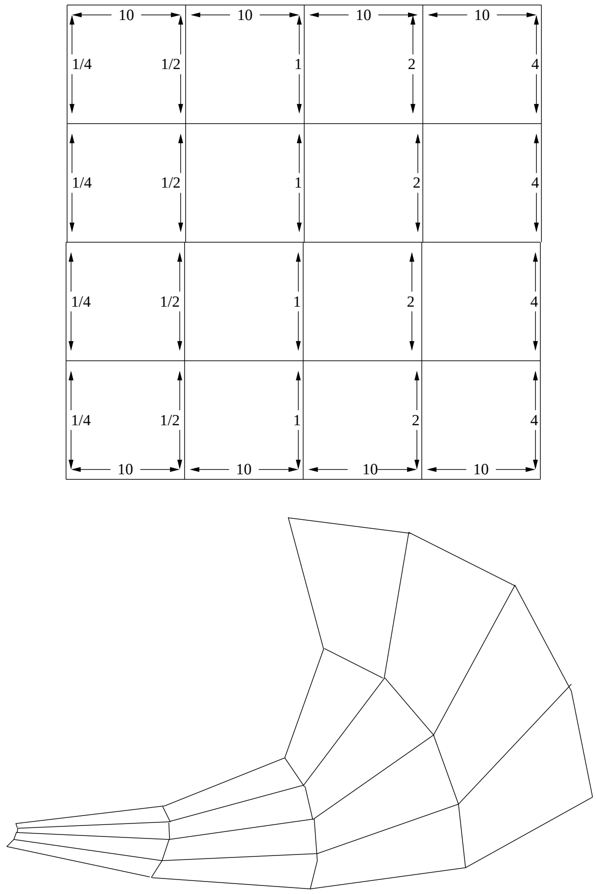



Figure 18: A network of rodsshowing the separations of the points and how it would roll up into a curved surface if constructed in 3 dimensions.

So far we have come up with several clues to the solution of the problem of reconciling gravity with Special Relativity into curved spacetime.

1. The strong equivalence principle – we should be able to remove the effects of gravity to an arbitrarily high degree by changing coordinates locally, i.e. within some small region of spacetime, to be those of a freely-falling non-rotating inertial observer.

2. In these new coordinates, freely moving particles follow straight lines at constant speeds.

We now discuss how these are signs that the correct formulation of gravity is as a geometric theory.

1. An important physical property in special relativity of two points $x^\mu$ and $x^\mu + \Delta x^\mu$ is the quantity $\Delta s^2 = \eta_{\mu\nu}\, \Delta x^\mu\, \Delta x^\nu$. However, the SEP only says that we can find coordinates around each point $x^\mu$ for which the laws of SR are approximately true. It may well be that around one point $P$, one set of coordinates ($x^\mu$, say) provide a LIF around $P$, but around another point $Q$, we need to take a different set of coordinates $x'^\mu$ to find a LIF around $Q$, and that the relationship between these coordinates is not just a Lorentz transformation, but

something quite general:

$$x'^{\mu} = x'^{\mu}(x^{\nu}) \ . \tag{37}$$

Let's now consider the value of $\Delta s^2$ for the point $Q$ with coordinates $x'^{\mu}$ and a nearby point with coordinates $x'^{\mu} + \Delta x'^{\mu}$. By the SEP and the fact that $x'^{\mu}$ are the coordinates of a LIF around $Q$, we know that

$$\Delta s^2 = \eta_{\mu\nu} \, \Delta x'^{\mu} \, \Delta x'^{\nu} \ . \tag{38}$$

However, we can also work out the coordinates of these two points in the coordinate system $x^{\mu}$, in which they are

$$x^{\mu} \quad \text{and} \quad x^{\mu} + \Delta x^{\mu} \ . \tag{39}$$

By Taylor expansion, these are related to the coordinates $x'^{\mu}$ by

$$x'^{\mu}(x^{\nu} + \Delta x^{\nu}) \approx x'^{\mu}(x^{\nu}) + \frac{\partial x'^{\mu}}{\partial x^{\nu}} \Delta x^{\nu} \ . \tag{40}$$

Hence the separation of the two points is

$$\Delta x'^{\mu} \approx \frac{\partial x'^{\mu}}{\partial x^{\nu}} \Delta x^{\nu} \ , \tag{41}$$

and hence the invariant $\Delta s^2$ is given by

$$\begin{aligned} \Delta s^2 &= \eta_{\mu\nu} \, \Delta x'^{\mu} \, \Delta x'^{\nu} = \eta_{\mu\nu} \, (\frac{\partial x'^{\mu}}{\partial x^{\rho}} \Delta x^{\rho})(\frac{\partial x'^{\nu}}{\partial x^{\sigma}} \Delta x^{\sigma}) = \left( \eta_{\mu\nu} \, \frac{\partial x'^{\mu}}{\partial x^{\rho}} \frac{\partial x'^{\nu}}{\partial x^{\sigma}} \right) \Delta x^{\rho} \, \Delta x^{\sigma} \\ &= g_{\rho\sigma} \, \Delta x^{\rho} \, \Delta x^{\sigma} \ . \end{aligned} \tag{42}$$

We see that in coordinates $x^{\mu}$, the invariant $\Delta s^2$ is no longer given in terms of the separation $\Delta x^{\mu}$ by the matrix $\eta_{\mu\nu}$, but by a more complicated matrix $g_{\mu\nu}$, about which we know nothing at the moment.

To somebody who knows differential geometry, this is clearly something called the *metric*, and we shall see how in general relativity Einstein's equations are differential equations for the metric, and how the metric determines the paths that particles follow.

2. The second idea, that in a LIF particles travel at approximately constant velocity, i.e. that a freely moving particle at point $Q$ has equation of motion $\ddot{x}'^{\mu} = 0$, can equally well be recognised by somebody who knows differential geometry as the equation of a *geodesic*, so that freely falling particles will follow geodesics in general relativity. In another coordinate system, this equation will not be so simple, and much of what we do later on will be to find the more complicated equations satisfied by a geodesic in some arbitrary coordinate system and then to solve these equations in special cases.

## 1.13 Exercises

### Short exercises

**1.1** The Newtonian Gravitational body outside an isolated spherically symmetric body of mass $M$ is $\Phi = -\frac{GM}{r}$. Show that the gravitational field is

$$\mathbf{g} = -\frac{GM}{r^2}\hat{\mathbf{r}} \,, \tag{43}$$

where $\hat{\mathbf{r}}$ is the unit vector in the direction of $\mathbf{r}$. (hint: you will need to calculate $\partial r/\partial x^i$. This can be done easily by first calculating $\partial(r^2)/\partial x^i$ and invoking the product rule)

**1.2** The gravitational potential inside a spherically symmetric body of mass $M$, radius $R$ and constant density $\rho$ is

$$\Phi = \frac{GMr^2}{2R^3} + \text{const.} \,. \tag{44}$$

Show that

(a) the gravitational field inside the body is $\mathbf{g} = -\frac{GM}{R^3} r\hat{\mathbf{r}}$.

(b) $\nabla^2\Phi = 4\pi G\rho$

**1.3** To find the GR and SR time dilation effects for a GPS satellite.

A GPS satellite typically orbits at an altitude of 26000 km and orbits the Earth in about 12 hours. The radius of the Earth is about 6400 km. The strength of the Earth's gravitational field at the surface is about 10 $ms^{-2}$.

(a) Calculate the Special Relativistic time dilation for the satellite relative to the Earth's surface.

(b) Calculate the Newtonian Potential difference between the satellite and the Earth's surface and hence find the gravitational time dilation.

(c) Which effect is larger? Do they have the same or opposite signs? What is the total time dilation over one day? Is this important for the use of the GPS satellites to find one's position accurately?

**1.4** Consider the path $x^\mu(u)$ in Minkowski space

$$ct = a\,\sinh u \,, \quad x = a\,\cosh u \,, \quad y = 0 \,, \quad z = 0 \,, \tag{45}$$

where $a$ is a positive constant and $u$ is a parameter.

(a) Sketch this path in the $t - x$ plane.

(b) Is the path timelike, spacelike or neither?

(c) Use equation (24) to find the proper time elapsed along the path starting from $u = 0$, as a function of $u$.

**Longer exercises**

**1.5 Orbits in Newtonian Gravity**

Consider a massive stationary object of mass $M$ fixed at the origin, and a particle moving in the x-y plane subject only to the Newtonian gravitational field of the massive object. If the polar coordinates of the particle are $(r, \theta)$ and $u = 1/r$, show that

$$\frac{\mathrm{d}^2 u}{\mathrm{d}\theta^2} + u = \frac{GM}{h^2} \ .$$

Find all the solutions of this equation and hence classify the possible Newtonian orbits of a particle in such a gravitational field.

Method:
(a) Consider the unit vectors $\hat{\mathbf{r}}$ and $\hat{\theta}$ given in Cartesian coordinates by

$$\hat{\mathbf{r}} = (\cos\theta, \sin\theta) \ , \quad \hat{\theta} = (-\sin\theta, \cos\theta) \ .$$

Show that

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{\mathbf{r}} = \dot{\theta}\,\hat{\theta} \ , \quad \frac{\mathrm{d}}{\mathrm{d}t}\hat{\theta} = -\dot{\theta}\,\hat{\mathbf{r}}$$

(b) Writing the vector $\mathbf{r} = r\,\hat{\mathbf{r}}$, show that

$$\dot{\mathbf{r}} = \dot{r}\,\hat{\mathbf{r}} \ + \ r\dot{\theta}\,\hat{\theta} \ , \quad \ddot{\mathbf{r}} = (\ddot{r} - r\dot{\theta}^2)\,\hat{\mathbf{r}} \ + \ (2\dot{r}\dot{\theta} + r\ddot{\theta})\,\hat{\theta}$$

(c) Newton's equations for the momentum $\mathbf{p}$ of a particle of mass $m > 0$ moving in the gravitational field of a stationary massive body of mass $M$ at the origin are $\ddot{\mathbf{r}} = \mathbf{g}$ where the gravitational field is given by equation (43). Show that for $m \neq 0$ (i.e. $\mathbf{p} = m\,\dot{\mathbf{r}}$) this reduces, in components, to the two equations

$$\ddot{r} - r\dot{\theta}^2 + \frac{GM}{r^2} = 0 \ , \quad 2\dot{r}\dot{\theta} + r\ddot{\theta} = 0$$

(d) Deduce that $h = r^2\dot{\theta}$ is a constant. ($mh$ is the angular momentum about the origin)

(e) By the substitution $r(t) = 1/u(\theta(t))$, show that $\dot{r} = -hu'$ , where $\dot{} \equiv \mathrm{d}/\mathrm{d}t$ and $' \equiv \mathrm{d}/\mathrm{d}\theta$, and further that $u$ satisfies

$$u'' + u = \frac{GM}{h^2} \ . \tag{46}$$

(f) Show that the general solution for the distance from the origin $r$, as a function of $\theta$, is

$$r = \frac{l}{1 + e\cos(\theta - \theta_0)} \ .$$

This is the equation of a conic section, $e$ is called the *eccentricity* of the orbit and the value of $e$ determines the qualitative nature of the orbit

What is the distance of closest approach to the origin? For which values of $e$ does the orbit extend to $r = \infty$? What are the shapes of the various orbits (e.g. $e = 0$ is a circle).

### 1.6 The relative acceleration in a spherical well.

The gravitational potential outside a single spherically symmetric massive body (such as an idealised version of the Earth) of mass $M$ is

$$\Phi = -\frac{GM}{r} \ .$$

Consider an observer and a set of nearby test-particles that are initially at rest and then are simultaneously released to fall freely in this gravitational field. Recall that the separation $\Delta^i$ of two nearby particles satisfies

$$\ddot{\Delta}^i = -\frac{\partial^2\Phi}{\partial x^i \partial x^j}\Delta^j$$

(a) Find the matrix $M$ with entries $M_{ij} = \partial^2\Phi/\partial x^i \partial x^j$.

(b) Show that the radius vector $\mathbf{r}$ is an eigenvector of the matrix $M$.

Hint: The components of $\mathbf{r}$ are $x^i$. Calculate the components of the matrix product $\mathbf{M}\cdot\mathbf{r}$ which are $(\mathbf{M}\cdot\mathbf{r})_i = \sum_j M_{ij}x^j$ and show that $(\mathbf{M}\cdot\mathbf{r})_i = -(2GM/r^3)x^i$.

(c) Show that any vector $\mathbf{a}$ orthogonal to $\mathbf{r}$ is also an eigenvector of $M$.

Hint: Calculate the components of the product $\mathbf{M}\cdot\mathbf{a}$ which are $(\mathbf{M}\cdot\mathbf{a})_i = \sum_j M_{ij}a^j$ and use the fact that if $\mathbf{a}\cdot\mathbf{r} = \sum_j x^i a^j = 0$ and show that $(\mathbf{M}\cdot\mathbf{a})_i = (GM/r^3)a^i$.

(d) Using these results describe the effect that the gravitational field has on the separations of the test particles from the central observer.

Hint: consider particles separated in the directions $\mathbf{r}$ and $\mathbf{a}$ and work out their accelerations using the results of parts (b) and (c).

(e) Use your knowledge of the gravitational field of a spherically symmetric massive body to show how the qualitative result could have been deduced more simply.

*These are known as gravitational tidal forces*

### 1.7 Accelerating paths in Minkowski space

Consider the path in exercise (1.4).

(a) Use equation (24) to find the proper time elapsed along the path between the times $t = -1$ and $t = 1$.

(b) Using equation (23) to find $\mathrm{d}\tau$ in term of $\mathrm{d}u$, find the four-velocity $U^\mu$ and four-acceleration $A^\mu$ along the path, where

$$U^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} \ , \quad A^\mu = \frac{\mathrm{d}U^\mu}{\mathrm{d}\tau} \ . \qquad (47)$$

(c) Show that $\eta_{\mu\nu}U^\mu A^\nu = 0$ along the path.

(d) Show that $A^2 = \eta_{\mu\nu}A^\mu A^\nu$ is a constant along the path. For which value of the parameter $a$ is the proper time along the path between $t = -1$ and $t = 1$ longest and what is the corresponding value of $A^2$?

**A very long exercise**

**1.8 Relativistic Doppler and Gravitational Redshift effects**

The relativistic Doppler effect related the frequency of light as seen in two frames moving with respect to each other. By considering a local inertial frame freely falling in spacetime and the change in its velocity due to the gravitational field, we can use the SEP to deduce the gravitational Doppler effect. We first find the relativistic Doppler effect of special Relativity.

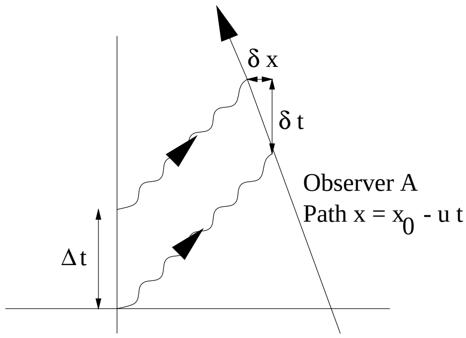
We recall the facts that

- light travels at the speed of light $c$,
- each photon has a particular frequency, $\nu$ say, which gives the period of oscillations of the electro-magnetic field, or time between successive wave-crests of the electro-magnetic waves, as $\Delta t = 1/\nu$.

Now consider the following set up: There is a source of photons of frequency $\nu$ which is static and at the spatial origin of a Lorentz frame $\mathcal{S}$ with coordinates $(t, x)$ and the photons are emitted to the right. (We shall ignore the transverse coordinates $\{y, z\}$ as they play no part in our considerations) The question is: what is the frequency of these photons as they are received by an observer $A$ moving at speed $u$ from the right?

(a) By considering the equations for the trajectories of two successive wave-crests, as in the figure below, or otherwise, show that the time difference $\delta t$ between the arrivals of subsequent wave-crests at $A$, and the spatial difference $\delta x$ between the positions that $A$ meets these wave-crests, *as measured in* $\mathcal{S}$ are given by

$$\delta t = \Delta t \left( \frac{1}{1 + (u/c)} \right) ,$$
$$\delta x = -\Delta t \left( \frac{v}{1 + (u/c)} \right)$$




(b) In classical physics, the time difference between the arrival of subsequent wave-crests as perceived by $A$ is $\delta t$. Show that the classical result for frequency $\nu'_{\text{classical}}$ of the photons as received by $A$ is

$$\nu'_{\text{classical}} = \frac{1}{\delta t} = (1 + \frac{u}{c})\, \nu \ .$$

(c) In Special Relativity, $A$'s clocks do not run at the same speed as the clocks in frame $\mathcal{S}$ but are related by the Lorentz transformations:

$$\begin{array}{ll} x' = \gamma(u)\,(x + ut) & x = \gamma(u)\,(x' - ut') \\ t' = \gamma(u)\,(t + ux/c^2) & t = \gamma(u)\,(t' - ux'/c^2) \end{array}$$

where $\gamma(u) = (1 - u^2/c^2)^{-1/2})$. Show that the difference in time and space coordinates $\delta t'$ and $\delta x'$ between the arrival of the successive wave-crests, in the frame $\mathcal{S}'$ of $A$, are given by

$$\delta t' = \Delta t\,\frac{1}{\gamma(u)\,(1 + (u/c))}\ , \quad \delta x' = 0\ ,$$

(d) Show that the correct formula $\nu'_{\mathrm{relativistic}}$ for the frequency of the photons as observed by $A$ is

$$\nu'_{\mathrm{relativistic}} = \frac{1}{\delta t'} = \gamma(u)\,(1 + \frac{u}{c})\,\nu\ .$$

We can now apply this result to find the gravitational red-shift using the Doppler shift formula of Special Relativity as follows: Consider the (usual) arrangement of a lift at height $z$ in a lift-shaft, with the lift stationary at time $t < 0$, and then dropped in a gravitational field of strength $g$ downward, so that it is in free-fall for $t > 0$. Suppose that at $t$=0 a photon of frequency $\nu$ starts travelling from the top of the lift to the bottom. Since the lift is in free-fall, the photon will not change frequency in the rest-frame of the lift.

(e) Use Newtonian mechanics to find how far the lift has fallen in a short time $\delta t$, the downward speed $\delta v$ that the lift has obtained in this time, the vertical displacement $\delta z$ of the photon in this time, and the frequency $\nu'$ of the photon as perceived by a stationary observer at the same height $z + \delta z$ as the photon. Using these results, show that

$$\frac{\nu'}{\nu} = \left(1 - \frac{g\,\delta z}{c^2}\right)\ ,$$

and hence that in this case

$$\frac{\nu'}{\nu} = \left(1 - \frac{\delta\Phi}{c^2}\right)\ .$$

Lecture notes for Spacetime Geometry and General Relativity

2015-2016

# Section 2: Introduction to Geometry

# 2 Some differential geometry

Since it will be useful, to begin with, not to stray too far from the the world we understand, we shall initially start with the theory of curved two-dimensional surfaces embedded in real three-space, or $\mathbb{R}^3$. However, we shall always be trying to find ways to describe the properties of the surface which do not use the explicit embedding, i.e. *intrinsic* properties of the surface rather than the *extrinsic* properties which it inherits from being embedded in $\mathbb{R}^3$.

## 2.1 Embeddings and coordinates

We shall describe a surface $\mathcal{S}$ embedded in $\mathbb{R}^3$ by drawing some arbitrary coordinate system $x^\alpha \equiv \{x^1, x^2\}$ on the surface, and then specifying the coordinates $\boldsymbol{\sigma}(x^\alpha)$ of the surface as $x^\alpha$ varies. Thus specifying values of $x^\alpha$ will specify a point in the surface.

We can use more than one different coordinate system to describe the same surface – three examples of different coordinate systems used to describe the unit sphere are (i) spherical polar coordinates (ii) a vertical projection and (iii) a stereographic projection. These are illustrated in figure 19

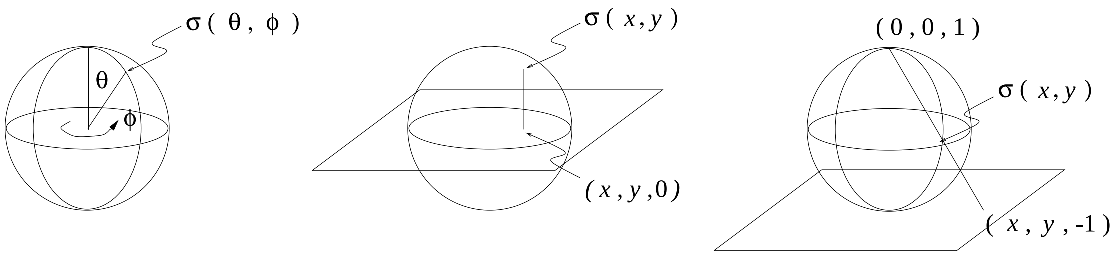



Fig 19(i): Spherical polar coordinates

Fig 19(ii): Vertical projection

Fig 19(iii): Stereographic projection.

The embeddings are given in case (i) by

$$\boldsymbol{\sigma}(\theta, \phi) = (\cos\phi \sin\theta, \sin\phi \sin\theta, \cos\theta) \;, \tag{48}$$

and in case (ii) by

$$\boldsymbol{\sigma}(x, y) = (x, y, \sqrt{1 - x^2 - y^2}\;) \;. \tag{49}$$

Case (iii) is covered on exercise sheet 2.

We shall require the coordinate map $\mathbb{R}^2 \to \mathcal{S}$ to be one–one, so none of the three coordinate systems we have presented covers the whole sphere in this fashion. For spherical polar coordinates, the North Pole corresponds to the whole line $\theta = 0$ and the South Pole to the whole line $\theta = \pi$; for the vertical projection we are limited to the upper hemisphere $z \geq 0$; for the stereographic projection, the whole sphere is covered *except* for the North Pole.

These coordinates do not cover the whole surface; the regions they do cover are known as *coordinate patches*.

The fact that we have not been able to cover the whole sphere with one coordinate patch is not just the result of a bad choice – it is possible to prove (far outside the scope of this course) that one needs at least two patches to cover a sphere.

As an example of how one can actually cover the sphere with two patches, we note that the stereographic projection can be used to cover all the points with $\theta \geq \theta_0$, and that a similar stereographic projection from the South Pole can be used to cover all the points with $\theta < \theta_1$. We can easily choose these such that the whole sphere is in either one or the other or both of the two coordinate patches, as shown in 20.

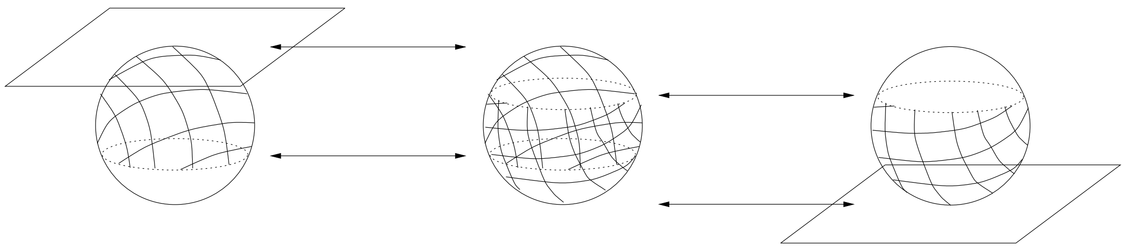


Fig 20a:
Stereographic coordinates on the upper hemisphere

Fig 20b:
The whole sphere covered by a combination of the two coordinate patches

Fig 20c:
Stereographic coordinates on the lower hemisphere

We could ask that the map

$$\boldsymbol{\sigma} : \mathbb{R}^2 \to \mathcal{S}$$

be continuous, differentiable, twice-differentiable, and so on. We could explore the possibilities of adding conditions one at a time, but since this an applied mathematics course we shall simply assume that this map can be differentiated as many times as we want. (This corresponds to the definition of a smooth manifold.)

Since we may have to consider two or more distinct sets of coordinates $x^\alpha$ and $x'^\alpha$ if we want to cover every point of a surface $\mathcal{S}$, we will need to consider changes of coordinates. Since we require the coordinate maps to be one–one, this means that we automatically get a map relating the coordinates in regions where they overlap, as in figure 21

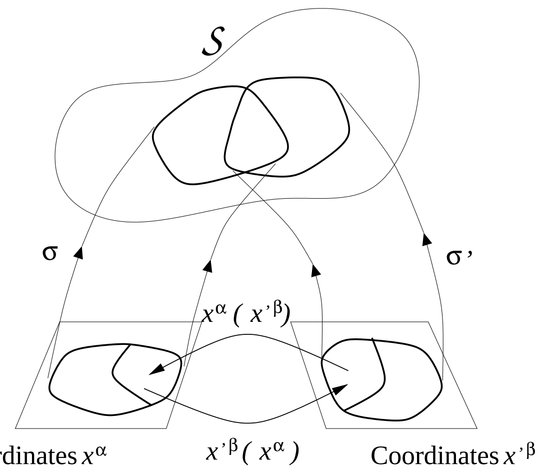



Figure 21

### 2.1.1 Curves

Another concept that we shall need is that of a curve – after all, particles will follow curves. Using the embedding of our surface into $\mathbb{R}^3$, we can specify a curve as a path $\boldsymbol{\gamma}(t)$. However, since we can parametrise the points in the surface by coordinates $x^\alpha$, this will give us a curve in the space of the $x^\alpha$ via $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(\, x^\alpha(t)\,)$ as in figure 22.

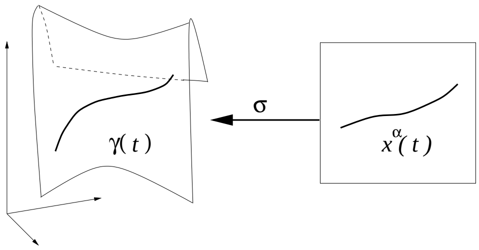



Figure 22

So, we can think of a curve either as $\boldsymbol{\gamma}(t)$ or $x^\alpha(t)$; the first uses the embedding $\boldsymbol{\sigma}$ into $\mathbb{R}^3$, while the second is an idea intrinsic to the coordinates $x^\alpha$.

### 2.1.2 Tangent vectors and the tangent plane

We should also like to be able to define vectors, since we have an intuitive idea that the velocity $\dot{\boldsymbol{\gamma}}(t)$ is a vector – indeed we shall call this the *tangent vector* to the curve $\boldsymbol{\gamma}$ at the point $\boldsymbol{\gamma}(t)$.

If we consider all the tangent vectors to all curves running through a point $\boldsymbol{\sigma}(x^\alpha)$ we find the remarkable fact that they all lie in a two-dimensional plane, called the *tangent plane* at that point. This is easy to prove, since we can use the chain rule to write

$$\begin{aligned} \dot{\boldsymbol{\gamma}}(t) &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\boldsymbol{\sigma}(x^\alpha(t))\right) \\ &= \dot{x}^\alpha \frac{\partial \boldsymbol{\sigma}}{\partial x^\alpha} \, . \end{aligned} \tag{50}$$

It is clear from this formula that any tangent vector at the point $\boldsymbol{\gamma}(t_0)$ can be written as a linear combination of two vectors

$$\boldsymbol{e}_1 = \frac{\partial \boldsymbol{\sigma}}{\partial x^1} \, , \quad \boldsymbol{e}_2 = \frac{\partial \boldsymbol{\sigma}}{\partial x^2} \, . \tag{51}$$

Conversely, given a linear combination of these two vectors

$$\boldsymbol{v} = v^1 \boldsymbol{e}_1 \; + \; v^2 \boldsymbol{e}_2 \equiv v^\alpha \boldsymbol{e}_\alpha \, , \tag{52}$$

then we can find a curve $\boldsymbol{\gamma}(t)$ such that $\boldsymbol{v}$ is the tangent vector $\dot{\boldsymbol{\gamma}}(t_0)$ at $\boldsymbol{\gamma}(t_0) = \boldsymbol{\sigma}(x_0^\alpha)$. An example of such a curve is

$$x^\alpha(t) = x_0^\alpha + (t - t_0)\, v^\alpha \, . \tag{53}$$

Clearly $x^\alpha(t_0) = x_0^\alpha$, and $\dot{x}^\alpha(t) = v^\alpha$, and hence

$$\dot{\boldsymbol{\gamma}}(t) = \dot{x}^\alpha \boldsymbol{e}_\alpha = v^\alpha \boldsymbol{e}_\alpha = \boldsymbol{v} \tag{54}$$

As a result, we see that the tangent vectors to all curves which pass through $\boldsymbol{\gamma}(t_0)$ span a two-dimensional plane with basis vectors $\boldsymbol{e}_1, \boldsymbol{e}_2$. This plane is called the *tangent plane*.

We can simply define a vector $\boldsymbol{v} = v^\alpha \boldsymbol{e}_\alpha$ in the tangent space at a point in the surface $\mathcal{S}$ by giving its components $v^\alpha$.

However, a change of coordinates for $\mathcal{S}$ will, in general, change the components of a vector, and so if we are going to specify a vector by its components $v^\alpha$ in one coordinate system, we also need to say what these components are in any other coordinate system.

### 2.1.3 Vector transformation law

To find how a the components $v^\alpha$ of a vector in coordinates $x^\alpha$ are related to the components of the same vector $v'^\alpha$ in the coordinate system $x'^\alpha$, we need to be able to relate the basis vectors $\boldsymbol{e}_\alpha$ and $\boldsymbol{e}'_\alpha$ of the tangent plane at that point in the two coordinate systems.

This is straightforward since we can again use the chain rule in the definition of the basis vectors. If the two embeddings are $\boldsymbol{\sigma} : x^\alpha \to \mathcal{S}$ and $\boldsymbol{\sigma}' : x'^\alpha \to \mathcal{S}$, then we have

$$\begin{aligned} &\boldsymbol{\sigma}(x) = \boldsymbol{\sigma}'(x'(x)) \;, \\ &\boldsymbol{e}_\alpha = \frac{\partial \boldsymbol{\sigma}(x)}{\partial x^\alpha} = \frac{\partial}{\partial x^\alpha}\big(\boldsymbol{\sigma}'(x'(x))\big) = \frac{\partial x'^\beta}{\partial x^\alpha}\frac{\partial \boldsymbol{\sigma}'(x')}{\partial x'^\beta} = \frac{\partial x'^\beta}{\partial x^\alpha}\,\boldsymbol{e}'_\beta \;. \end{aligned} \tag{55}$$

Similarly, we find

$$\boldsymbol{e}'_\alpha = \frac{\partial x^\beta}{\partial x'^\alpha}\,\boldsymbol{e}_\beta \;. \tag{56}$$

Using this, we have

$$\begin{cases} \boldsymbol{v} = v'^\alpha\,\boldsymbol{e}'_\alpha = v'^\alpha\,\dfrac{\partial x^\beta}{\partial x'^\alpha}\,\boldsymbol{e}_\beta = v^\beta\,\boldsymbol{e}_\beta \;, \\ \boldsymbol{v} = v^\alpha\,\boldsymbol{e}_\alpha = v^\alpha\,\dfrac{\partial x'^\beta}{\partial x^\alpha}\,\boldsymbol{e}'_\beta = v'^\beta\,\boldsymbol{e}'_\beta \;, \end{cases} \tag{57}$$

and hence we find the transformation laws

$$\boxed{\begin{aligned} v^\beta &= \frac{\partial x^\beta}{\partial x'^\alpha}\,v'^\alpha \;, \\ v'^\beta &= \frac{\partial x'^\beta}{\partial x^\alpha}\,v^\alpha \;. \end{aligned}} \tag{58}$$

From now on we shall use this as our *definition* of a vector: a vector is given by a set of components $v^\alpha$ in a coordinate system $x^\alpha$ such that the components $v'^\alpha$ in another coordinate system $x'^\alpha$ are related by (58).

### 2.1.4 Vector dot products, the Length of a vector and the metric

An important quantity we will want to define is the inner, or dot, product of two vectors. If the vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are defined on a surface embedded in $\mathbb{R}^3$ then given as

$$\boldsymbol{u} = u^\alpha \boldsymbol{e}_\alpha \ , \ \ \boldsymbol{v} = v^\beta \boldsymbol{e}_\beta \ ,$$

and their dot product is

$$\boldsymbol{u} \cdot \boldsymbol{v} = (u^\alpha \boldsymbol{e}_\alpha) \cdot (v^\beta \boldsymbol{e}_\beta) = u^\alpha v^\beta (\boldsymbol{e}_\alpha \cdot \boldsymbol{e}_\beta) \ .$$

Likewise, we can find the length of a vector in terms of the components $v^\alpha$. Again, in the case of a surface embedded in $\mathbb{R}^3$ this is easy to find since

$$v^2 = |\boldsymbol{v}|^2 = \boldsymbol{v} \cdot \boldsymbol{v} = (\, v^\alpha \, \boldsymbol{e}_\alpha \,) \cdot \left( v^\beta \, \boldsymbol{e}_\beta \right) = (\, \boldsymbol{e}_\alpha \cdot \boldsymbol{e}_\beta \,) \; v^\alpha \, v^\beta \ . \tag{59}$$

In both these expressions, the inner product of two vectors and the length of a vector still depend explicitly on the embedding through the matrix $(\boldsymbol{e}_\alpha \cdot \boldsymbol{e}_\beta)$. It is important to introduce a new symbol for this matrix since we shall want to apply our understanding of geometry to circumstances in which it makes no sense to think of an embedding — we want to apply it to the whole of spacetime, and we certainly don't usually think of spacetime as being embedded in some larger space. So, if we want to eliminate all mention of the embedding, but still need to know the value of $(\, \boldsymbol{e}_\alpha \cdot \boldsymbol{e}_\beta \,)$, the best we can do is to give this matrix a name, the *metric*, and say that we shall specify the lengths of vectors in terms of such a metric which is defined at all points of the surface. So, we introduce the symbol $g_{\alpha\beta}$ for the symmetric matrix

$$(\, \boldsymbol{e}_\alpha \cdot \boldsymbol{e}_\beta \,) = g_{\alpha\beta} \ ,$$

and this matrix is known as the *metric*. In terms of the metric, the inner product of two vectors $u^\alpha$ and $v^\alpha$ at a point $\mathcal{P}$, and the length of a vector $v^\alpha$ at a point $\mathcal{P}$ are given in terms of the metric at that point $g_{\alpha\beta}$ as

$$\boldsymbol{u} \cdot \boldsymbol{v} = g_{\alpha\beta} \, u^\alpha \, v^\beta \ , \ \ v^2 = g_{\alpha\beta} \, v^\alpha \, v^\beta \ . \tag{60}$$

• Note that $g_{\alpha\beta}$ is a *symmetric* matrix, that is, that $g_{\alpha\beta} = g_{\beta\alpha}$ for all $\alpha$ and $\beta$.

We might wonder how many other objects that are defined in terms of the embedding we will need to find names for to entirely eliminate the embedding from our discussions — the remarkable and pleasant fact is that we will need no more. If we specify the metric at all points of our surface, then we have completely specified all the properties of the surface that we need.

Example: the metric on a sphere in spherical polar coordinates

To work out $g_{\alpha\beta}$ for the simple example of a unit sphere we first recall the map $\boldsymbol{\sigma}$ given in eqn (48),

$$\boldsymbol{\sigma} = (\, \sin\theta \cos\phi \, , \, \sin\theta \sin\phi \, , \, \cos\theta \,) \ ,$$

from which we find

$$\boldsymbol{e}_\theta \equiv \frac{\partial \boldsymbol{\sigma}}{\partial \theta} = (\, \cos\theta \cos\phi \, , \, \cos\theta \sin\phi \, , \, -\sin\theta \,) \ , \tag{61}$$

$$\boldsymbol{e}_\phi \equiv \frac{\partial \boldsymbol{\sigma}}{\partial \phi} = (\, -\sin\theta \sin\phi \, , \, \sin\theta \cos\phi \, , \quad 0 \quad) \ , \tag{62}$$

and hence

$$g_{\theta\,\theta} = \frac{\partial \boldsymbol{\sigma}}{\partial\,\theta} \cdot \frac{\partial \boldsymbol{\sigma}}{\partial\,\theta} = 1\ ,$$
$$g_{\theta\,\phi} = \frac{\partial \boldsymbol{\sigma}}{\partial\,\theta} \cdot \frac{\partial \boldsymbol{\sigma}}{\partial\,\phi} = 0\ ,$$
$$g_{\phi\,\phi} = \frac{\partial \boldsymbol{\sigma}}{\partial\,\phi} \cdot \frac{\partial \boldsymbol{\sigma}}{\partial\,\phi} = \sin^2\theta\ .$$

Or, in matrix form,

$$g_{\alpha\beta} = \begin{pmatrix} 1 & 0 \\ 0 & \sin^2\theta \end{pmatrix}\ . \tag{63}$$

• Note that it is usual to use the coordinates themselves (here $\theta$ and $\phi$) as the indices of vectors, the metric etc., so that one writes $\boldsymbol{e}_\theta$ instead of $\boldsymbol{e}_1$.

Having said that we shall give the lengths of vectors in terms of the metric, we need to know how the components of the metric change when we change coordinate systems. Again this is easy using our formulae for the changes of the basis vectors (55) and (56):

$$g'_{\alpha\beta} = \boldsymbol{e}'_\alpha \cdot \boldsymbol{e}'_\beta = \left( \frac{\partial x^\gamma}{\partial x'^\alpha}\,\boldsymbol{e}_\gamma \right) \cdot \left( \frac{\partial x^\delta}{\partial x'^\beta}\,\boldsymbol{e}_\delta \right) = \frac{\partial x^\gamma}{\partial x'^\alpha}\,\frac{\partial x^\delta}{\partial x'^\beta}\,(\,\boldsymbol{e}_\gamma \cdot \boldsymbol{e}_\delta\,) = \frac{\partial x^\gamma}{\partial x'^\alpha}\,\frac{\partial x^\delta}{\partial x'^\beta}\,g_{\gamma\delta}\ , \tag{64}$$

so that combining this with the corresponding formula giving $g_{\alpha\beta}$ in terms of $g'_{\alpha\beta}$, we have altogether

$$\boxed{\begin{aligned} g'_{\alpha\beta} &= \frac{\partial x^\gamma}{\partial x'^\alpha}\,\frac{\partial x^\delta}{\partial x'^\beta}\,g_{\gamma\delta}\ , \\ g_{\alpha\beta} &= \frac{\partial x'^\gamma}{\partial x^\alpha}\,\frac{\partial x'^\delta}{\partial x^\beta}\,g'_{\gamma\delta}\ . \end{aligned}} \tag{65}$$

• We mention here that the transformation laws of vectors (58) and of the metric (65) are the transformation laws of objects known as *tensors.* We shall have a lot more to say on this subject later on.

### 2.1.5 The signature of a metric

By changing coordinates it is possible to change the appearance of the metric drastically, so that it can be very hard to recognise a surface one knows when presented with the metric in a strange coordinate system.

However, there are several invariants, i.e. combinations of the metric and its derivatives which do not change under coordinate transformations, and we shall come to some of these later, but the simplest invariant is something called the *signature.*

The metric is a real symmetric matrix, and so it has real eigenvalues and can be diagonalised. The signature is simply the number $n^+$ of positive eigenvalues and the number $n^-$ of negative eigenvalues of the metric. It can be written $(n^+, n^-)$ or $(\underbrace{+\,+\cdots+}_{n^+},\underbrace{-\,-\cdots-}_{n^-})$.

In polar coordinates, the unit sphere in two-dimensions has metric

$$g_{\alpha\beta} = \begin{pmatrix} 1 & 0 \\ 0 & \sin^2\theta \end{pmatrix} ,$$

which has two positive eigenvalues and no negative eigenvalues and so has signature $(2, 0)$ or $(++)$.

The spacetime of special relativity is known as *Minkowski space* and has the constant metric

$$\eta_{\alpha\beta} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} ,$$

which has one positive eigenvalue and three negative eigenvalues and so has signature $(1, 3)$ or $(+ - --)$.

Note that the relation between the old and new metrics (65) can be written in matrix form as

$$g' = L^T g L ,$$

where $L^\alpha{}_\beta = \partial x^\alpha / \partial x'^\beta$. If on looks just at a fixed point, one can always find a change of coordinates which will leave the new metric $g'$ diagonal with $+1$ amd $-1$ on the diagonal. This cannot of course be done simultaneously at all points unless the metric is that of a flat surface.

### 2.1.6 The inverse metric and the Kronecker $\delta$

We shall also find it very convenient to have a special symbol for the inverse of the metric. Here the metric is a $2 \times 2$ matrix, and so its inverse will also be a $2 \times 2$ matrix. There is a standard notation for the *inverse metric* as it is known, and that is

$$g^{\alpha\beta} ,$$

i.e. the same symbol $g$ as the metric but with the two indices written 'upstairs' rather than 'downstairs'.

There is also a special symbol for the unit matrix, and that is the *Kronecker delta*, $\delta^\alpha_\beta$, with components

$$\delta^\alpha_\beta = \begin{cases} 1 & \alpha{=}\beta , \\ 0 & \alpha{\neq}\beta . \end{cases} \tag{66}$$

In terms of these two new symbols, the equation that $g^{\alpha\beta}$ is the inverse matrix of the matrix $g_{\alpha\beta}$ is written

$$g^{\alpha\beta}\, g_{\beta\gamma} = \delta^\alpha_\gamma . \tag{67}$$

## 2.2 Geodesics

We know by the SEP that in curved space-time that particles will move along curves that are generalisations of straight-lines, and so we would like to have a concept for curves in curved surfaces analogous to straight-lines in flat space. Such curves are known as *geodesics*.

If we consider two curves connecting points $\boldsymbol{a}$ and $\boldsymbol{b}$ in flat two-dimensional space, one straight and the other curved, then there are many ways of characterising these properties of 'straight' and 'curved', but two we shall use are

- A straight-line has the shortest length of any curve between $\boldsymbol{a}$ and $\boldsymbol{b}$
- The tangent vector to a straight line is always pointing in the same direction, i.e. the tangent vector $\dot{\boldsymbol{\gamma}}(t_0)$ at any one point $\boldsymbol{\gamma}(t_0)$ is parallel to the tangent vector $\dot{\boldsymbol{\gamma}}(t)$ at all other points $\boldsymbol{\gamma}(t)$.

This is exemplified in figure 23.

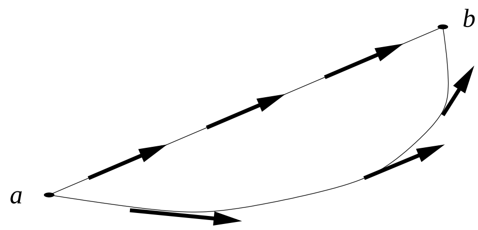



Figure 23

We shall find ways to apply both of these ideas to curved surfaces, and find that they agree, First we shall find the equations for a curve of shortest length, and to do that we need the formula for the length of a curve joining two points.

### 2.2.1 The length of a curve

If we consider a curve $\boldsymbol{\gamma}(t)$ and two neighbouring points parametrised by $t$ and $t + \delta t$, then the vector joining them is

$$\boldsymbol{\gamma}(t+\delta t) - \boldsymbol{\gamma}(t) = \dot{\boldsymbol{\gamma}}(t)\,\delta t \;+\; O(\delta t^2) = \boldsymbol{e}_\alpha \dot{x}^\alpha(t)\,\delta t \;+\; O(\delta t^2)\;.$$

In general, this vector will not lie in the tangent plane to the surface, but will tend to as $\delta t$ goes to zero.

Hence, if we think in terms of points points $x^\alpha(t)$ and $x^\alpha(t + \mathrm{d}t)$ which are infinitesimally close, the length of the curve $\mathrm{d}s$ joining these points is given as

$$\mathrm{d}s^2 \;=\; g_{\alpha\beta}\,(\,\dot{x}^\alpha(t)\,\mathrm{d}t\,)(\,\dot{x}^\beta(t)\,\mathrm{d}t\,) \;=\; g_{\alpha\beta}\,\dot{x}^\alpha(t)\,\dot{x}^\beta(t)\;\mathrm{d}t^2\;. \tag{68}$$

So, the total length of the curve is given by

$$l = \int_{t_0}^{t_1} \mathrm{d}s = \int_{t_0}^{t_1} \sqrt{g_{\alpha\beta}\,\dot{x}^\alpha\,\dot{x}^\beta}\;\mathrm{d}t\;. \tag{69}$$

- Note that if we write $\dot{x}^\alpha \mathrm{d}t = \mathrm{d}x^\alpha$, then equation (68) becomes

$$\mathrm{d}s^2 = g_{\alpha\beta}\,\mathrm{d}x^\alpha\,\mathrm{d}x^\beta \;, \tag{70}$$

which is often the form in which a metric is presented. This way of presenting the metric has the advantage that it is often much shorter since the metric may contain a lot of zeroes. For example, two equivalent ways of presenting the Minkowski metric $\eta_{\alpha\beta}$ are

$$g_{\alpha\beta} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \;, \qquad \mathrm{d}s^2 = (\mathrm{d}x^0)^2 - (\mathrm{d}x^1)^2 - (\mathrm{d}x^2)^2 - (\mathrm{d}x^3)^2 \;. \tag{71}$$

### 2.2.2 The geodesic equations from the variational principle

Given the formula (69) for the length of a curve, we see that it can be written in the form

$$l = \int L(x,\dot{x})\;\mathrm{d}t \;, \qquad \text{where} \qquad L(x,\dot{x}) = \sqrt{g_{\alpha\beta}(x)\;\dot{x}^\alpha\;\dot{x}^\beta}\;\;. \tag{72}$$

The equations for the curve $x^\alpha(t)$ which minimises this integral are the Euler-Lagrange equations,

$$\mathcal{EL}[\,L\,] \;\equiv\; \frac{\partial L}{\partial x^\gamma} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{x}^\gamma}\right) \;=\; 0 \;. \tag{73}$$

Let's take these two terms separately:

$$\begin{aligned} \frac{\partial L}{\partial x^\gamma} &= \frac{\partial}{\partial x^\gamma}\left(\sqrt{g_{\alpha\beta}\,\dot{x}^\alpha\,\dot{x}^\beta}\,\right) \\ &= \frac{1}{2}\,\frac{1}{\sqrt{g_{\delta\epsilon}\,\dot{x}^\delta\,\dot{x}^\epsilon}}\;\frac{\partial}{\partial x^\gamma}\Big(\,g_{\alpha\beta}\,\dot{x}^\alpha\,\dot{x}^\beta\,\Big) \\ &= \frac{1}{2L}\,\frac{\partial g_{\alpha\beta}}{\partial x^\gamma}\,\dot{x}^\alpha\,\dot{x}^\beta \;. \end{aligned} \tag{74}$$

Note that in the second line we have had to use new dummy indices $\delta$ and $\epsilon$ since $\alpha$ and $\beta$ have already appeared, and the summation convention is that each index can appear at most twice. Similarly, we have

$$\begin{aligned} \frac{\partial L}{\partial \dot{x}^\gamma} &= \frac{\partial}{\partial \dot{x}^\gamma}\left(\sqrt{g_{\alpha\beta}\,\dot{x}^\alpha\,\dot{x}^\beta}\,\right) \\ &= \frac{1}{2}\,\frac{1}{\sqrt{g_{\delta\epsilon}\,\dot{x}^\delta\,\dot{x}^\epsilon}}\;\frac{\partial}{\partial \dot{x}^\gamma}\Big(\,g_{\alpha\beta}\,\dot{x}^\alpha\,\dot{x}^\beta\,\Big) \\ &= \frac{1}{2L}\,\Big(\,g_{\gamma\beta}\dot{x}^\beta \;+\; g_{\alpha\gamma}\dot{x}^\alpha\,\Big) \;. \end{aligned} \tag{75}$$

To go from the second line to the third line, we used

$$\frac{\partial \dot{x}^\alpha}{\partial \dot{x}^\gamma} = \delta^\alpha_\gamma \;, \qquad \text{and} \qquad g_{\alpha\beta}\,\delta^\alpha_\gamma = g_{\gamma\beta} \;. \tag{76}$$

We now work out the second term to be

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{x}^\gamma}\right) &= -\frac{\dot{L}}{2L^2}\left(g_{\gamma\beta}\dot{x}^\beta + g_{\alpha\gamma}\dot{x}^\alpha\right) + \frac{1}{2L}\frac{\mathrm{d}}{\mathrm{d}t}\left(g_{\gamma\beta}\dot{x}^\beta + g_{\alpha\gamma}\dot{x}^\alpha\right)\\ &= -\frac{\dot{L}}{2L^2}\left(g_{\gamma\beta}\dot{x}^\beta + g_{\alpha\gamma}\dot{x}^\alpha\right) + \frac{1}{2L}\left(\frac{\partial g_{\gamma\beta}}{\partial x^\alpha}\dot{x}^\alpha\,\dot{x}^\beta + g_{\gamma\beta}\ddot{x}^\beta + \frac{\partial g_{\alpha\gamma}}{\partial x^\beta}\dot{x}^\beta\,\dot{x}^\alpha + g_{\alpha\gamma}\ddot{x}^\alpha\right)\end{aligned}$$

To derive the terms with derivatives of the metric we used

$$\frac{\mathrm{d}}{\mathrm{d}t}g_{\epsilon\zeta}(x) = \dot{x}^\eta\,\frac{\partial}{\partial x^\eta}\,g_{\epsilon\zeta}(x)\ ,$$

and replaced the dummy index $\eta$ by $\alpha$ in the first such expression and by $\beta$ in the second.

Putting this altogether and multiplying through by $L$, we have the Euler-Lagrange equations for $L$ in the form

$$\frac{1}{2}\left(\frac{\partial g_{\gamma\beta}}{\partial x^\alpha}\dot{x}^\alpha\,\dot{x}^\beta + g_{\gamma\beta}\ddot{x}^\beta + \frac{\partial g_{\alpha\gamma}}{\partial x^\beta}\dot{x}^\beta\,\dot{x}^\alpha + g_{\alpha\gamma}\ddot{x}^\alpha - \frac{\partial g_{\alpha\beta}}{\partial x^\gamma}\dot{x}^\alpha\,\dot{x}^\beta\right) = \frac{\dot{L}}{2L}\left(g_{\gamma\beta}\dot{x}^\beta + g_{\alpha\gamma}\dot{x}^\alpha\right)\ . \tag{77}$$

We would be much happier if we had this in the form $\ddot{x}^\delta + \cdots = \cdots$. To put it in this form, we can multiply both sides by $g^{\delta\gamma}$. Consider the second term on the left hand side; this gives

$$g^{\delta\gamma}\,g_{\gamma\beta}\,\ddot{x}^\beta \;=\; \delta^\delta_\beta\,\ddot{x}^\beta \;=\; \ddot{x}^\delta\ , \tag{78}$$

where we used eqns (67) and (66). Similarly, the fourth term on the left hand side gives

$$g^{\delta\gamma}\,g_{\alpha\gamma}\,\ddot{x}^\alpha \;=\; g^{\delta\gamma}\,g_{\gamma\alpha}\,\ddot{x}^\alpha \;=\; \delta^\delta_\alpha\,\ddot{x}^\alpha \;=\; \ddot{x}^\delta\ , \tag{79}$$

where in the first substitution we used the fact that the metric $g_{\alpha\gamma}$ is symmetric in $\alpha$ and $\gamma$. Finally, the remaining terms on the left hand side cannot be simplified further while the terms on the right hand side become

$$g^{\delta\gamma}g_{\gamma\beta}\dot{x}^\beta = \dot{x}^\delta\ , \quad \text{and} \quad g^{\delta\gamma}g_{\alpha\gamma}\dot{x}^\alpha = \dot{x}^\delta\ , \tag{80}$$

so the the final form for the Euler-Lagrange equations for $L$ is

$$\ddot{x}^\delta + \frac{1}{2}\,g^{\delta\gamma}\left(\frac{\partial g_{\gamma\beta}}{\partial x^\alpha} + \frac{\partial g_{\alpha\gamma}}{\partial x^\beta} - \frac{\partial g_{\alpha\beta}}{\partial x^\gamma}\right)\dot{x}^\alpha\,\dot{x}^\beta = \frac{\dot{L}}{L}\,\dot{x}^\delta\ . \tag{81}$$

The particular combination of derivatives of the metric appearing on the left hand side is very important and will appear very frequently in the rest of this course and has a special symbol $\Gamma^\delta_{\alpha\beta}$, and a special name, the *Christoffel symbol* or the *Levi-Civita Connection*:

$$\boxed{\Gamma^\delta_{\alpha\beta} \;=\; \frac{1}{2}\,g^{\delta\gamma}\left(\frac{\partial g_{\gamma\beta}}{\partial x^\alpha} + \frac{\partial g_{\alpha\gamma}}{\partial x^\beta} - \frac{\partial g_{\alpha\beta}}{\partial x^\gamma}\right)}\ . \tag{82}$$

In terms of the connection, the Euler-Lagrange equations for $L$, that is the equations satisfied by a geodesic with an arbitrary parametrisation $x^\alpha(t)$, are

$$\ddot{x}^\delta + \Gamma^\delta_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta = \frac{\dot{L}}{L}\,\dot{x}^\delta\ . \tag{83}$$

Some more comments:

- The Christoffel symbols $\Gamma^\alpha_{\beta\gamma}$ are *symmetric* in $\beta$ and $\gamma$, that is for all $\alpha$, $\beta$ and $\gamma$

$$\Gamma^\alpha_{\beta\gamma} = \Gamma^\alpha_{\gamma\beta} \ .$$

- The length $\mathrm{d}s$ of a small piece of a curve is given by $L\,\mathrm{d}t$, so that if $L$ is a constant, the parameter $t$ parametrising the curve is proportional to the length of the curve. In this case, $t$ is called an *affine parameter*. If $t$ is an affine parameter, then $\dot{L} = 0$ and the general geodesic equation (83) becomes

$$\boxed{\ddot{x}^\alpha + \Gamma^\alpha_{\beta\gamma}\,\dot{x}^\beta\,\dot{x}^\gamma = 0} \tag{84}$$

  We say that eqn. (84) is the geodesic equation for an affinely parametrised geodesic.

- One general result on the solutions to the equations (84) that comes from the theory of differential equations is that a solution is uniquely specified by giving the initial conditions $x^\alpha(0)$ and $\dot{x}^\alpha(0)$. In other words:

  **Theorem:** Given a point $X^\alpha$ and vector $V$, then there is a unique geodesic $x^\alpha(t)$ going through that point with that vector as its tangent vector, i.e. such that $x^\alpha(0) = X^\alpha$ and $\dot{x}^\alpha(0) = V^\alpha$.

- The reason that affine parametrisations are important is that the proper time along a particle path is an affine parameter for that path. The 4–velocity $U^\mu = \mathrm{d}x^\mu/\mathrm{d}\tau$ of a massive particle in special relativity takes a special form in the rest-frame of the particle – namely $(1, 0, 0, 0)$. Hence the length-squared of this vector which is given in special relativity by

$$U^2 = \eta_{\mu\nu}\,U^\mu\,U^\nu = c^2 \ ,$$

  is a constant. By the SEP, local statements such as this have to remain true in general relativity, so that in GR where $\eta_{\mu\nu}$ gets replaced by $g_{\mu\nu}$, this will remain the same constant,

$$U^2 = g_{\mu\nu}\,U^\mu\,U^\nu = c^2 \ ,$$

  and so $L = \sqrt{U^2}$ is a constant, $\dot{L} = 0$, and the proper time $\tau$ is an affine parameter of a particle path in GR.

Although eqn. (84) is very concise, the components of the connection are quite time-consuming to work out from the formula (82) — as we shall see in exercise set 3 — so it is worthwhile learning a convenient trick which gives the equations of an affinely parametrised geodesic (84) directly and from which the components of the connection can be read very easily. This is the subject of the next section.

### 2.2.3 Another variational principle

Let's consider what we get from applying the variational principle to the modified integral

$$\int f(L)\,\mathrm{d}t\,, \tag{85}$$

where $f(L)$ is some arbitrary function. We get

$$\begin{aligned}
\mathcal{EL}[\,f(L)\,] &= \frac{\partial}{\partial x^\gamma}\,(f(L)) - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial}{\partial \dot{x}^\gamma} f(L)\right) \\
&= f'(L)\,\frac{\partial L}{\partial x^\gamma} - \frac{\mathrm{d}}{\mathrm{d}t}\left(f'(L)\frac{\partial L}{\partial \dot{x}^\gamma}\right) \\
&= f'(L)\,\frac{\partial L}{\partial x^\gamma} - \frac{\mathrm{d}}{\mathrm{d}t}\,(\,f'(L)\,)\,\frac{\partial L}{\partial \dot{x}^\gamma} - f'(L)\,\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{x}^\gamma}\right) \\
&= f'(L)\left\{\frac{\partial L}{\partial x^\gamma} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{x}^\gamma}\right)\right\} - f''(L)\,\dot{L}\,\frac{\partial L}{\partial \dot{x}^\gamma} \\
&= f'(L)\,\mathcal{EL}[\,L\,] \;-\; f''(L)\,\dot{L}\,\frac{\partial L}{\partial \dot{x}^\gamma}
\end{aligned} \tag{86}$$

So, we see that if $f'(L) \neq 0$, ie $f$ is not just a constant, then

$$\left\{\begin{array}{c} \dot{L} = 0 \\ \text{and} \\ \mathcal{EL}[\,f(L)\,] = 0 \end{array}\right\} \quad \Rightarrow \quad \mathcal{EL}[\,L\,] = 0\,.$$

This means that if we restrict ourselves to affinely parametrised geodesics, then we can work out the equations using the Euler-Lagrange equations for any function $f(L)$ we like, in particular one that makes the equations simple to work out. Recalling that

$$L = \sqrt{g_{\alpha\beta}\,\dot{x}^\alpha\,\dot{x}^\beta}\,,$$

it is clear that a good function to think of is $f(L) = L^2$, since this will get rid of the square root. Let's call this by the name $\mathcal{L}$:

$$\mathcal{L} = L^2 = g_{\alpha\beta}\,\dot{x}^\alpha\,\dot{x}^\beta\,.$$

So, let's work out the Euler-Lagrange equations for $\mathcal{L}$,

$$\mathcal{EL}[\,\mathcal{L}\,] = \frac{\partial \mathcal{L}}{\partial x^\gamma} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathcal{L}}{\partial \dot{x}^\gamma}\right)\,.$$

Again, it makes sense to deal with the two terms in the Euler-Lagrange equations separately. The first term is

$$\frac{\partial \mathcal{L}}{\partial x^\gamma} = \frac{\partial}{\partial x^\gamma}\Big(\,g_{\alpha\beta}\,\dot{x}^\alpha\,\dot{x}^\beta\,\Big) = \frac{\partial g_{\alpha\beta}}{\partial x^\gamma}\,\dot{x}^\alpha\,\dot{x}^\beta\,. \tag{87}$$

Similarly, we have

$$\frac{\partial \mathcal{L}}{\partial \dot{x}^\gamma} = \frac{\partial}{\partial \dot{x}^\gamma}\Big(\,g_{\alpha\beta}\,\dot{x}^\alpha\,\dot{x}^\beta\,\Big) = g_{\gamma\beta}\dot{x}^\beta \;+\; g_{\alpha\gamma}\dot{x}^\alpha\,, \tag{88}$$

so that the second term is

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathcal{L}}{\partial \dot{x}^\gamma}\right) = \frac{\mathrm{d}}{\mathrm{d}t}\Big(\,g_{\gamma\beta}\dot{x}^\beta \;+\; g_{\alpha\gamma}\dot{x}^\alpha\,\Big) = \frac{\partial g_{\gamma\beta}}{\partial x^\alpha}\dot{x}^\alpha\,\dot{x}^\beta + g_{\gamma\beta}\ddot{x}^\beta + \frac{\partial g_{\alpha\gamma}}{\partial x^\beta}\dot{x}^\beta\,\dot{x}^\alpha + g_{\alpha\gamma}\ddot{x}^\alpha \tag{89}$$

Putting this altogether we have the Euler-Lagrange equations for $\mathcal{L}$ in the form

$$\frac{\partial g_{\gamma\beta}}{\partial x^\alpha}\dot{x}^\alpha\,\dot{x}^\beta + g_{\gamma\beta}\ddot{x}^\beta + \frac{\partial g_{\alpha\gamma}}{\partial x^\beta}\dot{x}^\beta\,\dot{x}^\alpha + g_{\alpha\gamma}\ddot{x}^\alpha - \frac{\partial g_{\alpha\beta}}{\partial x^\gamma}\dot{x}^\alpha\,\dot{x}^\beta = 0 \tag{90}$$

Again, this is simpler if we can multiply both sides by $g^{\delta\gamma}$ to get the final form for the Euler-Lagrange equations for $\mathcal{L}$:

$$\ddot{x}^\delta + \Gamma^\delta_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta = 0\ , \tag{91}$$

i.e., the Euler-Lagrange equations for $\mathcal{L} = L^2$ are *exactly* the equations for an affinely parametrised geodesic.

It turns out, not surprisingly, that $\dot{L} = 0$ is a direct consequence of eqn. (91); if $x^\alpha(t)$ satisfies the equations for an affinely parametrised geodesic, then $t$ is an affine parameter. This means that in the particular case of $f(L) = \mathcal{L} = L^2$, we have the result that

$$\Big\{\ \mathcal{EL}[\ \mathcal{L}\ ] = 0\ \Big\} \quad \Rightarrow \quad \left\{ \begin{array}{c} \dot{L} = 0 \\ \text{and} \\ \mathcal{EL}[\ \mathcal{L}\ ] = 0 \end{array} \right\} \quad \Rightarrow \quad \mathcal{EL}[\ L\ ] = 0\ ,$$

and all that we need to do to find the equations of an affinely parametrised geodesic is to find the Euler-Lagrange equations of

$$\int \mathcal{L}\ \mathrm{d}t = \int g_{\alpha\beta}\,\dot{x}^\alpha\,\dot{x}^\beta\ \mathrm{d}t\ .$$

- The first real gain we have made is that it is usually quite easy to put the metric directly into this formula and find the Euler-Lagrange equations. Writing the metric as $\mathrm{d}s^2$ and the Lagrangian as $\mathcal{L}$, we see

  $$\mathrm{d}s^2 = g_{\alpha\beta}\,\mathrm{d}x^\alpha\,\mathrm{d}x^\beta\ , \qquad \mathcal{L} = g_{\alpha\beta}\,\dot{x}^\alpha\,\dot{x}^\beta\ ,$$

  and we see that given the metric in the form $\mathrm{d}s^2 = \ldots$, it is very easy to write down the function $\mathcal{L}$: one just has to replace the differentials $\mathrm{d}x^\alpha$ by $\dot{x}^\alpha$, and then work out the Euler-Lagrange equations for the function that you get.

- The second real gain is that having found the Euler-Lagrange equations, one can then put them in the form $\ddot{x}^\alpha + \cdots = 0$, and just read off the components of the connection.

The best way to understand how this works is to try an example, so let's try to find the equations for an affinely parametrised geodesic and the components of the connection for the general two-dimensional diagonal metric

$$g_{\alpha\beta} = \begin{pmatrix} A(x,y) & 0 \\ 0 & B(x,y) \end{pmatrix} \quad , \quad \mathrm{d}s^2 = A\,\mathrm{d}x^2 \;+\; B\,\mathrm{d}y^2 \; . \tag{92}$$

From the second form, substituting $\mathrm{d}x^\alpha \to \dot{x}^\alpha$, we immediately find

$$\mathcal{L} = A\,\dot{x}^2 \;+\; B\,\dot{y}^2 \; ,$$

The Euler-Lagrange equation for $x$ for this function is very straightforward to find:

$$\begin{aligned} 0 \;=\; \frac{\partial \mathcal{L}}{\partial x} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial \dot{x}} \;&=\; \frac{\partial A}{\partial x}\,\dot{x}^2 + \frac{\partial B}{\partial x}\,\dot{y}^2 - \frac{\mathrm{d}}{\mathrm{d}t}\left( 2A\dot{x} \right) \\ &=\; \frac{\partial A}{\partial x}\,\dot{x}^2 + \frac{\partial B}{\partial x}\,\dot{y}^2 - 2A\ddot{x} - 2\frac{\partial A}{\partial x}\dot{x}^2 - 2\frac{\partial A}{\partial y}\dot{x}\dot{y} \\ &=\; -2A\ddot{x} - \frac{\partial A}{\partial x}\,\dot{x}^2 - 2\frac{\partial A}{\partial y}\,\dot{x}\dot{y} + \frac{\partial B}{\partial x}\,\dot{y}^2 \; . \end{aligned} \tag{93}$$

Writing this in the canonical form $\ddot{x} + \ldots = 0$, we have

$$\ddot{x} + \frac{1}{2A}\frac{\partial A}{\partial x}\,\dot{x}^2 + \frac{1}{A}\frac{\partial A}{\partial y}\,\dot{x}\dot{y} - \frac{1}{2A}\frac{\partial B}{\partial x}\,\dot{y}^2 = 0 \; . \tag{94}$$

If we write out the geodesic equation in components underneath this,

$$\ddot{x} + \Gamma^x_{xx}\,\dot{x}^2 + (\Gamma^x_{xy} \;+\; \Gamma^x_{yx})\,\dot{x}\dot{y} + \Gamma^x_{yy}\,\dot{y}^2 = 0 \; , \tag{95}$$

we see that we can very easily read off the connection components as

$$\Gamma^x_{xx} = \frac{1}{2A}\,\frac{\partial A}{\partial x} \; , \qquad \Gamma^x_{xy} = \Gamma^x_{yx} = \frac{1}{2A}\,\frac{\partial A}{\partial y} \; , \qquad \Gamma^x_{yy} = -\frac{1}{2A}\,\frac{\partial B}{\partial x} \; . \tag{96}$$

### 2.2.4 Example: geodesics on a sphere

As a first concrete example, let's look at a sphere of radius 1, find the corresponding geodesic equations in spherical polar coordinates, and see what the geodesics are. We have already worked out the metric on the unit sphere to be

$$\mathrm{d}s^2 \;=\; \mathrm{d}\theta^2 \;+\; \sin^2\!\theta\;\mathrm{d}\phi^2 \; . \tag{97}$$

Using the substitution $\mathrm{d}x^\alpha \to \dot{x}^\alpha$, we find the function $\mathcal{L}$ to be

$$\mathcal{L} \;=\; \left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2 \;=\; \dot{\theta}^2 \;+\; \sin^2\!\theta\;\dot{\phi}^2 \; . \tag{98}$$

The Euler-Lagrange equations for this function are:

$$\begin{aligned} &\ddot{\theta} \;=\; \sin\theta\,\cos\theta\,\dot{\phi}^2 \; , \\ &\left\{ \begin{aligned} &\ddot{\phi} \;=\; -2\,\dot{\theta}\,\dot{\phi}\,\cot\theta \\ &\frac{\mathrm{d}}{\mathrm{d}t}\left(\sin^2\!\theta\,\dot{\phi}\right) = 0 \end{aligned} \right. \end{aligned} \tag{99}$$

We shall leave it to next week to work out the general solution to these equations.

The equations for the geodesics on a sphere are rather fearsome looking equations at first sight, but we can try some curves which we believe should be geodesics, and check if they work. Our normal understanding of the sphere is that the geodesics are the great circles, of which simple examples are the meridians ($\phi$ constant) and the equator ($\theta = \pi/2$). Let's check these now:

• The meridians:

These have $\phi = \phi_0$, a constant, so that $\dot{\phi} = \ddot{\phi} = 0$. The two equations (99) now become

$$\ddot{\theta} = 0 \ , \quad 0 = 0 \ , \tag{100}$$

so that this is a geodesic with affine parameter $t$ provided that $\theta = \theta_0 + u\,t$.

• The equator:

We can now try a curve of constant $\theta = \theta_0$, so that $\dot{\theta} = \ddot{\theta} = 0$. Equations (99) now read

$$0 = (\dot{\phi})^2 \sin\theta_0 \cos\theta_0 \ , \quad \ddot{\phi} = 0 \ . \tag{101}$$

These have solutions in the following three cases:

1. $\dot{\phi} = 0$.

   Thus $\phi$ is also a constant and the whole geodesic is simply the point $\theta = \theta_0, \phi = \phi_0$. This is not a very interesting curve.

2. $\sin\theta_0 = 0$ and $\ddot{\phi} = 0$.

   In this case we need $\theta_0 = 0$ or $\theta_0 = \pi$, which mean that the whole geodesics are simple the North or South Poles of the sphere.

3. $\cos\theta_0 = 0$ and $\ddot{\phi} = 0$.

   This is the interesting case for which $\theta_0 = \pi/2$, and we find that the equator is a geodesic and $t$ is an affine parameter provided that $\phi = \phi_0 + u\,t$.

We can in fact go further to find all the geodesics on a sphere, as we do next.

### 2.2.5 The general geodesics on a sphere

The first point to note is that we do not really need to have an affine parametrisation $(\theta(t), \phi(t))$ of the geodesics – we can happily parametrise them by the longitude $\phi$ and hence we shall try to solve for the function $\theta(\phi)$.

To derive the equations for $\theta(\phi)$ from (99), we put $\theta = \theta(\phi(t))$ and find

$$\dot{\theta} = \theta' \, \dot{\phi} \; , \quad \ddot{\theta} = \theta'' \, (\dot{\phi})^2 + \theta' \, \ddot{\phi} \; .$$

Substituting these into the equation for $\ddot{\theta}$, and substituting $\ddot{\phi}$ the $\phi$ geodesic equation, we find

$$\ddot{\theta} - \sin\theta\cos\theta(\dot{\phi})^2 = (\dot{\phi})^2 \left(\theta'' - 2\cot\theta(\theta')^2 - \sin\theta\cos\theta\right) = 0 \; . \tag{102}$$

This equation is nonlinear in $\theta'$, but by changing variables $\theta \to u(\theta)$ we can try to remove the nonlinearity. This is indeed possible with (e.g.) $u = \cot\theta$, in which case the equation for $u(\theta)$ becomes

$$\theta'' - 2\cot\theta(\theta')^2 - \sin\theta\cos\theta = -\sin^2(\theta)(u'' + u) = 0 \; ,$$

or just

$$u'' + u = 0 \; , \tag{103}$$

with the general solution

$$u = A \, \sin(\phi - \phi_0) \; , \tag{104}$$

Since sin takes values between 1 and $-1$, it makes sense to write this equation in the form

$$\cot\theta = \cot\alpha \, \sin(\phi - \phi_0) \; . \tag{105}$$

We can now examine this solution. To find where the geodesic cuts the equator we look for the points at which $\theta = \pi/2$, i.e. $\cot\theta = 0$; for $\cot\alpha \neq 0$, this means $\phi = \phi_0$ and $\phi = \phi_0 + \pi$. Similarly, the maximal values of $\cot\theta$ occur at $\phi = \phi_0 + \pi/2$ and $\phi_0 + 3\pi/2$. Putting this together, we see that the general geodesic we have found is a great circle making an angle $\pi/2 - \alpha$ with the equator and passing through it at the points $\phi = \phi_0$ and $\phi = \phi_0 + \pi$, as in figure 24.

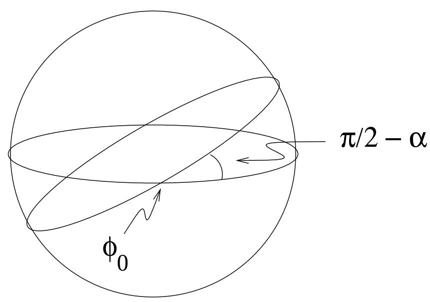



Figure 24

## 2.3 Parallel transport

The second idea we had to find an equation for geodesics is to try to generalise the idea that the tangent vector of a straight line is always pointing in the same direction. An immediate problem with phrasing the notion that the tangent vectors at all points along a curve are parallel is that, for a general curved surface, the tangent vectors at different points lie in different tangent planes. For example, if we consider a sphere, then we have already seen that the equator is a geodesic, but the tangent vector at $\phi = 0$ does not even lie in the tangent plane at $\phi = \pi/2$, and is exactly opposite to the tangent vector at $\phi = \pi$, as is shown in figure 25.

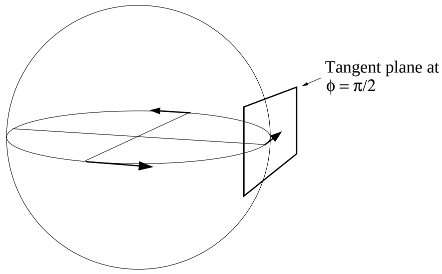



Figure 25

So, we see that since the tangent vector to a curve at one point does not even lie in the tangent plane at another point, there is no hope of finding an **absolute** concept of parallel for curved surfaces.

However, it is possible to replace this absolute idea of parallel by the weaker idea of *parallel transport*. This is based on the idea that at each point on a surface, if we look close enough it will look like a plane, and on the plane we have an idea of parallel. In this way, if we are given a curve, we can move a vector along a curve so that at each point it is roughly being kept pointing in the same direction.

There is a very physical way to imagine this: If we take a curve on the plane, we can cut out a small piece of the plane on either side of the curve and try to paste this onto the surface. If you try it, you will see that there is essentially only one way to do it. Now draw on the flat plane lots of parallel copies of the vector we want to parallel-transport. Then, when the strip of paper is pasted on the surface, at each point we will have a copy of the vector, and it seems natural that this is how to do parallel-transport.

Although we haven't worked out the equations of parallel transport, we can try an example since we know that the tangent vectors to geodesics are 'parallel transported' in this way, so any other vector parallel transported along a geodesic should always make the same angle with the tangent vector.

We already know that the geodesics on a sphere are the great circles, so that our straight-lines are mapped onto great circles on the sphere. So, let us consider a closed path on the sphere made up of three segments of great circles. We shall start at a point on the equator and move a distance $(\pi R/2)$ due east. Then we turn to the left and move due north the same distance to reach the north pole. Finally, turning again an angle $\pi/2$ to the left we shall travel the same distance due south to arrive at the point at which we started. Let's now suppose that initially we carry with us a vector pointing due east. After carrying this vector around our closed loop, always trying to keep it parallel, we find that in fact it has ended up pointing due north, i.e. it is at right-angles to the vector we first started out with, as shown in figure 26.

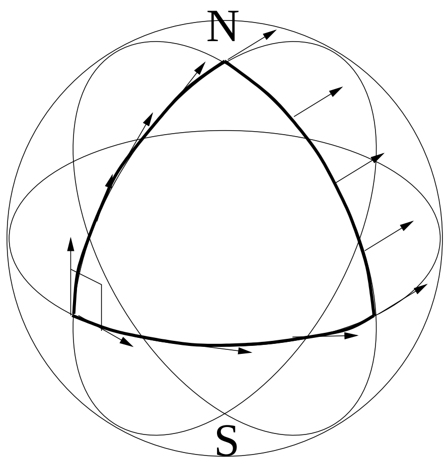


Figure 26

This should show that it will be a pointless task trying to define an **absolute** idea of parallel for vectors in the tangent planes attached to the points of a sphere.

However, *parallel transport* is a useful concept and we shall now find the mathematical formulation of this concept.

The basic idea we shall try to use is that parallel transport means parallel in the plane. On the plane, the metric in Cartesian coordinates is

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 \; . \tag{106}$$

In this metric, the equation for the parallel transport of a vector is simply the equation that the Cartesian components of the vector do not change:

$$\dot{N}^\alpha = 0 \; .$$

If we change to new coordinates $x'^\alpha$ which are related linearly to the old coordinates

$$x^\alpha = M^\alpha{}_\beta \, x'^\beta \; ,$$

with a constant matrix $M^\alpha{}_\beta$, then the metric in the new coordinates is also a constant

$$\mathrm{d}s^2 = \left(\delta_{\alpha\beta} \, M^\alpha{}_\gamma \, M^\beta{}_\delta\right) \mathrm{d}x'^\gamma \, \mathrm{d}x'^\delta \; ,$$

and the components of the vector in the new coordinates are related to the old by

$$N^\alpha = \frac{\partial x^\alpha}{\partial x'^\beta} \, N^\beta = M^\alpha{}_\beta \, N'^\beta \; ,$$

and hence the equation for parallel transport which are

$$\dot{N}^\alpha = M^\alpha{}_\beta \, \dot{N}'^\beta = 0 \; ,$$

become, when multiplied by the inverse of the constant matrix $M$,

$$\dot{N}'^{\alpha} = 0$$

in the new coordinates.

The reason for finding the form of the equations of parallel transport for a constant metric is that it is simpler for us to find a change of coordinates which will make a metric as close to constant as possible at a point, than it is to find a change which will actually make the metric close to (106). We now try to find such a change of coordinates which will render the metric as close to constant as possible.

On a general surface we have a non-constant metric,

$$\mathrm{d}s^2 = g_{\alpha\beta}(x)\,\mathrm{d}x^\alpha\,\mathrm{d}x^\beta \ ,$$

where the Taylor expansion of the metric about the origin is

$$g_{\alpha\beta}(x) = g_{\alpha\beta}(0) \ + \ x^\gamma \frac{\partial g_{\alpha\beta}}{\partial x^\gamma}(0) \ + \ \frac{1}{2} x^\delta x^\gamma \frac{\partial^2 g_{\alpha\beta}}{\partial x^\delta \partial x^\gamma}(0) \ + \ O(x^3) \ .$$

We cannot hope to make all the non-constant terms in the Taylor expansion of the metric vanish, but by a suitable change of coordinates we might be able to make enough of the derivatives vanish to enable us to take the equation for parallel transport of a vector *at the point* $x = 0$ to be simply given by the formula for the plane:

$$\dot{N}'^{\alpha} = 0 \ .$$

Let us consider the new coordinates $x'^\mu(x)$, and as we are only interested in the properties at the single point $x = 0$, we only need to think about the Taylor expansion of $x'^\mu$:

$$\begin{aligned} x'^\alpha &= x^\alpha \ + \ C^\alpha_{\beta\gamma} x^\beta x^\gamma \ + O(x^3) \ , \\ x^\alpha &= x'^\alpha \ - \ C^\alpha_{\beta\gamma} x'^\beta x'^\gamma \ + O(x'^3) \ . \end{aligned} \qquad (107)$$

(You can check that each of these equations implies the other). We note here that since $x^\beta x^\gamma$ is symmetric in $\beta$ and $\gamma$, we can safely assume that the constant coefficients $C^\alpha_{\beta\gamma}$ are also symmetric in $\beta$ and $\gamma$.

In order to find the components of the metric $g'_{\alpha\beta}$ and the vector $N'^\alpha$ in the new coordinates, we shall need the derivatives of these coordinate transformations close to the point $x = x' = 0$:

$$\begin{aligned} \frac{\partial x'^\alpha}{\partial x^\delta} &= \delta^\alpha_\delta + C^\alpha_{\beta\gamma}(\delta^\beta_\delta x^\gamma + x^\beta \delta^\gamma_\delta) + O(x^2) \\ &= \delta^\alpha_\delta + 2C^\alpha_{\delta\beta} x^\beta + O(x^2) \ , \\ \frac{\partial x^\alpha}{\partial x'^\delta} &= \delta^\alpha_\delta - 2C^\alpha_{\delta\beta} x'^\beta + O(x'^2) \ , \end{aligned} \qquad (108)$$

where we used the symmetry of $C$ in the lower two indices to simplify the expressions a little.

Using the second of these two expressions, we can find the Taylor expansion of the metric $g'(x')$ about $x' = 0$:

$$\begin{aligned} g'_{\gamma\delta}(x') &= \frac{\partial x^\alpha}{\partial x'^\gamma} \frac{\partial x^\beta}{\partial x'^\delta} g_{\alpha\beta}(x) \\ &= \left(\delta^\alpha_\gamma - 2C^\alpha_{\gamma\epsilon} x'^\epsilon + O(x'^2)\right)\left(\delta^\beta_\delta - 2C^\beta_{\delta\eta} x'^\eta + O(x'^2)\right)\left(g_{\alpha\beta}(0) + x^\zeta \frac{\partial g_{\alpha\beta}}{\partial x^\zeta}(0) + O(x^2)\right) \\ &= g_{\gamma\delta}(0) + x^\epsilon \left(\frac{\partial g_{\gamma\delta}}{\partial x^\epsilon}(0) - 2g_{\alpha\delta} C^\alpha_{\gamma\epsilon} - 2g_{\gamma\beta} C^\beta_{\delta\epsilon}\right) + O(x^2) \ , \end{aligned} \qquad (109)$$

where we have used the fact that $x = x' + O(x'^2)$ to turn all the $x'$s in this expression into $x$s.

Looking at this expression, we see that we can make the first order terms vanish, so that the leading corrections to the metric $g'$ *at the point* $x = 0$ are quadratic, provided the constants $C$ satisfy

$$\frac{\partial g_{\gamma\delta}}{\partial x^\epsilon}(0) = 2g_{\alpha\delta}\, C^\alpha_{\gamma\epsilon} + 2g_{\gamma\alpha}\, C^\alpha_{\delta\epsilon}\ . \tag{110}$$

We now have to try to solve this equation for $C$. This doesn't look very promising – there are two copies of $C$ on the right hand side of (110), multiplied by different matrices – but it turns out quite possible to solve this equation for $C$, provided you assume that $C^\alpha_{\beta\gamma}$ is symmetric in $\beta$ and $\gamma$ (and provided that you know what the answer is to begin with.)

The trick is to consider the combination

$$\Gamma^\alpha_{\beta\gamma}(0) = \frac{1}{2} g^{\alpha\delta}(0) \left( \frac{\partial g_{\beta\alpha}}{\partial x^\gamma}(0) + \frac{\partial g_{\gamma\alpha}}{\partial x^\beta}(0) - \frac{\partial g_{\beta\gamma}}{\partial x^\alpha}(0) \right)\ . \tag{111}$$

Putting the expression (110) for the partial derivatives if $x$ into eqn. (111), this equation just reduces to

$$\Gamma^\alpha_{\beta\gamma}(0) = 2C^\alpha_{\beta\gamma}\ . \tag{112}$$

Now we simply assume that the equations for parallel transport in coordinates $x'^\alpha$ *at the point* $x' = x = 0$ are

$$\dot{N}'^\alpha(0) = 0\ .$$

Using

$$N'^\alpha(x') = \frac{\partial x'^\alpha}{\partial x^\beta}\, N^\beta(x) = N^\alpha + \Gamma^\alpha_{\beta\gamma}(0) N^\beta x^\gamma + O(x^2)\ ,$$

we find that

$$0 = \dot{N}'^\alpha(x') = \dot{N}^\alpha(x) + \Gamma^\alpha_{\beta\gamma}(0) N^\beta(x) \dot{x}^\gamma + O(x)$$

or *at the point* $x = x' = 0$:

$$0 = \dot{N}^\alpha + \Gamma^\alpha_{\beta\gamma} N^\beta \dot{x}^\gamma$$

Since the coordinates $x^\alpha$ are completely general, there is nothing special about the point $x = 0$, so we can just try to apply this equation at all points along the curve $x^\alpha(t)$. We want this equation to be true at all points and in all coordinate systems, and it is not clear that this will be the case, but it is in fact the case as we shall prove when we come to discuss tensors. For the moment, we state that the equation for the components of a vector $N^\alpha(\ x(t)\ )$ to be parallel-transported along the curve $x^\alpha(t)$ are:

$$\boxed{\dot{N}^\alpha + \Gamma^\alpha_{\beta\gamma} N^\beta \dot{x}^\gamma = 0} \tag{113}$$

There is a special notation for the combination which appears in the equation of parallel transport, and that is

$$\boxed{\frac{\mathrm{D}N^\alpha}{\mathrm{d}t} \equiv \frac{\mathrm{d}N^\alpha}{\mathrm{d}t} + \Gamma^\alpha_{\beta\gamma} \frac{\mathrm{d}x^\beta}{\mathrm{d}t} N^\gamma} \tag{114}$$

This is called the *absolute derivative* of the vector $N^\alpha$ with respect to $t$ along the curve $x^\alpha(t)$.

We now make some observations.

- Since the metric $g'_{\alpha\beta}$ is constant to order $x^2$, this means that that the first derivative vanishes *at the point* $x' = 0$

$$\frac{\partial g'_{\alpha\beta}}{\partial x'^{\gamma}}(0) = 0 \; ,$$

  and hence the connection $\Gamma'^{\alpha}_{\beta\gamma}$ also vanishes at that point, since it is given simply in terms of the metric and its first derivative at that point, i.e. in these special coordinates

$$\Gamma'^{\alpha}_{\beta\gamma}(0) = 0 \; . \tag{115}$$

  Coordinates $x'$ such that eqn. (115) holds are called *normal coordinates* at the point $x' = 0$. They are very important in later discussions as they make many expressions much simpler.

- It is very important to note that the transformation $x'^{\alpha}(x^{\beta})$ which makes $\Gamma'^{\alpha}_{\beta\gamma}(0) = 0$ is *not unique*. The higher terms in the Taylor expansion of $x'^{\mu}$ are not fixed, and so there are an infinite number of different coordinate systems $x'^{\mu}$ for which $\Gamma'^{\alpha}_{\beta\gamma}(0) = 0$. We have simply exhibited one of them here.

- We still have to show that this equation makes sense – i.e. if a vector satisfies this equation in one coordinate system, does it satisfy it in all? The answer is yes, but we will defer this to next week.

### 2.3.1 Parallel transport on a sphere

First we can read off the Christoffel symbols for the unit sphere from the geodesic eqns (99)

$$\boxed{\begin{array}{lll} \Gamma^{\theta}_{\theta\theta} = 0 \; , & \Gamma^{\theta}_{\theta\phi} = \Gamma^{\theta}_{\phi\theta} = 0 \; , & \Gamma^{\theta}_{\phi\phi} = -\sin\theta\,\cos\theta \; , \\ \Gamma^{\phi}_{\phi\phi} = 0 \; , & \Gamma^{\phi}_{\theta\phi} = \Gamma^{\phi}_{\phi\theta} = \cot\theta \; , & \Gamma^{\phi}_{\theta\theta} = 0 \; . \end{array}} \tag{116}$$

We can now write down the parallel transport equations for the components of a vector $N^{\alpha}$, remembering that we shall replace the coordinates $x^{\alpha}$ by the angles

$$x^{\theta} \equiv \theta \; , \;\; x^{\phi} \equiv \phi \; ,$$

the parallel transport equations (113) become:

The $\theta$ equation:

$$\begin{array}{llllllll} \dot N^{\theta} & + \; ( & \Gamma^{\theta}_{\theta\theta} N^{\theta}\dot x^{\theta} & + \; \Gamma^{\theta}_{\theta\phi} N^{\theta}\dot x^{\phi} & + \; \Gamma^{\theta}_{\phi\theta} N^{\phi}\dot x^{\theta} & + & \Gamma^{\theta}_{\phi\phi} N^{\phi}\dot x^{\phi} & ) = 0 \\ \dot N^{\theta} & + \; ( & 0 & + \; 0 & + \; 0 & + & -\sin\theta\,\cos\theta\, N^{\phi}\dot\phi & ) = 0 \end{array} \tag{117}$$

and the $\phi$ equation:

$$\begin{array}{llllllll} \dot N^{\phi} & + \; ( & \Gamma^{\phi}_{\theta\theta} N^{\theta}\dot x^{\theta} & + \; \Gamma^{\phi}_{\theta\phi} N^{\theta}\dot x^{\phi} & + \; \Gamma^{\phi}_{\phi\theta} N^{\phi}\dot x^{\theta} & + & \Gamma^{\phi}_{\phi\phi} N^{\phi}\dot x^{\phi} & ) = 0 \\ \dot N^{\phi} & + \; ( & 0 & + \; \cot\theta\, N^{\theta}\dot\phi & + \; \cot\theta\, N^{\phi}\dot\theta & + & 0 & ) = 0 \end{array} \tag{118}$$

Writing these two equations out,

$$\begin{cases} \dot{N}^\theta = \sin\theta\,\cos\theta\;N^\phi\,\dot{\phi}\;, \\ \dot{N}^\phi = -\cot\theta\,(\,N^\theta\,\dot{\phi} + N^\phi\,\dot{\theta}\,)\;. \end{cases} \tag{119}$$

The first thing to note is that these equations cannot be solved – we need to specify the path $(\theta(t), \phi(t))$. As an example, let us consider the closed path in figure 27:

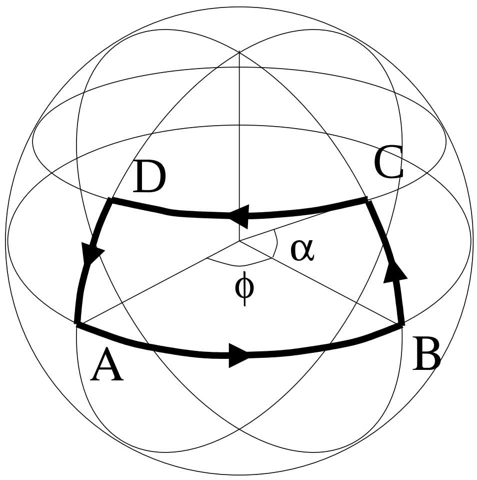



Figure 27

This path is formed of four segments,

1. $A \to B$. Here $\theta = \pi/2$ is constant and $\phi$ varies between 0 and $\phi$.
2. $B \to C$. Here $\phi$ is constant and $\theta$ varies between $\pi/2$ and $\pi/2 - \alpha$.
3. $C \to D$. Here $\theta = \pi/2 - \alpha$ is constant and $\phi$ varies between $\phi$ and 0.
4. $D \to A$. Here $\phi$ is constant and $\theta$ varies between $\pi/2 - \alpha$ and $\pi/2$.

To find how a vector transforms as it is parallel-transported along this closed path, we therefore only need to consider transport along paths with either $\theta$ constant or $\phi$ constant.

### 2.3.2 Parallel transport along $\theta$ =constant.

We can parametrise such a path by $\theta = \theta_0$ and $\phi = t$ so that $\dot{\theta} = 0$ and $\dot{\phi} = 1$ and $\mathrm{d}/\mathrm{d}t = \mathrm{d}/\mathrm{d}\phi$ The parallel transport equations (119) now become

$$\begin{cases} \dot{N}^\theta = \frac{\mathrm{d}N^\theta}{\mathrm{d}\phi} = \sin\theta_0\,\cos\theta_0\,N^\phi\;, \\ \dot{N}^\phi = \frac{\mathrm{d}N^\phi}{\mathrm{d}\phi} = -\cot\theta_0\,N^\theta\;. \end{cases} \tag{120}$$

We can turn these two coupled first order differential equations into a single second order differential equation by differentiating $\mathrm{d}N^\theta/\mathrm{d}\phi$ again and eliminating $N^\phi$ to get

$$\frac{\mathrm{d}^2 N^\theta}{\mathrm{d}\,\phi^2} = -\cos^2\theta_0\,N^\theta\;, \tag{121}$$

with the simple solution (finding $N^\phi$ by substituting into the first equation of (120)

$$\begin{aligned} N^\theta(\phi) &= A\cos(\cos\theta_0\,\phi) \;+\; B\sin(\cos\theta_0\,\phi)\;, \\ N^\phi(\phi) &= -\frac{A}{\sin\theta_0}\sin(\cos\theta_0\,\phi) \;+\; \frac{B}{\sin\theta_0}\cos(\cos\theta_0\,\phi)\;. \end{aligned} \tag{122}$$

The easiest way to see the effects of the total transport is to write the effect of a small curve of fixed $\theta$ or fixed $\phi$ in matrix form, and then multiply the matrices. From eqns (122) we find the effect of transport along a curve of fixed $\theta = \theta_0$ from $\phi_0$ to $\phi_1$ is

$$\begin{aligned} \begin{pmatrix} N^\theta(\phi_1) \\ N^\phi(\phi_1) \end{pmatrix} &= \begin{pmatrix} \cos(\,(\phi_1-\phi_0)\cos\theta_0\,) & \sin\theta_0\,\sin(\,(\phi_1-\phi_0)\cos\theta_0\,) \\ -\dfrac{\sin(\,(\phi_1-\phi_0)\cos\theta_0\,)}{\sin\theta_0} & \cos(\,(\phi_1-\phi_0)\cos\theta_0\,) \end{pmatrix} \begin{pmatrix} N^\theta(\phi_0) \\ N^\phi(\phi_0) \end{pmatrix} \\ &= L_{\theta_0}(\phi_1,\phi_0) \begin{pmatrix} N^\theta(\phi_0) \\ N^\phi(\phi_0) \end{pmatrix}\;, \end{aligned}$$

which defines the matrix $L_{\theta_0}(\phi_1,\phi_0)$.

### 2.3.3 Parallel transport along $\phi$ =constant

We can now repeat this for curves of $\phi = \phi_0$ and $\theta = t$ where $t$ runs between 0 and $\pi$, so that $\dot\phi = 0$ and $\dot\theta = 1$ and $\mathrm{d}/\mathrm{d}t = \mathrm{d}/\mathrm{d}\theta$. The parallel transport equations (119) are now even simpler than before,

$$\begin{cases} \dot N^\theta = \frac{\mathrm{d}N^\theta}{\mathrm{d}\,\theta} = 0\;, \\ \dot N^\phi = \frac{\mathrm{d}N^\phi}{\mathrm{d}\,\theta} = -\cot\theta\,N^\phi\;, \end{cases} \tag{123}$$

with the solutions

$$N^\theta(\theta) = A\;,\quad N^\phi(\theta) = \frac{B}{\sin(\theta)}\;. \tag{124}$$

Similarly, from eqns (124) the effect of transport along a curve of fixed $\phi = \phi_0$ from $\theta = \theta_0$ to $\theta = \theta_1$ is

$$\begin{pmatrix} N^\theta(\theta_1) \\ N^\phi(\theta_1) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \sin(\theta_0)/\sin(\theta_1) \end{pmatrix} \begin{pmatrix} N^\theta(\theta_0) \\ N^\phi(\theta_0) \end{pmatrix} = M_{\phi_0}(\theta_1,\theta_0) \begin{pmatrix} N^\theta(\theta_0) \\ N^\phi(\theta_0) \end{pmatrix}\;,$$

which defines the matrix $M_{\phi_0}(\theta_1,\theta_0)$.

### 2.3.4 The effect of the circuit considered in Figure 27

We can now work out the effect on a vector $N^\alpha$ of parallel transporting it along the closed curve in figure 27 by multiplying the matrices for the four segments, to get

$$\begin{aligned} &\begin{pmatrix} N^\theta \\ N^\phi \end{pmatrix}_{\text{final}} \\ &= M_0(\pi/2, \pi/2-\alpha)\;L_{\pi/2-\alpha}(0,\phi)\;M_\phi(\pi/2-\alpha,\pi/2)\;L_{\pi/2}(\phi,0)\;\begin{pmatrix} N^\theta \\ N^\phi \end{pmatrix}_{\text{initial}} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & \cos\alpha \end{pmatrix} \begin{pmatrix} \cos(\phi\sin\alpha) & -\cos\alpha\sin(\phi\sin\alpha) \\ \dfrac{\sin(\phi\sin\alpha)}{\cos\alpha} & \cos(\phi\sin\alpha) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1/\cos\alpha \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} N^\theta \\ N^\phi \end{pmatrix}_{\text{initial}} \\ &= \begin{pmatrix} \cos(\phi\sin\alpha) & -\sin(\phi\sin\alpha) \\ \sin(\phi\sin\alpha) & \cos(\phi\sin\alpha) \end{pmatrix} \begin{pmatrix} N^\theta \\ N^\phi \end{pmatrix}_{\text{initial}}\;. \end{aligned}$$

We see that the net effect is a rotation through an angle $\phi \sin\alpha$, which is exactly the area enclosed by the closed path.

This is an example of a general result that the result of parallel transporting a vector around a closed path on a two-dimensional surface is to rotate it through an angle equal to the integral of the *Gaussian Curvature* over the area enclosed by the path. The Gaussian curvature of the unit sphere is one, so the angle of rotation is exactly equal to the area enclosed in this case.

### 2.3.5 Length of a vector preserved under parallel transport

Although it was implicit in our derivation of the equations of parallel transport that the length of the vector is invariant – indeed the equation is $\dot{N}^\alpha = 0$ in normal coordinates at that point – we still have to check this. The length-squared of a vector $N^\alpha$ is

$$N^2 = g_{\alpha\beta}(x)\, N^\alpha\, N^\beta \ ,$$

so that

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\Big(N^2\Big) &= \frac{\mathrm{d}}{\mathrm{d}t}\Big(g_{\alpha\beta}(x)\Big)\, N^\alpha\, N^\beta \ + \ g_{\alpha\beta}\, \dot{N}^\alpha N^\beta \ + \ g_{\alpha\beta}\, N^\alpha\, \dot{N}^\beta \\ &= \frac{\partial g_{\alpha\beta}}{\partial x^\gamma}\, \dot{x}^\gamma\, N^\alpha N^\beta \ - \ g_{\alpha\beta}\, \Gamma^\alpha_{\gamma\delta}\, \dot{x}^\gamma\, N^\delta N^\beta \ - \ g_{\alpha\beta}\, N^\alpha\, \Gamma^\beta_{\gamma\delta}\, \dot{x}^\gamma N^\delta \ . \end{aligned}$$

Let's look at this more closely.

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(N^2\Big) = \frac{\partial g_{\alpha\beta}}{\partial x^\gamma}\, \dot{x}^\gamma\, N^\alpha N^\beta \ - \ g_{\alpha\beta}\, \Gamma^\alpha_{\gamma\delta}\, \dot{x}^\gamma\, N^\delta N^\beta \ - \ g_{\alpha\beta}\, N^\alpha\, \Gamma^\beta_{\gamma\delta}\, \dot{x}^\gamma N^\delta \ . \tag{125}$$

Each term in this expression has two $N$s and one $\dot{x}$ in it, and we would like to simplify it by pulling these outside. However, we can't do it immediately as they do not have the right indices. This is an example of a case where we shall have to do some relabelling of dummy indices in order to simplify an expression. (Remember that each dummy index represents a sum and we can change the name of the dummy index without altering the value of the expression.)

So, let's choose to pull out a factor of $N^\alpha\, N^\beta\, \dot{x}^\gamma$. This is already present in the first term so we need change nothing there. In the second term we have $N^\delta\, N^\beta\, \dot{x}^\gamma$, so that we shall have to change the dummy index $\delta$ to $\alpha$ – but now we notice that we are already using the dummy index $\alpha$ in the second term, so we shall also have to change that, to $\epsilon$ say, otherwise we will be breaking the rules of the summation convention.

Similarly, the third term contains $N^\alpha\, N^\delta\, \dot{x}^\gamma$, so we shall have to change the dummy indices $\delta$ to $\beta$, and $\beta$ to $\epsilon$.

Putting all this together, we arrive at the factorised expression

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(N^2\Big) = N^\alpha N^\beta\, \dot{x}^\gamma \left( \frac{\partial g_{\alpha\beta}}{\partial x^\gamma} \ - \ g_{\epsilon\beta}\, \Gamma^\epsilon_{\gamma\alpha} \ - \ g_{\alpha\epsilon}\, \Gamma^\epsilon_{\gamma\beta} \right) \ . \tag{126}$$

It's not obvious, but the term in brackets is identically zero, and the simplest way to see this is to substitute the expression (82) for $\Gamma^{\alpha}_{\beta\gamma}$ and collect terms:

$$\begin{aligned}&\frac{\partial g_{\alpha\beta}}{\partial x^\gamma} \;-\; g_{\epsilon\beta}\,\Gamma^{\epsilon}_{\gamma\alpha} \;-\; g_{\alpha\epsilon}\,\Gamma^{\epsilon}_{\gamma\beta}\\ =\;&\frac{\partial g_{\alpha\beta}}{\partial x^\gamma} \;-\; g_{\epsilon\beta}\,\frac{1}{2}\,g^{\epsilon\delta}\left(\frac{\partial g_{\delta\gamma}}{\partial x^\alpha}+\frac{\partial g_{\delta\alpha}}{\partial x^\gamma}-\frac{\partial g_{\gamma\alpha}}{\partial x^\delta}\right) \;-\; g_{\alpha\epsilon}\,\frac{1}{2}\,g^{\epsilon\delta}\left(\frac{\partial g_{\delta\gamma}}{\partial x^\beta}+\frac{\partial g_{\delta\beta}}{\partial x^\gamma}-\frac{\partial g_{\gamma\beta}}{\partial x^\delta}\right)\end{aligned}$$

In the second term in this expression, we have $g_{\epsilon\beta}\,g^{\epsilon\delta}$ which is equal to $\delta^{\delta}_{\beta}$. When we multiply anything with an index $X_\delta$ by this, we have $g_{\epsilon\beta}\,g^{\epsilon\delta}\,X_\delta = \delta^{\delta}_{\beta}\,X_\delta = X_\beta$, so the net contribution of the second term is the expression in brackets but with $\delta$ replaced by $\beta$. Similarly, the third term gives the expression in brackets with $\delta$ replaced by $\alpha$. The results is:

$$= \frac{\partial g_{\alpha\beta}}{\partial x^\gamma} \;-\; \frac{1}{2}\left(\frac{\partial g_{\beta\gamma}}{\partial x^\alpha}+\frac{\partial g_{\beta\alpha}}{\partial x^\gamma}-\frac{\partial g_{\gamma\alpha}}{\partial x^\beta}\right) \;-\; \frac{1}{2}\left(\frac{\partial g_{\alpha\gamma}}{\partial x^\beta}+\frac{\partial g_{\alpha\beta}}{\partial x^\gamma}-\frac{\partial g_{\gamma\beta}}{\partial x^\alpha}\right)\,,$$

and remembering that $g_{\alpha\beta} = g_{\beta\alpha}$, we find that all the terms cancel and the whole expression is zero:

$$\boxed{\frac{\partial g_{\alpha\beta}}{\partial x^\gamma} \;-\; g_{\epsilon\beta}\,\Gamma^{\epsilon}_{\gamma\alpha} \;-\; g_{\alpha\epsilon}\,\Gamma^{\epsilon}_{\gamma\beta} = 0\,,} \tag{127}$$

and hence the length of a vector $N^\alpha$ is preserved under parallel transport:

$$\frac{\mathrm{d}}{\mathrm{d}t}(N^2) = 0\,.$$

### 2.3.6 Geodesic equation again

Having studied parallel transport, we can now go back to the idea which motivated this study in the first place – that of geodesics. If you remember from section 2.2, we first introduced parallel transport since we wanted a mathematical formulation of the idea that a geodesic is a curve for which the tangent vector always remains parallel to the curve.

Since the tangent vector to a curve $x^\alpha(t)$ has components $\dot{x}^\alpha(t)$, we find the equation for a geodesic is simply the equation for the parallel transport of a vector $N^\alpha(t)$ along a curve $x^\alpha(t)$ –

$$\dot{N}^\alpha + \Gamma^{\alpha}_{\beta\gamma}\,N^\beta\,\dot{x}^\gamma = 0 \tag{128}$$

– but with the vector $N^\alpha(t)$ replaced by the tangent vector $\dot{x}^\alpha$, that is

$$\ddot{x}^\alpha + \Gamma^{\alpha}_{\beta\gamma}\,\dot{x}^\beta\,\dot{x}^\gamma = 0 \tag{129}$$

Notice that we have not found the general equation for a geodesic (81), but instead the equations for an affinely parametrised geodesic (84). This is because the parallel transport equations have the property of preserving the length of the vector being transported. For the application to geodesics, we are transporting the tangent vector, and hence our geodesics will have a tangent vector of constant length – i.e. they will be affinely parametrised geodesics.

### 2.3.7 Parallel transport and changes of coordinates

Finally, let us reconsider the question of whether the equations of parallel transport make sense irrespective of the coordinate system - if they depend on the coordinates we use, then they are hardly a very geometric concept.

We recall the combination (114) appearing in the equation of parallel transport, that is

$$\boxed{\frac{\mathrm{D}N^\alpha}{\mathrm{d}t} \equiv \frac{\mathrm{d}N^\alpha}{\mathrm{d}t} + \Gamma^\alpha_{\beta\gamma}\frac{\mathrm{d}x^\beta}{\mathrm{d}t}N^\gamma}$$

The question we have to ask ourselves then is whether a solution of

$$\frac{\mathrm{D}N^\alpha}{\mathrm{d}t} = 0$$

in one coordinate system $x^\alpha$ is the same as the solution to

$$\frac{\mathrm{D}N'^\alpha}{\mathrm{d}t} = 0$$

in another $x'^\alpha$. This reduces to working out the transformation properties of $\mathrm{D}N^\alpha/\mathrm{d}t$, and we will find that it is a vector.

Vectors have the property that if they are zero in one coordinate system, then they are zero in all, and this property, which follows from their simple transformation properties under coordinate transformations, is one of the properties of the more general class of objects called *Tensors.*

Tensors are useful for the two reasons that they transform simply under coordinate transformations (and hence if they are zero in one system they are zero in all – a very useful property if we are trying to formulate systems of equations) and secondly that Einstein's equations are naturally formulated in terms of tensors. For that reason we study them next.

## 2.4 Exercises

### Shorter Exercises

**2.1**

Consider the following embedding $\boldsymbol{\sigma}$ of a paraboloid into $\mathbb{R}^3$ and the path $\boldsymbol{\gamma}(t)$ in this paraboloid

$$\boldsymbol{\sigma}(x,y) = (x, x^2+y^2, y) \ , \quad \boldsymbol{\gamma}(t) = (\cos(t), 1, \sin(t)) \ .$$

Find the parametrisation of this path in terms of the coordinates, ie find $x(t)$ and $y(t)$ such that $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(x(t), y(t))$.

Calculate the vectors $\boldsymbol{e}_x$ and $\boldsymbol{e}_y$ at the point $\boldsymbol{\gamma}(0)$.

Calculate the tangent vector $\dot{\boldsymbol{\gamma}}(0)$.

Show that $\dot{\boldsymbol{\gamma}}(0)$ is a linear combination of $\boldsymbol{e}_x$ and $\boldsymbol{e}_y$ calculated at $\boldsymbol{\gamma}(0)$.

Repeat these calculations for the point $\boldsymbol{\gamma}(\pi/2)$.

**2.2**

Consider the unit sphere in $\mathbb{R}^3$.

The region $x > 0, y > 0, z > 0$ of this sphere $\mathbb{R}^3$ can be found from the two different embeddings $(x^1, x^2) \mapsto \boldsymbol{\sigma}(x^1, x^2)$ and $(x'^1, x'^2) \mapsto \boldsymbol{\sigma}'(x'^1, x'^2)$ defined by

$$\boldsymbol{\sigma}(x^1, x^2) = (x^1, x^2, \sqrt{1-(x^1)^2-(x^2)^2}) \ ,$$

$$\boldsymbol{\sigma}'(x'^1, x'^2) = (x'^1, \sqrt{1-(x'^1)^2-(x'^2)^2}, x'^2) \ .$$

Find the maps between the coordinate systems

$$x^\alpha \mapsto x'^\alpha(x) \ , \quad x'^\alpha \mapsto x^\alpha(x') \ ,$$

ie find $(x'^1, x'^2)$ in terms of $(x^1, x^2)$ and vice versa.

Find the two matrices

$$\frac{\partial x^\alpha}{\partial x'^\beta} \ , \quad \frac{\partial x'^\alpha}{\partial x^\beta} \ . \tag{130}$$

Show that their matrix product is the identity matrix, ie they are the matrix inverse of each other.

**2.3**

Consider two coordinate systems $x^\alpha = (u, v)$ and $x'^\alpha = (r, \theta)$ for a surface related by $u = r\cos\theta \ , \ \ v = r\sin\theta$ .

(a) Find the matrices

$$\frac{\partial x^\alpha}{\partial x'^\beta} = \begin{pmatrix} \partial u/\partial r & \partial u/\partial\theta \\ \partial v/\partial r & \partial v/\partial\theta \end{pmatrix} \ , \qquad \frac{\partial x'^\alpha}{\partial x^\beta} = \begin{pmatrix} \partial r/\partial u & \partial r/\partial v \\ \partial\theta/\partial u & \partial\theta/\partial v \end{pmatrix} \ . \tag{131}$$

(b Consider the point $P$ with coordinates $(u = 0, v = 1)$. Let $\mathbf{a}$ be the vector defined at $P$ with components $a^u = 1, a^v = 2$ in the coordinate system $(u, v)$.

Find the coordinates of $P$ in the coordinate system $(r, \theta)$ and the components $a^r$ and $a^\theta$ of $\mathbf{a}$ in those coordinates.

(c) Consider the point $Q$ with coordinates $(r = 2, \theta = \pi/3)$. Let $\mathbf{b}$ be the vector defined at $Q$ with components $b^r = 0, b^\theta = 1$ in the coordinate system $(r, \theta)$.

Find the coordinates of $Q$ in the coordinate system $(u, v)$ and the components $b^u$ and $b^v$ of $\mathbf{b}$ in those coordinates.

**2.4**

Consider the metric $g_{\alpha\beta}$ and vectors $\mathbf{a}$ and $\mathbf{b}$ which have components

$$g_{\alpha\beta} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \ , \ a^\alpha = \begin{pmatrix} 1 \\ 3 \end{pmatrix} \ , \ b^\alpha = \begin{pmatrix} 2 \\ -2 \end{pmatrix} \ . \tag{132}$$

(a) Find the components of $a_\alpha = g_{\alpha\beta} a^\beta$ and $b_\alpha = g_{\alpha\beta} b^\beta$.

(b) Use the results from part (a) to find $g_{\alpha\beta} a^\alpha a^\beta$, $g_{\alpha\beta} a^\alpha b^\beta$, and $g_{\alpha\beta} b^\alpha b^\beta$.

**2.5**

What are the signatures of the following metrics?

$$(a) \ g_{\alpha\beta} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \ , \ (b) \ g_{\alpha\beta} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \ , \ (c) \ g_{\alpha\beta} = \begin{pmatrix} 1 & 2 \\ 2 & 0 \end{pmatrix} \ . \tag{133}$$

**2.6**

Consider the $2 \times 2$ matrix

$$\epsilon_{\alpha\beta} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \ .$$

(a) Show that for all $v^\alpha$

$$\epsilon_{\alpha\beta} \, v^\alpha \, v^\beta = 0$$

(b) Suppose that $M^{\alpha\beta}$ is a symmetric matrix, i.e. that $M^{\alpha\beta} = M^{\beta\alpha}$, show, by writing out the expression in terms of the components of $M^{\alpha\beta}$ that

$$\epsilon_{\alpha\beta} M^{\alpha\beta} = 0 \ .$$

(c) Show the same result but without writing out the expression in terms of $M^{11}$, $M^{12}$ etc

**2.7**

A paraboloid in three dimensional Euclidean space (coordinates $x, y, z$) is given by the equations

$$x = u \cos\phi \ , \ \ y = u \sin\phi \ , \ \ z = \frac{u^2}{2} \ , \qquad u \ge 0 \ , \ 0 \le \phi < 2\pi \ .$$

Show that the metric induced on the paraboloid is given by

$$ds^2 = (1 + u^2) \, du^2 \ + \ u^2 \, d\phi^2 \ .$$

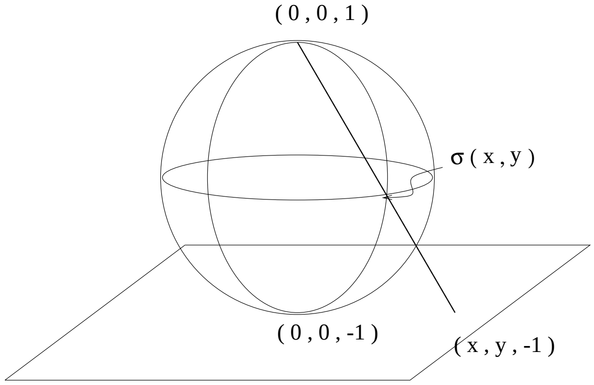
**2.8**

Find the Christoffel symbols for the plane in polar coordinates using the variational principle.

**Longer Exercises**

**2.9 Alternative coordinates for the sphere**

A sphere of radius one centred on the origin is tangent to the plane $z = -1$. Every straight line through the north pole $(0, 0, 1)$ that intersects the plane $z = -1$ also intersects the sphere at a unqiue point. If the coordinates of the point on the plane are $(x, y, -1)$, then we label the corresponding point on the sphere by these coordinates $(x, y)$. This is shown in the following figure:




Find the embedding $\boldsymbol{\sigma}(x, y)$, the basis vectors $\mathbf{e}_\alpha$ and the metric for this embedding.

**2.10 The *Pseudosphere* (or the surface of revolution generated by the *Tractrix*) – a surface with constant negative Gaussian curvature.**

The pseudosphere is a surface embedded in three dimensional Euclidean space given by the equations

$$\boldsymbol{\sigma} = \{e^u \cos\theta\ ,\ e^u \sin\theta\ ,\ \sqrt{1 - e^{2u}} - \cosh^{-1}(e^{-u})\,\} \qquad -\infty < u \leq 0\ ,\ \ 0 \leq \theta < 2\pi\ .$$

a) Show that the metric is

$$\mathrm{d}s^2 = \mathrm{d}u^2 + e^{2u}\mathrm{d}\theta^2\ .$$

b) Find the Christoffel symbols for this metric directly using the formula

$$\Gamma^\alpha_{\beta\gamma} = \frac{1}{2} g^{\alpha\delta}\ (\,\partial_\gamma g_{\delta\beta}\, +\, \partial_\beta g_{\delta\gamma}\, -\, \partial_\delta g_{\beta\gamma}\,)$$

and also using a variational principle.

Show that the **affine** geodesic equations are

$$\begin{aligned} \ddot{u} &= e^{2u}\,\dot{\theta}\,\dot{\theta} \\ \ddot{\theta} &= -2\,\dot{u}\,\dot{\theta} \end{aligned}$$

and hence show that the curves $\theta = \text{const.}$ are geodesics on the pseudosphere.

c) Find the Gaussian curvature $K$ of the pseudosphere using the formula

$$K = -G^{-1/2} \frac{\partial^2}{\partial u^2} \left( G^{1/2} \right) ,$$

valid for metrics of the form

$$g_{\alpha\beta} = \begin{pmatrix} 1 & 0 \\ 0 & G(u) \end{pmatrix}_{\alpha\beta} .$$

d) Consider the geodesics

$$\boldsymbol{\gamma}^{(\theta)}(t) = \{e^{u(t)} \cos\theta, e^{u(t)} \sin\theta, \sqrt{1 - e^{2u(t)}} - \cosh^{-1}(e^{-u(t)})\} , \tag{134}$$

where $\theta$ is a constant. Find the parametrisation $u(t)$ such that these geodesics are unit speed, i.e.

$$|\,\dot{\boldsymbol{\gamma}}\,|^2 = 1 .$$

e) Consider two the family of geodesics $\boldsymbol{\gamma}^{(\theta)}(t)$ of the form (134). Show that the deviation $\boldsymbol{\delta}$ of the metrics defined by

$$\boldsymbol{\delta} = \frac{\partial}{\partial\theta} \boldsymbol{\gamma}^{(\theta)}(t)$$

satisfies the geodesic deviation equation in two-dimensions:

$$\ddot{\boldsymbol{\delta}} + K\,\boldsymbol{\delta} = 0 .$$

## 2.11 The contribution of the spatial geometry to the gravitational deflection of light.

The geometry of a cross-section of space through a massive body is given by *Flamm's Paraboloid* with metric

$$\mathrm{d}s^2 = \frac{\mathrm{d}r^2}{1 - \frac{2m}{r}} + r^2 \mathrm{d}\theta^2 .$$

We shall find the deflection of a geodesic as it passes the body assuming that it start from infinity at $\theta = 0$ with "impact parameter" $R$, that is, if undeflected, it would pass a distance $R$ from the body.

(i) Let $t$ be an affine parameter along a geodesic. Show that the affine geodesic equation for $\theta$ leads to

$$h \equiv r^2 \dot{\theta} ,$$

being constant.

(ii) Consider the change of variables from $t$ to $\theta$ and let $' \equiv \mathrm{d}/\mathrm{d}\theta$. Further, define $u$ as $u \equiv 1/r$. In terms of $u$, the boundary conditions for the geodesic are $u(0) = 0$ and $u'(0) = 1/R$. Assuming that the length of the tangent vector to a geodesic is constant, show that

$$\frac{(u')^2}{(1 - 2mu)} + u^2 = \frac{1}{R^2} , \tag{135}$$

(iii) By finding an equation for $(u')^2$ and differentiating it, or otherwise, show that

$$u'' + u = -\frac{m}{R^2} + 3mu^2 \ , \tag{136}$$

(iv) Assume that we can make an expansion of the geodesic in powers of $(m/R)$ as

$$u = \frac{1}{R}\left(\sin\theta + u_1\frac{m}{R} + u_2\frac{m^2}{R^2} + \dots\right) \ .$$

By equating coefficients of $m/R^2$ in equation (136) show that $u_1$ satisfies

$$u_1'' + u_1 = \frac{1}{2} - \frac{3}{2}\cos(2\theta) \ .$$

(v) Assuming the initial conditions that $u(0) = 0$ and $u'(0) = 1/R$, show that

$$u = \frac{\sin\theta}{R} + \frac{m}{2\,R^2}(1 + \cos(2\theta) - 2\cos(\theta)) + \dots \ .$$

(vi) Assuming that $u(\pi + \Delta) = 0$, and that $\Delta$ is small, show that the total deflection of the orbit is approximately

$$\Delta \simeq \frac{2m}{R} \ .$$

This is exactly **one half** of the total deflection predicted by Einstein's theory.

**2.12 A question on embedded surfaces, Christoffel symbols and parallel transport.**

A paraboloid in three dimensional Euclidean space (coordinates $x, y, z$) is given by the equations

$$x = u\cos\phi \ , \quad y = u\sin\phi \ , \quad z = \frac{u^2}{2} \ , \qquad u \geq 0 \ , \ 0 \leq \phi < 2\pi \ .$$

a) Show that the metric induced on the paraboloid is given by

$$ds^2 = (1 + u^2)\,du^2 \ + \ u^2\,d\phi^2 \ .$$

b) Find the Christoffel symbols for this metric

$$\left[\text{You may use:} \quad \Gamma^\alpha_{\beta\gamma} = \frac{1}{2}\,g^{\alpha\delta}\,\left(\partial_\gamma g_{\delta\beta} \ + \ \partial_\beta g_{\delta\gamma} \ - \ \partial_\delta g_{\beta\gamma}\right)\right]$$

c) Solve the equations of parallel transport

$$\dot{n}^\alpha \ + \ \Gamma^\alpha_{\beta\gamma}\,n^\beta\,\dot{x}^\gamma = 0 \ ,$$

for the curve $u = u_0$ where $u_0$ is a positive constant, with the initial conditions $n^u = 1, n^\phi = 0$ at $\phi = 0$.

d) Verify that the length of the vector is preserved under parallel transport.

**2.13 Some results on geodesics.**

Consider a geodesic $x^\mu(\lambda)$ with affine parameter $\lambda$, satisfying

$$\frac{\mathrm{D}}{\mathrm{d}\lambda}\left[\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\right] = 0 \ , \tag{1}$$

(a) Derive the form of the geodesic equation for $x^\mu(\kappa)$, where the curve is re-parametrised by $\lambda \to \kappa(\lambda)$.

(b) Show that the geodesic equations has the form

$$\frac{\mathrm{D}}{\mathrm{d}\kappa}\left[\frac{\mathrm{d}x^\mu}{\mathrm{d}\kappa}\right] = 0 \ ,$$

if and only if $\kappa = A\,\lambda + B$.

(c) Show that any curve $x^\mu(\kappa)$ satisfying

$$\frac{\mathrm{D}}{\mathrm{d}\kappa}\left[\frac{\mathrm{d}x^\mu}{\mathrm{d}\kappa}\right] = f(\kappa)\,\frac{\mathrm{d}x^\mu}{\mathrm{d}\kappa} \ ,$$

for some function $f(\kappa)$ is a geodesic, and find a transformation $\kappa \to \kappa(\lambda)$ which brings it into the form (1).

**2.14 A result on metrics**

Two metrics $g_{\mu\nu}$ and $\hat{g}_{\mu\nu}$ are said to be *conformally related* if

$$\hat{g}_{\mu\nu} = \Omega^2\, g_{\mu\nu} \ ,$$

for some scalar function $\Omega = \Omega(x)$

Show that their Christoffel symbols are related by

$$\hat{\Gamma}^\mu_{\nu\rho} = \Gamma^\mu_{\nu\rho} + \Omega^{-1}\left(\delta^\mu_\nu \partial_\rho \Omega \ + \ \delta^\mu_\rho \partial_\nu \Omega \ - \ g^{\mu\sigma}\, g_{\nu\rho}\, \partial_\sigma \Omega\right) \ .$$

Lecture notes for Spacetime Geometry and General Relativity

2015-2016

# Section 3: Tensors

# 3 Tensors

Scalars and vectors are the simplest examples of *tensors*. Tensors are multi-index objects (including zero and one index) which are defined by the fact that they obey definite transformation rules under change of coordinates. There are other more elegant definitions, but this is the one we shall use in this course.

Tensors are objects with indices, which we shall simply call 'up' indices and 'down' indices, and if a tensor has $m$ up indices and $n$ down, we shall say it is of type $\left[{m \atop n}\right]$.

The correct names for the 'up' indices is *contravariant*, meaning varying in the *opposite* way to a set of basis vectors, and *covariant* for the 'down' indices, meaning varying in the *same* way as a set of basis vectors, but we will not use these names much, if at all, in this course

We have already seen examples of at least three sorts of tensors:

1. Scalars (such as the length-squared of a vectors) have no indices are and are tensors of type $\left[{0 \atop 0}\right]$
2. Vectors with components $U^a$ are tensors of type $\left[{1 \atop 0}\right]$
3. The metric with components $g_{ab}$ is a tensor of type $\left[{0 \atop 2}\right]$

However, be warned that not everything with indices is a tensor - for example the Christoffel symbols $\Gamma^a_{bc}$ are **not** the components of a tensor.

• It's also time to say a few words on the types of letters used for indices. Although this is clearly not important mathematically, it is often helpful to use certain sorts of indices for certain sorts of roles. In this course we shall use the following sorts of indices with the following ranges in the following cases:

| indices | range | situation |
|---|---|---|
| $\alpha, \beta, \gamma \ldots$ | $1, 2$ | Two-dimensional surfaces |
| $i, j, k \ldots$ | $1, 2, 3$ | Three-dimensional spaces of signature $(+\ +\ +)$ |
| $\lambda, \mu, \nu \ldots$ | $0, 1, 2, 3$ | Four-dimensional spacetimes of signature $(+\ -\ -\ -)$ |
| $a, b, c, d \ldots$ | $1, 2, \ldots n$ | General spaces of arbitrary dimension |

We have already seen two examples of tensor transformation laws, those of vectors and of the metric:

$$\begin{aligned} U'^a &= \frac{\partial x'^a}{\partial x^b} U^b \\ g'_{ab} &= \frac{\partial x^c}{\partial x'^a} \frac{\partial x^d}{\partial x'^b} g_{cd} \end{aligned}$$

These are both examples of the general transformation law for a tensor of type $\left[{m \atop n}\right]$, which is:

**Definition:**

A tensor $A$ of type $\left[{m \atop n}\right]$ is an object which, in any coordinate system, has components with $m$ indices 'up' and $n$ indices 'down', e.g.

$$A^{a_1 a_2 \ldots a_m}_{b_1 b_2 \ldots b_n} \ .$$

The components will change under changes of coordinates, and the rules is that if the components in coordinates $x^a$ are written $A^{a_1 a_2 \ldots a_m}_{b_1 b_2 \ldots b_n}$, and those in coordinate system $x'^a$ are $A'^{a_1 a_2 \ldots a_m}_{b_1 b_2 \ldots b_n}$, then these are related by

$$A'^{a_1 a_2 \ldots a_m}_{b_1 b_2 \ldots b_n} = \underbrace{\left(\frac{\partial x'^{a_1}}{\partial x^{c_1}}\right)\left(\frac{\partial x'^{a_2}}{\partial x^{c_2}}\right)\cdots\left(\frac{\partial x'^{a_m}}{\partial x^{c_m}}\right)}_{m \text{ matrices}} \underbrace{\left(\frac{\partial x^{d_1}}{\partial x'^{b_1}}\right)\left(\frac{\partial x^{d_2}}{\partial x'^{b_2}}\right)\cdots\left(\frac{\partial x^{d_n}}{\partial x'^{b_n}}\right)}_{n \text{ matrices}} A^{c_1 c_2 \ldots c_m}_{d_1 d_2 \ldots d_n} \tag{137}$$

Conversely, if the components $A^{a_1 a_2 \ldots a_m}_{b_1 b_2 \ldots b_n}$ in coordinate system $x^a$ and those in any other coordinate system $x'^b$, $A'^{a_1 a_2 \ldots a_m}_{b_1 b_2 \ldots b_n}$, are related by (137), then these are the components of a type $\left[{m \atop n}\right]$ tensor.

- One very important property of the two matrices in (137) is that they are the matrix inverses of each other:

$$\frac{\partial x'^a}{\partial x^b}\frac{\partial x^b}{\partial x'^c} = \frac{\partial x^a}{\partial x'^b}\frac{\partial x'^b}{\partial x^c} = \delta^a_c \ . \tag{138}$$

To prove this, consider the equation:

$$\frac{\partial x^a}{\partial x^c} = \delta^a_c \ .$$

However, if we consider $x^a$ to be a function of $x'^b$, then the chain rule gives

$$\delta^a_c = \frac{\partial}{\partial x^c} x^a(x'^b) = \frac{\partial x^a}{\partial x'^b}\frac{\partial x'^b}{\partial x^c} \ ,$$

as required. Since the inverse of a matrix is unique, this proves both of eqns. (138).

## 3.1 Properties of tensors

- Tensors of type $\left[{m \atop n}\right]$ form a vector space, that is one can take arbitrary linear multiples of tensors of the same type, add them, and the result is still a tensor. For example, if $A^a$ and $B^a$ are vectors, then $A^a + B^a$ is also a vector. However, it makes **no sense** to try to add tensors of different types: $A^a + B_{bc}$ is not a sensible object; if you find you have an expression containing something of this sort, then most probably you have made a mistake.

- One can take products of tensors in the following way: If

$$A^{a_1 \ldots a_m}_{b_1 \ldots b_n} \quad , \quad B^{c_1 \ldots c_p}_{d_1 \ldots d_q} \ ,$$

are tensors of types $\left[{m \atop n}\right]$ and $\left[{p \atop q}\right]$ respectively, then the object

$$C^{a_1 \ldots a_m c_1 \ldots c_q}_{b_1 \ldots b_n d_1 \ldots d_q} = A^{a_1 \ldots a_m}_{b_1 \ldots b_n} \, B^{c_1 \ldots c_p}_{d_1 \ldots d_q} \ ,$$

is a tensor of type $\left[{m+p \atop n+q}\right]$.

- One can also form new tensors from old by *contraction* on a pair of indices. If an tensor of type $\left[{m \atop n}\right]$ has an 'up' index $a$ and a 'down' index $b$, then one can take a sum over these two indices and the result will be a tensor of type $\left[{m-1 \atop n-1}\right]$.

  For example, if $A^{ab}_{cd}$ is a tensor of type $\left[{2 \atop 2}\right]$, then the following are all components of tensors of type $\left[{1 \atop 1}\right]$:

$$A^{ac}_{bc} \ , \quad A^{ca}_{bc} \ , \quad A^{ac}_{cb} \ , \quad A^{ca}_{cb} \ .$$

  One cannot however 'contract' over two 'up' indices or two 'down' indices. The result may make sense under the summation convention but it will **not be a tensor**.

  This contraction procedure works because of the property (138). For example, consider the contraction of a type $\left[{1 \atop 0}\right]$ tensor $A^a$ with a type $\left[{0 \atop 1}\right]$ tensor $B_a$:

$$C = B_a \, A^a \ .$$

  The resulting expression has no indices and so if it is really a type $\left[{0 \atop 0}\right]$ tensor, i.e. a scalar, then it should be the same in all coordinate systems.

  So, let us verify this:

$$C' = B'_a \, A'^a = (\, B_b \, \frac{\partial x^b}{\partial x'^a} \,)(\, \frac{\partial x'^a}{\partial x^c} \, A^c \,) = B_b \, (\, \frac{\partial x^b}{\partial x'^a} \, \frac{\partial x'^a}{\partial x^c} \,) A^c = B_b \, \delta^b_c \, A^c = B_b \, A^b = C \ .$$

There are two other very important properties of tensors. One is the quotient rule which we deal with in the next section. The other is the fact that

- If a tensor is zero in one coordinate system then it is zero in all coordinate systems

  This is evident from the transformation law for tensors (137), which we can write in the symbolic way

$$A'^{\cdots}_{\cdots} = (\partial x'/\partial x)^{\cdot}_{\cdot} \ldots (\partial x/\partial x')^{\cdot}_{\cdot} A^{\cdots}_{\cdots} \ .$$

  Since this is linear in $A$, if all the components of $A$ are zero, then so are all the components of $A'$.

### 3.1.1 Quotient rule

Our definition of a tensor is that it is something that transforms according to the rule (137). Sometimes this is straightforward to verify, but there is often a simpler way, and that is to use the quotient rule.

For example, if you know that for **any** vector $U^a$, then the expression

$$W = U^a V_a \tag{139}$$

is a scalar, then this implies that $V_a$ is a type $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ tensor. The fact that $U^a$ must be **any** vector is essential, and leads to a simple proof.

Since $W$ is a scalar and $U^a$ is a vector, that means that in another coordinate system $x'^a$,

$$W' = W \ , \quad U'^a = \frac{\partial x'^a}{\partial x^b} U^b \ , \quad U^a = \frac{\partial x^a}{\partial x'^b} U'^b \ . \tag{140}$$

However, by considering equation (139) in the new coordinate system, we have

$$U'^a V'_a = W' = W = U^a V_a = \frac{\partial x^a}{\partial x'^b} U'^b V_a \ . \tag{141}$$

Now, since $U^a$, and consequently $U'^a$ is arbitrary, we are free to consider the special vector which $U' = (1, 0, 0, ...)$ which has $U'^1 = 1$, and all other components zero. Then, eqn. (141) becomes

$$V'_1 = \frac{\partial x^a}{\partial x'^1} V_a \ . \tag{142}$$

Similarly, by choosing $U' = (0, 1, 0, 0...)$ so that $U'^2 = 1$, and all other components zero, we find

$$V'_2 = \frac{\partial x^a}{\partial x'^2} V_a \ , \tag{143}$$

and so on. As a result, we find that

$$V'_b = \frac{\partial x^a}{\partial x'^b} V_a \tag{144}$$

for all $b$, and so $V_a$ transforms like a type $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ tensor, and hence **is** a type $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ tensor.

This leads to the general result:

**The Quotient Rule:** If for **any** type $\left[\begin{smallmatrix}m+n\\p+q\end{smallmatrix}\right]$ tensor $A^{a_1\, a_2\, ...\, a_m\, b_1\, b_2 ...\, b_n}_{c_1\, c_2\, ...\, c_p\, d_1\, d_2 ...\, d_q}$, the expression

$$C^{b_1\, b_2 ...\, b_n\, e_1\, e_2 ...\, e_r}_{d_1\, d_2 ...\, d_q\, f_1\, f_2 ...\, f_s} = A^{a_1\, a_2\, ...\, a_m\, b_1\, b_2 ...\, b_n}_{c_1\, c_2\, ...\, c_p\, d_1\, d_2 ...\, d_q} \ B^{c_1\, c_2\, ...\, c_p\, e_1\, e_2 ...\, e_r}_{a_1\, a_2\, ...\, a_m\, f_1\, f_2 ...\, f_s} \tag{145}$$

is a type $\left[\begin{smallmatrix}n+r\\q+s\end{smallmatrix}\right]$ tensor, then $B^{c_1\, c_2\, ...\, c_p\, e_1\, e_2 ...\, e_r}_{a_1\, a_2\, ...\, a_m\, f_1\, f_2 ...\, f_s}$ is a type $\left[\begin{smallmatrix}p+r\\m+s\end{smallmatrix}\right]$ tensor.

### 3.1.2 Raising and lowering indices

We now introduce the very important idea of raising and lowering indices. Given a vector $U^a$, we can define a $\left[{0 \atop 1}\right]$ $U_b$ by

$$U_b = g_{ba}\, U^a \, .$$

Since $g^{ab}$ and $g_{bc}$ are inverse, we equally have

$$U^a = g^{ab}\, U_b \, .$$

Since $U_a$ and $U^a$ are so simply related, we think of them as essentially the same thing; of course their components have different values, but the relation between them is so straightforward that it introduces no confusion to call them by the same letter $U$.

The same is true of more general tensors - if we have a $\left[{1 \atop 2}\right]$ tensor $T^a{}_{bc}$, we can *raise* and *lower* the indices at will using the metric and its inverse. Of course, if we are going to move indices up and down we have to keep track of where the indices are, and for that reason we keep the indices in order, leaving spaces where appropriate:

$$T^{abc} = g^{bd}\, g^{ce}\, T^a{}_{de} \, , \quad T_{abc} = g_{ad}\, T^d{}_{bc} \, , \quad T_a{}^b{}_c = g_{ad}\, g^{be}\, T^d{}_{ec} \, , \quad \text{etc.} \tag{146}$$

In each case we think of '$T$' as being essentially the same thing with the various incarnations in eqns. (146) transforming as type $\left[{3 \atop 0}\right]$, $\left[{0 \atop 3}\right]$ and $\left[{1 \atop 2}\right]$ tensors respectively.

## 3.2 Differentiation of tensors

We now come to an important topic - the differentiation of tensors. The simplest example we can think of is the derivative of a scalar field $\phi$:

$$V_a = \frac{\partial \phi}{\partial x^a} = \partial_a\, \phi \, .$$

(Note that it is often very convenient to use the simplified notation $\partial_a$ for $\partial/\partial x^a$, $\partial'_a$ for $\partial/\partial x'^a$ etc.)

It is easy to show directly that this is a type $\left[{0 \atop 1}\right]$ tensor, or a 1–form as it is otherwise known:

$$V'_a = \frac{\partial \phi}{\partial x'^a} = \frac{\partial x^b}{\partial x'^a}\frac{\partial \phi}{\partial x^b} = \frac{\partial x^b}{\partial x'^a} V_b \, .$$

This is indeed the transformation law of a type $\left[{0 \atop 1}\right]$ tensor,

We can now try to differentiate a vector, and we shall find that the usual partial derivative of the components does not give us the components of a tensor. Suppose $U^a$ are the components of a vector, then $\partial_a U^b$ transforms as

$$\partial'_a\, U'^b = \frac{\partial}{\partial x'_a} U'^b = \frac{\partial x^c}{\partial x'^a}\frac{\partial}{\partial x^c}\left(\frac{\partial x'^b}{\partial x^d}\, U^d\right) = \frac{\partial x^c}{\partial x'^a}\frac{\partial x'^b}{\partial x^d}\,(\,\partial_c U^d\,) + \frac{\partial x^c}{\partial x'^a}\frac{\partial^2 x'^b}{\partial x^c \partial x^d}\, U^d \, . \tag{147}$$

The first term in (147) on its own gives exactly the correct transformation law for a type $\left[{1 \atop 1}\right]$ tensor, but the presence of the second term as well shows that $\partial_a U^b$ are not the components of a type $\left[{1 \atop 1}\right]$ tensor.

The question is: is there a sensible way to define the derivative of a tensor in such a way that we get a tensor back again? The answer is yes, it is called a covariant derivative.

### 3.2.1 Covariant Derivatives

A covariant derivative is written $\nabla$, (or in components $\nabla_a$) and must satisfy certain properties. These are simple generalisations of the properties of the partial derivative $\partial_a$ in Cartesian coordinates in flat space. These are

- $\nabla$ is a linear operator, and other such elementary properties
- $\nabla$ maps type $\left[{m \atop n}\right]$ tensors to type $\left[{m \atop n+1}\right]$ tensors.
- $\nabla$ obeys the Leibniz or product rule
- On scalar fields, $\nabla = \partial$. Since $\nabla\phi$ is a type $\left[{0 \atop 1}\right]$ tensor, it has components $\nabla_a\phi$, and so we require $\nabla_a\phi = \partial_a\phi$

This is sufficient to show that the covariant derivative is determined by a set of components, which are known as the components of a connection. In other words, the covariant derivative is defined by a connection which is itself defined by a set of components.

In principle there is a certain arbitrariness in defining a covariant derivative, but we shall require certain properties which pin it down uniquely. We'll show how this works in section 3.2.6, but for now we'll just 'guess' the answer and show how it works.

We already have some idea of how we might add something to $\partial_a U^b$ to give a type $\left[{1 \atop 1}\right]$ tensor since we guessed that the correct derivative to use for the equation of parallel transport was the absolute derivative,

$$\frac{\mathrm{D}U^a}{\mathrm{d}t} = \frac{\mathrm{d}U^a}{\mathrm{d}t} + \Gamma^a_{bc}\dot{x}^b U^c = \dot{x}^b\left(\partial_b U^a + \Gamma^a_{bc}U^c\right) \ .$$

The absolute derivative $\mathrm{D}U^a/\mathrm{d}t$ will transform as a vector for all $\dot{x}^b$ if and only if

$$\partial_b U^a + \Gamma^a_{bc}U^c \ ,$$

transforms as a type $\left[{1 \atop 1}\right]$ tensor.

This is indeed the case, and there is a special notation for this combination,

$$\boxed{\nabla_a U^b = \partial_a U^b + \Gamma^b_{ac}U^c} \tag{148}$$

which is known as the *covariant derivative* of $U^b$.

To check that this really does define a $\left[{1 \atop 1}\right]$ tensor we need to find the transformation properties of the Christoffel symbols. This calculation is about as lengthy as any we shall do in this course.

### 3.2.2 The transformation law of $\Gamma^a_{bc}$ (not examinable)

From the definitions,

$$\Gamma^a_{bc} = \frac{1}{2} g^{ad} \left( \partial_b g_{dc} + \partial_c g_{db} - \partial_d g_{bc} \right) \ , \quad \Gamma'^a_{bc} = \frac{1}{2} g'^{ad} \left( \partial'_b \, g'_{dc} + \partial'_c \, g'_{db} - \partial'_d \, g'_{bc} \right) \ , \tag{149}$$

where

$$\partial'_b = \frac{\partial}{\partial x'^b} = \frac{\partial x^a}{\partial x'^b} \frac{\partial}{\partial x^a} \ , \quad g'_{ab} = \frac{\partial x^c}{\partial x'^a} \frac{\partial x^d}{\partial x'^b} \, g_{cd} \ , \quad g'^{ab} = \frac{\partial x'^a}{\partial x^c} \frac{\partial x'^b}{\partial x^d} \, g^{cd} \ . \tag{150}$$

We can now evaluate one of the terms appearing in the bracket in the definition of $\Gamma'^a_{bc}$:

$$\begin{aligned}
\partial'_b \, g'_{dc} &= \partial'_b \left( \frac{\partial x^h}{\partial x'^d} \frac{\partial x^k}{\partial x'^c} \, g_{hk} \right) \\
&= \left( \partial'_b \frac{\partial x^h}{\partial x'^d} \right) \frac{\partial x^k}{\partial x'^c} \, g_{hk} + \frac{\partial x^h}{\partial x'^d} \left( \partial'_b \frac{\partial x^k}{\partial x'^c} \right) g_{hk} + \frac{\partial x^h}{\partial x'^d} \frac{\partial x^k}{\partial x'^c} \left( \partial'_b \, g_{hk} \right) \\
&= \frac{\partial^2 x^h}{\partial x'^d \partial x'^b} \frac{\partial x^k}{\partial x'^c} \, g_{hk} + \frac{\partial x^h}{\partial x'^d} \frac{\partial^2 x^k}{\partial x'^c \partial x'^b} \, g_{hk} + \frac{\partial x^h}{\partial x'^d} \frac{\partial x^k}{\partial x'^c} \frac{\partial x^j}{\partial x'^b} \left( \partial_j \, g_{hk} \right) \ .
\end{aligned} \tag{151}$$

Rather than work out $\Gamma'^a_{bc}$ all in one go, it's easier to split it into two parts – one ($\text{term}_1$) not containing any terms in $\frac{\partial^2 x^a}{\partial x'^b \partial x'^c}$ from (151), and the the other part ($\text{term}_2$) which does contain $\frac{\partial^2 x^a}{\partial x'^b \partial x'^c}$ terms. The first term is

$$\begin{aligned}
&\text{term}_1 \\
=\ & \frac{1}{2} \frac{\partial x'^a}{\partial x^m} \frac{\partial x'^d}{\partial x^n} \, g^{mn} \left( \frac{\partial x^h}{\partial x'^d} \frac{\partial x^k}{\partial x'^c} \frac{\partial x^j}{\partial x'^b} \left( \partial_j \, g_{hk} \right) + \frac{\partial x^h}{\partial x'^d} \frac{\partial x^k}{\partial x'^b} \frac{\partial x^j}{\partial x'^c} \left( \partial_j \, g_{hk} \right) - \frac{\partial x^h}{\partial x'^b} \frac{\partial x^k}{\partial x'^c} \frac{\partial x^j}{\partial x'^d} \left( \partial_j \, g_{hk} \right) \right) \\
=\ & \frac{1}{2} \frac{\partial x'^a}{\partial x^m} \frac{\partial x^k}{\partial x'^c} \frac{\partial x^j}{\partial x'^b} \, g^{mn} \partial_j \, g_{hk} \left( \frac{\partial x^h}{\partial x'^d} \frac{\partial x'^d}{\partial x^n} \right) + \frac{1}{2} \frac{\partial x'^a}{\partial x^m} \frac{\partial x^k}{\partial x'^b} \frac{\partial x^j}{\partial x'^c} \, g^{mn} \partial_j \, g_{hk} \left( \frac{\partial x^h}{\partial x'^d} \frac{\partial x'^d}{\partial x^n} \right) \\
& - \frac{1}{2} \frac{\partial x'^a}{\partial x^m} \frac{\partial x^k}{\partial x'^c} \frac{\partial x^h}{\partial x'^b} \, g^{mn} \partial_j \, g_{hk} \left( \frac{\partial x^j}{\partial x'^d} \frac{\partial x'^d}{\partial x^n} \right) \\
=\ & \frac{1}{2} \frac{\partial x'^a}{\partial x^m} \frac{\partial x^k}{\partial x'^c} \frac{\partial x^j}{\partial x'^b} \, g^{mn} \partial_j \, g_{hk} \, \delta^h_n + \frac{1}{2} \frac{\partial x'^a}{\partial x^m} \frac{\partial x^k}{\partial x'^b} \frac{\partial x^j}{\partial x'^c} \, g^{mn} \partial_j \, g_{hk} \, \delta^h_n - \frac{1}{2} \frac{\partial x'^a}{\partial x^m} \frac{\partial x^k}{\partial x'^c} \frac{\partial x^h}{\partial x'^b} \, g^{mn} \partial_j \, g_{hk} \, \delta^j_n \\
=\ & \frac{\partial x'^a}{\partial x^m} \frac{\partial x^k}{\partial x'^c} \frac{\partial x^j}{\partial x'^b} \left[ \frac{1}{2} g^{mn} \left( \partial_j \, g_{nk} + \partial_k \, g_{nj} - \partial_n \, g_{jk} \right) \right] \\
=\ & \frac{\partial x'^a}{\partial x^m} \frac{\partial x^j}{\partial x'^b} \frac{\partial x^k}{\partial x'^c} \, \Gamma^m_{jk}
\end{aligned} \tag{152}$$

This is the correct transformation property for a type $\left[ {1 \atop 2} \right]$ tensor, but we have also to include the term with the double derivatives:

$$\begin{aligned}
\text{term}_2 = \ & \frac{1}{2} g'^{ad} \Big( \frac{\partial^2 x^h}{\partial x'^d \partial x'^b} \frac{\partial x^k}{\partial x'^c} \, g_{hk} + \frac{\partial x^h}{\partial x'^d} \frac{\partial^2 x^k}{\partial x'^c \partial x'^b} \, g_{hk} + \frac{\partial^2 x^h}{\partial x'^d \partial x'^c} \frac{\partial x^k}{\partial x'^b} \, g_{hk} \\
& + \frac{\partial x^h}{\partial x'^d} \frac{\partial^2 x^k}{\partial x'^b \partial x'^c} \, g_{hk} - \frac{\partial^2 x^h}{\partial x'^b \partial x'^d} \frac{\partial x^k}{\partial x'^c} \, g_{hk} - \frac{\partial x^h}{\partial x'^b} \frac{\partial^2 x^k}{\partial x'^c \partial x'^d} \, g_{hk} \Big)
\end{aligned} \tag{153}$$

If we use the fact that $\frac{\partial^2 x^a}{\partial x'^b \partial x'^c} = \frac{\partial^2 x^a}{\partial x'^c \partial x'^b}$ and do some suitable relabelling, then the 1st term in (153) cancels with the 5th, the 3rd with the 6th, and the 2nd and 4th are equal, giving

$$\text{term}_2 = g'^{ad} \frac{\partial x^h}{\partial x'^d} \frac{\partial^2 x^k}{\partial x'^c \partial x'^b} \, g_{hk}$$

$$\begin{aligned}
&= \left(\frac{\partial x'^a}{\partial x^m}\frac{\partial x'^d}{\partial x^n}\,g^{mn}\right)\frac{\partial x^h}{\partial x'^d}\,\frac{\partial^2 x^k}{\partial x'^c \partial x'^b}\,g_{hk} \\
&= \frac{\partial x'^a}{\partial x^m}\,g^{mn}\,\frac{\partial^2 x^k}{\partial x'^c \partial x'^b}\,g_{hk}\left(\frac{\partial x^h}{\partial x'^d}\,\frac{\partial x'^d}{\partial x^n}\right) \\
&= \frac{\partial x'^a}{\partial x^m}\,g^{mn}\,\frac{\partial^2 x^k}{\partial x'^c \partial x'^b}\,g_{hk}\delta^h_n \\
&= \frac{\partial x'^a}{\partial x^m}\,\frac{\partial^2 x^k}{\partial x'^c \partial x'^b}\left(g_{hk}\,g^{mh}\right) \\
&= \frac{\partial x'^a}{\partial x^m}\,\frac{\partial^2 x^k}{\partial x'^c \partial x'^b}\left(\delta^m_k\right) \\
&= \frac{\partial x'^a}{\partial x^m}\,\frac{\partial^2 x^m}{\partial x'^c \partial x'^b} = \frac{\partial x'^a}{\partial x^m}\,\frac{\partial^2 x^m}{\partial x'^c \partial x'^b} \qquad (154)
\end{aligned}$$

Combining these two terms, we have the final result

$$\boxed{\Gamma'^a_{bc} = \frac{\partial x'^a}{\partial x^m}\frac{\partial x^j}{\partial x'^b}\frac{\partial x^k}{\partial x'^c}\,\Gamma^m_{jk} \; + \; \frac{\partial x'^a}{\partial x^m}\,\frac{\partial^2 x^m}{\partial x'^c \partial x'^b}\ .} \qquad (155)$$

To restate it, this is **not** the transformation law of a tensor. Since $\Gamma^a_{bc}$ plays such an important role, this is instead called the transformation law of a connection (remember that $\Gamma^a_{bc}$ is also called a connection). The second piece in (155) is zero if and only if

$$\frac{\partial^2 x^m}{\partial x'^c \partial x'^b} = 0\ ,$$

in which case the coordinate transformation is called *locally affine*.

We can also use the property (138) to rewrite (155) in terms of the second derivatives of $x'$ rather than $x$. Starting from

$$\begin{aligned}
0 &= \frac{\partial}{\partial x'^d}(\delta^a_c) = \frac{\partial}{\partial x'^d}\Big(\frac{\partial x'^a}{\partial x^b}\,\frac{\partial x^b}{\partial x'^c}\Big) = \Big(\frac{\partial}{\partial x'^d}\,\frac{\partial x'^a}{\partial x^b}\Big)\frac{\partial x^b}{\partial x'^c} + \frac{\partial x'^a}{\partial x^b}\Big(\frac{\partial}{\partial x'^d}\,\frac{\partial x^b}{\partial x'^c}\Big) \\
&= \Big(\frac{\partial x^e}{\partial x'^d}\frac{\partial}{\partial x^e}\,\frac{\partial x'^a}{\partial x^b}\Big)\frac{\partial x^b}{\partial x'^c} + \frac{\partial x'^a}{\partial x^b}\Big(\frac{\partial}{\partial x'^d}\,\frac{\partial x^b}{\partial x'^c}\Big) = \frac{\partial x^e}{\partial x'^d}\,\frac{\partial x^b}{\partial x'^c}\Big(\frac{\partial^2 x'^a}{\partial x^e \partial x^b}\Big) + \frac{\partial x'^a}{\partial x^b}\Big(\frac{\partial^2 x^b}{\partial x'^c \partial x'^d}\Big)
\end{aligned}$$

we can rewrite (155) as

$$\boxed{\Gamma'^a_{bc} = \frac{\partial x'^a}{\partial x^m}\frac{\partial x^j}{\partial x'^b}\frac{\partial x^k}{\partial x'^c}\,\Gamma^m_{jk} \; - \; \frac{\partial x^j}{\partial x'^b}\frac{\partial x^k}{\partial x'^c}\,\frac{\partial^2 x'^a}{\partial x^j \partial x^k}\ .} \qquad (156)$$

We can now combine this result with eqn. (147) to show that eqn. (148) does indeed define a type $\left[{}^1_1\right]$ tensor.

### 3.2.3 The covariant derivative of a vector (not examinable)

We now combine the transformation laws (147) and (156) to find that

$$\begin{aligned}
\nabla'_b U'^a &= \partial'_b U'^a + \Gamma'^a{}_{bc} U'^c \\
&= \left( \frac{\partial x^c}{\partial x'^b} \frac{\partial'^a}{\partial x^d} (\, \partial x_c U^d \,) + \frac{\partial x^c}{\partial x'^b} \frac{\partial^2 x'^a}{\partial^c \partial x^d} U^d \right) \\
&\quad + \left( \frac{\partial x'^a}{\partial x^m} \frac{\partial x^j}{\partial x'^b} \frac{\partial x^k}{\partial x'^c} \Gamma^m_{jk} - \frac{\partial x^j}{\partial x'^b} \frac{\partial x^k}{\partial x'^c} \frac{\partial^2 x'^a}{\partial x^j \partial x^k} \right) \left( \frac{\partial x'^c}{\partial x^l} U^l \right) \\
&= \frac{\partial x^c}{\partial x'^b} \frac{\partial'^a}{\partial x^d} (\, \partial_c U^d \,) + \frac{\partial x^c}{\partial x'^b} \frac{\partial^2 x'^a}{\partial^c \partial x^d} U^d \\
&\quad + \frac{\partial x'^a}{\partial x^m} \frac{\partial x^j}{\partial x'^b} ( \frac{\partial x^k}{\partial x'^c} \frac{\partial x'^c}{\partial x^l} ) \, \Gamma^m_{jk} \, U^l - \frac{\partial x^j}{\partial x'^b} ( \frac{\partial x^k}{\partial x'^c} \frac{\partial x'^c}{\partial x^l} ) \frac{\partial^2 x'^a}{\partial x^j \partial x^k} U^l \\
&= \frac{\partial x^c}{\partial x'^b} \frac{\partial'^a}{\partial x^d} (\, \partial_c U^d \,) + \frac{\partial x'^a}{\partial x^m} \frac{\partial x^j}{\partial x'^b} (\delta^k_l) \, \Gamma^m_{jk} \, U^l + \frac{\partial x^c}{\partial x'^b} \frac{\partial^2 x'^a}{\partial^c \partial x^d} U^d - \frac{\partial x^j}{\partial x'^b} (\delta^k_l) \frac{\partial^2 x'^a}{\partial x^j \partial x^k} U^l \\
&= \frac{\partial x^c}{\partial x'^b} \frac{\partial'^a}{\partial x^d} (\, \partial_c U^d \,) + \frac{\partial x'^a}{\partial x^m} \frac{\partial x^j}{\partial x'^b} \, \Gamma^m_{jk} \, U^k + \frac{\partial x^c}{\partial x'^b} \frac{\partial^2 x'^a}{\partial^c \partial x^d} U^d - \frac{\partial x^j}{\partial x'^b} \frac{\partial^2 x'^a}{\partial x^j \partial x^k} U^k \\
&= \frac{\partial x^c}{\partial x'^b} \frac{\partial'^a}{\partial x^d} \nabla_c U^d \,,
\end{aligned}$$

where we did some relabelling in the last line to cancel the two terms with second derivatives of the metric and to combine the terms which make up $\nabla_c U^d$.

This shows that the covariant derivative of a vector as defined in eqn. (148) is indeed a type $\left[{1 \atop 1}\right]$ tensor.

### 3.2.4 Covariant derivatives of arbitrary tensors

We can now try to extend the definition of the covariant derivative to more general tensors. So far we have the action on vectors, and we can say that the action on scalars is given simply by the partial derivative,

$$\nabla_a \phi = \partial_a \phi \,, \quad \nabla_a U^b = \partial_a U^b + \Gamma^b_{ac} U^c \,. \tag{157}$$

How can we define the action on 1–forms, that is on type $\left[{0 \atop 1}\right]$ tensors?

Well, we would like the covariant derivative to satisfy the Leibniz rule, that is we would like the action of $\nabla_a$ on the scalar given by the contraction of a vector and a 1–form $U^a V_a$ to obey both the rules for the derivative of a scalar, i.e.

$$\nabla_a(\, U^b V_b \,) = \partial_a(\, U^b V_b \,) = (\, \partial_a U^b \,) \, V_b + U^b \, (\, \partial_a V_b \,) \,, \tag{158}$$

and the Leibniz rules, that is

$$\nabla_a(\, U^b V_b \,) = (\, \nabla_a U^b \,) \, V_b + U^b \, (\, \nabla_a V_b \,) = (\, \partial_a U^b + \Gamma^b_{ac} U^c \,) V_b + U^b \, (\, \nabla_a V_b \,) \tag{159}$$

Putting these two equal, we find

$$U^b \, (\, \nabla_a V_b - \partial_a V_b + \Gamma^c_{ab} V_c \,) = 0 \,,$$

If this is to be true for all $U^b$, then the term in brackets must be identically zero, and so we find the form for the covariant derivative of a 1–form to be

$$\boxed{\nabla_a V_b = \partial_a V_b \; - \; \Gamma^c_{ab} V_c} \tag{160}$$

In exactly the same way we can find the expressions for the action of $\nabla_a$ on general type $\left[{m \atop n}\right]$ tensors. The result is that each 'up' index behaves like a vector index, and each 'down' index like a type $\left[{0 \atop 1}\right]$ index, as follows:

$$\nabla_a A^{b_1 b_2 \ldots b_m}_{c_1 c_2 \ldots c_n} = \partial_a A^{b_1 b_2 \ldots b_m}_{c_1 c_2 \ldots c_n} + \underbrace{\Gamma^{b_1}_{a\,a'} A^{a' b_2 \ldots b_m}_{c_1 c_2 \ldots c_n} + \cdots + \Gamma^{b_m}_{a\,a'} A^{b_1 b_2 \ldots a'}_{c_1 c_2 \ldots c_n}}_{m \text{ terms}} - \underbrace{\Gamma^{a'}_{a\,c_1} A^{b_1 b_2 \ldots b_m}_{a' c_2 \ldots c_n} - \cdots - \Gamma^{a'}_{a\,c_n} A^{b_1 b_2 \ldots b_m}_{c_1 c_2 \ldots a'}}_{n \text{ terms}} \tag{161}$$

- Note that the covariant derivative takes an especially simple form in normal coordinates at a point $\mathcal{P}$, i.e. coordinates in which $\Gamma^a_{bc}(\mathcal{P}) = 0$. In such coordinates, the covariant derivative of some tensor $x^{\cdots}{}_{\cdots}$ *at the point* $\mathcal{P}$ is simply

$$\nabla_a X^{\cdots}{}_{\cdots}(\mathcal{P}) = \partial_a X^{\cdots}{}_{\cdots}(\mathcal{P}) \ .$$

- Given the formulae (161) for the covariant derivative of general tensors, we can define the absolute derivative along a curve $x^a(\lambda)$ of general tensors $X$ by

$$\frac{\mathrm{D}}{\mathrm{d}\lambda} X^{\cdots}{}_{\cdots} = \frac{\mathrm{d}x^a}{\mathrm{d}\lambda} \nabla_a X^{\cdots}{}_{\cdots}$$

### 3.2.5 Covariant derivative of the metric

An important fact is that the action of taking covariant derivatives commutes with the action of raising and lowering indices, so that it does not matter if we raise or lower and index before or after taking the covariant derivative of an expression - we get the same answer, i.e. if we define

$$X_a{}^b = \nabla_a(V^b) \ ,$$

then we require that the following two equations be true

$$X_{ab} = g_{bc}\, X_a{}^c \ , \quad \text{and} \quad X_{ab} = \nabla_a(V_b) \ .$$

Expanding out the second form, and using the fact that $\nabla_a$ obeys the Leibniz rule,

$$\nabla_a(V_b) = \nabla_a(\, g_{bc}\, V^c\,) = \nabla_a(g_{bc})V^c + g_{bc}\,\nabla_a V^c = \nabla_a(g_{bc})V^c + g_{bc}\, X_a{}^c \ ,$$

we see that they can only both be true if the covariant derivative of the metric $\nabla_a g_{bc}$ is identically zero. This is indeed the case, as we can work out quite easily.

The metric is a $\left[{0 \atop 2}\right]$ tensor, so from (161), we have

$$\nabla_a\, g_{bc} = \partial_a g_{bc} - \Gamma^{a'}_{ab}\, g_{a'\,c} - \Gamma^{a'}_{ac}\, g_{b\,a'} \ . \tag{162}$$

However, we have already worked out this expression in equation (127) (in section 2.3.5 where we worked out that the length of a vector is constant under parallel transport) and in which we found that exactly this combination vanished identically:

$$\partial_a g_{bc} - \Gamma^{a'}_{ab}\, g_{a'\,c} - \Gamma^{a'}_{ac}\, g_{b\,a'} = 0$$

and hence we find that the metric is covariantly constant, i.e.

$$\boxed{\nabla_a\, g_{bc} = 0 \ .} \tag{163}$$

An important example of this is that we can raise and lower indices before and after taking the absolute derivative, and the answer is the same. Consider the 4–velocity $U^\mu$ of a particle, and the acceleration $A^\mu$ defined to be the absolute derivative of the velocity with respect to the proper time:

$$A^\mu = \frac{\mathrm{D}U^\mu}{\mathrm{d}\tau} = U^\nu \nabla_\nu\, U^\mu \ .$$

Then, we have

$$A_\mu = g_{\mu\nu} A^\nu = g_{\mu\nu}\,(U^\sigma \nabla_\sigma\, U^\nu) = U^\sigma \nabla_\sigma\ (g_{\mu\nu} U^\nu) - U^\sigma\, (\nabla_\sigma g_{\mu\nu})\ U^\nu = U^\sigma \nabla_\sigma\ (U_\mu) - 0 = \frac{\mathrm{D}U_\mu}{\mathrm{d}\tau} \ ,$$

so that the following four equations, which *a priori* are not necessarily consistent, are in fact consistent:

$$\boxed{U_\mu = g_{\mu\nu} U^\nu \ , \ \ A_\mu = g_{\mu\nu} A^\nu \ , \ \ A^\mu = \frac{\mathrm{D}U^\mu}{\mathrm{d}\tau} \ , \ \ A_\mu = \frac{\mathrm{D}U_\mu}{\mathrm{d}\tau} \ .} \tag{164}$$

### 3.2.6 The form of a general covariant derivative (not examinable)

As mentioned before, the requirements listed at the start of section 3.2.1 do not fix the form of the covariant derivative entirely. We now find what it is that determines a general covariant derivative that satisfies this list of properties.

A simple way to do this is to introduce a basis of vectors $e_{(a)}$ and a dual basis of type $\left[{1 \atop 1}\right]$ tensors, $e^{(a)}$. We can take our basis of vectors to have particularly simple components in coordinates $x^a$, for example $e_{(1)} = (1, 0, 0, ...0)$ $e_{(2)} = (0, 1, 0, ..., 0)$ etc, i.e. $e^b_{(a)} = \delta^b_a$.

It is important to realise that if we change coordinates these vectors will still be a basis of the space of vectors, but will not have especially simple components. In a different coordinate system,

$$e'_{(a)}{}^b = \frac{\partial x'^b}{\partial x^c} e^c_{(a)} = \frac{\partial x'^b}{\partial x^c} \delta^c_a = \frac{\partial x'^b}{\partial x^a} . \tag{165}$$

Secondly, these vectors may look like unit vectors, but it is easy to see that they are not. Their inner products are given by

$$e_{(a)} \cdot e_{(b)} = g_{cd}\, e^c_{(a)}\, e^d_{(b)} = g_{cd}\, \delta^c_a\, \delta^d_b = g_{ad} .$$

Using these basis vectors, we can write any vector as a sum of basis vectors in the form

$$V = V^{(a)} e_{(a)} , \tag{166}$$

where $V^{(a)}$ are scalar functions which happen to have the same values as the components of $V$ in our coordinate system $x^a$, since taking components of (166),

$$V^a = \left(V^{(b)} e_{(b)}\right)^a = V^{(b)} e^a_{(b)} = V^{(b)} \delta^a_b = V^{(a)} .$$

When we change coordinates, equation (166) remains true - the functions $V^{(a)}$ expressing the decomposition of $V$ into basis vectors do not change. When we take components of (166) in a different coordinate system, the components of the basis vectors are different, not the scalar functions $V^{(a)}$.

$$V'^a = V^{(b)} e'_{(b)}{}^a = V^{(b)} \frac{\partial x'^a}{\partial x^b} .$$

We can equally well choose a basis of type $\left[{0 \atop 1}\right]$ tensors (also known as 1-forms) to be the 1-forms which have equally simple components in our coordinate system $x^a$, i.e.

$$e^{(1)} = (1, 0, 0, ....) , \ \ e^{(2)} = (0, 1, 0, ....) , \ \text{etc} \quad \text{i.e.} \ e^{(a)}_b = \delta^a_b .$$

These are dual to the basis of vectors we choose before since the contraction of the 1-form $e^{(a)}$ with the vector $e_{(a)}$ is

$$e^{(a)}_b\, e^b_{(c)} = \delta^a_c , \tag{167}$$

which is true in *any* coordinate system.

We can now use the bases $e_{(a)}$ and $e^{(b)}$ to write general tensors. Let's start with type $\left[{1 \atop 1}\right]$ tensors. Then there is a way of making a type $\left[{1 \atop 1}\right]$ tensor out of a type $\left[{1 \atop 0}\right]$ tensor and a type $\left[{0 \atop 1}\right]$ tensor, and that is called the **tensor product**, which is denoted $\otimes$. In general, the tensor product takes a vector space $\mathcal{U}$ of dimension $M$ and a vector space $\mathcal{V}$ of dimension $N$, and gives you a vector space

$$\mathcal{U} \otimes \mathcal{V} \ ,$$

of dimension $M \times N$. If $u_i$ are a basis of $\mathcal{U}$ and $v_j$ of $\mathcal{V}$ then a basis of $\mathcal{U} \otimes \mathcal{V}$ is given by

$$u_i \otimes v_j \ .$$

The product $\otimes$ obeys the following laws:

- If $a \in \mathcal{U}$ and $f \in \mathcal{V}$ and $A, C \in \mathbb{R}$, then

$$(Aa) \otimes (Cf) = AC(a \otimes f) \in \mathcal{U} \otimes \mathcal{V}$$

- If $a, b \in \mathcal{U}$ and $f, g \in \mathcal{V}$ and $A, B, C, D \in \mathbb{R}$, then

$$(Aa + Bb) \otimes (Cf + Gg) = AC(a \otimes f) + BC(b \otimes f) + AG(a \otimes g) + BG(b \otimes g) \ .$$

We can now apply this to the vector space of type $\left[{1 \atop 0}\right]$ tensors and type $\left[{0 \atop 1}\right]$ tensors. Since $e_{(a)}$ are a basis for type $\left[{1 \atop 0}\right]$ tensors, and $e^{(b)}$ are a basis for type $\left[{0 \atop 1}\right]$ tensors, this means that

$$e_{(a)} \otimes e^{(b)} \ ,$$

are a basis for type $\left[{1 \atop 1}\right]$ tensors. We can thus write any type $\left[{1 \atop 1}\right]$ tensor $T$ as a linear combination of these basis tensors,

$$T = \sum_{a=1}^{n} \sum_{b=1}^{n} \left( \ T^{(a)}{}_{(b)} \ e_{(a)} \otimes e^{(b)} \ \right) \ .$$

In this way, $T$ has a meaning independent of any coordinate system In exactly the same way,

$$e_{(a_1)} \otimes e_{(a_2)} \otimes \cdots e_{(a_m)} \otimes e^{(b_1)} \otimes e^{(b_2)} \otimes \cdots e^{(b_n)} \ ,$$

is a type $\left[{m \atop n}\right]$ tensor, and we write a type $\left[{m \atop n}\right]$ tensor $T$ in coordinate free notation by writing it as a sum of basis tensors:

$$T = T^{(a_1)\cdots(a_m)}{}_{(b_1)\cdots(b_n)} \ e_{(a_1)} \otimes e_{(a_2)} \otimes \cdots e_{(a_m)} \otimes e^{(b_1)} \otimes e^{(b_2)} \otimes \cdots e^{(b_n)} \ . \tag{168}$$

Since we want the covariant derivative to map vectors (i.e. type $\left[{1 \atop 0}\right]$ tensors) to type $\left[{1 \atop 1}\right]$ tensors, the covariant derivative of the basis vectors $e_{(a)}$ must be expressible in terms of the basis of type $\left[{1 \atop 1}\right]$ tensors, so that there must be an equation of the type

$$\nabla e_{(a)} = \gamma^b_{ca} e_{(b)} \otimes e^{(c)} \ ,$$

or in components,

$$\nabla_c (e^b_{(a)}) = \left(\nabla e_{(a)}\right)^b_c = \gamma^b_{ca} \ .$$

This means we can now work out the form of the covariant derivative of an arbitrary vector using (166):

$$\nabla V = \nabla(V^{(a)} e_{(a)}) = \nabla(V^{(a)}) e_{(a)} + V^{(a)}(\nabla e_{(a)}) = \partial(V^{(a)}) e_{(a)} + V^{(a)}(\gamma^b_{ca} e_{(b)} \otimes e^{(c)}) \ .$$

In components, this means

$$\nabla_a V^b = \partial_a V^b + \gamma^b_{ac} V^c \ .$$

Thus the covariant derivative of a vector is uniquely determined by its action on the basis vectors.

We can also work out the covariant derivative of an arbitrary 1 form by working out the covariant derivative of the basis 1-forms, which we can determine from differentiating (167) in the same way as in section 3.2.4. We find that the covariant derivative of a 1–form with components $w_b$ is given by

$$\nabla_a \, w_b = \partial_a \, w_b - \gamma^c_{ab} \, w_c \ .$$

In this way we can find the covariant derivative of *any* tensors by using the expansion (168) and the form of the covariant derivative of a scalar function, a basis vector and a basis 1-form. The covariant derivative is entirely fixed by the functions $\gamma^a_{bc}$.

This is still very general, and includes a much larger class of covariant derivatives than the single one in which we are interested. There are two simple requirements which will fix these functions $\gamma^a_{bc}$ to be the functions $\Gamma^a_{bc}$ which we have already found.

These requirements are that covariant derivatives acting on scalar fields commute, and that covariant differentiation commutes with raising and lowering indices.

We can act with two covariant derivatives $\nabla_a$ and $\nabla_b$ on a scalar field in two orders, and compare the answers:

$$\nabla_a \nabla_b \phi - \nabla_b \nabla_a \phi = \partial_a \partial_b \phi - \gamma^c_{ab} \nabla_c \phi - \partial_b \partial_a \phi + \gamma^c_{ba} \nabla_c \phi = -(\gamma^c_{ab} - \gamma^c_{ba}) \nabla_c \phi \ ,$$

since partial derivatives commute.

We can now apply the quotient theorem. Since $\phi$ is an arbitrary scalar function, the value of $\nabla_c \phi$ at a point is arbitrary, and so for an arbitrary value of the tensor $\nabla_c \phi$ we know that $(\gamma^c_{ab} - \gamma^c_{ba}) \nabla_c \phi$ is a type $\left[{0 \atop 2}\right]$ tensor, and hence the combination $(\gamma^c_{ab} - \gamma^c_{ba})$ must be the components of a type $\left[{1 \atop 2}\right]$ tensor.

This is known as the *torsion* of the connection, and is written $T^a{}_{bc}$,

$$T^a{}_{bc} = (\gamma^a_{bc} - \gamma^a_{cb}) \ ,$$

and the result for the commutator of the covariant derivative on a scalar field is

$$[\, \nabla_a \, , \, \nabla_b \,] \, \phi = -T^c{}_{ab} \, \nabla_c \phi \ ,$$

An alternative proof that $T^a{}_{bc}$ is a tensor is to use the tensor transformation law for $[\nabla_a, \nabla_b]\phi]$ and choose $\phi = x'^g$, one of the coordinates, so that:

$$\nabla_a{}' \phi = \nabla_a{}' x'^g = \frac{\partial x'^g}{\partial x'^a} = \delta^g_a \ , \quad \nabla_a \phi = \nabla_a x'^g = \frac{\partial x'^g}{\partial x^a} \ .$$

This then means

$$T'^g{}_{ab} = T'^c{}_{ab} \nabla_c{}'(x'^g) = T'^c{}_{ab} \nabla_c{}' \phi = \frac{\partial x^d}{\partial x'^a} \frac{\partial x^e}{\partial x'^b} (T^f{}_{de} \nabla_f \phi) = \frac{\partial x^d}{\partial x'^a} \frac{\partial x^e}{\partial x'^b} \frac{\partial x'^g}{\partial x^f} T^f{}_{de} \ .$$

which is exactly the required tensor transformation law for the torsion tensor.

There are theories of gravity which include a torsion term, but Einstein's theory does not. We shall simply require that the torsion vanishes, i.e. that

$$\gamma^c_{ab} = \gamma^c_{ba} \,.$$

Such a connection is called *symmetric.*

The second requirement, that covariant differentiation commute with raising and lowering indices boils down to the requirement that the covariant derivative of the metric be zero, as shown in section 3.2.5. Since the metric is a type $\left[{0 \atop 2}\right]$ tensor, its covariant derivative is

$$\nabla_a g_{bc} = \partial_a g_{bc} \;-\; \gamma^d_{ab}\, g_{dc} \;-\; \gamma^d_{ac}\, g_{bd} \,. \tag{169}$$

Hence if (169) holds, we have the following three identities

$$\begin{aligned} \partial_a g_{bc} &= \gamma^d_{ab}\, g_{cd} \;+\; \gamma^d_{ac}\, g_{db} \;, \\ \partial_c g_{ba} &= \gamma^d_{cb}\, g_{ad} \;+\; \gamma^d_{ca}\, g_{db} \;, \\ -\partial_b g_{ac} &= -\gamma^d_{ba}\, g_{cd} \;-\; \gamma^d_{bc}\, g_{da} \,. \end{aligned}$$

Adding these three together we have

$$\begin{aligned} & \partial_a g_{bc} + \partial_c g_{ba} - \partial_b g_{ac} \\ =\; & \gamma^d_{ab}\, g_{dc} \;+\; \gamma^d_{ac}\, g_{bd} \;+\; \gamma^d_{cb}\, g_{da} \;+\; \gamma^d_{ca}\, g_{bd} \;-\; \gamma^d_{ba}\, g_{dc} \;-\; \gamma^d_{bc}\, g_{ad} \\ =\; & g_{dc}\left(\gamma^d_{ab} \;-\; \gamma^d_{ba}\right) \;+\; g_{ad}\left(\gamma^d_{cb} \;-\; \gamma^d_{bc}\right) \;+\; g_{bd}\left(\gamma^d_{ac} \;+\; \gamma^d_{ca}\right) \\ =\; & g_{dc}T^d{}_{ab} \;+\; g_{ad}T^d{}_{cb} \;+\; g_{bd}T^d{}_{ca} \;+\; 2g_{bd}\gamma^d_{ac} \end{aligned}$$

Hence, premultiplying by $g^{be}$, we find that the left hand side becomes the Christoffel symbols, and the right hand side a combination of $\gamma^a_{bc}$ and $T^a{}_{bc}$,

$$\Gamma^e_{ac} \;=\; \tfrac{1}{2}\left(T_{ca}{}^e + T_{ac}{}^e + T^e{}_{ca}\right) + \gamma^e_{ac} \tag{170}$$

The upshot is that requiring that the connection be torsion-free (symmetric, i.e. one for which $T^a{}_{bc} = 0$) and that the metric be covariantly constant results in a unique answer: the Levi-Civita, or metric, connection, for which $\Gamma^a_{bc}$ is given by the Christoffel symbols.

### 3.3 The Riemann tensor

There is a property of second derivatives of vectors which we haven't looked at yet, and that is that the actions of two covariant derivatives on vector fields do not commute, i.e. in general

$$\nabla_a \nabla_b U^c \neq \nabla_b \nabla_a U^c .$$

To show this, it's easiest go in several stages, so we shall first work out $\nabla_b V^c$ and then $\nabla_a \nabla_b V^c$, and then take the difference $\nabla_a \nabla_b V^c - \nabla_b \nabla_a V^c$. Firstly,

$$\nabla_b V^c = \partial_b V^c + \Gamma^c_{bd} V^d .$$

Next,

$$\begin{aligned}
&\nabla_a \left(\nabla_b V^c\right) \\
=\ & \partial_a \left(\nabla_b V^c\right) - \Gamma^d_{ab} \nabla_d V^c + \Gamma^c_{ad} \nabla_b V^d \\
=\ & \partial_a \partial_b V^c + \partial_a \left(\Gamma^c_{bd} V^d\right) - \Gamma^d_{ab} \nabla_d V^c + \Gamma^c_{ad} \nabla_b V^d \\
=\ & \partial_a \partial_b V^c + \left(\partial_a \Gamma^c_{bd}\right) V^d + \Gamma^c_{bd}\, \partial_a V^d - \Gamma^d_{ab} \partial_d V^c - \Gamma^d_{ab} \Gamma^c_{de} V^e + \Gamma^c_{ad}\, \partial_b V^d + \Gamma^c_{ad} \Gamma^d_{be} V^e
\end{aligned}$$

We can now subtract $\nabla_b \nabla_a V^c$ from $\nabla_a \nabla_b V^c$:

$$\begin{aligned}
&\nabla_a \left(\nabla_b V^c\right) - \nabla_b \left(\nabla_a V^c\right) \\
=\ & \partial_a \partial_b V^c + \left(\partial_a \Gamma^c_{bd}\right) V^d + \Gamma^c_{bd}\, \partial_a V^d - \Gamma^d_{ab}\, \partial_d V^c - \Gamma^d_{ab} \Gamma^c_{de} V^e + \Gamma^c_{ad}\, \partial_b V^d + \Gamma^c_{ad} \Gamma^d_{be} V^e \\
-\ & \partial_b \partial_a V^c - \left(\partial_b \Gamma^c_{ad}\right) V^d - \Gamma^c_{ad}\, \partial_b V^d + \Gamma^d_{ba}\, \partial_d V^c + \Gamma^d_{ba} \Gamma^c_{de} V^e - \Gamma^c_{bd}\, \partial_a V^d - \Gamma^c_{bd} \Gamma^d_{ae} V^e
\end{aligned} \tag{171}$$

Now, the terms in (171) in two partial derivatives of $V^c$

$$\partial_a \partial_b V^c - \partial_b \partial_a V^c ,$$

cancel, as do the terms in one partial derivative,

$$\begin{aligned}
\Gamma^c_{bd}\, \partial_a V^d &- \Gamma^c_{bd}\, \partial_a V^d , \\
- \Gamma^d_{ab}\, \partial_d V^c &+ \Gamma^d_{ba}\, \partial_d V^c ,
\end{aligned}$$

and

$$\Gamma^c_{ad}\, \partial_b V^d - \Gamma^c_{ad}\, \partial_b V^d .$$

This leaves only the terms linear in $V^c$, which, after some rearrangement of the indices gives the final answer

$$\nabla_a \nabla_b V^c - \nabla_b \nabla_a V^c = \left( \partial_a \Gamma^c_{be} - \partial_b \Gamma^c_{ae} + \Gamma^c_{ad} \Gamma^d_{be} - \Gamma^c_{bd} \Gamma^d_{ae} \right) V^e \tag{172}$$

However, since the only terms left are those linear in $V^e$, we can apply the quotient theorem: the left-hand side of (172) is clearly a tensor for any $V^c$, so the right-hand side must be a tensor for any $V^e$, and hence the terms in brackets on the right-hand side must also be a tensor – and from counting the indices, it is a type $\left[{1 \atop 3}\right]$ tensor. This tensor is so important that is has a special name, it is the *Riemann Tensor*, and is denoted ${R^c}_{eab}$:

$$\boxed{{R^c}_{eab} = \partial_a \Gamma^c_{be} - \partial_b \Gamma^c_{ae} + \Gamma^c_{ad} \Gamma^d_{be} - \Gamma^c_{bd} \Gamma^d_{ae} .} \tag{173}$$

(n.b. this formula will always be given to you in exams!)

• It turns out that the Riemann tensor is the only true measure of curvature – a manifold is flat if and only if all the components $R^a{}_{bcd}$ are zero, but it is beyond the scope of the course to prove this.

It is often more useful to study the Riemann tensor with all its indices 'down', which is given (as one might expect) by

$$R_{abcd} = g_{ae}\, R^e{}_{bcd}\ . \tag{174}$$

There are two further tensors which derive from the Riemann tensor, and these are the *Ricci Tensor*, a type $\left[{0 \atop 2}\right]$ tensor denoted by $R_{ab}$, and the *Ricci Scalar*, denoted by $R$, and which are defined by

$$\boxed{\begin{aligned} R_{bd} &= R^c{}_{bcd} = g^{ac}\, R_{abcd}\ , \\ R &= R^a{}_a = g^{ab}\, R_{ab}\ . \end{aligned}} \tag{175}$$

Also, just as one can give the action of $\nabla_a$ on any type of tensor in terms of $\partial_a$ and $\Gamma^a_{bc}$, so one can give the commutator $(\nabla_a\nabla_b - \nabla_b\nabla_a)X$ on any tensor just in terms of $R^a{}_{bcd}$. If $A^{c_1\,c_2\ldots c_m}_{d_1\,d_2\ldots d_n}$ is type $\left[{m \atop n}\right]$ tensor, then (you don't need to learn this!)

$$\begin{aligned} \nabla_a\nabla_b\, A^{c_1\,c_2\ldots c_m}_{d_1\,d_2\ldots d_n} - \nabla_b\nabla_a\, A^{c_1\,c_2\ldots c_m}_{d_1\,d_2\ldots d_n} &= \underbrace{R^{c_1}{}_{c'ab}\, A^{c'\,c_2\ldots c_m}_{d_1\,d_2\ldots d_n} + \cdots + R^{c_m}{}_{c'ab}\, A^{c_1\,c_2\ldots c'}_{d_1\,d_2\ldots d_n}}_{m \text{ terms}} \\ &\quad - \underbrace{R^{d'}{}_{d_1ab}\, A^{c_1\,c_2\ldots c_m}_{d'\,d_2\ldots d_n} - \cdots - R^{d'}{}_{d_nab}\, A^{c_1\,c_2\ldots c_m}_{d_1\,d_2\ldots d'}}_{n \text{ terms}}\ . \end{aligned}$$

The Riemann tensor looks especially nice in normal coordinates since $\Gamma^a_{bc}$ vanishes in these, so we find in normal coordinates at a point $\mathcal{P}$,

$$R^a{}_{bcd}(\mathcal{P}) = \partial_c\Gamma^a_{db}(\mathcal{P}) - \partial_d\Gamma^a_{cb}(\mathcal{P})\ , \tag{176}$$

$$R_{abcd}(\mathcal{P}) = \frac{1}{2}\left(\partial_c\partial_b g_{ad} - \partial_c\partial_a g_{bd} - \partial_d\partial_b g_{ac} + \partial_d\partial_a g_{bc}\right)\Big|_{\mathcal{P}}\ . \tag{177}$$

Before we return to the geodesic deviation equation, we shall list some properties of the Riemann tensor. These are fairly easy to prove using the expression for the Riemann tensor in normal coordinates, but there is also a nicer and more elegant way which is treated in section

- Antisymmetry in the last two indices; this is immediate from the definition.

$$R^a{}_{bcd} = -R^a{}_{bdc}\ . \tag{178}$$

- Invariant when the first pair and the second pair of indices are swapped; immediate from the expression (177) for $R_{abcd}$ in normal coordinates.

$$R_{abcd} = R_{cdab}\ . \tag{179}$$

- The first Bianchi Identity; again this can be shown easily using normal coordinates:

$$R^a{}_{bcd} + R^a{}_{dbc} + R^a{}_{cdb} = 0\ . \tag{180}$$

- The second Bianchi identity – this is a differential equation satisfied identically by the Riemann tensor. This is rather more tedious to show using the definition of the Riemann tensor, even using normal coordinates. There is another method which provides a little more insight and is more "geometric", which we leave to the next section [Note: this was not part of the level 6 course in 2012-13 but is in 2014-15]

$$\nabla_e R^a{}_{bcd} + \nabla_c R^a{}_{bde} + \nabla_d R^a{}_{bec} = 0 \,. \tag{181}$$

### 3.3.1 Proofs of the Bianchi Identities using normal coordinates

**Throughout this section we shall assume that we are using the Levi-Civita connection**

The proof of the first Bianchi identity is immediate from the expression for $R^a{}_{bcd}$ in terms of the connection. It is easier in normal coordinates since the full expression,

$$R^a{}_{bcd} = \partial_c \Gamma^a_{db} \,-\, \partial_d \Gamma^a_{cb} \,+\, \Gamma^a_{cf}\Gamma^f_{db} \,-\, \Gamma^a_{df}\Gamma^f_{cb} \,,$$

becomes, in normal coordinates at the point $\mathcal{P}$:

$$R^a{}_{bcd}(\,\mathcal{P}\,) = \partial_c \Gamma^a_{db}(\,\mathcal{P}\,) \,-\, \partial_d \Gamma^a_{cb}(\,\mathcal{P}\,) \,.$$

We then make the combination (using normal coordinates at $\mathcal{P}$)

$$\begin{array}{rcl} R^a{}_{bcd}(\mathcal{P}) & = & \partial_c\Gamma^a_{db}(\mathcal{P}) \,-\, \partial_d\Gamma^a_{cb}(\mathcal{P}) \\ +\, R^a{}_{cdb}(\mathcal{P}) & & \partial_d\Gamma^a_{bc}(\mathcal{P}) \,-\, \partial_b\Gamma^a_{dc}(\mathcal{P}) \\ +\, R^a{}_{dbc}(\mathcal{P}) & & \partial_b\Gamma^a_{cd}(\mathcal{P}) \,-\, \partial_c\Gamma^a_{bd}(\mathcal{P}) \end{array}$$

The terms on the left hand side than cancel in pairs, remembering that the Christoffel symbols are symmetric in the lower two indices.

Having shown that $R^a{}_{bcd}(\mathcal{P}) + R^a{}_{cdb}(\mathcal{P}) + R^a{}_{dbc}(\mathcal{P}) = 0$ in normal coordinates at $\mathcal{P}$ we can now use the fact that $R^a{}_{bcd}$ is a tensor to say that this must be true in all coordinate systems at that point, and hence is true in general.

The proof of the second Bianchi identity is just as simple. We have

$$\nabla_a R^b{}_{cde} = \partial_a R^b{}_{cde} + \Gamma^b_{af} R^f{}_{cde} - \Gamma^f_{ac} R^b{}_{fde} - \Gamma^f_{ad} R^b{}_{cfe} - \Gamma^f_{ae} R^b{}_{cdf} \,.$$

and so in normal coordinates at $\mathcal{P}$ again there are only two remaining terms:

$$\begin{array}{rcl} \nabla_a R^b{}_{cde}(\mathcal{P}) & = & \partial_a R^b{}_{cde}|_{\mathcal{P}} = \partial_a(\partial_d\Gamma^b_{ec} \,-\, \partial_e\Gamma^b_{dc} \,+\, \Gamma^b_{df}\Gamma^f_{ec} \,-\, \Gamma^b_{ef}\Gamma^f_{dc})|_{\mathcal{P}} \\ & = & \partial_a\partial_d\Gamma^b_{ec}(\mathcal{P}) \,-\, \partial_a\partial_e\Gamma^b_{dc}(\mathcal{P}) \,+\, \partial_a\Gamma^b_{df}(\mathcal{P})\Gamma^f_{ec}(\mathcal{P}) \,+\, \Gamma^b_{df}(\mathcal{P})\partial_a\Gamma^f_{ec}(\mathcal{P}) \\ & & -\,\partial_a\Gamma^b_{ef}(\mathcal{P})\Gamma^f_{dc}(\mathcal{P}) \,-\, \Gamma^b_{ef}(\mathcal{P})\partial_a\Gamma^f_{dc}(\mathcal{P}) \\ & = & \partial_a\partial_d\Gamma^b_{ec}(\mathcal{P}) \,-\, \partial_a\partial_e\Gamma^b_{dc}(\mathcal{P}) \end{array} \tag{182}$$

Hence, putting these together, the six terms again cancel in pairs:

$$\begin{array}{rcl} \nabla_a R^b{}_{cde}(\mathcal{P}) & = & \partial_a\partial_d\Gamma^b_{ec}(\mathcal{P}) \,-\, \partial_a\partial_e\Gamma^b_{dc}(\mathcal{P}) \\ +\,\nabla_d R^b{}_{cea}(\mathcal{P}) & & \partial_d\partial_e\Gamma^b_{ac}(\mathcal{P}) \,-\, \partial_d\partial_a\Gamma^b_{ec}(\mathcal{P}) \\ +\,\nabla_e R^b{}_{cad}(\mathcal{P}) & & \partial_e\partial_a\Gamma^b_{dc}(\mathcal{P}) \,-\, \partial_e\partial_d\Gamma^b_{ac}(\mathcal{P}) \end{array}$$

Since the expression vanishes in normal coordinates, it must vanish in all and the proof is complete.

### 3.3.2 Proofs of the Bianchi identities using the Jacobi identity

The proofs of the Bianchi identities using normal coordinates, while being easy, do not really give us any insight into 'why' this equation must be true, or where it comes from, or why we then might expect the 2nd Bianchi identity to hold. There is another proof based on taking derivatives of an arbitrary scalar field which does lend us more insight of this sort.

Let us consider the equation which defined ${R^a}_{bcd}$ from taking the commutator of the covariant derivative on a 1–form:

$$[\,\nabla_a\,,\,\nabla_b\,]\,U_c = \nabla_a\nabla_b U_c - \nabla_b\nabla_a U_c = -{R^d}_{cab}\,U_d\;.$$

Then we find that for all 1–forms $U_a$,

$$-\,({R^a}_{bcd} + {R^a}_{cdb} + {R^a}_{dbc})\,U_a = (\nabla_c\nabla_d - \nabla_d\nabla_c)U_b + (\nabla_d\nabla_b - \nabla_b\nabla_d)U_c + (\nabla_b\nabla_c - \nabla_c\nabla_b)U_d\;.$$

This is not very useful as it currently stands since the covariant derivatives are each acting on a 1–form with a different index. However, if we take the 1–form $U_a$ to be given by the covariant derivative of some scalar field $\phi$, we then get that

$$\begin{aligned}
&-\,({R^a}_{bcd} + {R^a}_{cdb} + {R^a}_{dbc})\,\nabla_a\phi \\
=\;& (\nabla_c\nabla_d - \nabla_d\nabla_c)\nabla_b\,\phi + (\nabla_d\nabla_b - \nabla_b\nabla_d)\nabla_c\,\phi + (\nabla_b\nabla_c - \nabla_c\nabla_b)\nabla_d\,\phi \qquad (183)\\
=\;& (\nabla_c\nabla_d\nabla_b - \nabla_d\nabla_c\nabla_b + \nabla_d\nabla_b\nabla_c - \nabla_b\nabla_d\nabla_c + \nabla_b\nabla_c\nabla_d - \nabla_c\nabla_b\nabla_d)\,\phi \\
=\;& \nabla_c(\nabla_d\nabla_b - \nabla_b\nabla_d)\,\phi + \nabla_d(\nabla_b\nabla_c - \nabla_b\nabla_c)\,\phi + \nabla_b(\nabla_c\nabla_d - \nabla_c\nabla_d)\,\phi \qquad (184)\\
=\;& 0\,,
\end{aligned}$$

since we have already shown that for the Levi-Civita connection given by the Christoffel symbols the torsion vanishes, and so the covariant derivative commutes on scalar fields, i.e. that $[\nabla_a, \nabla_b]\phi = 0$ for all $\phi$. So, for all $\phi$, we have

$$({R^a}_{bcd} + {R^a}_{cdb} + {R^a}_{dbc})\,\nabla_a\phi = 0\;. \qquad (185)$$

However, the scalar field is arbitrary, so we can take it to be the coordinate $x^e$, in which case

$$\nabla_a x^e = \partial_a x^e = \delta^e_a\;,$$

and substituting this into eqn. (185) we recover the Bianchi identity.

At the moment this gives no further insight into the reason for the existence of the Bianchi identity, but let us consider the reason this argument worked, which is that the following two expressions coming from lines (183) and (184) are actually identical,

$$[\nabla_c, \nabla_d]\nabla_b + [\nabla_d, \nabla_b]\nabla_c + [\nabla_b, \nabla_c]\nabla_d \equiv \nabla_c[\nabla_d, \nabla_b] + \nabla_d[\nabla_b, \nabla_c] + \nabla_b[\nabla_c, \nabla_d]\;. \qquad (186)$$

We can further compress this by noting that if we bring the right hand side over to the left, the whole expression can be written as a sum of commutators of commutators:

$$\boxed{[[\nabla_c, \nabla_d], \nabla_b] + [[\nabla_d, \nabla_b], \nabla_c] + [[\nabla_b, \nabla_c], \nabla_d] \equiv 0\;.} \qquad (187)$$

This equation has a special name, the Jacobi Identity, and it holds for all operators $\nabla_a$ whose product is associative, which is certainly the case for us.

So, the first Bianchi identity arises simply as the result of applying the Jacobi identity to a scalar field.

In exactly the same way, the second Bianchi identity arises as the result of applying the Jacobi identity to a vector field, which we do in the next section.

—**The second Bianchi identity**

We shall use the ideas of the previous section by saying that the result of applying both sides of (186) to a vector field give the same result, i.e.

$$([\nabla_a,\nabla_b]\nabla_c + [\nabla_b,\nabla_c]\nabla_a + [\nabla_c,\nabla_a]\nabla_b)\, U^d = (\nabla_a[\nabla_b,\nabla_c] + \nabla_b[\nabla_c,\nabla_a] + \nabla_c[\nabla_a,\nabla_b])\, U^d\ . \tag{188}$$

On the right hand side we have

$$\nabla_a[\nabla_b,\nabla_c]U^d + \nabla_b[\nabla_c,\nabla_a]U^d + \nabla_c[\nabla_a,\nabla_b]U^d\ .$$

When we applied the same differential operator to a scalar field we found this vanished identically, but since

$$\nabla_b\nabla_c U^d - \nabla_c\nabla_b U^d = R^d{}_{ebc}U^e\ ,$$

we find instead that

$$\nabla_a(\nabla_b\nabla_c U^d - \nabla_c\nabla_b U^d) = (\nabla_a R^d{}_{ebc})U^e + R^d{}_{ebc}(\nabla_a U^e)\ .$$

So, taking the three pairings together we get in total

$$(\nabla_a R^d{}_{ebc} + \nabla_b R^d{}_{eca} + \nabla_c R^d{}_{eab})U^e + (R^d{}_{ebc}(\nabla_a U^e) + R^d{}_{eca}(\nabla_b U^e) + R^d{}_{eab}(\nabla_c U^e))\ . \tag{189}$$

So, while this does not itself vanish, it does at least contain the left-hand side of the second Bianchi identity in the first term.

Looking at the left hand side of (188), the first term is

$$[\nabla_a,\nabla_b]\nabla_c U^d = -R^e{}_{cab}\nabla_e U^d + R^d{}_{eab}\nabla_c U^e\ ,$$

using the standard rules for the commutator of covariant derivatives on a type $\left[{1 \atop 1}\right]$ tensor, and the total of the three pairs is

$$-(R^e{}_{cab} + R^e{}_{abc} + R^e{}_{bca})\,\nabla_e U^d + (R^d{}_{ebc}(\nabla_a U^e) + R^d{}_{eca}(\nabla_b U^e) + R^d{}_{eab}(\nabla_c U^e))\ . \tag{190}$$

The first term in brackets vanishes identically using the first Bianchi identity, and the second term is exactly the second term in (189), so that when we equate (189) and (190), we end up with

$$(\nabla_a R^d{}_{ebc} + \nabla_b R^d{}_{eca} + \nabla_c R^d{}_{eab})U^e = 0\ , \tag{191}$$

for an arbitrary vector field $U^e$, and hence, since we can choose $U^e$ to be any vector we like, this means that the term in brackets must be zero,

$$\nabla_a R^d{}_{ebc} + \nabla_b R^d{}_{eca} + \nabla_c R^d{}_{eab} = 0\ , \tag{192}$$

i.e. the second Bianchi identity.

### 3.3.3 The contracted Bianchi identity

We shall find that we need an expression for $\nabla^a R_{ab}$. We can find this by contracting over two pairs of free indices in the 2nd Bianchi identity.

Depending which indices one contracts over, one will either get the required identity or $0 = 0$. We shall follow a correct path.

Multiplying eqn. (192) by $\delta^b_d\, g^{ae}$, we have

$$\begin{aligned} & \delta^b_d\, g^{ae}(\nabla_a R^d{}_{ebc} + \nabla_b R^d{}_{eca} + \nabla_c R^d{}_{eab}) \\ = \; & \nabla^e R_{ec} + \nabla_b R^b_c + \nabla_c R^d{}_{ead} g^{ea} \\ = \; & 2\nabla^e R_{ec} - \nabla_c R^d_{eda} g^{ea} \\ = \; & 2\nabla^e R_{ec} - \nabla_c R_{ea} g^{ea} \\ = \; & 2\nabla^e R_{ec} - \nabla_c R \\ = \; & 0 \end{aligned} \tag{193}$$

In other words, we have found that

$$\nabla^a R_{ab} = \frac{1}{2}\nabla_b(R) \; .$$

Rewriting this as

$$\nabla^a(R_{ab} - \frac{R}{2} g_{ab}) = 0 \; ,$$

we have shown that the *Einstein tensor* $G_{ab}$ defined as

$$G_{ab} = R_{ab} - \frac{R}{2} g_{ab} \; ,$$

satisfies

$$\nabla^a G_{ab} = 0 \; .$$

## 3.4 Geodesic deviation

In exercise 1.1 we considered two particle paths $\boldsymbol{x}(t)$ and $\boldsymbol{y}(t)$ and the difference of their coordinates $\boldsymbol{x}(t) - \boldsymbol{y}(t)$. This made sense in the context of Newtonian gravity in flat Euclidean space, but if we try to think about two nearby geodesics $x^a(\lambda)$ and $y^a(\lambda)$ in a curved space then the coordinate difference of two nearby geodesics is not a vector, indeed it does not have any nice transformation properties at all. However, there is a genuine vector which does measure the displacement of two nearby geodesics, but to define this we have to consider a whole family of neighbouring geodesics.

Let $x^a(\lambda, \mu)$ be a family of affinely parametrised geodesics, that is, for each value $\mu_0$ say of $\mu$, the curve $x^a(\lambda, \mu_0)$ is an affinely parametrised geodesic, i.e.

$$\frac{\mathrm{D}}{\mathrm{d}\lambda}\frac{\mathrm{d}x^a(\lambda, \mu_0)}{\mathrm{d}\lambda} = 0 \; .$$

Fixing $\mu$ to be $\mu_0$, we can define the tangent vector $U^a$ to be the curve of constant $\mu$,

$$U^a(\lambda, \mu_0) = \frac{\partial x^a}{\partial \lambda} \; .$$

In terms of this vector, the fact that each curve of fixed $\mu$ is an affinely parametrised geodesic can be written

$$\frac{\mathrm{D}}{\mathrm{d}\lambda}U^a = U^b \nabla_b U^a = 0 \; . \tag{194}$$

We are interested in the separation at parameter value $\lambda_0$ between two neighbouring geodesics, say those with parameter values $\mu$ and $\mu + \delta\mu$. The coordinate difference is given by

$$x^a(\lambda_0, \mu + \delta\mu) - x^a(\lambda_0, \mu) = \delta\mu \frac{\partial x^a(\lambda_0, \mu)}{\partial \mu} \; + \; O(\delta\mu^2) \; . \tag{195}$$

If we think of a curve of fixed $\lambda_0$ in its own right, then

$$V^a(\lambda_0, \mu) = \frac{\partial x^a(\lambda_0, \mu)}{\partial \mu} \; , \tag{196}$$

is the tangent vector to this curve, and hence is also a genuine vector. Mathematically $V^a$ is a much better measure of the separation of the two geodesics than the coordinate difference is – for a start it **is** a vector, i.e. a $\left[{1 \atop 0}\right]$ tensor, whereas the coordinate difference has no special properties at all under coordinate changes.

Finally, in figure 28 we show the general picture of the two parameter family of curves, and the vectors $U^a$ and $V^a$.

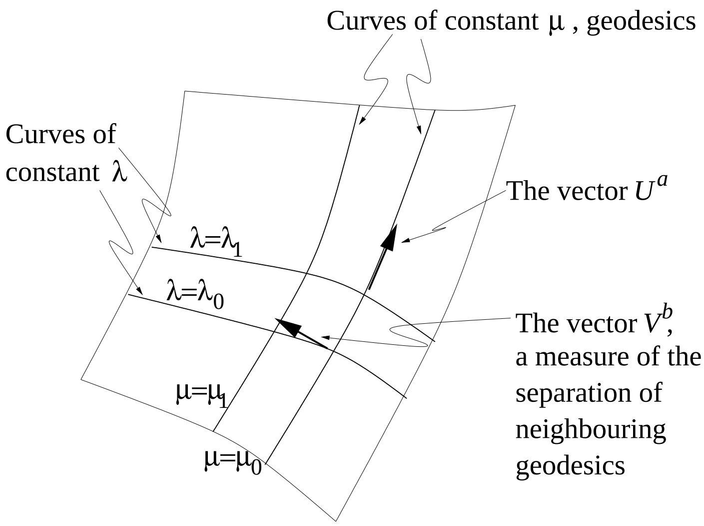



Figure 28: set up for calculation of geodesic deviation

Since $V^a$ is a vector, we should like the geodesic deviation equation to be an equation for a vector, and so it is the second absolute derivative of $V^a$ with respect to $\lambda$ that we shall evaluate, i.e.

$$\frac{\mathrm{D}}{\mathrm{d}\lambda}\frac{\mathrm{D}}{\mathrm{d}\lambda}V^c = U^a \nabla_a \left(U^b \nabla_b \, V^c\right) \; ,$$

remembering that $\mathrm{D}/\mathrm{d}\lambda \equiv U^a \nabla_a$.

As a last preparation we need to prove the following lemma:

**Lemma:** If $x^a(\lambda, \mu)$ are a two-parameter family set of points, with the vectors $U^a$ and $V^a$ defined to be the tangents to curves of fixed $\mu$ and fixed $\lambda$ respectively, i.e.

$$U^a = \frac{\partial x^a}{\partial \lambda} \; , \qquad V^a = \frac{\partial x^a}{\partial \mu} \; ,$$

then

$$\boxed{\frac{\mathrm{D}U^a}{\mathrm{d}\mu} = \frac{\mathrm{D}V^a}{\mathrm{d}\lambda} \; ,} \qquad \text{or, equivalently} \qquad \boxed{V^b \nabla_b \, U^a = U^b \nabla_b \, V^a \; .} \tag{197}$$

**Proof:** By direct calculation,

$$\begin{aligned} \frac{\mathrm{D}U^a}{\mathrm{d}\mu} - \frac{\mathrm{D}V^a}{\mathrm{d}\lambda} &= \left( \frac{\partial U^a}{\partial \mu} + \Gamma^a_{bc} U^b V^c \right) - \left( \frac{\partial V^a}{\partial \lambda} + \Gamma^a_{bc} V^b U^c \right) \\ &= \left( \frac{\partial^2 x^a}{\partial \mu \partial \lambda} - \frac{\partial^2 x^a}{\partial \lambda \partial \mu} \right) + U^b V^c \left( \Gamma^a_{bc} - \Gamma^a_{cb} \right) \\ &= 0 \; , \end{aligned} \tag{198}$$

where we have used the fact that partial derivatives commute, did some small relabelling $b \leftrightarrow c$ in the very last expression and finally used the fact that $\Gamma^a_{bc} = \Gamma^a_{cb}$.

Now we can derive an expression for the geodesic deviation as follows:

$$\begin{aligned} \frac{\mathrm{D}}{\mathrm{d}\lambda} \frac{\mathrm{D}}{\mathrm{d}\lambda} V^a &= \frac{\mathrm{D}}{\mathrm{d}\lambda} \frac{\mathrm{D}}{\mathrm{d}\mu} U^a && \text{using eqn. (197)} \\ &= \frac{\mathrm{D}}{\mathrm{d}\lambda} \frac{\mathrm{D}}{\mathrm{d}\mu} U^a - \frac{\mathrm{D}}{\mathrm{d}\mu} \frac{\mathrm{D}}{\mathrm{d}\lambda} U^a && \text{using eqn. (194)} \\ &= U^b \nabla_b (V^c \nabla_c U^a) - V^b \nabla_b (U^c \nabla_c U^a) && \text{definition} \\ &= (U^b \nabla_b V^c) \nabla_c U^a + U^b V^c (\nabla_b \nabla_c U^a) \\ &\quad - (V^b \nabla_b U^c) \nabla_c U^a - V^b U^c (\nabla_b \nabla_c U^a) && \text{using Leibniz rule} \\ &= \left( U^b \nabla_b V^c - V^b \nabla_b U^c \right) \nabla_c U^a \\ &\quad + (U^b V^c - U^c V^b) \nabla_b \nabla_c U^a && \text{relabelling} \\ &= 0 + U^b V^c (\nabla_b \nabla_c U^a - \nabla_c \nabla_b U^a) && b \leftrightarrow c \text{ in last term} \\ &= U^b V^c (\nabla_b \nabla_c U^a - \nabla_c \nabla_b U^a) \end{aligned}$$

The term in brackets is simply the commutator of two covariant derivatives on a vector, and so is give in terms of the Riemann tensor, and so the geodesic deviation equation is

$$\boxed{\frac{\mathrm{D}}{\mathrm{d}\lambda} \frac{\mathrm{D}}{\mathrm{d}\lambda} V^a = R^a{}_{bcd} \, U^b \, U^c \, V^d \; .} \tag{199}$$

We will compare this with the equation we found for Newtonian gravity in the next section.

## 3.5 Exercises

### Shorter exercises

**3.1**
Consider the Euclidean plane in Cartesian and polar coordinates, as in exercise (2.3).

(a) In Cartesian coordinates, the tensor $F$ has components

$$F_{\alpha\beta} = \begin{pmatrix} 0 & a \\ -a & 0 \end{pmatrix} . \tag{200}$$

Find the components $F'_{\alpha\beta}$ in polar coordinates and show that $F'_{\alpha\beta} = -F'_{\beta\alpha}$.

(b) In Cartesian coordinates, the tensor $C$ has components

$$C_{\alpha\beta} = \begin{pmatrix} a & b \\ b & a \end{pmatrix} . \tag{201}$$

Find the components $C'_{\alpha\beta}$ in polar coordinates and show that $C'_{\alpha\beta} = C'_{\beta\alpha}$.

Show that $C'_{rr} \neq C'_{\theta\theta}$ but $g^{\alpha\beta}C_{\alpha\beta} = g'^{\alpha\beta}C'_{\alpha\beta}$

**3.2**
Suppose that for any type $\left[{0 \atop 1}\right]$ tensor $V_a$ the quantity $W = V_a U^a$ is a scalar. Show that this implies that $U^a$ is a vector.

**3.3**
(a) Consider the sphere with coordinates $\{\theta, \phi\}$ and metric $\mathrm{d}s^2 = \mathrm{d}\theta^2 + \sin(\theta)^2 \mathrm{d}\phi^2$. Consider the vector $U^\alpha = (1, 0)$. Find the components of $\nabla_\alpha U^\beta$.

The Christoffel symbols for this metric are

$$\begin{aligned} \Gamma^\theta_{\theta\theta} &= 0 , \quad \Gamma^\theta_{\theta\phi} = \Gamma^\theta_{\phi\theta} = 0 , \quad \Gamma^\theta_{\phi\phi} = -\sin\theta \cos\theta , \\ \Gamma^\phi_{\phi\phi} &= 0 , \quad \Gamma^\phi_{\theta\phi} = \Gamma^\phi_{\phi\theta} = \cot\theta , \quad \Gamma^\phi_{\theta\theta} = 0 . \end{aligned}$$

(b) Consider the plane with polar coordinates $\{r, \theta\}$ and metric $\mathrm{d}s^2 = \mathrm{d}r^2 + r^2 \mathrm{d}\theta^2$. Consider the vector $V^\alpha = (\cos(\theta), -\sin(\theta)/r)$. Find the components of $\nabla_\alpha V^\beta$.

The Christoffel symbols for this metric are

$$\begin{aligned} \Gamma^r_{rr} &= 0 , \quad \Gamma^r_{r\theta} = \Gamma^r_{\theta r} = 0 , \quad \Gamma^r_{\theta\theta} = -r , \\ \Gamma^\theta_{rr} &= 0 , \quad \Gamma^\theta_{r\theta} = \Gamma^\theta_{\theta r} = \tfrac{1}{r} , \quad \Gamma^\theta_{\theta\theta} = 0 . \end{aligned}$$

**3.4**
Consider the tensor $U_{\mu\nu}$ defined by

$$U_{\mu\nu} = \nabla_\mu \nabla_\nu f - g_{\mu\nu} \nabla^\sigma \nabla_\sigma f ,$$

where $f$ is a scalar function on spacetime. Show that

$$(a) \qquad \nabla^\mu U_{\mu\nu} = \nabla^\mu [\nabla_\mu, \nabla_\nu] f + [\nabla_\mu, \nabla_\nu] \nabla^\mu f .$$

$$(b) \qquad \nabla^\mu U_{\mu\nu} = R_\nu{}^\sigma \nabla_\sigma f .$$

**Longer exercises**

### 3.5 Normal coordinates

The coordinates $\tilde{x}$ are *Normal coordinates at the point* $\tilde{x} = 0$ if either of the two equivalent conditions is true:

$$\tilde{\Gamma}^a_{bc}(0) = 0 \quad \Leftrightarrow \quad \tilde{\partial}_a \, \tilde{g}_{bc}(0) = 0 \; .$$

Given an arbitrary set of coordinates $x^a$, it is possible to find coordinates $\tilde{x}^a$ which are normal coordinates at the point $x^a = 0$ using these two conditions in a rather simple way

1. Let $\xi^a(\lambda)$ be a geodesic and $\lambda$ an affine parameter along the geodesic, and let $\xi^a(0) = 0$ so that the geodesic passes through the point $x^a = 0$.

   We define (in a region around $x^a = 0$ new coordinates $\tilde{x}^a$ as

   $$\tilde{x}^a = x^a + \frac{1}{2} C^a_{bc} x^b x^c \; ,$$

   where $C^a_{bc}$ are constants satisfying $C^a_{bc} = C^a_{cb}$.

   (a) Write down the equation satisfied by $\xi^a(\lambda)$

   (b) Write down the coordinates $\tilde{\xi}^a(\lambda)$ of the geodesic in the coordinates $\tilde{x}^a$.

   (c) Write down the equation satisfied by $\tilde{\xi}^a(\lambda)$

   (d) By considering the equation satisfied by $\tilde{\xi}^a(\lambda)$ at $\lambda = 0$, find $C^a_{bc}$ under the assumption that $\tilde{\Gamma}^a_{bc}(0) = 0$.

2. Let $\tilde{x}^a$ be coordinates satisfying (in some small region about $x^a = 0$)

   $$x^a = \tilde{x}^a + \frac{1}{2} D^a_{bc} \tilde{x}^b \tilde{x}^c \; ,$$

   where $D^a_{bc}$ are constants satisfying $D^a_{bc} = D^a_{cb}$.

   (a) Find the matrix $\frac{\partial x^c}{\partial \tilde{x}^d}$, and its value at $x^a = 0$.

   (b) Show that

   $$\tilde{\partial}_c \tilde{g}_{ab}(0) \; = \; g_{bd}(0) D^d_{ac} \; + \; g_{ad}(0) D^d_{bc} \; + \; \partial_c g_{ab}(0) \; .$$

   (c) Show that $\tilde{\partial}_c \tilde{g}_{ab}(0) = 0$ implies that $D^a_{bc} = -\Gamma^a_{bc}(0)$.

If we assume that we have a more general relationship between the coordinates $x^a$ and $\tilde{x}^a$ than those in 1. and 2. above, then these arguments instead fix the next-to-leading terms in the Taylor expansions $x^a = \tilde{x}^a + \ldots$ and $\tilde{x}^a = x^a + \ldots$.

### 3.6. The Lie derivative

Given a vector $k^a$, there is an operation which maps $\left[{m \atop n}\right]$ tensors to $\left[{m \atop n}\right]$ tensors, known as the Lie derivative, and written $\mathcal{L}_k$. It obeys the Leibniz or product rule.

On a vector $V^b$ this takes the form

$$\mathcal{L}_k V^a = k^b \nabla_b V^a \ - \ V^b \nabla_b k^a \ .$$

On a type $\left[{0 \atop 2}\right]$ tensor $\omega_{bc}$ it takes the form

$$\mathcal{L}_k \omega_{ab} = k^c \nabla_c \omega_{ab} \ + \ \omega_{cb} \nabla_a k^c \ + \ \omega_{ac} \nabla_b k^c \ .$$

Show that

$$\begin{array}{lrcl} \text{(i)} & \mathcal{L}_k V^a & = & k^b \partial_b V^a \ - \ V^b \partial_b k^a \ , \\ \text{(ii)} & \mathcal{L}_k \omega_{ab} & = & k^c \partial_c \omega_{ab} \ + \ \omega_{cb} \partial_a k^c \ + \ \omega_{ac} \partial_b k^c \ , \\ \text{(iii)} & \mathcal{L}_k (\omega_{ab} V^a V^b) & = & k^c \partial_c (\omega_{ab} V^a V^b) \ . \end{array}$$

(It is a general result that the Lie derivative is independent of the metric tensor)

**3.7** A ‘Killing vector’ is a vector $V^\mu$ which satisfies Killing’s equation:

$$\nabla_\mu V_\nu + \nabla_\nu V_\mu = 0 \ .$$

Given two vectors $U^\mu$ and $V^\mu$, their **commutator** is defined as $W^\mu \equiv [U, V]^\mu$ where

$$[U, V]^\mu = U^\nu \nabla_\nu V^\mu - V^\nu \nabla_\nu U^\mu \ .$$

Show that the commutator of two Killing vectors is also a Killing vector

**3.8** Let $T_{\mu\nu}$ be a symmetric tensor satisfying

$$\nabla^\mu T_{\mu\nu} = 0 \ ,$$

and let $k^\mu$ satisfy Killing’s equation,

$$\nabla_\mu k_\nu \ + \ \nabla_\nu k_\mu \ = \ 0 \ .$$

Show that

$$\nabla^\mu \ ( T_{\mu\nu} k^\nu ) = 0$$

### 3.9 The equations of a relativistic fluid.

The motion of a relativistic fluid is characterised by its velocity field $U^\mu$ which is time-like everywhere and normalised so that $U^\mu\, U_\mu = 1$.

(a) Show that the acceleration $A^\mu = U^\nu\, \nabla_\nu\, U^\mu$ is orthogonal to the velocity $U^\mu$

(b) Show that the tensor $h^\mu_\nu = \delta^\mu_\nu - U^\mu\, U_\nu$ satisfies

$$\begin{array}{lrcl} \text{(i)} & h^\mu_\nu\, U^\nu & = & 0\ , \\ \text{(ii)} & h^\mu_\nu\, h^\nu_\rho & = & h^\mu_\rho\ , \\ \text{(iii)} & h^\mu_\mu & = & 3\ . \end{array}$$

(c) Assume that space-time has a tensor

$$T^{\mu\nu} = (\rho + p)\, U^\mu\, U^\nu \ - \ g^{\mu\nu}\, p$$

where $\rho$ and $p$ are scalar functions.

Show that the continuity equation $\nabla_\mu\, T^{\mu\nu} = 0$ leads to the relativistic continuity equation

$$U^\mu\, \nabla_\mu \rho \ + \ (p + \rho)\, \nabla_\mu\, U^\mu = 0\ .$$

(d) Show that if $p = 0$, i.e. the fluid is a collection of dust particles following trajectories $x^\mu(\tau)$ with $U^\mu = \mathrm{D}x^\mu/\mathrm{D}\tau$, then the particles follow geodesics, that is that

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

### 3.10 The equations of the Electromagnetic Field

The electro-magnetic field strength is a $\left[{0 \atop 2}\right]$ tensor with components $F_{\mu\nu}$ which are given in terms of the components $A_\mu$ of a $\left[{0 \atop 1}\right]$ tensor field by

$$F_{\mu\nu} = \nabla_\mu A_\nu - \nabla_\nu A_\mu\ ,$$

and which, in the absence of charged matter, satisfy

$$\nabla^\mu F_{\mu\nu} = 0\ .$$

Furthermore, in the absence of charged matter, the energy-momentum tensor of the electro-magnetic field is also a $\left[{0 \atop 2}\right]$ tensor with components $T_{\mu\nu}$ given by

$$T_{\mu\nu} = \frac{1}{4\pi}\left\{F_{\mu\sigma}F_{\nu\tau}g^{\sigma\tau} - \frac{1}{4}g_{\mu\nu}F_{\sigma\tau}F^{\sigma\tau}\right\}\ .$$

Show that, in the absence of charged matter, that

(i) $\nabla_\mu F_{\nu\rho} + \nabla_\rho F_{\mu\nu} + \nabla_\nu F_{\rho\mu} = 0$.

(ii) $F^{\mu\sigma}\nabla_\mu F_{\nu\sigma} = (1/2)F^{\mu\sigma}\,(\nabla_\mu F_{\nu\sigma} + \nabla_\sigma F_{\mu\nu})$.

(iii) $\nabla^\mu T_{\mu\nu} = 0$.

Lecture notes for Spacetime Geometry and General Relativity

2015-2016

# Section 4: Spacetime and Physics

# 4 Spacetime and physics

In this section we shall first motivate Einstein's equations by comparing geodesic deviation in spacetime with the Newtonian analogue. We then breifly discuss the two classes of solutions we shall investigate. We next consider in some detail what it means for a spacetime to be static, and then deduce some properties of a static spacetime and show how Einstein's equations imply that a particle moving slowly in a weak static field wil be seen to obey Newtonian Gravity for a potential which is given in terms of the metric.

## 4.1 Particle paths and geometry

The path of a particle is given by a curve $x^\mu(\lambda)$ parametrised by $\lambda$. The tangent vector to the path, $\mathrm{d}x^\mu/\mathrm{d}\lambda$ satisfies

$$g_{\mu\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda} > 0 \quad , \qquad \text{for a massive particle,} \tag{202}$$

$$g_{\mu\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda} = 0 \quad , \qquad \text{for a massless particle.} \tag{203}$$

The proper time along a path - the time which the particle moving along the path would itself experience to have elapsed - is given by

$$\int \mathrm{d}\tau \ , \quad \text{where} \quad c^2\,\mathrm{d}\tau^2 = \mathrm{d}s^2 = g_{\mu\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda}\,\mathrm{d}\lambda^2 \ .$$

If the particle is massless then this is zero, but if the particle is massive, then we can take the proper time along the path as the parameter along the path. Since the metric and the proper time are related by $\mathrm{d}s^2 = c^2\mathrm{d}\tau^2$, In this case we have

$$U^\mu\,U^\nu\,g_{\mu\nu} = \left(\frac{\mathrm{d}s}{\mathrm{d}\tau}\right)^2 = c^2 \ , \qquad \text{where} \qquad U^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} \ ,$$

or in the usual choice of units in which $c = 1$,

$$U^\mu\,U_\mu = 1 \ .$$

We shall henceforth take units in which $c = 1$, unless stated otherwise. If the particle is falling freely, it follows a geodesic, which is a curve satisfying the equation

$$\frac{\mathrm{D}}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda} \equiv \frac{\mathrm{d}^2x^\mu}{\mathrm{d}\lambda^2} + \Gamma^\mu_{\rho\sigma}\frac{\mathrm{d}x^\rho}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\sigma}{\mathrm{d}\lambda} = f\,\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda} \ ,$$

for some $f$. If $f$ is identically zero, then the parameter $\lambda$ along the path is an *affine* parameter, and the equation of such a path is

$$\frac{\mathrm{D}}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda} = 0 \ .$$

For a massive particle, any affine parameter is linearly related to the proper time along the path (see the exercises).

## 4.2 Einstein's equations from geodesic deviation

It is important to remember we cannot *derive* Einstein's equations – it is only possible to guess them, and see if they actually describe the universe well or not. So, this section is not a derivation of Einstein's equations, but rather provides inspiration for making an educated guess.

We have the formula for geodesic variation:

$$\frac{\mathrm{D}}{\mathrm{d}\lambda}\frac{\mathrm{D}}{\mathrm{d}\lambda}V^\mu = R^\mu{}_{\rho\sigma\tau}\,U^\rho\,U^\sigma\,V^\tau\ . \tag{199}$$

However, we have already found (in question sheet 1) that the separation $\boldsymbol{\xi}(t)$ of the paths of two particles falling under Newtonian gravity also satisfies a second order differential equation:

$$\ddot{\xi}^i = -\xi^j\,\frac{\partial^2\Phi}{\partial x^i \partial x^j}\ , \tag{204}$$

where $\Phi$ is the Newtonian potential. These two equations should be equivalent in the weak-metric, slow-motion limit, in which case $V^i \sim \xi^i$ and $\mathrm{D}/\mathrm{d}\lambda \sim \mathrm{d}/\mathrm{d}t$. Taking only the spatial components of (199), we have

$$\frac{\mathrm{D}}{\mathrm{d}\lambda}\frac{\mathrm{D}}{\mathrm{d}\lambda}V^i = R^i{}_{\rho\sigma\tau}\,U^\rho\,U^\sigma\,V^\tau\ . \tag{205}$$

Let's also *assume* that at the point in question, the separation $V^a$ is purely spatial, i.e. $V^0 = 0$, then (205) becomes

$$\frac{\mathrm{D}}{\mathrm{d}\lambda}\frac{\mathrm{D}}{\mathrm{d}\lambda}V^i = R^i{}_{\rho\sigma j}\,U^\rho\,U^\sigma\,V^j\ . \tag{206}$$

Now, comparing (204) and (206) suggests the identification

$$\frac{\partial^2\Phi}{\partial x^i \partial x^j} \sim -R^i{}_{\rho\sigma j}\,U^\rho\,U^\sigma = R^i{}_{\rho j\sigma}\,U^\rho\,U^\sigma \tag{207}$$

Finally, if we consider the frame in which the two particles following the geodesics are instantaneously at rest, i.e. for which $U^0 = 1, U^i = 0$, then (207) simplifies to

$$\frac{\partial^2\Phi}{\partial x^i \partial x^j} \sim R^i{}_{0j0} \tag{208}$$

So far so good, but the aim is to find a differential equation for the metric which reduces to the equations of Newtonian gravity. To do this, we recall that the potential is itself determined at any time by the distribution of matter at that time – given by the mass-density $\rho$ – through the equation

$$\boldsymbol{\nabla}^2\Phi = (4\pi G)\,\rho\ , \tag{209}$$

where $G$ is Newton's constant. An alternative way of writing this is

$$\sum_{i=1}^{3}\frac{\partial^2\Phi}{\partial x^i \partial x^i} = (4\pi G)\,\rho\ , \tag{210}$$

which, looking at eqn. (208), suggests that

$$\sum_{i=1}^{3} R^i{}_{0i0} \sim (4\pi G)\,\rho\ . \tag{211}$$

This is not yet in quite the form we would like, since it is not an equation between tensors. The first simplification we can make is to notice that ${R^0}_{000} = 0$, and hence that

$$\sum_{i=1}^{3} {R^i}_{0i0} = \sum_{\mu=0}^{3} {R^\mu}_{0\mu0} = R_{00} \ , \tag{212}$$

so that eqn. (211) now reads

$$R_{00} \sim (4\pi G)\, \rho \ . \tag{213}$$

The next simplification we shall make is to assume that *at the point in question* there is no matter, i.e. let's consider only those parts of spacetime essentially free of matter (the gaps between planets, between galaxies,...) in which case, from eqn. (213) we have

$$R_{00} \sim 0 \ . \tag{214}$$

This is still not yet a tensor equation, but we can recall where the '0' indices came from in eqn. (208), by assuming that the the particles moving along the geodesics were instantaneously at rest. If we were to relax this restriction, then eqn. (214) should become

$$R_{\mu\nu}\, U^\mu\, U^\nu \sim 0 \ . \tag{215}$$

However, it is hard to see how the equation for the metric can depend on the velocity of *test particles* moving in that metric, and so eqn. (215) should hold for any (timelike) vector $U^\mu$. The only way to ensure this is to require that *all* the components of the Ricci tensor are zero. These equations are exactly Einstein's equations in the absence of matter:

**Einstein's equations in the absence of matter**

$$\boxed{R_{\mu\nu} = 0 \ .} \tag{216}$$

You should remember that this discussion is *not* a careful mathematical derivation. There is little point in being very worried if some of the steps are not entirely justifiable, since we are trying to find *new* equations which are not equivalent to Newton's equations – and so we can't have equalities all the way through. The real test is whether the new equations (216) actually describe the universe as it is seen, or not – and the answer is, that to the experimental accuracy that can be obtained today, they do.

### 4.2.1 Einstein's equations in the presence of matter

As we were careful to state, (216) should only be expected to hold in the absence of matter. Einstein's equations in the presence of matter are only a little more complicated to find, and we shall need the results later when we come to Cosmology – cosmology attempts to describe the large-scale evolution of the universe, and there is certainly an appreciable amount of matter in the universe! So, let us go back to equation (213), which was the last place we kept the matter density $\rho$ in our considerations:

$$R_{00} \sim (4\pi G)\,\rho\ . \tag{213}$$

Now, although we haven't studied it in any great detail, we have mentioned that $\rho$ is not itself a scalar, but only part of a tensor – $T_{\mu\nu}$, the energy–momentum tensor, given for a fluid of pressure $p$, density $\rho$ and 4–velocity $U^\mu$ by

$$T_{\mu\nu} = (p+\rho)\,U_\mu\,U_\nu - p\,g_{\mu\nu}\ .$$

Assuming that the metric is approximately the flat metric $\eta$, and that the fluid is instantaneously at rest, we find $T_{00} \sim \rho$, which suggests that eqn. (213) should read

$$R_{00} - (4\pi G)\,T_{00} \sim 0\ . \tag{217}$$

This immediately suggests the tensor equation

$$R_{\mu\nu} = (4\pi G)\,T_{\mu\nu}\ . \tag{218}$$

as the generalisation of eqn. (216) points with matter – but this is **WRONG**.

The reason this is wrong is that $T_{\mu\nu}$ *always* satisfies the identity

$$\nabla_\mu T^{\mu\nu} = 0\ , \tag{219}$$

the relativistic version of the continuity equation. So, for eqn. (218) to be consistent with eqn. (219), we should need $R_{\mu\nu}$ to satisfy

$$\nabla_\mu R^{\mu\nu} = 0\ , \tag{220}$$

and this is not in general the case. Requiring this to be true would make a very strong restriction on the sorts of metrics that are allowed – far stronger than is physically the case.

The solution to this dilemma is that there is *another* possible way to generalise eqn. (213) to a full tensor equation, and that is[8]

$$R_{\mu\nu} - \frac{R}{2}\,g_{\mu\nu} = (8\pi G)\,T_{\mu\nu}\ , \tag{221}$$

The combination on the left-hand side of eqn. (221) has a special name, the Einstein tensor, which is denoted by $G_{\mu\nu}$,

$$G_{\mu\nu} = R_{\mu\nu} - \frac{R}{2}\,g_{\mu\nu}\ , \tag{222}$$

[8] To check that this is compatible with eqn. (213), multiply both sides by $g^{\mu\nu}$, which will give $R = (8\pi G)(\rho - 3p)$, and then neglecting $p$ and substituting this back into (221) and taking the 00 component does indeed give eqn. (213)

and which always satisfies

$$\nabla_\mu G^{\mu\nu} = 0 \,, \tag{223}$$

thanks to the Bianchi identity (181). So, we can state

**Einstein's equations in the presence of matter**

$$\boxed{G_{\mu\nu} = (8\pi G)\, T_{\mu\nu} \,.} \tag{224}$$

#### 4.2.2 The energy-momentum tensor

There is an energy-momentum tensor, a type $\left[{0 \atop 2}\right]$ tensor, with components $T_{\mu\nu}$, associated to each type of matter. The component $T_{00} = \mathcal{H}$ is the *energy density* at that point, and $T_{i0} = \mathcal{P}_i$ are the components of the *momentum density.*

The momentum density $T_{i0}$ can also be interpreted as the flux of energy density in the $i$-th direction, and similarly the remaining components $T_{ij}$ are the flux of the $j$-th component of the momentum density in the $i$-th direction.

For example, for a fluid of density $\rho$ and pressure $p$, the energy-momentum tensor is given by

$$T_{\mu\nu} = U_\mu\, U_\nu\, (p + \rho) - p\, g_{\mu\nu} \,, \tag{225}$$

where $U_\mu$ is the $\left[{0 \atop 1}\right]$ tensor defined in terms of the 4–velocity $U^\nu$ of the fluid by $U_\mu = g_{\mu\nu}\, U^\nu$.

For a second example, the energy momentum tensor of the electro-magnetic field is given in terms of the field strength $F_{\mu\nu}$ by

$$T_{\mu\nu} = \frac{1}{4\pi} \left[ F_{\mu\sigma}\, F_{\nu\tau}\, g^{\sigma\tau} - \frac{1}{4} g_{\mu\nu}\, F_{\sigma\tau}\, F_{\rho\lambda}\, g^{\sigma\rho}\, g^{\tau\lambda} \right] \,, \tag{226}$$

where $g^{\mu\nu}$ is the inverse of the metric $g_{\mu\nu}$.

#### 4.2.3 The cosmological constant

In fact, Einstein was not happy with equations (224) since, as we shall see when we come to deal with cosmology, they do not allow a static universe to exist with matter in it. At the time Einstein wrote down these equation it was believed that the universe was indeed unchanging with time, and so Einstein generalised his equations slightly by adding a new term which allows such static solutions to exist. He added a term[9] proportional to a *cosmological constant* $\Lambda$, so that the new equations read

$$\boxed{G_{\mu\nu} - \Lambda g_{\mu\nu} = (8\pi G)\, T_{\mu\nu} \,.} \tag{227}$$

It was only later that the expansion of the universe was discovered, and since then people have spent a long time trying to find reasons why $\Lambda$ should be zero. Most recently however, it seems that it should not in fact be zero, but instead quite large. Whether or not such a term exists will in the end be decided by the results of careful astronomical measurements and at the moment the question is still open.

[9] This term is possible since $g_{\mu\nu}$ is covariantly constant, and does not contradict $\nabla^\mu G_{\mu\nu} = \nabla^\mu T_{\mu\nu} = 0$

### 4.2.4 Solutions of Einstein's equations

Having come this far, we are finally in a position to consider what sort of equations Einstein's equations are, and what sort of solutions we can find.

For a region of the universe with matter in it, there are two equations to solve,

$$\begin{aligned} G_{\mu\nu} - \Lambda g_{\mu\nu} &= (8\pi G)\, T_{\mu\nu} \\ \nabla_\mu T^{\mu\nu} &= 0 \end{aligned} \tag{228}$$

We can think of the first equation as a differential equation for the metric given the distribution of matter as summarised by the energy-momentum tensor, and the second equation as an equation for how the matter should move because of the curvature of spacetime.

This can be summarised by saying that matter tells spacetime how to curve, and spacetime tells matter how to move.

The problem is that both these equations are nasty non-linear equations, and almost impossible to solve. We shall in fact only be able to solve them in two cases -

• Firstly for the spacetime around a single, spherically symmetric, static body, in the case $\Lambda = 0$. Outside the body there is only empty space, and so in this region $T_{\mu\nu} = 0$, and we only have to solve $R_{\mu\nu} = 0$. There is a unique solution for the metric around such a body and it only depends on the mass and is the Schwarzschild solution, given by

$$\mathrm{d}s^2 = (1 - \frac{2m}{r})\mathrm{d}t^2 - (1 - \frac{2m}{r})^{-1}\mathrm{d}r^2 - r^2(\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2) \; .$$

Just as for Newtonian gravity, we do not need to know how to solve for the metric inside the massive body where it will be complicated and depend in a nasty way on the matter distribution in the body. We can solve for the Newtonian gravitational potential outside a massive body without knowing the details of the mass inside, and in the same way we can solve for the metric outside a spherically symmetric, static body and get this result.

• Secondly, we can solve for certain sorts of cosmological model in which the matter is uniformly distributed throughout space, and in which at any given time the universe looks the same at all points and in all directions. We leave this topic to the end of the course.

We turn to a discussion of general static spacetimes to get a feel for what will happen in the Schwarzschild spacetime before embarking on a detailed investigation of this spacetime.

## 4.3 Stationary and static metrics

Before we consider the Schwarzschild solution, it is first important to clarify exactly what we mean by a 'static' metric, and to find out which properties are generic to such a metric.

While we may think that an 'unchanging universe' has a definite physical meaning, it turns out that there is a subtle distinction to be made between 'stationary' metrics and 'static' metrics, which are both possible interpretations of an 'unchanging spacetime'.

In Newtonian gravity there is no difference between the gravitational potentials of a planet that is not moving and a planet that is spinning on its axis, but it turns out that there is general relativity. In some way, the spinning planet 'drags' spacetime around with it, so that the light-cones get bent over in the direction that the planet is spinning, as is sketched in figure 29.

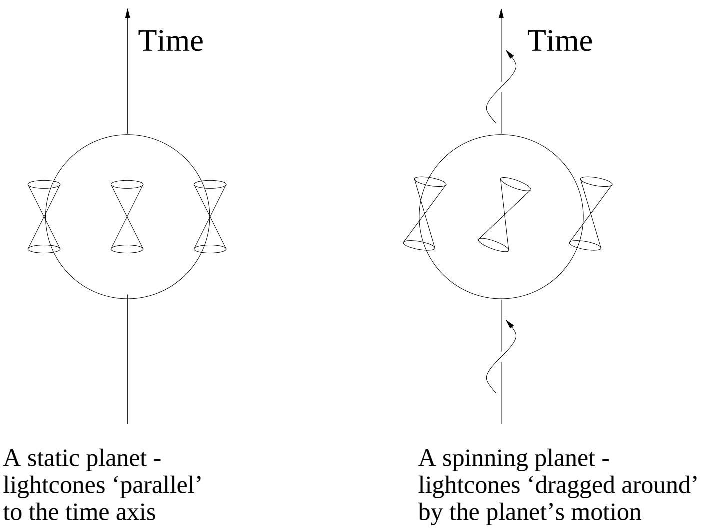



Figure 29

A planet that is spinning but otherwise not moving is an example of a *stationary* spacetime, and a planet that is not moving and not spinning, an example of a *static* spacetime. If we keep in mind these two possibilities, this will give us a good idea of the differences between the two sorts of spacetimes.

We can characterise static and stationary spacetimes in both physical and mathematical ways, and since static metrics are a special case of stationary metrics first, we'll discuss stationary metrics first.

### 4.3.1 Stationary metrics

Let's consider a spacetime with coordinates $\{t, x^1, x^2, x^3\}$. Then a natural requirement is that the metric be independent of $t$. This is called a stationary metric:

**Definition 1:** A stationary metric is one such that the metric $g_{\mu\nu}$ is independent of $t \equiv x^0$.

It turns out that this mathematical definition is completely equivalent to the following physical definition:

**Definition 2:** A spacetime has a stationary metric if one can find a 'good time coordinate' $t$ such that:

2.i) The proper time increments $\mathrm{d}\tau$ at a fixed point are a constant (independent of $t$) multiple of the coordinate time increments $\mathrm{d}t$ at that point.

2.ii) The coordinate time taken for light to traverse any given path $x^i(\lambda)$ is independent of $t$.

We shall now show the equivalence of these two definitions. The method will be to show that 2.i) is equivalent to $g_{00}$ independent of $t$, and that 2.ii) (when combined with 2.i)) is equivalent to $g_{0i}$ and $g_{ij}$ being independent of $t$, where we split $g_{\mu\nu}$ up into $g_{00}, g_{i0} = g_{0i}$ and $g_{ij}$ so that we can write

$$\mathrm{d}s^2 = g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu = g_{00}\,\mathrm{d}t^2 + 2g_{0i}\,\mathrm{d}t\,\mathrm{d}x^i + g_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j\ . \tag{229}$$

Let us consider 2.i) first.

The proper time increment is related to the metric by $\mathrm{d}s^2 = c^2\mathrm{d}\tau^2$. If we consider a fixed point $\mathcal{P}$, then clearly for the 'path' joining $\mathcal{P}$ at time $t$ and at time $t + \mathrm{d}t$, the spatial increments $\mathrm{d}x^i$ are zero, and so the proper time between $(t, x^i(\mathcal{P}))$ and $(t + \mathrm{d}t, x^i(\mathcal{P}))$ given by (229) becomes,

$$\mathrm{d}\tau^2 = \frac{1}{c^2}\mathrm{d}s^2 = \frac{g_{00}(\mathcal{P})}{c^2}\,\mathrm{d}t^2\ . \tag{230}$$

If $\mathrm{d}\tau$ is a constant multiple of $\mathrm{d}t$, then clearly $g_{00}(\mathcal{P})$ must be constant, and conversely if $g_{00}$ is independent of $t$, then $\mathrm{d}\tau$ is a constant multiple of $\mathrm{d}t$.

Now we can consider 2.ii)

The coordinate time taken for light to traverse a curve $x^i(\lambda)$ is given by

$$\Delta t = \int_{\lambda_0}^{\lambda_1} \frac{\mathrm{d}t}{\mathrm{d}\lambda}\,\mathrm{d}\lambda\ . \tag{231}$$

If this is to be independent of time for any path, then this will be true if the time taken to traverse the infinitesimal path segment from $\lambda$ to $\lambda + \mathrm{d}\lambda$, i.e. from $x^i(\lambda)$

$$x^i(\lambda + \mathrm{d}\lambda) = x^i(\lambda) + \frac{\mathrm{d}x^i}{\mathrm{d}\lambda}\mathrm{d}\lambda = x^i(\lambda) + \mathrm{d}x^i\ ,$$

is also independent of time. However, condition 2.ii) explicitly concerns the path $x^i(\lambda)$ of a photon[10] and we know that for photon paths,

$$g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu = 0\ .$$

Inserting this into eqn (229), this means that d$t$ satisfies

$$g_{00}\mathrm{d}t^2 + \mathrm{d}t\left(2g_{0i}\frac{\mathrm{d}x^i}{\mathrm{d}\lambda}\mathrm{d}\lambda\right) + \left(g_{ij}\frac{\mathrm{d}x^i}{\mathrm{d}\lambda}\frac{\mathrm{d}x^j}{\mathrm{d}\lambda}\right)\mathrm{d}\lambda^2 = 0\ . \tag{232}$$

Solving this equation, we find that there are potentially two solutions d$t^\pm$ for the time taken to traverse this path, being given by

$$\begin{aligned}\mathrm{d}t^\pm &= \frac{1}{2g_{00}}\left(-2g_{0i}\frac{\mathrm{d}x^i}{\mathrm{d}\lambda}\mathrm{d}\lambda \pm \sqrt{4(g_{0i}\frac{\mathrm{d}x^i}{\mathrm{d}\lambda}\mathrm{d}\lambda)^2 - 4g_{00}g_{ij}\frac{\mathrm{d}x^i}{\mathrm{d}\lambda}\frac{\mathrm{d}x^j}{\mathrm{d}\lambda}\mathrm{d}\lambda^2}\right) \\ &= \left[\frac{1}{2g_{00}}\left(-2g_{0i}\frac{\mathrm{d}x^i}{\mathrm{d}\lambda} \pm \sqrt{4(g_{0i}\frac{\mathrm{d}x^i}{\mathrm{d}\lambda})^2 - 4g_{00}g_{ij}\frac{\mathrm{d}x^i}{\mathrm{d}\lambda}\frac{\mathrm{d}x^j}{\mathrm{d}\lambda}}\right)\right]\mathrm{d}\lambda\end{aligned} \tag{233}$$

If we substitute e.g. $g_{\mu\nu} = \eta_{\mu\nu}$, we see that d$t^+$ is positive and d$t^-$ is negative, so we can interpret these two solutions as the time taken to traverse the path from $x^i$ to $x^i + \mathrm{d}\lambda(\mathrm{d}x^i/\mathrm{d}\lambda)$ in the positive and negative senses, as in figure 30.

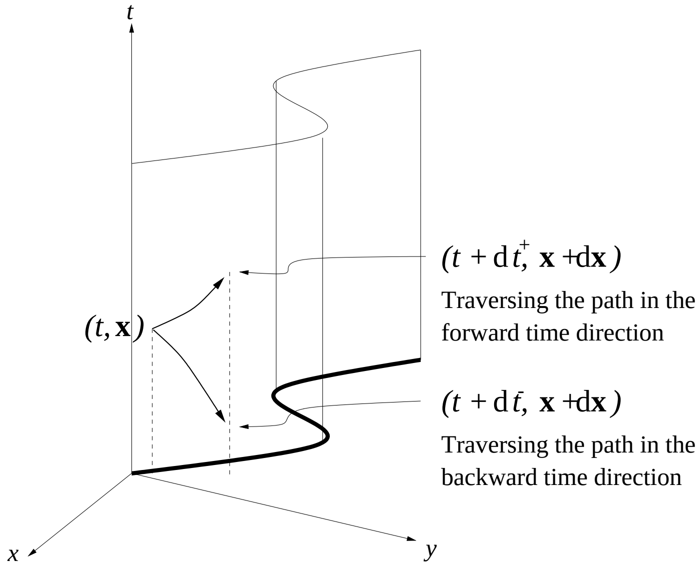



Figure 30

These should each be independent of $t$ separately, and hence so should their sum

$$\mathrm{d}t^+ + \mathrm{d}t^- = \left[-\frac{2g_{0i}}{g_{00}}\frac{\mathrm{d}x^i}{\mathrm{d}\lambda}\right]\mathrm{d}\lambda\ , \tag{234}$$

and their product

$$\mathrm{d}t^+\,\mathrm{d}t^- = \left[\frac{g_{ij}}{g_{00}}\frac{\mathrm{d}x^i}{\mathrm{d}\lambda}\frac{\mathrm{d}x^j}{\mathrm{d}\lambda}\right]\mathrm{d}\lambda^2\ . \tag{235}$$

[10]Don't worry that the path is not necessarily a geodesic – we can accelerate photons just as easily as massive particles – just consider the path that light takes inside a fibre-optic cable. This isn't (necessarily) a straight line, yet the light manages to take it somehow

Now, let us consider what we can deduce from eqns. (234) and (235).

So far, the path $x^i(\lambda)$ has been completely arbitrary, so we can easily choose

$$\frac{\mathrm{d}x^i}{\mathrm{d}\lambda} = (1,0,0) \ , \tag{236}$$

in which case eqns. (234) and (235) become

$$\mathrm{d}t^+ + \mathrm{d}t^- = -\frac{2g_{01}}{g_{00}}\,\mathrm{d}\lambda \ , \quad \mathrm{d}t^+\,\mathrm{d}t^- = \frac{g_{11}}{g_{00}}\,\mathrm{d}\lambda^2 \ . \tag{237}$$

Clearly, these are to be independent of $t$ if and only if $g_{01}$ and $g_{11}$ are independent of $t$.

Similarly, by taking $\frac{\mathrm{d}x^i}{\mathrm{d}\lambda} = (0,1,0)$ and $\frac{\mathrm{d}x^i}{\mathrm{d}\lambda} = (0,0,1)$ in turn, we find that $\mathrm{d}t^\pm$ for these paths are independent of $t$ if and only if $g_{02}, g_{22}, g_{03}$ and $g_{33}$ are independent of $t$.

Finally, taking

$$\frac{\mathrm{d}x^i}{\mathrm{d}\lambda} = (1,1,0) \ , \tag{238}$$

we find

$$\mathrm{d}t^+\,\mathrm{d}t^- = \frac{g_{11} + 2g_{12} + g_{22}}{g_{00}}\,\mathrm{d}\lambda^2 \ , \tag{239}$$

and having already shown that $g_{11}$ and $g_{22}$ are independent of $t$, we find that for this path, $\mathrm{d}t^\pm$ are independent of $t$ if and only if $g_{12}$ is independent of $t$.

Similarly, by taking $\frac{\mathrm{d}x^i}{\mathrm{d}\lambda} = (0,1,1)$ and $\frac{\mathrm{d}x^i}{\mathrm{d}\lambda} = (1,0,1)$, we can complete the argument by showing that $g_{23}$ and $g_{13}$ are independent of $t$.

This completes the proof that the two definitions of a stationary spacetime, 1 and 2, are completely equivalent.

### 4.3.2 Static metrics

The physical definition of a static metric includes one extra condition:

**Definition 3:** A spacetime has a static metric if

3.i) is is a stationary spacetime, and further

3.ii) the coordinate times taken for light to traverse any path in one direction and in the opposite direction are equal.

It turns out that there is again a mathematical definition which is completely equivalent to the physical definition:

**Definition 4:** A static metric is one such that

4.i) $g_{\mu\nu}$ is independent of $t$

4.ii) $g_{0i} = 0$ for $i = 1,2,3$.

We shall show that these are equivalent in an exactly analogous method to that we applied to the two definitions on stationary metrics, by considering the time taken for light to traverse a small segment of the path $x^i(\lambda)$ between $\lambda$ to $\lambda + \mathrm{d}\lambda$, in either direction.

In showing that the two definitions of a stationary metric were equivalent, we already worked out the coordinate time taken for light to traverse a small path from $x^i$ to $x^i + (\mathrm{d}x^i/\mathrm{d}\lambda)\mathrm{d}\lambda$ – these being $\mathrm{d}t^+$ and $|\mathrm{d}t^-| = -\mathrm{d}t^-$. We recall the expressions for these from eqn (109):

$$|\mathrm{d}t^+| = \ \mathrm{d}t^+ \ = \left[\frac{1}{2g_{00}}\left(-2g_{0i}\frac{\mathrm{d}x^i}{\mathrm{d}\lambda} + \sqrt{4(g_{0i}\frac{\mathrm{d}x^i}{\mathrm{d}\lambda})^2 - 4g_{ij}\frac{\mathrm{d}x^i}{\mathrm{d}\lambda}\frac{\mathrm{d}x^j}{\mathrm{d}\lambda}}\right)\right]\mathrm{d}\lambda \tag{240}$$

$$|\mathrm{d}t^-| = -\mathrm{d}t^- \ = \left[\frac{1}{2g_{00}}\left(\ 2g_{0i}\frac{\mathrm{d}x^i}{\mathrm{d}\lambda} + \sqrt{4(g_{0i}\frac{\mathrm{d}x^i}{\mathrm{d}\lambda})^2 - 4g_{ij}\frac{\mathrm{d}x^i}{\mathrm{d}\lambda}\frac{\mathrm{d}x^j}{\mathrm{d}\lambda}}\right)\right]\mathrm{d}\lambda \tag{241}$$

Clearly these times are equal if and only if their difference,

$$|\mathrm{d}t^+| - |\mathrm{d}t^-| = -2\,\frac{g_{0i}}{g_{00}}\,\frac{\mathrm{d}x^i}{\mathrm{d}\lambda}\ ,$$

is zero. Hence, since for a static metric we require this to be zero for *any* path $x^i(\lambda)$, we see that the time taken to traverse the path in both directions is the same if and only if

$$g_{0i} = 0\ . \tag{242}$$

### 4.3.3 Form of $g_{00}$ in a stationary metric

One of the implication of the physical conditions on a stationary metric is that the frequency of light does not change, **as measured relative to coordinate time**. This is easy to see when you consider the paths in spacetime of two subsequent wave-crests of light, as in figure 31.

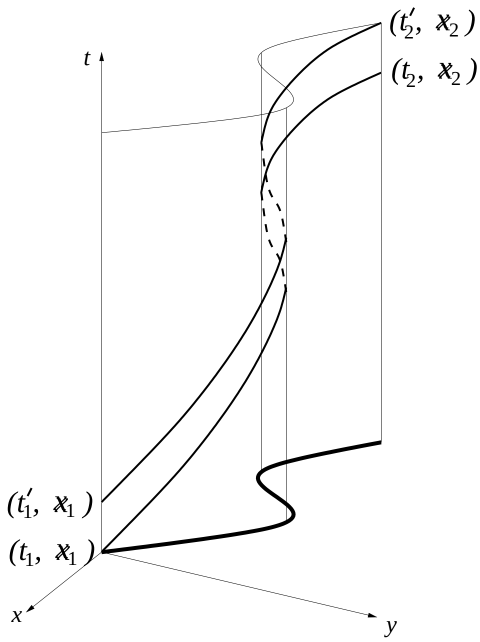



Figure 31

Condition 2.ii) says that the coordinate time taken for light to traverse the path is the same, i.e. that

$$t_2 - t_1 = t_2' - t_1'\ ,$$

but a trivial rearrangement shows that the coordinate time between successive wave-crests is the same at the beginning and end of the path:

$$\Rightarrow t_1' - t_1 = t_2' - t_2 \ .$$

However, the **physical frequency of light** is measured relative to the proper time at $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, and when the light arrives at $\boldsymbol{x}_2$ its physical frequency is not necessarily the same as that when it left $\boldsymbol{x}_1$. Condition 2.i) says that the proper time elapsed at a point is proportional to the coordinate time elapsed, with the constant of proportionality given from the metric by

$$\mathrm{d}s^2 = g_{00}\,\mathrm{d}t^2 + 2g_{0i}\,\mathrm{d}t\,\mathrm{d}x^i + g_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j \ ,$$

so that if the coordinates $x^i$ are unchanged, we have

$$\mathrm{d}s^2\big|_{\mathrm{d}x^i=0} = g_{00}\,\mathrm{d}t^2 \ ,$$

and hence the proper time differences at $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ between the successive wave-crests are given by $\Delta\tau_i$ as

$$\Delta\tau_i = \sqrt{g_{00}(\boldsymbol{x}_i)}\ \Delta t_i = \sqrt{g_{00}(\boldsymbol{x}_i)}\ (t_i' - t_i) \ .$$

Hence the *physical* frequencies of the light at the points $\boldsymbol{x}_i$, being given by $\Delta\tau_i$, are related by

$$\begin{aligned} \nu_2 &= \frac{1}{\Delta\tau_2} = \frac{1}{\sqrt{g_{00}(\boldsymbol{x}_2)}\ (t_2' - t_2)} = \frac{1}{\sqrt{g_{00}(\boldsymbol{x}_2)}}\,\frac{1}{(t_1' - t_1)} = \frac{1}{\sqrt{g_{00}(\boldsymbol{x}_2)}}\,\frac{\sqrt{g_{00}(\boldsymbol{x}_1)}}{\Delta\tau_1} \\ &= \sqrt{\frac{g_{00}(\boldsymbol{x}_1)}{g_{00}(\boldsymbol{x}_2)}}\,\nu_1 \ . \end{aligned} \tag{243}$$

However, we saw in exercise 3, sheet 1, that the SEP implies that (for weak fields) the frequencies are related by

$$\frac{\nu_2}{\nu_1} = \exp\left(-\frac{\Phi(\boldsymbol{x}_2) - \Phi(\boldsymbol{x}_1)}{c^2}\right) \ , \tag{244}$$

where $\Phi$ is the Newtonian gravitational potential. Equating (243) and (244), we find that $g_{00}$ is given, up to a constant, by

$$g_{00}(t, \boldsymbol{x}) = c^2\ \exp\left(\frac{2\Phi(\boldsymbol{x})}{c^2}\right) \ . \tag{245}$$

The constant factor has been chosen to be $c^2$, so that the dimensions of $g_{00}$ are correct, and so that for $\Phi = 0$ we recover the flat space answer $g_{00}|_{\text{flatspace}} = \eta_{00} = c^2$.

In summary, one can find coordinates in which a static metric takes the form

$$\mathrm{d}s^2 = \exp\left(\frac{2\Phi}{c^2}\right) c^2\mathrm{d}t^2 + g_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j \ ,$$

with $\Phi$ and $g_{ij}$ independent of $t$, or in units in which $c = 1$,

$$\boxed{\mathrm{d}s^2 = \exp(2\Phi)\,\mathrm{d}t^2 + g_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j \ ,} \tag{246}$$

## 4.4 Slow motion in a static metric

We have found that consideration of gravitational time dilation and the SEP meant that a static metric could be written as (246)

$$\mathrm{d}s^2 = e^{2\Phi}\,\mathrm{d}t^2 + g_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j \ ,$$

with $\Phi$ and $g_{ij}$ independent of $t$. We can now show that in the limiting case of *slow motion* in a *weak field*, the Geodesic equation reduces to the Newtonian equations of motion. If we did not recover Newton's gravity in this limiting case, then something would be seriously wrong, as for most purposes Newton's gravity does seem to work quite well on Earth, for slow speeds (much less than the speed of light) and weak fields (such as the Earth's gravitational field). To show that this is the case, we need to come up with some mathematical formulation of the words 'slow motion' and 'weak field'. Since we are using units in which $c = 1$, then we shall understand these words to mean the following:

• **Weak field:** If the field is weak, then we shall assume that the potential is also small,

$$|\Phi| \ll 1 \ , \tag{247}$$

so that, for example,

$$\exp(2\Phi) \sim 1 + 2\Phi + \ldots \ ,$$

Since the field is weak, space should not be distorted too much. so that if we write the spatial part of the metric as

$$g_{ij} = -\delta_{ij} - h_{ij} \ ,$$

then we shall assume that both $h_{ij}$ and the first derivatives of $h_{ij}$ are small, i.e.

$$|h_{ij}| \ll 1 \ , \quad |\partial_i h_{jk}| \ll 1 \ . \tag{248}$$

Together, (247) and (248) comprise the mathematical formulation of 'weak field'.

• **Slow motion:** If $x^\mu(\tau)$ is the path of a particle, then the motion is 'slow' if

$$|\mathrm{d}x^i/\mathrm{d}\tau|^2 = (\delta_{ij} + h_{ij})\,\frac{\mathrm{d}x^i}{\mathrm{d}\tau}\,\frac{\mathrm{d}x^j}{\mathrm{d}\tau} \sim \delta_{ij}\,\frac{\mathrm{d}x^i}{\mathrm{d}\tau}\,\frac{\mathrm{d}x^j}{\mathrm{d}\tau} \ll 1 \ . \tag{249}$$

The motion of a massive particle always satisfies $g_{\mu\nu}\,(\mathrm{d}x^\mu/\mathrm{d}\tau)(\mathrm{d}x^\nu/\mathrm{d}\tau) = 1$, so that

$$1 = g_{\mu\nu}\,(\mathrm{d}x^\mu/\mathrm{d}\tau)(\mathrm{d}x^\nu/\mathrm{d}\tau) = \exp(2\Phi)\left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^2 - (\delta_{ij} + h_{ij})\,\frac{\mathrm{d}x^i}{\mathrm{d}\tau}\,\frac{\mathrm{d}x^j}{\mathrm{d}\tau} \sim \left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^2 \ ,$$

i.e.

$$\left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^2 \sim 1 \tag{250}$$

Together, (249) and (250) comprise the mathematical formulation of 'slow motion'.

The geodesic equations are

$$\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\tau^2} + \Gamma^\mu_{\nu\rho} \frac{\mathrm{d}x^\nu}{\mathrm{d}\tau} \frac{\mathrm{d}x^\rho}{\mathrm{d}\tau} = 0 \,, \tag{251}$$

whereas Newton's equations are

$$\frac{\mathrm{d}^2 x^i}{\mathrm{d}t^2} = -\partial_i \Phi \,. \tag{252}$$

Since Newton's equations are equations for the *spatial* coordinates $x^i$ as a function of $t$, and we will use $t \sim \tau$, we only need consider the restriction of the geodesic equations (251) to $\mu$ being a spatial index $i = 1, 2, 3$. As usual, there are (at least) two ways to find these geodesic equations. One method is to directly evaluate $\Gamma^i_{jk}$ and substitute it into the spatial components $i = 1, 2, 3$ of eqn. (251). The first simplification we can make is to notice that since $\mathrm{d}t/\mathrm{d}\tau \sim 1$, we can also replace $\mathrm{d}/\mathrm{d}\tau$ by $\mathrm{d}/\mathrm{d}t$ in (251), so that the spatial geodesic equations become

$$\frac{\mathrm{d}^2 x^i}{\mathrm{d}t^2} + \Gamma^i_{\nu\rho} \frac{\mathrm{d}x^\nu}{\mathrm{d}t} \frac{\mathrm{d}x^\rho}{\mathrm{d}t} \sim 0 \,, \tag{253}$$

Also, since $|\mathrm{d}x^i/\mathrm{d}t| \ll 1$, we can essentially ignore the contributions in the sums over $\nu$ and $\rho$ from the spatial components, and the dominant term in the sums in eqn (253) are given from $\nu{=}0$ and $\rho{=}0$. The geodesic equations now become

$$\frac{\mathrm{d}^2 x^i}{\mathrm{d}t^2} + \Gamma^i_{00} \sim 0 \,, \tag{254}$$

We only need to work out $\Gamma^i_{00}$, which is given by the definition as

$$\Gamma^i_{00} = \frac{1}{2} g^{ij} \left(\partial_0 g_{j0} + \partial_0 g_{j0} - \partial_j g_{00}\right) \sim -\frac{1}{2} \delta^{ij} \left(0 + 0 - \partial_j \left(1 + 2\Phi\right)\right) = \partial_i \Phi \,. \tag{255}$$

Substituting (255) into eqn (254), we indeed find that the geodesic equations reduce, in the case of slow motion in a static metric, to Newton's equations, (252),

$$\frac{\mathrm{d}^2 x^i}{\mathrm{d}t^2} = -\Gamma^i_{00} + \ldots = -\partial_i \Phi + \ldots \,.$$

The other way is to use the Euler-Lagrange equations for

$$\mathcal{L} \;=\; \exp(2\Phi) \frac{\mathrm{d}t}{\mathrm{d}\tau}^{2} - (\delta_{ij} + h_{ij}) \frac{\mathrm{d}x^i}{\mathrm{d}\tau} \frac{\mathrm{d}x^j}{\mathrm{d}\tau} \;\;.$$

The Euler-Lagrange equations for $x^i$ are

$$\begin{aligned}
0 \;&=\; \frac{\mathrm{d}}{\mathrm{d}\tau} \frac{\partial \mathcal{L}}{\partial x'^i} - \frac{\partial \mathcal{L}}{\partial x^i} \\
&= \frac{\mathrm{d}}{\mathrm{d}\tau} (-2(\delta_{ij} + h_{ij}) x'^j) - 2\partial_i \Phi (\exp(2\Phi) t'^2) + \partial_i h_{jk}\, x'^j\, x'^k \\
&= -2(x'^k \partial_k h_{ij}) x'^j - 2(\delta_{ij} + h_{ij}) x''^j - 2\partial_i \Phi \, \exp(2\Phi)\, t'^2 + \partial_i h_{jk}\, x'^j\, x'^k \\
&\sim -2\delta_{ij} \ddot{x}^j - 2\partial_i \Phi + \ldots \;,
\end{aligned}$$

where in the last step we used eqns. (247), (248), (249) and (250) to drop the terms in $h_{ij}$ and $\partial_i h_{jk}$, and to replace $\exp(2\Phi)$ by 1 and $x''^i$ by $\ddot{x}^j$. These are indeed Newton's equations:

$$-2(\,\delta_{ij} \ddot{x}^j + \partial_i \Phi) \sim 0 \;\;\Leftrightarrow\;\; \ddot{x}^i \sim -\partial_i \Phi$$

## 4.5 Exercises

### Short Exercises

**4.1. Static or Stationary?** Consider the following metrics.

In each case say for which values of $(t, x)$ the path $x$ = constant, ie a path parametrised as $(t(\lambda), x(\lambda)) = (\lambda, x_0)$, is timelike, spacelike or null.

In each case say for which values of $(t, x)$ the path $t$ = constant, ie a path parametrised as $(t(\lambda), x(\lambda)) = (t_0, \lambda)$, is timelike, spacelike or null.

For those cases for which $t$ is a timelike coordinate, ie for which the path $x$ = constant is a timelike path, state whether the metric is static, stationary or neither in the appropriate region of spacetime.

$$(i) \qquad \mathrm{d}s^2 = (1 - \tfrac{2m}{x})\mathrm{d}t^2 - \frac{\mathrm{d}x^2}{(1 - \frac{2m}{x})}$$

$$(ii) \qquad \mathrm{d}s^2 = (1 - \tfrac{2m}{x})\mathrm{d}t^2 - 2\mathrm{d}t\,\mathrm{d}x$$

$$(iii) \qquad \mathrm{d}s^2 = \frac{\mathrm{d}t^2}{(\frac{2m}{t} - 1)} - (\tfrac{2m}{t} - 1)\mathrm{d}x^2$$

**4.2. Null geodesics in some simple metrics**

For each of the metrics in question 4.1, sketch the patterns of null geodesics (paths that light can follow).

For each of the metrics in question 4.1, find the exact solutions of the null geodesics by solving the equations $\mathrm{d}s^2 = 0$.

**Longer Exercises**

**4.3. Motion in a static field (part of an exam question from 2001).**

In coordinates $(t, x^i)$ in which the speed of light is 1, the metric of a static spacetime can be written in the form

$$ds^2 = e^{2\varphi}\, dt^2 \; - \; h_{ij}\, dx^i\, dx^j \; ,$$

where $\varphi$ and $h_{ij}$ are independent of $t$.

(ii) Let $x^\mu(\tau)$ be the path of a freely falling massive particle with $\tau$ the proper time along the path. Furthermore, the particle is instantaneously at rest in these coordinates at $\tau = \tau_0$, i.e.

$$\frac{dx^i}{d\tau}(\tau_0) = 0 \; .$$

Show that at $\tau = \tau_0$

$$\text{(i)} \qquad \frac{dt}{d\tau} \; = \; e^{-\varphi}$$

$$\text{(ii)} \qquad h_{ij}\, \frac{d^2 x^j}{d\tau^2} \; = \; -\frac{\partial}{\partial x^i}\, \varphi$$

**4.4. Symmetries of a stationary metric**

A vector $\xi^\mu$ defines a symmetry of a metric if it satisfies Killing's equation:

$$\nabla_\mu \xi_\nu + \nabla_\nu \xi_\mu = 0 \; ,$$

- Show that Killing's equation is equivalent to the equation

$$\xi^\sigma \partial_\sigma (g_{\mu\nu}) + g_{\sigma\nu} \partial_\mu \xi^\sigma + g_{\mu\sigma} \partial_\nu \xi^\sigma = 0 \; .$$

- Show that if the 4–vector

$$T^\mu = (1, 0, 0, 0) \; ,$$

satisfies Killing's equation, then the metric is independent of $t \equiv x^0$.

Hence show that in a stationary metric

$$U^\mu T_\mu \; ,$$

is constant along an affinely parametrised geodesic $x^\mu(\lambda)$, where

$$U^\mu = \frac{\partial x^\mu}{\partial \lambda} \; .$$

**4.5. Null geodesics in Nordstrom's theory.**

In Nordstrom's theory of gravity, the metric has the form

$$g_{\mu\nu} = \exp(2\Phi)\,\eta_{\mu\nu}\ .$$

Using the results of questions 2.13 and 2.14, show that null (light-like) geodesics undergo no gravitational deflection in Nordstrom's theory.

**A very long exercise**

**4.6. Geodesics in Nordstrom's theory and the falling lift - an extended discussion**

Consider the metric in Nordstrom's theory in Cartesian coordinates $(t, x^i) = (t, x, y, z)$:

$$\mathrm{d}s^2 = e^{2\varphi/c^2}\,(c^2\mathrm{d}t^2 - \mathrm{d}x^i\mathrm{d}x^i)\ .$$

Let $x^\mu(\lambda)$ be an affinely parametrised geodesic. Show that

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(e^{2\varphi/c^2}\frac{\mathrm{d}t}{\mathrm{d}\lambda}\right) &= 0\\
\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(e^{2\varphi/c^2}\frac{\mathrm{d}x^i}{\mathrm{d}\lambda}\right) &= -\partial_i\varphi\, e^{2\varphi/c^2}\left(\left(\frac{\mathrm{d}t}{\mathrm{d}\lambda}\right)^2 - \frac{1}{c^2}\frac{\mathrm{d}x^k}{\mathrm{d}\lambda}\frac{\mathrm{d}x^k}{\mathrm{d}\lambda}\right)
\end{aligned}$$

We can now consider the two cases of (1) paths of particles and light and (2) 'straight lines'

1. In the first case, we can take $t$ to be increasing along the geodesic, and so take

   $$e^{2\varphi/c^2}\frac{\mathrm{d}t}{\mathrm{d}\lambda} = 1\ .$$

   Show that in this case the affine geodesic equations reduce to

   $$\frac{\mathrm{d}^2x^i}{\mathrm{d}t^2} = -\partial_i\varphi\,(1 - \frac{v^2}{c^2})\ ,$$

   where $v^i = \mathrm{d}x^i/\mathrm{d}t$.

   Hence we see that for 'slow' particles, the gravitational deflection is almost exactly given by Newton's equations, but for particles moving at the speed of light there is no deflection

2. In the second case, we can take $t = 0$. If we define a new (non-affine) parameter $\mu$ by

   $$e^{2\varphi/c^2}\frac{\mathrm{d}\mu}{\mathrm{d}\lambda} = 1\ ,$$

   show that the equations for a spacelike geodesic are

   $$\frac{\mathrm{d}^2x^i}{\mathrm{d}\mu^2} = \frac{1}{c^2}\partial_i\varphi\left(\frac{\mathrm{d}x^k}{\mathrm{d}\mu}\frac{\mathrm{d}x^k}{\mathrm{d}\mu}\right)\ .$$

We can now consider the special case of a constant gravitational field for which $\varphi = gz$. Consider a spacelike geodesic in the $x$—$z$ plane (i.e. we take $y = 0$).

Show that the equations reduce to

$$\frac{\mathrm{d}^2 z}{\mathrm{d}x^2} = \frac{g}{c^2}\left(1 + \left(\frac{\mathrm{d}z}{\mathrm{d}x}\right)^2\right) ,$$

and by considering the equation for $\mathrm{d}z/\mathrm{d}x$, or otherwise, show that we can find a solution in the form

$$z = \frac{c^2}{g} \log\left|\sec\left(\frac{gx}{c^2}\right)\right| .$$

We can find these results in approximate form by using normal coordinates at the point $x^\mu = 0$, $\tilde{x}^\mu$, in which the geodesic equations (at the point $x^\mu = 0$) are simply

$$\frac{\mathrm{d}^2\tilde{x}^\mu}{\mathrm{d}\lambda^2} = 0 .$$

1. Consider the simple case of $\varphi = gz$ and find all the Christoffel symbols for the metric

$$\mathrm{d}s^2 = e^{2\varphi/c^2}\left(c^2\mathrm{d}t^2 - \mathrm{d}x^i\mathrm{d}x^i\right) .$$

2. Define $\tilde{x}^\mu$ by

$$x^\mu = \tilde{x}^\mu + \frac{1}{2}\Gamma^\mu_{\nu\sigma}(0)\tilde{x}^\nu\tilde{x}^\sigma .$$

   Show that the timelike and null geodesics moving in the $x$–$z$ plane passing through $x^\mu = 0$ take the form

$$x \simeq vt\ , y \simeq 0\ , z \simeq -\frac{1}{2}gt^2\,(1 - \frac{v^2}{c^2}) .$$

3. Show that the spacelike geodesic plane passing through $x^\mu = 0$ along the $x$-axis satisfies

$$z \simeq \frac{1}{2}gx^2/c^2 .$$

4. Verify that this is consistent with the exact solution found earlier.

Lecture notes for Spacetime Geometry and General Relativity

2015-2016

## Section 5: The Schwarzschild solution

# 5 The Schwarzschild solution

## 5.1 Spherically symmetric static metrics

With coordinates $t, x^i$, a static metric has the form

$$\mathrm{d}s^2 = g_{00}\mathrm{d}t^2 \ + \ g_{ij}\mathrm{d}x^i\mathrm{d}x^j \ , \tag{256}$$

with $g_{00}$ and $g_{ij}$ independent of $t$. How does spherical symmetry further restrict this metric? This is up to us, as we have to choose coordinates and then try to turn the vague idea of 'spherically symmetric' into something mathematical[11].

Supposing the metric is spherically symmetric about the point $x^i = 0$, let us try to take spherical polar coordinates $\{r, \theta, \phi\}$. Then we should like $r$ to be some kind of radial coordinate and $\theta$ and $\phi$ to measure angles, just as we are used to. We should also like all distances to be invariant under rotations about the origin.

Firstly, the cross terms $g_{\theta r}, g_{\theta\phi}$ and $g_{\phi r}$ must be zero[12], just as the metric being static demands $g_{0,i} = 0$.

Secondly, we would like the radial distance to be independent of the direction, and hence $g_{rr}$ must be independent of $\theta$ and $\phi$ (as well as independent of $t$).

Finally, the metric on each surface of constant $r$ should be the metric of a sphere, i.e.

$$c\,(\mathrm{d}\theta^2 + \sin^2\!\theta\,\mathrm{d}\phi^2)\ ,$$

for some function $c$ which only depends on $r$.

Hence, the general static spherically symmetric metric takes the form

$$\mathrm{d}s^2 = a(r)\mathrm{d}t^2 \ - \ \left(b(r)\mathrm{d}r^2 \ + \ c(r)\,(\mathrm{d}\theta^2 + \sin^2\!\theta\,\mathrm{d}\phi^2)\right)\ . \tag{257}$$

While $\{\theta, \phi\}$ will describe the direction of a point as viewed from the origin, we have to make a choice as to the role $r$ plays:

1. We can choose $r$ to measure the radial distance from the origin, so that a point at coordinate $r$ is distance $r$ in the radial direction from the origin. In this case $b(r) = 1$.
2. We can choose $r$ to measure the area of a 2–sphere of coordinate $r$, so that the area is exactly $4\pi r^2$. In this case $c(r) = r^2$

What we **cannot** do is ask both 1. and 2. to be true. This will mean that the 3–space of constant $t$ is flat, and this is a very strong restriction on the metric which will end up enforcing the whole metric to be flat.

It is conventional (because it makes the equations easier) to choose option 2., so that our choice of spherically symmetric static metric is [with $a(r) \equiv \exp A(r)$ and $b(r) \equiv \exp B(r)$ for future convenience]

$$\mathrm{d}s^2 = \mathrm{e}^{A(r)}\mathrm{d}t^2 \ - \ \mathrm{e}^{B(r)}\mathrm{d}r^2 \ - \ r^2\,(\mathrm{d}\theta^2 + \sin^2\!\theta\,\mathrm{d}\phi^2)\ . \tag{258}$$

[11] A rigorous and general treatment of symmetries involves a discussion of 'Killing vectors'. These are vectors $k_\mu$ which satisfy $\nabla_\mu k_\nu + \nabla_\nu k_\mu = 0$, and each such vector implies the existence of a corresponding symmetry. We shall explore this a little in the exercise sheets, and in this section we adopt a more pedestrian approach

[12] Otherwise the distance from the point with coordinates $(r, \theta, \phi)$ to a point $(r', \theta', \phi')$ would not be the same as that from $(r, \theta', \phi')$ to a point $(r', \theta, \phi)$

The next step is to find the components of the Ricci tensor $R_{\mu\nu}$ for this metric and solve eqn. (254). We do this in the next section.

### 5.2 Solving the Einstein equations for a spherically symmetric flat metric

The components of the Ricci tensor for the metric (258) (discussed on exercise sheet 5) are

$$
\begin{aligned}
R_{tt} &= \mathrm{e}^{A-B}\left(\frac{1}{2}A'' - \frac{1}{4}A'B' + \frac{1}{4}(A')^2 + \frac{A'}{r}\right) , && (259)\\
R_{rr} &= -\frac{1}{2}A'' + \frac{1}{4}A'B' - \frac{1}{4}(A')^2 + \frac{B'}{r} , && (260)\\
R_{\theta\theta} &= -\mathrm{e}^{-B}\left(1 + \frac{r}{2}(A' - B')\right) + 1 , && (261)\\
R_{\phi\phi} &= R_{\theta\theta}\sin^2\theta , && (262)
\end{aligned}
$$

where primes denote differentiation with respect to $r$. All other components vanish. Now we have to solve eqn. (254), i.e. set the four expressions (259) – (262) to zero.

The first observation is that $R_{\phi\phi}$ is a multiple of $R_{\theta\theta}$, and so we shall forget about it. Next, we take the following combination of $R_{tt}$ and $R_{rr}$

$$
\begin{aligned}
0 &= \mathrm{e}^{B-A}\, R_{tt} + R_{rr}\\
&= \frac{A' + B'}{r} . && (263)
\end{aligned}
$$

Solving this equation, we find

$$A = -B + k , \qquad (264)$$

where $k$ is a constant. By rescaling the $t$ coordinate appropriately, $t \to \mathrm{e}^{\lambda}t$, so that

$$\mathrm{e}^{B}\,\mathrm{d}t^2 \to e^{B+2\lambda}\,\mathrm{d}t^2 ,$$

we can set $k$ to zero, and have

$$A = -B . \qquad (265)$$

Next we substitute (265) into $R_{\theta\theta}$:

$$
\begin{aligned}
0 &= R_{\theta\theta}\\
&= -\mathrm{e}^{A}\left(1 + rA'\right) + 1\\
&= -\frac{\mathrm{d}}{\mathrm{d}r}\left(r\mathrm{e}^{A}\right) + 1 , && (266)
\end{aligned}
$$

and hence

$$
\begin{aligned}
\tfrac{\mathrm{d}}{\mathrm{d}r}\left(r\mathrm{e}^{A}\right) &= 1 ,\\
\Rightarrow \quad r\mathrm{e}^{A} &= r + k \qquad \text{For some constant } k ,\\
\Rightarrow \quad \mathrm{e}^{A} &= \left(1 + \tfrac{k}{r}\right) . && (267)
\end{aligned}
$$

Substituting (267) into the metric, we find

$$\mathrm{d}s^2 = \left(1 + \frac{k}{r}\right)\mathrm{d}t^2 - \frac{\mathrm{d}r^2}{\left(1 + \frac{k}{r}\right)} - r^2\left(\mathrm{d}\theta^2 + \sin^2\!\theta\,\mathrm{d}\phi^2\right) . \qquad (268)$$

(We have not yet checked that $R_{tt} = 0$ and $R_{rr} = 0$ separately, but can check that this is the case by substituting (265) and (267) into (259) and (260)).

Finally we have to identify the constant $k$. Back in section 3.2.1, we found that for a static metric, $g_{00}$ was approximately given by (118), i.e.

$$g_{00} \sim \exp(2\Phi) ,$$

(where we have set the speed of light $c = 1$). In our case of a spherically symmetric distribution of mass, we know both the exact general relativistic $g_{00}$ and the Newtonian potential $\Phi$:

$$\Phi = -\frac{GM}{r} , \quad g_{00} = \left(1 + \frac{k}{r}\right)$$

Expanding out $\exp 2\Phi$ for large $r$,

$$\exp 2\Phi = \exp(-\frac{2GM}{r}) = 1 - \frac{2GM}{r} + O(\frac{1}{r})^2 ,$$

Comparing this with $g_{00}$, we find that $k = -2GM$, and that conventionally, we denote $GM$ by $m$ (so that $m$ is the mass in units in which $G$, Newton's constant, is 1).

This metric is known as the **Schwarzschild Metric**:

$$\boxed{\mathrm{d}s^2 = \left(1 - \tfrac{2m}{r}\right)\mathrm{d}t^2 - \frac{\mathrm{d}r^2}{\left(1 - \tfrac{2m}{r}\right)} - r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right) .} \tag{269}$$

This metric is the metric in the regions with no matter of a universe containing a single spherically symmetric mass distribution , e.g. a planet, a spherical galaxy, or (most famously) a black hole, i.e it is the metric outside the planet/galaxy/black hole.

We also expect that it is a good description of spacetime in the region of a single spherically symmetric mass distribution, far from any other masses – e.g. it should describe quite well the gravitational field around the sun, neglecting the influence of the planets.

We know that to extract the most information from this metric, we should solve the geodesic equations to find the paths of test particles and photons, and that is exactly what we shall do next.

However, we can start to understand the nature of this metric by looking at (a) the gravitational time-dilation (b) the gravitational red-shift, and (c) the pattern of light cones. We do these next.

## 5.3 Gravitational time dilation

Gravitational time-dilation is the effect whereby clocks close to massive bodies seem to run slow compared to those far away. This effect is directly related to the gravitational redshift as the time between successive wave-crests of a light ray changes as the light moves in a gravitational field (see sections 1.8 and 3.2).

However, let us work out the relation between proper time and coordinate time for static observers with $r > 2m$. The proper time interval $\mathrm{d}\tau$ along a path $x^\mu(\tau)$ is given (for the speed of light being 1) by

$$d\tau^2 = g_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} = g_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} ,$$

and hence for a stationary path $r, \theta, \phi$ constant, we have

$$\begin{aligned} d\tau &= \sqrt{g_{00}}\, dt = \sqrt{\left(1 - \tfrac{2m}{r}\right)}\, dt \\ \Rightarrow \quad (\tau - \tau_0) &= \int_{\tau_0}^{\tau} d\tau \\ &= \int_{t_0}^{t} \frac{d\tau}{dt}\, dt \\ &= \int_{t_0}^{t} \sqrt{g_{00}}\, dt \\ &= \int_{t_0}^{t} \sqrt{\left(1 - \tfrac{2m}{r}\right)}\, dt \\ &= (t - t_0) \sqrt{\left(1 - \tfrac{2m}{r}\right)} \end{aligned} \tag{270}$$

If we set all clocks to zero at $t = 0$, then we can work out (a) the proper time shown on a clock at coordinate time $t = 1$ and (b) the coordinate time at which each stationary clock shows proper time $\tau = 1$. This is shown in figure 32

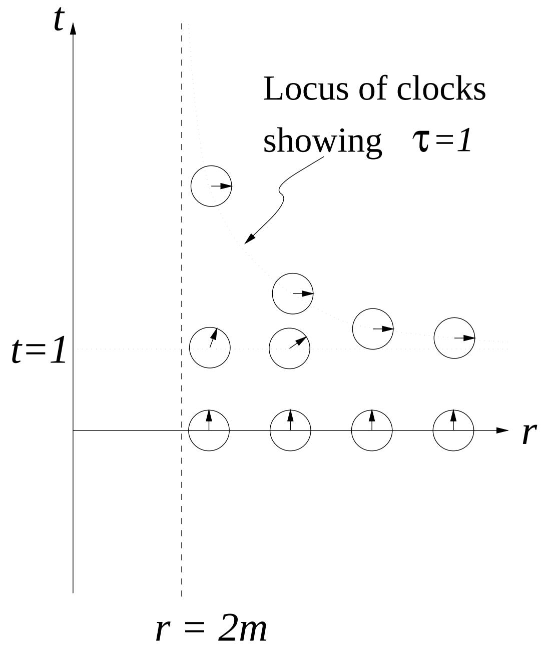



Figure 32:
Stationary clocks in the Schwarzschild metric

Clocks near $r = 2m$ appear to go slower and slower, until eventually time seems to come to a complete standstill at $r = 2m$.

However, on closer inspection, the 'path'

$$x^\mu(\lambda) = (\lambda, 2m, 0, 0) , \tag{271}$$

has tangent vector $\mathrm{d}x^\mu/\mathrm{d}\lambda$ of length zero:

$$\begin{aligned}\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda} &= (1,0,0,0) \\ g_{\mu\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda} &= g_{tt}.1.1 + g_{rr}.0.0 + g_{\theta\theta}.0.0 + g_{\phi\phi}.0.0 = (1-2m/r)|_{r=2m} = 0 \ ,\end{aligned}$$

and so the path (271) is in fact not a possible trajectory of a massive clock at all, but rather of a photon. This will be made clearer when we investigate the structure of light-cones next week.

## 5.4 Gravitational red-shift

The gravitational red-shift is also quite easy to work, since we have the general formula for a static metric in equation (116): if light of frequency $\nu_1$ is emitted at point $\boldsymbol{x}_1$, and received at point $\boldsymbol{x}_2$, then the frequency of the light received is given by

$$\nu_2 = \sqrt{\frac{g_{00}(\boldsymbol{x}_1)}{g_{00}(\boldsymbol{x}_2)}}\ \nu_1 \ . \tag{116}$$

For the Schwarzschild metric, $g_{00}$ depends only on $r$, so we have

$$\nu_2 = \sqrt{\frac{(1-2m/r_1)}{(1-2m/r_2)}}\ \nu_1 \ . \tag{272}$$

For $r_1, r_2 \gg 2m$, the effect is very small, but as soon as one of the two points approaches $r = 2m$, something very strange happens.

To examine what happens, let us consider two observers – one stationary at $r_1 \gg 2m$, and the other moving slowly radially from $r = r_1$ down to $r = 2m$, and both of them emitting light of a constant (as far as they are concerned) frequency $\nu$.

1. Let us first consider this from the point of view of the (slowly moving) observer, and let us denote the frequency of the light that the observer at $r$ observes by $\nu(r)$. This is given exactly by (272) as
$$\nu(r) = \sqrt{\frac{(1-2m/r_0)}{(1-2m/r)}}\ \nu \ .$$
As long as the observer is far from $2m$, the effect is small, with $\nu(r)$ approximately given in terms of the Newtonian potential $\Phi$
$$\nu(r) \sim \sqrt{1 + \frac{2m}{r} - \frac{2m}{r_0}}\ \nu \sim \left(1 + \frac{m}{r} - \frac{m}{r_0}\right) \nu = (1 - \Delta\Phi)\,\nu \ .$$
The effect of decreasing $r$ is to increase the frequency, so that the light emitted far away starts to get blue-shifted (the frequency of blue light is greater than that of red light).
As $r$ decreases close to $r = 2m$, the blue shift becomes very large until at $r = 2m$, the blue-shift diverges. One way to think about this is that the photons are 'falling' down the gravitational well and picking up energy, and the well is so steep that they reach $2m$ with infinite energy, as far as a static observer at $r = 2m$ would appreciate.
(This may seem paradoxical – how can a photon have infinite energy? The resolution is that it is not possible for any massive observer to be stationary at $r = 2m$.)

2. Now let us consider what the observer stationary at $r_0$ observes as the source moves slowly from $r_0$ down to $r = 2m$. Let us denote the frequency of the light that the observer at $r_0$ observes by $\tilde{\nu}(r)$. This is given exactly by (272) as
$$\tilde{\nu}(r) = \sqrt{\frac{(1-2m/r)}{(1-2m/r_0)}}\ \nu \ . \tag{273}$$

Again, as long as the observer is far from $2m$, the effect is small, with $\tilde{\nu}(r)$ approximately given in terms of the Newtonian potential $\Phi$

$$\tilde{\nu}(r) \sim \sqrt{1 - \frac{2m}{r} + \frac{2m}{r_0}}\ \nu \sim \left(1 - \frac{m}{r} + \frac{m}{r_0}\right) \nu = (1 - \Delta\Phi)\,\nu\ . \tag{274}$$

The effect of decreasing $r$ is to decrease the frequency, so that the light emitted close to the mass appears to be red-shifted to a far-away observer.

So, what happens for $r > 2m$ is that the frequency of the light starts to drop as $r$ approaches $2m$, and will pass from visible light to infra-red light, to micro-wave radiation, to radio waves, until eventually it is not detectable by any current technology.

However, the situation does not stop with the frequency just dropping: by (273), the frequency $\tilde{\nu}(r)$ becomes zero, at $r = 2m$ and *imaginary* for $r < 2m$.

How can the frequency become imaginary? The answer is that in fact *no* light from $r < 2m$ can ever reach an observer at $r_0 > 2m$, so the situation does not arrive, but we shall have to wait some time before seeing how this works out.

We can also ask whether an observer someone can actually cross $r = 2m$, and if so, how long does it take them to do so in coordinate time and in proper time. It turns out to be quite easy to show that it is *impossible* to cross $r = 2m$ in finite coordinate time, and only slightly harder to find out how long (in their own proper time) it takes for someone to fall through $r = 2m$.

## 5.5 The impossibility of crossing $r = 2m$ in finite coordinate time

The only equation we need here is that $\mathrm{d}s^2 > 0$ on a timelike curve. Since

$$\mathrm{d}s^2 = (1 - \frac{2m}{r})\mathrm{d}t^2 - \frac{\mathrm{d}r^2}{(1 - \frac{2m}{r})} - r^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2) > 0\ ,$$

this means that

$$\mathrm{d}t^2 > \frac{\mathrm{d}r^2}{(1 - 2m/r)^2}\ , \qquad \left|\frac{\mathrm{d}t}{\mathrm{d}r}\right| > \frac{\mathrm{d}r}{(1 - 2m/r)}$$

so that the coordinate time to move along any timelike path (geodesic or not) between $r_0$ and $r$ (with $r_0 > r$) must satisfy

$$\begin{aligned} \Delta t &= \int \mathrm{d}t = \int_r^{r_0} \left|\frac{\mathrm{d}t}{\mathrm{d}r}\right| \mathrm{d}r > \int_r^{r_0} \frac{\mathrm{d}r}{(1 - 2m/r)} = \int_{r^0}^{r_0} \mathrm{d}r \left(1 + \frac{2m}{r - 2m}\right) = [r + \log(r - 2m)]_r^{r_0} \\ &= (r_0 - r) + \log\left(\frac{r_0 - 2m}{r - 2m}\right)\ . \end{aligned} \tag{275}$$

From this expression we see that $\Delta t$ diverges as $r$ approaches $2m$, and that it is impossible to cross $r = 2m$ in finite coordinate time.

To find the proper time experienced by someone falling from $r_0$ to $r$, we will need to solve the geodesic equation for radial timelike geodesics, and so first we shall need to find the geodesic equations, which is the subject of the next section.

## 5.6 The geodesic equations in the Schwarzschild metric

Let's briefly recall how to find the geodesic equations from a variational principle.

1. To find the equations satisfied by a geodesic $x^\mu(\lambda)$, firstly construct

$$\mathcal{L} = g_{\mu\nu}\,\dot{x}^\mu\,\dot{x}^\nu\ , \tag{276}$$

   where $\dot{}\equiv \mathrm{d}/\mathrm{d}\lambda$.

2. Then, in terms of $\mathcal{L}$, the equations satisfied by a geodesic $x^\mu(\lambda)$ with $\lambda$ an affine parameter are

$$\begin{aligned} \frac{\partial\mathcal{L}}{\partial x^\mu} - \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{\partial\mathcal{L}}{\partial\dot{x}^\mu}\right) &= 0 \qquad \text{for } \mu=0,1,2,3 \\ \mathcal{L} &= \mathrm{C} = \begin{cases} 1 & \text{timelike geodesic, } \lambda{=}\tau\text{, the proper time} \\ 0 & \text{lightlike/null geodesic} \\ -1 & \text{spacelike geodesic, } \lambda \text{ the proper length} \end{cases} \end{aligned}$$

   The first four of these are the *Euler-Lagrange* equations for the variational principle based on the integral $\int\mathcal{L}\,\mathrm{d}\lambda$.

The first thing to note is that these five equations (one for each value of $\mu$, and then one for $\mathcal{L}$ being constant) are not independent. It is possible to solve any four of them and the fifth will then automatically be satisfied (up to a possible constant rescaling of the affine parameter $\lambda$).

We can now use apply this to the Schwarzschild metric (269). Working out $\mathcal{L}$, we get

$$\mathcal{L} = \left(1-\tfrac{2m}{r}\right)\dot{t}^2 - \frac{\dot{r}^2}{\left(1-\tfrac{2m}{r}\right)} - r^2\left(\dot{\theta}^2 + \sin^2\!\theta\,\dot{\phi}^2\right)\ . \tag{277}$$

and hence the first equation we have is

$$\left(1-\tfrac{2m}{r}\right)\dot{t}^2 - \frac{\dot{r}^2}{\left(1-\tfrac{2m}{r}\right)} - r^2\!\left(\dot{\theta}^2 + \sin^2\!\theta\,\dot{\phi}^2\right) = \mathrm{C} \qquad (= 0,\ 1 \text{ or } -1 \text{ as the case may be.}) \tag{278}$$

The remaining equations are the four Euler-Lagrange equations for $\mu = t, r, \theta\phi$ in turn:

$$-\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(2\!\left(1-\tfrac{2m}{r}\right)\dot{t}\right) = 0 \quad = \quad 0\ , \tag{279}$$

$$\frac{2m}{r^2}\left(\dot{t}^2 + \frac{\dot{r}^2}{\left(1-\tfrac{2m}{r}\right)^2}\right) - 2r\left(\dot{\theta}^2 + \sin^2\!\theta\,\dot{\phi}^2\right) + \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{2\dot{r}}{1-\tfrac{2m}{r}}\right) \quad = \quad 0\ , \tag{280}$$

$$-2r^2\sin\theta\,\cos\theta\,\dot{\phi}^2 + \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(2r^2\dot{\theta}\right) \quad = \quad 0\ , \tag{281}$$

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(2r^2\sin^2\!\theta\,\dot{\phi}\right) \quad = \quad 0\ . \tag{282}$$

Two of these can be integrated once directly, viz the equations for $t$ and $\phi$, which give

$$\left(1 - \tfrac{2m}{r}\right)\dot{t} \;=\; \kappa \qquad \text{constant}\,, \tag{283}$$

$$r^2\sin^2\!\theta\,\dot{\phi} \;=\; h \qquad \text{constant}\,. \tag{284}$$

These constants have very simple interpretations: $\kappa$ can be interpreted as the total energy (rest mass plus kinetic plus potential) of the particle, and $h$ as the angular momentum about the $\phi$ axis.

The next step depends on the details of the sort of motion in which we are interested – there are various tricks which are special to each situation and which make the calculations especially easy in that situation.

## 5.7 Planar motion

The most important result is that motion in a plane is possible in the first place – it is not necessarily obvious that a particle moving in a plane containing the massive body stays in that plane.

At any particular time $t_0$ a particle defines a plane which contains the vector from the particle to the origin, and the velocity of the particle. We *choose* coordinates so that this plane is the equatorial plane, $\theta = \pi/2$, and our initial conditions translate into

$$\theta(t_0) = \pi/2\,, \quad \dot{\theta}(t_0) = 0\,. \tag{285}$$

Now, rewriting eqn. (281) as an equation for $\theta$,

$$(2r^2)\,\ddot{\theta} + (4r\dot{r})\,\dot{\theta} - (r^2\dot{\phi}^2)\sin(2\theta) = 0\,, \tag{286}$$

we see that $\theta = \pi/2$ is a solution of this equation, and hence

$$\theta(t) = \pi/2 = \text{constant}\,, \tag{287}$$

for all time, which is the equation of the equatorial plane.

We have two remaining equations, (278) and (280), which are equivalent, and of these it is clear that (278) is simpler. Substituting (283), (284) and (287) into (278) we get

$$\frac{\kappa^2 - \dot{r}^2}{\left(1 - \frac{2m}{r}\right)} - \frac{h^2}{r^2} = \mathrm{C} = \begin{cases} 1 & \text{Timelike geodesics, } \lambda \text{ proper time} \\ 0 & \text{Lightlike geodesics} \\ -1 & \text{Spacelike geodesics, } \lambda \text{ proper distance} \end{cases}\,. \tag{288}$$

The discussion of planar motion can be now be divided into two parts – purely radial motion, for which the angular momentum $h$ is zero, and general planar motion for which $h$ is non-zero.

The difference is mainly technical – if $h$ is zero we shall be able to find the trajectory $x^\mu(\tau)$ as a function of the proper time.

However, if $h$ is non-zero, the orbit is generally too complicated for us to be able to do that, and instead we shall only be able to find geometric properties of the orbit. However these are all we need to know to find the formulae for the precession of the perihelion and of the gravitational deflection of light.

## 5.8 Planar motion and orbits

We now turn to the question of planar motion with non-zero angular momentum.

Let's first recall that, in the case of motion in the equatorial plane, the geodesic equations reduce to equations (285), (283), (284) and (288):

$$\begin{aligned}\theta &= \pi/2\,, \\ \left(1-\tfrac{2m}{r}\right)\dot{t} &= \kappa\,, \qquad \text{constant}\,, \\ r^2\sin^2\theta\,\dot{\phi} &= h\,, \qquad \text{constant}\,, \\ \frac{\kappa^2-\dot{r}^2}{\left(1-\frac{2m}{r}\right)}-\frac{h^2}{r^2} &= \mathrm{C} = \begin{cases}1 & \text{timelike geodesics} \\ 0 & \text{lightlike geodesics}\end{cases}\,,\end{aligned}$$

where $\dot{}\equiv \mathrm{d}/\mathrm{d}\tau$ for timelike geodesics and $\dot{}\equiv \mathrm{d}/\mathrm{d}\lambda$ for some affine parameter $\lambda$ for lightlike geodesics.

For some situations, we are only interested in the spatial path $r(\phi)$ of the geodesic, and the time or proper time dependence is not necessary – examples of such problems are the precession of the perihelion (how much per orbit?) and the deflection of light (what angle is the orbit bent through?)

In these cases, we are better off replacing the two equations (284) and (288) for $\dot{r}$ and $\dot{\phi}$, by a single equation for $\mathrm{d}r/\mathrm{d}\phi$. To do this we write (for $\theta=\pi/2$)

$$\dot{r}=\frac{\mathrm{d}}{\mathrm{d}\lambda}r(\phi)=\frac{\mathrm{d}\phi}{\mathrm{d}\lambda}\frac{\mathrm{d}r}{\mathrm{d}\phi}=\dot{\phi}\,r'=\frac{h}{r^2}\,r'\,, \tag{289}$$

where $'\equiv \mathrm{d}/\mathrm{d}\phi$.

The final simplification is that the equations turn out to be much simpler when expressed in terms of $u=1/r$ than in terms of $r$ – this is exactly the same as for Newtonian orbits.

In terms of $u$,

$$r'=\frac{\mathrm{d}}{\mathrm{d}\phi}\left(\frac{1}{u}\right)=-\frac{u'}{u^2}\,, \quad \Rightarrow \quad \dot{r}=\frac{h}{r^2}r'=hu^2r'=-hu'\,.$$

Hence eqn. (288) now reads

$$\frac{\kappa^2-(hu')^2}{1-2mu}-h^2u^2=c\,,$$

or, introducing a new notation $F(u)$,

$$\boxed{(u')^2=\left(\frac{\kappa^2-c}{h^2}\right)+\left(\frac{2m\mathrm{C}}{h^2}\right)u-u^2+2mu^3\equiv F(u)\,.} \tag{290}$$

Sometimes it is better to work with this equation for $(u')^2$, and sometimes it is better to differentiate this equation once to find an equation for $u''$,

$$u''=u'\frac{\mathrm{d}u'}{\mathrm{d}u}=\frac{1}{2}\frac{\mathrm{d}(u')^2}{\mathrm{d}u}=\frac{1}{2}\frac{\mathrm{d}F}{\mathrm{d}u}\,,$$

which becomes

$$\boxed{u''=\left(\frac{m\mathrm{C}}{h^2}\right)-u+3mu^2\,.} \tag{291}$$

## 5.9 Closed orbits

We can now examine the solutions for various closed orbits. We have the exact equation (290) for $u'$,

$$(u')^2 = \left(\frac{\kappa^2 - \mathrm{C}}{h^2}\right) \;+\; \left(\frac{2m\mathrm{C}}{h^2}\right) u \;-\; u^2 \;+\; 2mu^3 \;\;\equiv\;\; F(u) \;, \tag{292}$$

and this can be solved exactly to find $u(\phi)$, or rather $\phi(u)$ in terms of elliptic integrals –

$$\phi = \int_0^\phi \mathrm{d}\phi = \int_{u_0}^u \frac{\mathrm{d}u}{u'} = \int_{u_0}^u \frac{\mathrm{d}u}{\sqrt{F(u)}} \;,$$

– however this technique is beyond the scope of this course, although we shall explore some of the simple deductions one can make from eqn. (290) in the exercises.

Since we shall not solve eqn. (290) exactly, we shall have to settle for approximate solutions instead. It is fortunate for us that even these approximate solutions are in excellent agreement with the experimental data, and show that Einstein's theory does indeed describe the world much better than Newton's.

Also, to find approximate solutions, it is much better for us to find approximate solutions to the second order differential equation (291)

$$u'' = \left(\frac{m\mathrm{C}}{h^2}\right) \;-\; u \;+\; 3mu^2 \;. \tag{291}$$

than the first order equation (290)

It is a fact (as discussed in the exercises) that the equations for the Newtonian orbits $u(\phi)$ are identical to eqn. (291) but with the term $+3mu^2$ dropped.

Since we expect the deviation from the Newtonian orbit to be small, we can write

$$u = u_0 + \delta u \;, \tag{293}$$

where $u_0$ is the solution to the Newtonian equation and $\delta u$ is a perturbation which we hope remains small. Substituting this into (291), we find

$$u_0'' + \delta u'' = -u_0 - \delta u + 3m(u_0^2 + 2u_0\delta u + \delta u^2) \;,$$

and using the fact that $u_0$ is a solution of the Newtonian equations, we get

$$\delta u'' + \delta u = 3m(u_0^2 + 2u_0\delta u + \delta u^2) \;, \tag{294}$$

which is actually independent of the constant $c$.

This is an *exact* equation for $\delta u$, but we shall make the approximation that $\delta u$ is small, and so drop all but the first term on the right-hand side of eqn. (294), leaving

$$\delta u'' + \delta u = 3mu_0^2 \;, \tag{295}$$

as the solution for our first approximation to $\delta u$. We now have to choose which solution $u_0$ we want our answer $u = u_0 + \delta u + \ldots$ to be close to. Since we want to find the ways in which the solutions to Einstein's equations differ from those to Newton's, we shall start by solving for the Newtonian orbits $u(\phi)$.

## 5.10 The Newtonian orbits

We start from the second order linear equation (see the exercises)

$$u'' = \left(\frac{m\mathrm{C}}{h^2}\right) - u \,, \tag{296}$$

where C = 1 for the orbit of a massive body and C = 0 for the orbit of a massless body such as a photon.

Since this is such a simple a linear equation, we can solve it easily, to find

$$u = \left(\frac{m\mathrm{C}}{h^2}\right) + A\cos\phi + B\sin\phi \;.$$

Let's now consider separately the two cases of $C = 0$ and $C = 1$

If $C = 0$, then

$$u = A\cos\phi + B\sin\phi = C\cos(\phi - \phi_0) \;,$$

and hence the equation for $r$ is

$$r\cos(\phi - \phi_0) = R \,, \;\; \text{constant} \;, \tag{297}$$

which is the equation of a straight line which passes a distance $R$ from the origin, as figure 33 makes clear.

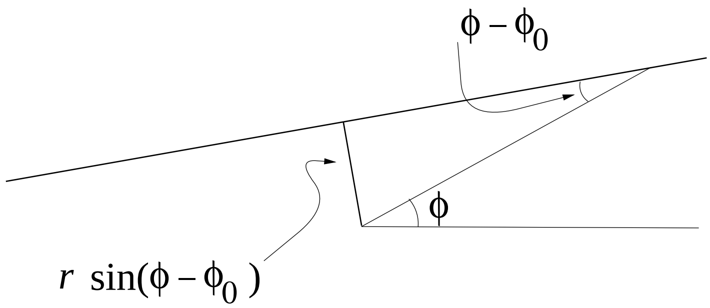



Figure 33:

If C = 1, then we can rewrite the solution for $u$ as

$$u = \frac{m}{h^2}\left(1 + e\cos(\phi - \phi_0)\right) \;, \tag{298}$$

which is the equation of a conic. $e$ is called the *eccentricity*, and the type of conic depends on the value of $e$ as in table 1

| conic | circle | ellipse | parabola | hyperbola |
|---|---|---|---|---|
| $e$ | 0 | $0 < e < 1$ | 1 | $1 < e$ |

Table 1: the eccentricities of the conics

We shall work out the corrections to the trajectories $u(\phi)$ for two cases – the deflection of light and for the precession of the perihelion of the orbit of a massive body.

## 5.11 The deflection of light

In the first case, we are interested in the corrections to the Newtonian path of the solutions of the equation (291) with C = 0, i.e.

$$u'' = -u + 3mu^2 \,. \tag{299}$$

The Newtonian limit of this equation is simply

$$u'' + u = 0 \,,$$

for which we shall take as solution

$$u_0 = \frac{\sin\phi}{R} \,, \tag{300}$$

i.e. the case $\phi_0 = \pi/2$ of the generic solution (297). In two dimension space with $x = r\cos(\phi), y = r\sin(\phi)$, this corresponds to the line $y = R$, as shown in figure 34

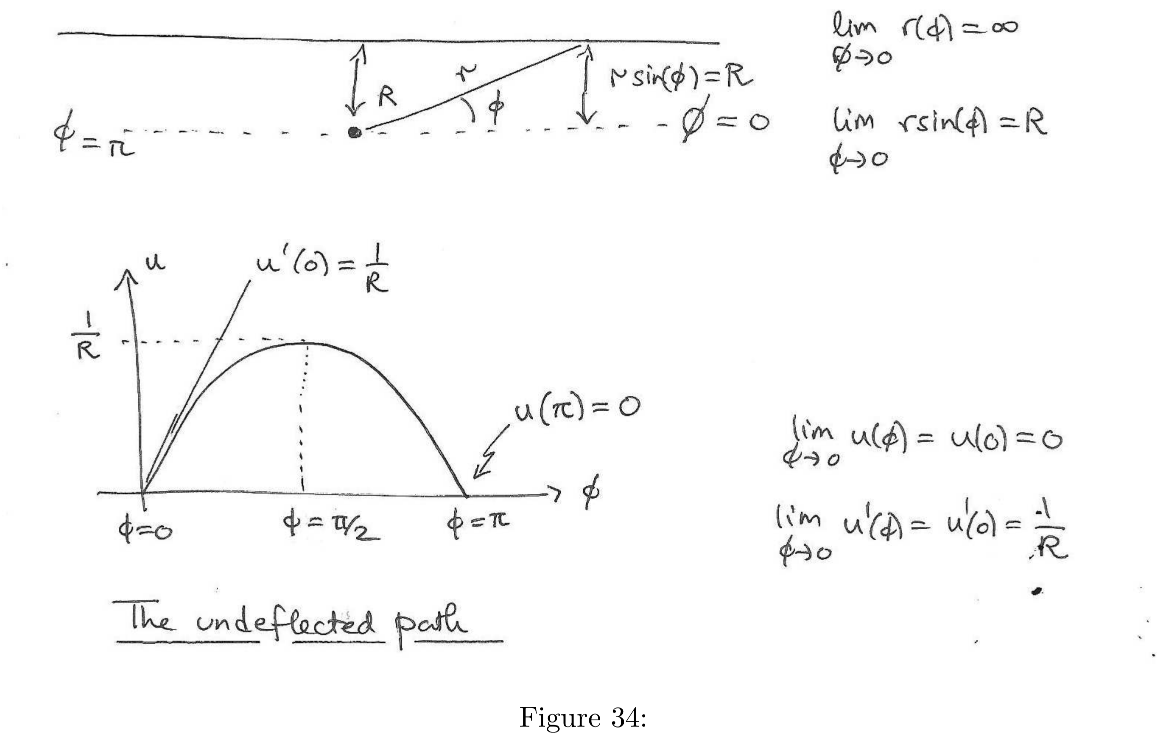



Figure 34:

Then we substitute this into (295) and get

$$\delta u'' + \delta u = 3mu_0^2 = \frac{3m}{R^2}\sin^2\phi = \frac{3m}{2R^2}\left(1 - \cos(2\phi)\,\right) \,. \tag{301}$$

By inspection, the solution to eqn. (301) should be of the form

$$\delta u = D\cos\phi + E\sin\phi + F + G\cos(2\phi) + H\sin(2\phi)\,. \tag{302}$$

We can fix $D$ and $E$ by requiring that the new solution still approximates the undeflected path as $\phi \to 0$. This gives the boundary conditions

$$u(0) = 0\;,\quad \lim_{\phi\to 0}\frac{u(\phi)}{\sin(\phi)} = u'(0) = 1/R\,.$$

In terms of the perturbation, we then just have to make the perturbation zero as $\theta \to 0$,

$$\delta u(0) = \delta u'(0) = 0\,.$$

Solving these conditions, we get

$$\delta u = \frac{m}{2R^2}\left(3 + \cos(2\phi) - 4\cos(\phi)\,\right)\;, \tag{303}$$

giving us the total answer

$$u = u_0 + \delta u = \frac{1}{R}\left(\sin\phi + \frac{m}{2R}(3 + \cos(2\phi) - 4\cos(\phi)\,)\,\right)\,. \tag{304}$$

What we now want to now is by how much the photon's trajectory has been deflected. The undeflected photon trajectory came in from infinity at $r = \infty$, i.e. $u = 0$, at $\phi = 0$ and went off to infinity at $\phi = \pi$.

By our choice of boundary condition $u(0) = 0$, The deflected photon trajectory still comes in from infinity at $\phi = 0$ but now it no longer goes off to infinity at $\phi = \pi/2$ as $\delta u(\pi/2) \neq 0$. Instead, it goes off to infinity at a slightly larger angel, $\phi = \pi + \Delta$, where $\Delta$ is the deflection of the path and which we assume is small This is shown in figure 35. To find the value of $\Delta$, we solve for $u(\pi + \Delta) = u_0(\pi + \Delta) + \delta u(\pi + \Delta) = 0$.

To find the value of $\Delta$, we can expand the right-hand side of (304)

$$\sin(\pi + \Delta) \sim -\Delta\;,\quad \cos(\pi + \Delta) \sim 1\;,\quad \cos(2\pi + 2\Delta) \sim 1\;,$$

we get

$$0 \sim \frac{1}{R}\left(-\Delta + \frac{m}{2R}(1 + 3 + 4\,)\,\right) = \frac{1}{R}\left(-\Delta + \frac{4m}{R}\right);. \tag{305}$$

This has the solution $\Delta = 4m/R$, so the total defection is $4m/R$.

We can now compare this with the results for light passing the Sun. The deflection of light past the Sun is most noticeable for photons which just scrape past the edge – and these can be seen only during an eclipse, when the light from the rest of the sun is blocked out by the moon. However, the moon is just the right size for these photons to get to us, and their deflection can be measured, and it is approximately $1.75''$ (seconds of arc) where 1 second of arc is $2\pi/(360^2)$ radians.

The data we need to work out the value of our prediction are the radius of the Sun (since this we are interested in photons which just miss the Sun), which is $R_{\rm Sun} = 6.96\;10^5$km, and the mass of the Sun in the correct units – since $(2m)$ is the Schwarzschild radius of the sun, we find that the correct units are kilometres, and the mass of the sun is $2m_{\rm Sun} = 2.95$km. Putting these values in gives a deflection of $1''.74$ seconds, in very good agreement for the approximate calculation we have made.

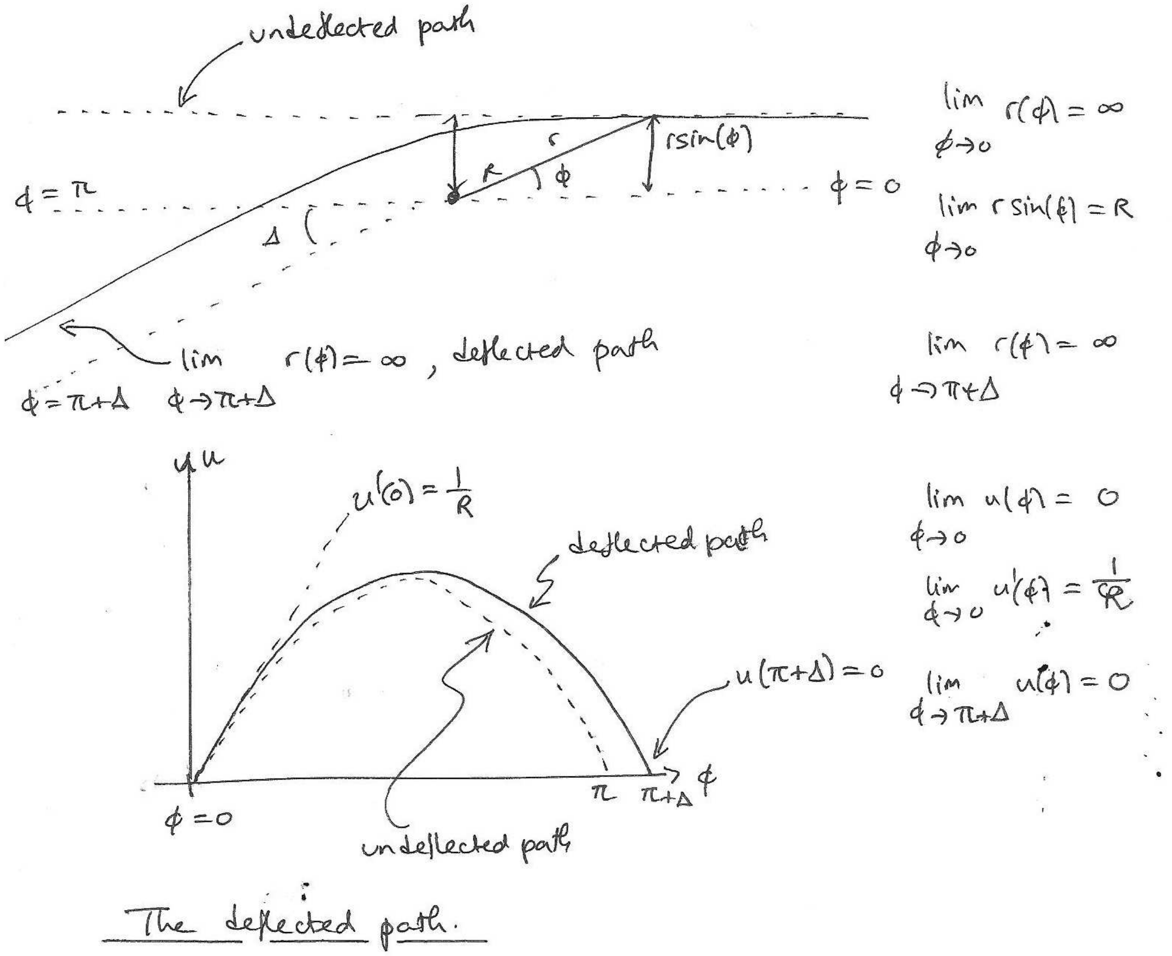



Figure 35:

### 5.11.1 Systematic expansion of the deflection angle

This previous calculation gives the leading term in the expansion of the deflection angle as a power series in $\epsilon = m/R$. It is possible to make a systematic expansion and so find a full power series for the deflection angle $\Delta = 4\epsilon + \ldots$. This is outside the course, but is included as an appendix after the exercises.

### 5.12 The precession of the perihelion

To work out the precession of the perihelion, we apply exactly the same method to the Newtonian orbit

$$u_0 = \frac{m}{h^2}(1 + e\cos\phi) \ , \tag{306}$$

which has boundary conditions fixing the point of closest approach (perigelion) to be at $\phi = 0$ and having a fixed value:

$$u_0(0) = \frac{m}{h^2}(1 + e) \ , \ \ u_0'(0) = 0 \ , \tag{307}$$

The differential equation for the full solution $u$ (291) with C = 1 is

$$u'' + u = \frac{m}{h^2} + 3mu^2 \ , \tag{308}$$

and we can choose the same boundary conditions, ie

$$u(0) = \frac{m}{h^2}(1 + e) \ , \ \ u'(0) = 0 \ , \tag{309}$$

So, setting $u = u_0 + \delta u$, we have eqn. (295) as the approximate equation for $\delta u$, that is

$$\begin{aligned} \delta u'' + \delta u &= 3mu_0^2 \\ &= \frac{3m^3}{h^4}\left(1 + 2e\cos\phi + e^2\cos^2\phi\right) \\ &= \frac{3m^3}{h^4}\left(1 + 2e\cos\phi + \frac{e^2}{2}(1 + \cos(2\phi))\right) \ , \end{aligned} \tag{310}$$

and the boundary conditions for $\delta u$ are

$$\delta u(0) = 0 \ , \ \ \delta u'(0) = 0 \ . \tag{311}$$

Guessing that $\delta u$ will be of the form

$$\delta u = A + B\phi + C\cos\phi + D\sin\phi + E\cos(2\phi) + F\sin(2\phi) + G\phi\sin\phi + H\phi\cos\phi \ , \tag{312}$$

and substituting this into eqn, (310), we find the solution

$$\delta u = \frac{m^3}{2h^4}\left(6 + 3e^2 - (6 + 2e^2)\cos(\phi) - e^2\cos(2\phi) + 6e\phi\sin(\phi)\right) \ . \tag{313}$$

What happens is that orbit no longer returns to the periheilon ($r = r_{min}, u = u_{max}$) at $\phi = 0$, so the solution no longer has period $T_0 = 2\pi$, but instead returns at a slightly different value, $\phi = \Delta$ correspomnding to a slightly longer period $T = 2\pi + \Delta$. This is shown on figure 36.

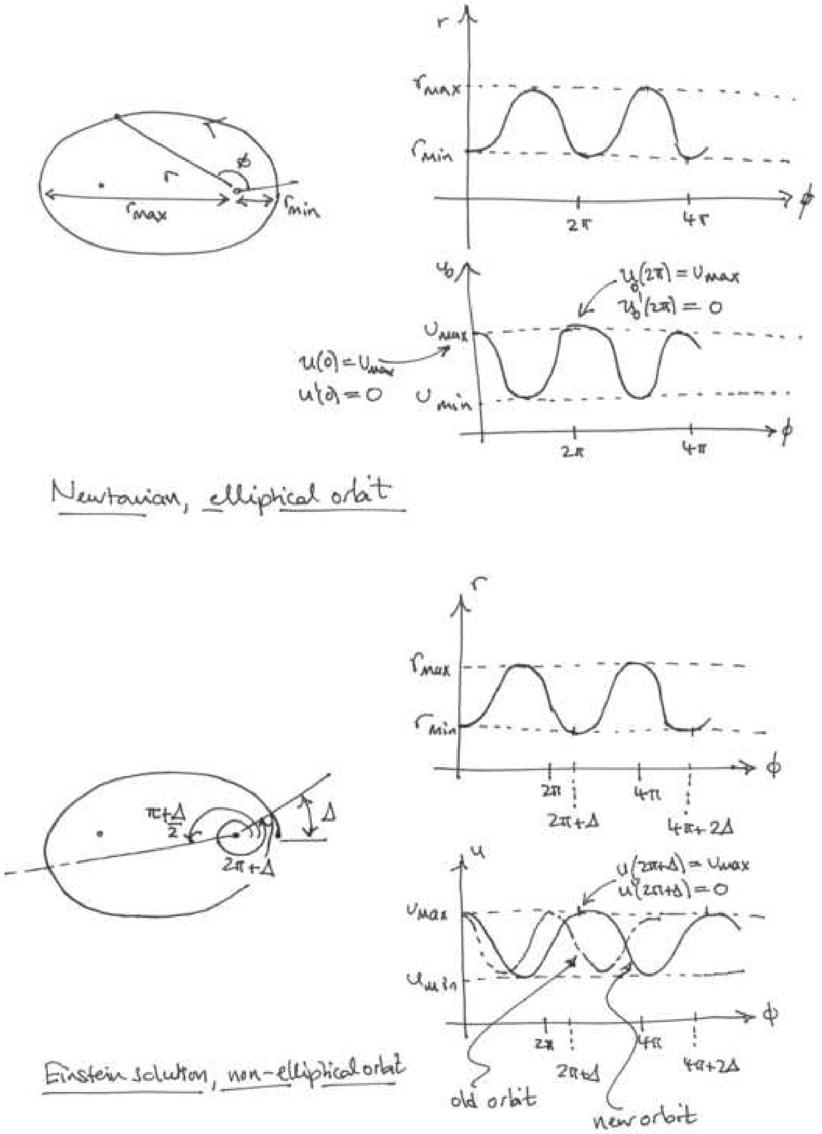

Newtonian, elliptical orbit
Einstein solution, non-elliptical orbit
old orbit
new orbit

Figure 36:

We can find $\Delta$ by requiring *either* of these two equivalent conditions -

$$u(2\pi+\Delta) = \frac{m}{h^2}(1+e)\ ,\ u'(2\pi+\Delta) = 0\ . \tag{314}$$

It turns out that it is practically more straightforward to solve the second equation. So, we have, making a Taylor expansion in $\Delta$,

$$\begin{aligned} 0 &= u'(2\pi+\Delta) = u_0'(2\pi+\Delta) + \delta u'(2\pi+\Delta) \\ &\approx u_0'(2\pi) + \Delta u_0''(2\pi) + \delta u'(2\pi) \\ &= -e\frac{m}{h^2}\Delta\ +\ 6e\pi\frac{m^3}{h^4} \\ &= e\frac{m}{h^2}(6\pi\frac{m^2}{h^2} - \Delta) \end{aligned} \tag{315}$$

so that

$$\Delta \approx 6\pi\frac{m^2}{h^2}\ . \tag{316}$$

Using Newtonian relations between $h$, $m$, $e$ and $a$, the *semi-major axis* of the elliptical orbit, this gives

$$\Delta\phi = \frac{6\pi m}{a(1-e^2)}\ . \tag{317}$$

For Mercury, we have $2m = 2.9$km (as before), $a = 5.8\ 10^{10}m$, and $e = 0.2$, and the period of the orbit is 88 days, giving (with this data) a total precession of $42''$ seconds in 100 years whereas the true answer is $43''$ per 100 years. Again the agreement using this simple approximation is very good.

Now, we can in fact see something strange – since there was a $\cos\phi$ term on the right-hand side of eqn. (310), and this is the resonant frequency of the differential operator acting on $\delta u$, we find a term in $\delta u$ which grows without bound – namely $(3m^3e/h^4)\phi\sin\phi$. So after some time our small perturbation $\delta u$ is no longer small, but large.

There is one way to solve this paradox, and that is to recognise that this term corresponds to a change in the period of the original solution, and that to leading order in $\phi$,

$$u = \frac{m}{h^2}\left(1 + e\cos\phi + \frac{3m^2e}{h^2}\phi\sin\phi + \dots\right) = \frac{m}{h^2}\left(1 + e\cos(\left[1 - \frac{3m^2}{h^2}\right]\phi) + \dots + \dots\right)\ , \tag{318}$$

We can now ask what the new period of the orbit is – by how much must $\phi$ change for $u$ to come back to the same value? If the change in $\phi$ on a whole orbit is $2\pi + \Delta\phi$, we need

$$\left[1 - \frac{3m^2}{h^2}\right](2\pi + \Delta\phi) = 2\pi \quad\Rightarrow\quad 2\pi + \Delta\phi = 2\pi + 2\pi\frac{3m^2}{h^2} + \dots\ . \tag{319}$$

Thus the extra amount through $\phi$ changes, the amount by which the orbit precesses is exactly what we found before:

$$\Delta\phi = \frac{6\pi m^2}{h^2}\ . \tag{320}$$

## 5.13 Radial geodesics

We can now go back to our original problem, which was to find out if it is possible for someone to 'fall freely' through $r = 2m$ in finite proper time. For this, we need to find the radial timelike geodesics and then find out if the proper time to reach $r = 2m$ is finite or not. For such geodesics, we can solve for $r(t)$ directly.

Going back, we recall eqns (283) and (284)

$$\dot{t} = \frac{\kappa}{1 - \frac{2m}{r}} \;, \quad \dot{\phi} = \frac{h}{r^2 \sin^2\theta}$$

If $\dot{\phi}$ is zero, then $h$ is zero, and so eqn (288) becomes

$$\frac{\kappa^2 - \dot{r}^2}{(1 - \frac{2m}{r})} = \mathrm{C}\,. \tag{321}$$

We can rewrite this as an eqn for $\dot{r} = \frac{\mathrm{d}r}{\mathrm{d}\lambda}$

$$\dot{r} = \pm\sqrt{\kappa^2 - \mathrm{C}\left(1 - \frac{2m}{r})\right)}\,, \tag{322}$$

and using (283), as an eqn for $\frac{\mathrm{d}r}{\mathrm{d}t} = \dot{r}/\dot{t}$ ,

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \pm\frac{1}{\kappa}\left(1 - \frac{2m}{r}\right)\sqrt{\kappa^2 - \mathrm{C}\left(1 - \frac{2m}{r}\right)}\,. \tag{323}$$

We can integrate (322) and (323) to find $\lambda(r)$ and $t(r)$, as

$$\lambda - \lambda_0 = \int_{\lambda_0}^{\lambda} d\lambda = \int_{r_0}^{r} \frac{\mathrm{d}r}{\dot{r}} = \int_{r_0}^{r} \frac{\pm\mathrm{d}r}{\sqrt{\kappa^2 - \mathrm{C}\left(1 - \frac{2m}{r}\right)}}\,, \tag{324}$$

$$t - t_0 = \int_{t_0}^{t} \mathrm{d}t = \int_{r_0}^{r} \frac{\mathrm{d}r}{\mathrm{d}r/\mathrm{d}t} = \int_{r_0}^{r} \frac{\pm\kappa\,\mathrm{d}r}{\left(1 - \frac{2m}{r}\right)\sqrt{\kappa^2 - \mathrm{C}\left(1 - \frac{2m}{r}\right)}}\,. \tag{325}$$

These integrals are not especially easy, but there are two simplifications we can make

1. In the case of radial photons, i.e. null geodesics, C = 0 and

$$\lambda - \lambda_0 = \int_{r_0}^{r} \pm\frac{\mathrm{d}r}{\kappa} = \pm\frac{(r-r_0)}{\kappa}\,, \tag{326}$$

$$t - t_0 = \int_{r_0}^{r} \pm\frac{\mathrm{d}r}{\left(1-\frac{2m}{r}\right)} = \pm\left[(r-r_0) + 2m\log\left|\frac{r-2m}{r_0-2m}\right|\right]\,. \tag{327}$$

2. In the case of massive particles, where C = 1, and $\lambda = \tau$ the proper time, we can make the assumption that the particle is *stationary at* $r = \infty$. From (321),

$$\dot{r}^2 = \kappa^2 - \left(1 - \frac{2m}{r}\right)$$

so

$$\dot{r}^2\big|_{r=\infty} = \kappa^2 - 1 = 0 \Rightarrow \kappa = 1$$

Taking C = 1, $\kappa = 1$ in (324) and (325), we find

$$\tau - \tau_0 = \int_{r_0}^{r} \pm\frac{\mathrm{d}r}{\sqrt{2m/r}} = \pm\frac{2}{3\sqrt{2m}}\left[r^{3/2} - {r_0}^{3/2}\right] \tag{328}$$

$$t - t_0 = \int_{r_0}^{r} \frac{\pm\mathrm{d}r}{\left(1-\frac{2m}{r}\right)\sqrt{\frac{2m}{r}}} = \pm\left[\frac{2}{3}\frac{r^{\frac{3}{2}} + 6mr^{\frac{1}{2}}}{\sqrt{2m}} + 2m\ln\left|\frac{\sqrt{r}+\sqrt{2m}}{\sqrt{r}-\sqrt{2m}}\right|\right]_{r_0}^{r} \tag{329}$$

(You needn't worry too much about this last expression, (329), it is just given here for completeness.)

In both cases, {C = 0} and {C = 1, $\kappa = 1$}, we see that $\lambda$ is finite as $r \to 2m$, but $t$ is infinite as $r \to 2m$.

So, although a photon or a massive particle may cross $r = 2m$ in finite affine parameter/ proper time, it never crosses $r = 2m$ in finite coordinate time $t$.

### 5.13.1 Radial lightlike geodesics

The lightlike radial geodesics can in fact be found without solving the geodesic equations. In general a lightlike path satisfies

$$\dot{x}^\mu \, \dot{x}^\nu \, g_{\mu\nu} = 0 \; .$$

For a radial path, $\theta$ and $\phi$ are constant so $\dot{\theta} = \dot{\phi} = 0$, and substituting these into the metric (269), we have

$$0 = \left(1 - \tfrac{2m}{r}\right) \dot{t}^2 - \frac{\dot{r}^2}{\left(1 - \tfrac{2m}{r}\right)} \; , \tag{330}$$

or

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \pm\left(1 - \tfrac{2m}{r}\right) \; . \tag{331}$$

The sign in eqn. (331) depends on whether the photon is out-moving or in-moving (does $r$ increase or decrease with $t$?) so that we take the '$-$' sign for in-moving and the '$+$' sign for out-moving. Integrating this equation, we have

$$\begin{aligned}
(t - t_0) &= \int_{t_0}^{t} \mathrm{d}t = \int_{r_0}^{r} \frac{\mathrm{d}r}{\dot{r}} = \int_{r_0}^{r} \frac{\mathrm{d}r}{\pm\left(1 - \frac{2m}{r}\right)} = \pm \int_{r_0}^{r} \mathrm{d}r \left(1 + \frac{2m}{r - 2m}\right) \\
&= \pm \Big[ r + 2m \log \big| r - 2m \big| \Big]_{r_0}^{r} \\
&= \pm \left( (r - r_0) + 2m \log \left| \frac{r - 2m}{r_0 - 2m} \right| \right) \; .
\end{aligned} \tag{332}$$

For $r \gg 2m$, we can ignore the log term as $\log r \ll r$ for large $r$, and so far from the massive body, photons follow straight lines

$$(t - t_0) \sim \pm(r - r_0) \; .$$

However, as $r$ gets closer to $2m$, the log term becomes more important, and we see that it is impossible for $r$ to reach $2m$, as this would imply $t \to \infty$. So, putting these pieces together, we can draw the paths followed by radial light rays for $r > 2m$:

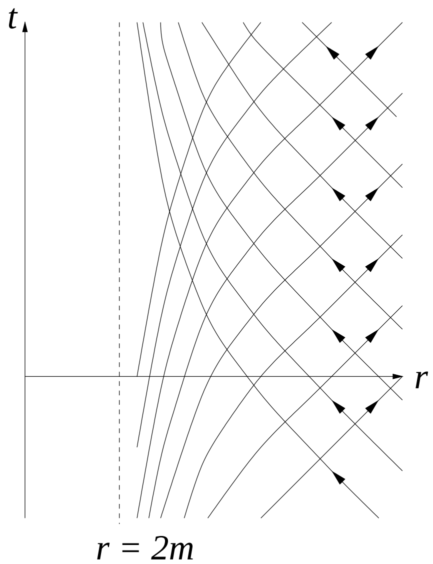



Figure 37:
Pattern of radial light rays for $r > 2m$

For large $r$ the light rays move at $\sim$45°, but for $r$ close to $2m$ they are 'squashed' together so that they never cross $r = 2m$ in finite time coordinate $t$.

## 5.14 Going beyond $r = 2m$ - Eddington Coordinates

Clearly we should use some new coordinates to investigate this region $r < 2m$, but what can we use? One method is to look at the physical problem - particles and photons falling past $r = 2m$ - and try to use that.

Consider some light rays moving radially inward. We can label any point by the values $(r, t)$ - but equally we can label the points by the value of $t$ at which the photon moving through that point crossed $r = r_0$.

A physical way to think of this is that each point $r < r_0$ is labelled by the coordinate time shown on a (non-standard) clock at $r = r_0$. What coordinate time will we see at any given point? - the time showing when the photons left $r = r_0$.

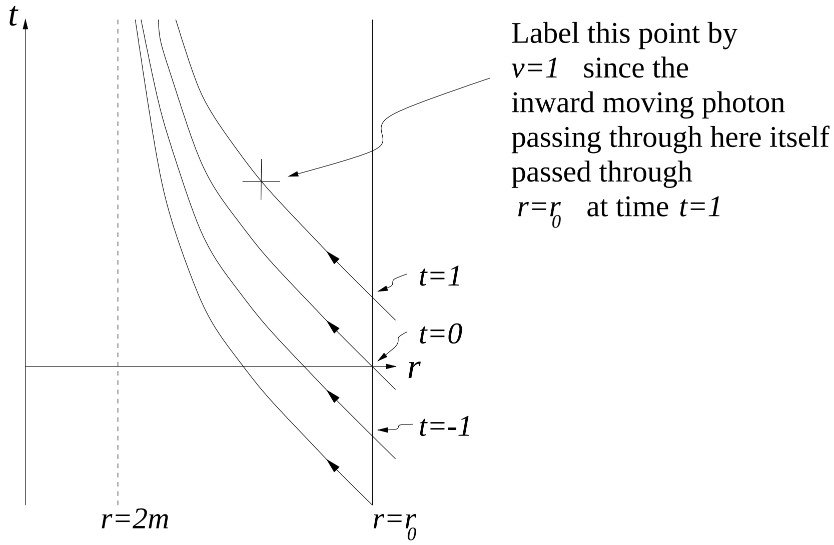



Figure 38:
The 'physical' definition of inward-moving Eddington coordinates in terms of inward moving photons

From eqn. (327) we know the relation between our old coordinates $(r, t)$ and our new labels - let us call them ($r$ and $v$).

If the inward-moving photon at $(r, t)$ passed through $r_0$ at time $t_0 = v$ then

$$(t - v) = -\left[(r - r_0) + 2m \ln\left|\frac{r - 2m}{r_0 - 2m}\right|\right] , \tag{333}$$

so that

$$t = v - \left[(r - r_0) + 2m \ln\left|\frac{r - 2m}{r_0 - 2m}\right|\right] , \tag{334}$$

and hence

$$\mathrm{d}t \quad = \quad \mathrm{d}v - \left(\mathrm{d}r + \frac{2m\,\mathrm{d}r}{r - 2m}\right) = \mathrm{d}v - \frac{\mathrm{d}r}{(1 - 2m/r)} . \tag{335}$$

To find the form of the metric in coordinates $(v, r, \theta, \phi)$ from the form (269) in coordinates $(t, r, \theta, \phi)$, we just have to substitute (335) into (269) to find

$$ds^2 = \left(1 - \frac{2m}{r}\right) dv^2 - 2\, dv\, dr - r^2(d\theta^2 + \sin^2\theta d\phi^2) \ . \tag{336}$$

We now see that (miraculously) the metric is no longer singular at $r = 2m$, and we can happily consider $r$ varying all the way down to zero.

We can also now draw patterns of light cones in the $(v - r)$ plane. It makes sense to have the inward falling light-rays at 45° to the $r$-axis, so we choose the vertical axis to be $(v - r)$. We find that the pattern of light rays moving in the $(v, r)$ plane is

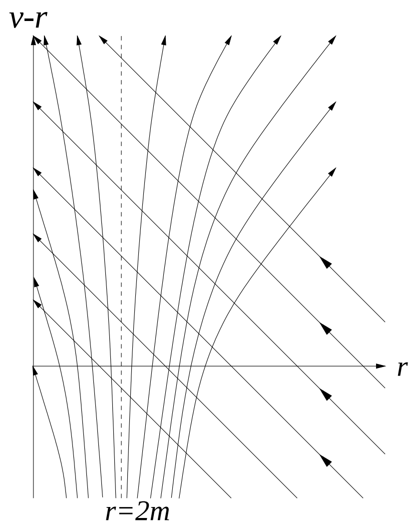



Figure 39:
The pattern of radial lightlike geodesics in coordinates $(v, r, \theta, \phi)$.

The coordinate $v$ is called an "Inward-moving Eddington coordinate", and with this coordinate we easily see from figure 40 that no light-rays with $r < 2m$ can ever escape, and all massive particles moving in $0 < r < 2m$ must hit $r = 0$.

How can it be that one cannot avoid hitting $r = 0$? To understand this, it is useful to look again at the Schwarzschild metric for $0 < r < 2m$. Now $(1 - 2m/r)$ is negative and $(2m/r - 1)$ positive, so if we want to write our metric in the form

$$ds^2 = +(\cdots)(dx^0)^2 - (\cdots)(dx^1)^2 - (\cdots)(dx^2)^2 - (\cdots)(dx^3)^2 \ ,$$

where each the of the terms $(\cdots)$ is positive, to show that the signature is explicitly $(+ - - -)$ and $x^0$ is a timelike coordinate while $x^1$, $x^2$ and $x^3$ are spacelike coordinates, we should find it natural to put the metric in the form

$$ds^2 = \frac{dr^2}{(2m/r - 1)} - (\frac{2m}{r} - 1)\, dt^2 - r^2 \left(d\theta^2 + \sin^2\theta\, d\phi^2\right) \ , \tag{337}$$

i.e. now $r$ is a timelike coordinate and $t$ is a spacelike coordinate. So, since $r$ is a timelike coordinate, $r = 0$ is like a fixed *time* in the future, and we all know that one cannot avoid the future!

To make this clearer, we can look at the pattern of 'radial' lightlike geodesics for $0 < r < 2m$, and with the correct interpretation of $r$ as a time coordinate and $t$ as a space coordinate, we find figure 5.14:

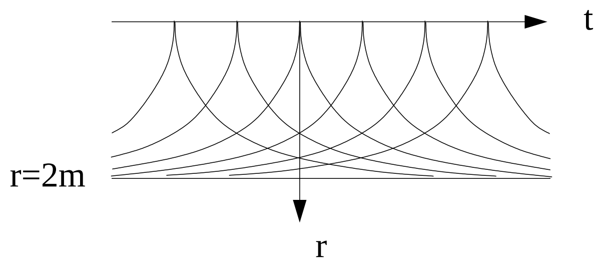



Figure 5.14:
Pattern of 'radial' light rays for $0 < r < 2m$ in coordinates $(t, r, \theta, \phi)$.

Given this picture, it is clear that $r = 0$ is a 'time' in the 'future' of any observer who happens to find themselves in $0 < r < 2m$.

In the exercises, we see that an observer at $r = r_0 < 2m$ must hit $r = 0$ in a finite maximal proper time, and can do nothing to avoid this.

Finally we have to ask the question, what happens at $r = 0$? The metric (337) is again singular at $r = 0$, just as it is at $r = 2m$. However, we have found that we can 'cross' the barrier $r = 2m$ by changing coordinates from $(t, r, \theta, \phi)$ to $(v, r, \theta, \phi)$. Is it possible that we can remove the singularity in the metric at $r = 0$ by a similar change of coordinates?

The answer is no, it is not possible. It is beyond the scope of the course to show this, but one can make a *scalar* quantity from the Riemann tensor which diverges at $r = 0$. Since this is a scalar quantity, it must be the same in all coordinate systems, and hence there can be no coordinate system for which this quantity is finite at $r = 0$. So $r = 0$ really is the end of spacetime.

However, our extension of the Schwarzschild spacetime to include this region with $r < 2m$ has been asymmetric in time - we could equally have tried to extend the spacetime by using coordinates based on outward moving photons. I.e., just as we used in-moving photons to label points to arrive at figure 39, we can label points by *out-moving photons.* If the point $(r, t)$ is now labelled $(r, u)$ where $u$ is the coordinate time at which the out-moving photon passing through $(r, t)$ reaches $(r_0, u)$, we have from (327):

$$t - u = + \left[ (r - r_0) + 2m \ln \left| \frac{r - 2m}{r_0 - 2m} \right| \right] , \quad (338)$$

$$\Rightarrow \quad \mathrm{d}t = \mathrm{d}u - \frac{\mathrm{d}r}{(1 - 2m/r)} . \quad (339)$$

which gives the new metric

$$\mathrm{d}s^2 = \left(1 - \frac{2m}{r}\right) \mathrm{d}u^2 + 2\,\mathrm{d}u\,\mathrm{d}r - r^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2) . \quad (340)$$

With these coordinates the pattern of light rays moving in the $(u, r)$ plane is

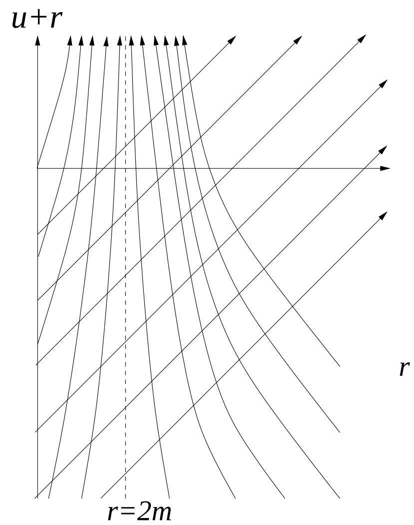



Figure 41:
The pattern of radial lightlike geodesics in coordinates $(u, r, \theta, \phi)$.

Now all the photons in $0 < r < 2m$ *leave* $r = 0$. So, if we label the region of space-time with $r < 2m$ in figure 41 by $I_1$ (for inside-region 1) and the region with $r < 2m$ in figure 39 by $I_2$ (for inside-region 2) and the region $r > 2m$ as $O_1$ (for outside-region 1), we see that choosing coordinates $(v, r, \theta, \phi)$ we can cover both $I_2$ and $O_1$, and $(u, r, \theta, \phi)$ we can cover both $I_1$ and $O_1$.

The natural question that remains is then: Is there a *single* set of coordinates which can cover all three regions? The answer took a long time to be found, and strictly speaking is *outside the range of courses 333Y and 334Z* but we summarise it on the next page.

### 5.15 Kruskal Coordinates

First, we change coordinates from $(t, r, \theta, \phi)$ to $(u, v, \theta, \phi)$, which gives

$$ds^2 = \left(1 - \frac{2m}{r}\right) dudv - r^2(d\theta^2 + sin^2\theta d\phi^2)$$

and then re-scale $u$ and $v$ to give $U$ and $V$ as

$$U = -e^{-u/4m} , \quad V = e^{v/4m}$$

and finally make linear combinations of these

$$T = \frac{U + V}{2} , \quad X = \frac{V - U}{2}$$

(these are known as 'Kruskal Coordinates'). Then $r$ = constant corresponds to hyperbolic $T^2 - X^2$ = constant. $t$ = constant corresponds to straight lines $\frac{T}{X}$ = constant, and the whole space-time has the form (drawing $T$ and $X$ only) on which we have easily identified the regions $I_1$, $I_2$ and $O_1$ covered by figures 39 and 41 - but there is even an extra part, $O_2$, which we have never encountered before, and which is just like a copy of $O_1$!

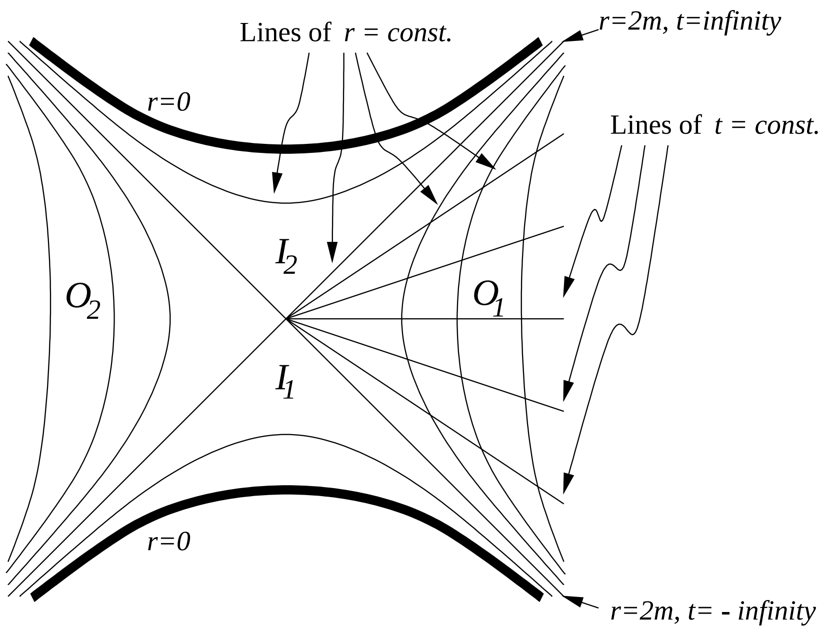



Figure 42: The fully extended Schwarzschild spacetime in Kruskal coordinates $(T, X, \theta, \phi)$.

Now some comments:

- In figures 42 and 43, all light rays are at 45° to the horizontal, so we see that no light can reach $I_1$ from $O_1$, or $O_1$ from $I_2$.
- the simple line $r = 2m$ in figure 37 has become two lines.
- the point $r = 0$ has become two lines, which are *singularities* of the metric. All matter (light, particles, whatever) entering $I_2$ will hit $r = 0$ and be destroyed - there is no more space-time after $r = 0$. The singularity $r = 0$ in $I_2$ is in the *future* of any observer in $I_2$, and so cannot be avoided - it is not possible to avoid the future by running sideways!

### 5.16 Gravitational collapse

So, we have found coordinates $(T, X, \theta, \phi)$ in which it is possible to extend the Schwarzschild spacetime with $r > 2m$ to include the four regions $I_1, I_2, O_1$ and $O_2$. However, it is clear that this solution, while being mathematically elegant, is not very physical.

In particular, it is hard to see how there can be a sensible physical interpretation of the points $r = 0$ inside $I_1$. This region may be a source of timelike and lightlike geodesics, but there is no 'cause' for this strange part of spacetime.

In the gravitational collapse of a massive star, it is possible for a region similar to $I_2$ to form, but without a corresponding region $I_1$. How is this possible?

The explanation is that we have only solved for the metric of spacetime *outside* the regions with matter. Inside the matter of the star that is collapsing we have not yet found the metric. While the surface of the star lies at a radial coordinate greater than $2m$, the spacetime is entirely regular and there is no need for a singularity. However, if the star decreases in size under its own gravitational attraction, then it is possible for the surface to fall inside $r = 2m$, and in that case the Schwarzschild solution will be valid for $r$ down to values smaller than $2m$, and it is then inevitable for a singularity akin to $r = 0$ to form.

This is illustrated in figure 43

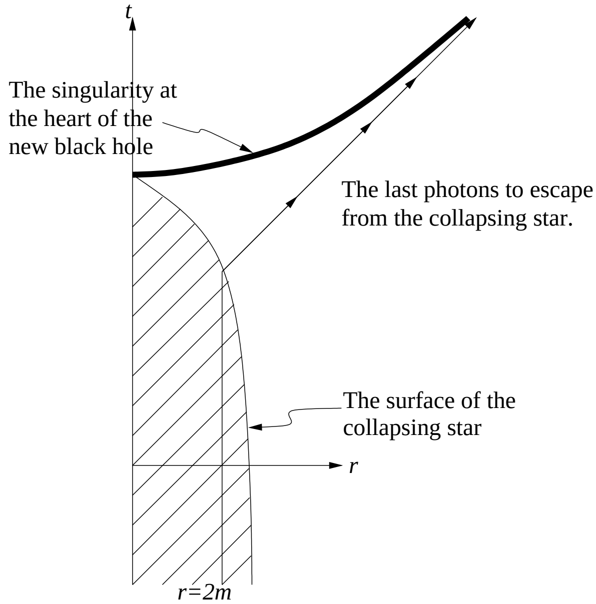



Figure 43:
A simplified diagram showing how a singularity forms when a star collapses so that its surface lies inside $r = 2m$.

## 5.17 Short Exercises

In these questions, take the Schwarzschild metric to be

$$\mathrm{d}s^2 = (1 - \tfrac{2m}{r})\mathrm{d}t^2 - \frac{\mathrm{d}r^2}{1 - \frac{2m}{r}} - r^2\mathrm{d}\theta^2 - r^2\sin^2\theta\,\mathrm{d}\phi^2 \ ,$$

**5.1: Circular paths**

Consider a circular path in the Schwarzschild metric with coordinates $(t, r, \theta, \phi)$ given by $x^\mu(\lambda) = (\lambda, r_0, \pi/2, \omega\,\lambda)$

Find the values of $\omega$ for which the path is (i) timelike, (ii) lightlike or null, (iii) spacelike.

**5.2: Outward lightlike radial geodesic**

Consider a lightlike radial outward-moving geodesic. This means a path with $\mathrm{d}s^2 = 0$ and $\theta$ and $\phi$ both constant and $\mathrm{d}r/\mathrm{d}t > 0$.

Show that on such a path $\mathrm{d}r/\mathrm{d}t = (1 - \frac{2m}{r})$.

Suppose the geodesic has an affine parameter $\lambda$. Show that $\mathrm{d}t/\mathrm{d}\lambda = \kappa/(1 - \frac{2m}{r})$ for some constant $\kappa$.

Show that $\mathrm{d}r/\mathrm{d}\lambda$ is a constant - that is, $r$ is an affine parameter.

**5.3: radial spacelike paths**

Consider the path $x^\mu(r) = (t_0, r, \theta_0, \phi_0)$, ie a path that is purely radial at a constant coordinate time $t_0$.

Show that this path is spacelike for $r > 2m$.

Find an integral expression for the proper distance along the path between $r_0$ and $r_1$.

Show that the event horizon of a black hole located at $r = 2m$ is a finite proper distance along a radial path from a point outside the black hole.

**Longer Exercises**

**5.4 A rotationally invariant, static, metric.**

The general form of the metric for a spherically symmetric, static space-time can be taken to be

$$\mathrm{d}s^2 = \mathrm{e}^A \mathrm{d}t^2 - \mathrm{e}^B \mathrm{d}r^2 - r^2(\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2),$$

where $A$ and $B$ are functions of $r$ only.

a) Calculate the Christoffel symbols $\Gamma^\mu_{\nu\rho}$ for this metric.

(the only non-vanishing Christoffel symbols are:

$$\Gamma^t_{rt} = \Gamma^t_{tr}\ ;\ \Gamma^r_{tt}\ ;\ \Gamma^r_{rr}\ ;\ \Gamma^r_{\theta\theta}\ ;\ \Gamma^r_{\phi\phi}\ ;\ \Gamma^\theta_{r\theta} = \Gamma^\theta_{\theta r}\ ;\ \Gamma^\theta_{\phi\phi}\ ;\ \Gamma^\phi_{r\phi} = \Gamma^\phi_{\phi r}\ ;\ \Gamma^\phi_{\phi\theta} = \Gamma^\phi_{\theta\phi}\ .)$$

b) Show that the non-vanishing components of the Ricci tensor $R_{\mu\nu}$ are

$$\begin{aligned} R_{tt} &= \mathrm{e}^{A-B}\left(\tfrac{1}{2}A'' - \tfrac{1}{4}A'B' + \tfrac{1}{4}(A')^2 + \tfrac{A'}{r}\right), \\ R_{rr} &= -\tfrac{1}{2}A'' + \tfrac{1}{4}A'B' - \tfrac{1}{4}(A')^2 + \tfrac{B'}{r}, \\ R_{\theta\theta} &= -\mathrm{e}^{-B}\left(1 + \tfrac{r}{2}(A' - B')\right) + 1, \\ R_{\phi\phi} &= R_{\theta\theta}\sin^2\theta \end{aligned}$$

where primes denote differentiation with respect to $r$.

[All other components vanish since the metric is diagonal.]

**5.5 The inevitability of reaching $r = 0$.**

Consider an observer who finds themselves inside the event horizon of a black hole at some radial coordinate $r_0 < 2m$. Show that there is a maximal proper time $\Delta\tau_{\max}$ within which this observer must reach $r = 0$, irrespective of their actions.

Hint – remember that $\mathrm{d}s^2 > 0$ for a timelike path and hence find a bound on $|\mathrm{d}\tau/\mathrm{d}r|$.

**5.6 A typical exam question on the Schwarzschild metric.**

The metric of a static spacetime with coordinates $x^\mu$ in which the speed of light is 1 may be written in the form

$$\mathrm{d}s^2 = g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu = g_{00}\,\mathrm{d}t^2\ +\ g_{ij}\,\mathrm{d}x^i\mathrm{d}x^j\ ,$$

where $t = x^0$ and where $g_{\mu\nu}$ is independent of $t$.

(a) State the physical conditions that a "good time coordinate" must satisfy for the metric to take this form.

In coordinates $(t, r, \theta, \phi)$ in which the speed of light is 1, the Schwarzschild metric may be written as

$$\mathrm{d}s^2 = \left(1 - \frac{a}{r}\right) \mathrm{d}t^2 \; - \; \frac{\mathrm{d}r^2}{\left(1 - \frac{a}{r}\right)} \; - \; r^2 \left( \mathrm{d}\theta^2 \; + \; \sin^2\theta \, \mathrm{d}\phi^2 \right) \; .$$

(b) A photon moving freely in the Schwarzschild spacetime moves radially outwards from a position $(0, r_0, \pi/2, 0)$ with $r_0 > a$ along a trajectory with coordinates $(t(\lambda), r(\lambda), \pi/2, 0)$, with $\lambda$ some parametrisation of the path.

Find the coordinate time taken for the photon to reach the radial coordinate $r_1$, and give the relation between the frequency of the photon as observed by a stationary observer at $r = r_0$ to that observed by a stationary observer at $r = r_1$.

(c) A massive particle moves non-freely along a circular path of radius $r = R$ in the equatorial plane $\theta = \pi/2$.

Show that the coordinate angular frequency $\omega(t) = \mathrm{d}\phi/\mathrm{d}t$ of the particle satisfies

$$\left| \omega \right| < \frac{1}{R} \left(1 - \frac{a}{R}\right)^{1/2}$$

**5.7 Symmetries of the Schwarzschild metric**

recall from sheet 4 that a vector $\xi^\mu$ defines a symmetry of a metric if it satisfies Killing's equation:

$$\nabla_\mu \xi_\nu + \nabla_\nu \xi_\mu = 0 \; ,$$

which is equivalent to the equation

$$\xi^\sigma \partial_\sigma (g_{\mu\nu}) + g_{\sigma\nu} \partial_\mu \xi^\sigma + g_{\mu\sigma} \partial_\nu \xi^\sigma = 0 \; .$$

• Show that the 4–vector

$$\xi^\mu = (0, 0, 0, \alpha) \; ,$$

satisfies Killing's equation for the Schwarzschild metric, and hence that

$$U^\mu \xi_\mu \; ,$$

is constant along an affinely parametrised geodesic $x^\mu(\lambda)$, where

$$U^\mu = \frac{\partial x^\mu}{\partial \lambda} \; .$$

(A very hard question: find the most general solution to Killing's equation in the Schwarzschild metric)

### 5.8 The qualitative nature of orbits in the Schwarzschild metric

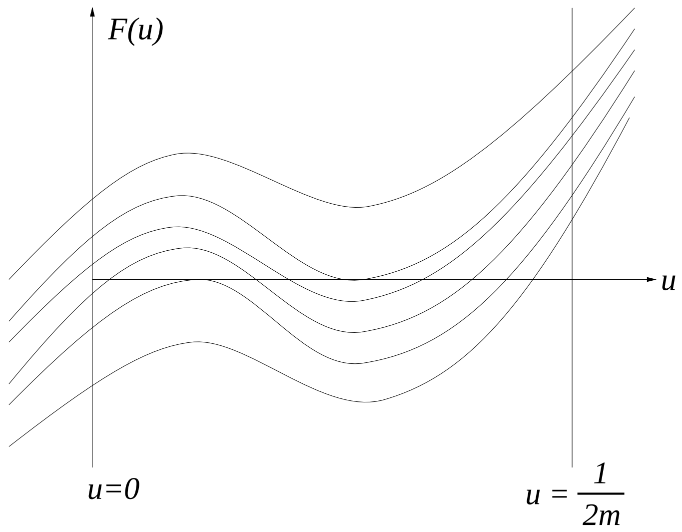
Consider the function $F(u)$

$$F(u) = (u')^2 = \left(\frac{\kappa^2 - \mathrm{C}}{h^2}\right) + \left(\frac{2m\mathrm{C}}{h^2}\right) u - u^2 + 2mu^3 ,$$

where $r = 1/u$ is the radial coordinate along a geodesic in the Schwarzschild metric lying in the plane $\theta = \pi/2$. Show that

$$F'(0) \geq 0 , \quad F(1/2m) > 0 , \quad F'(1/2m) > 0 .$$

Describe the qualitative forms of the geodesics for which the values of the constants $\kappa, m, \mathrm{C}$ and $h$ are such that $F(u)$ takes one of the six qualitative forms



### 5.9 Most of question 1, 1998 exam.

The metric of a static spacetime with coordinates $x^\mu$ in which the speed of light is 1 may be written in the form

$$\mathrm{d}s^2 = g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu = g_{00}\,\mathrm{d}t^2 \;+\; g_{ij}\,\mathrm{d}x^i\mathrm{d}x^j \;,$$

where $t = x^0$ and where $g_{\mu\nu}$ is independent of $t$.

(a) State the physical conditions that a "good time coordinate" must satisfy for the metric to take this form.

(b) State the condition on $\mathrm{d}s$ for a photon to be able to travel from $x^\mu$ to $x^\mu + \mathrm{d}x^\mu$.

(c) Hence show that the coordinate time taken for a photon to traverse a path $x^i(\lambda)$ from $x^i(\lambda_0)$ to $x^i(\lambda_1)$ is

$$\Delta t = \int_{\lambda_0}^{\lambda_1} \sqrt{-\frac{g_{ij}}{g_{00}}\frac{\mathrm{d}x^i}{\mathrm{d}\lambda}\frac{\mathrm{d}x^j}{\mathrm{d}\lambda}}\;\mathrm{d}\lambda\;.$$

(d) Suppose that light of frequency $\nu_0$ is emitted at by a static source at position $x^i(\lambda_0)$. By showing that $\Delta t$ is independent of $t$, or otherwise, find the frequency $\nu_1$ of the photon as received by a static observer at position $x^i(\lambda_1)$.

Lecture notes for Spacetime Geometry and General Relativity

2015-2016

Section 6: Cosmology

# 6 Cosmology

Cosmology attempts to describe the universe as a 4-dimensional space-time with a metric

$$ds^2 = g_{\mu\nu}\, dx^\mu\, dx^\nu \tag{341}$$

and with matter, the whole satisfying Einstein's equations. Without any more assumptions this is a hopeless task, so we look to the universe to see what it is like to see if that helps us in any way.

The most striking feature is that the universe looks roughly the same in all directions. At some times this was thought to mean that we occupy a central position in the universe, with everything else arranged symmetrically around us. However the modern view is that it is quite unlikely that we are anywhere special and so we assume:

> The Cosmological Principle:
> At all times and at all places the universe has appeared the same in all directions - "isotropic".

What does the universe look like, once we acknowledge it is the same in all directions? Here we encounter a problem that it is hard to measure distance, and our measures of distance are in great part based on a particular model of the history of the universe. There is a second feature that the further away light has come, the longer it has taken to reach us, (it travels at the speed of light), so looking into the distance is also looking back in time. What we see is a universe that has changed - close to us the universe is 'like it is now' - with galaxies, stars forming and stars dying. Further away we see galaxies forming, and further than that there is the background radiation which was emitted at a time before galaxies were formed.

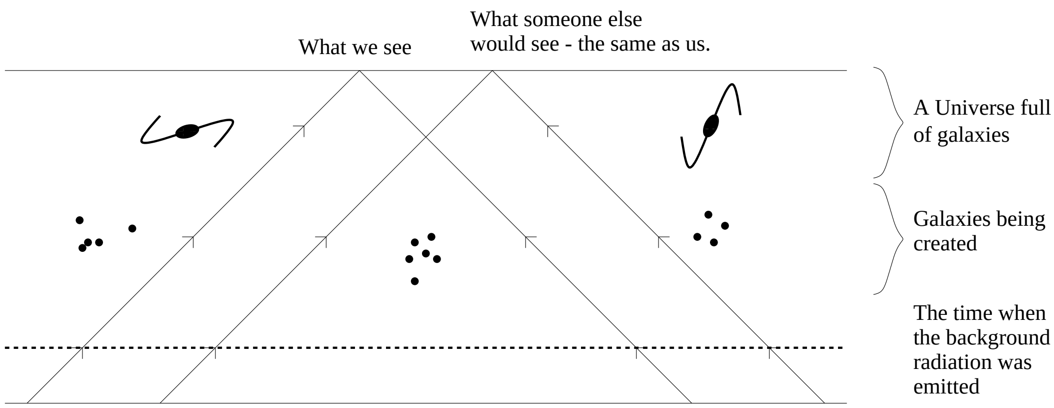



Figure 44: Cosmological time - defined so that 'stationary' observers see the same cosmological picture at the same cosmological time

Since at all places at all times the universe has looked the same in all directions, this means that each stationary observer has seen the same history (sequence of events) and so there is a standard measure of time, called 'cosmological time'. We choose the scale of this time so that it is the measure of proper time for each stationary observer. This means we choose coordinates $\{t, x^i\}$ for which stationary observers have constant values of $x^i$ and in which coordinates $g_{00} = 1$ so that $\mathrm{d}s^2 = \mathrm{d}t^2$ for stationary observers. This means the metric takes the form

$$\mathrm{d}s^2 = \mathrm{d}t^2 + 2g_{0i}\,\mathrm{d}t\,\mathrm{d}x^i + g_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j \ . \tag{342}$$

[This is *very* different from a *static metric*, for which $g_{00}$ is not a constant]. Now, $g_{0i}$ is a three-vector, which singles out a direction at each point, but the universe must be the same in each direction, so $g_{0i} = 0$ and

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

where $\mathrm{d}\Sigma^2$ is the metric of a 3-dimensional manifold.

This 3-dimensional manifold has a Ricci scalar $R^{(3)}$. If this scalar were not constant, then $\partial_i R^{(3)}$ would pick out a direction at each point - but there are no special directions so $R^{(3)}$ must be a constant. We pull out a scale factor $R^2(t)$ [not the same as $R^{(3)}$!] from $\mathrm{d}\Sigma^2$ so that

$$\mathrm{d}\Sigma^2 = R^2(t)\,\mathrm{d}\sigma^2 \ , \tag{343}$$

and $\mathrm{d}\sigma^2$ is the metric of a space of constant Gaussian curvature $0, +1$ or $-1$ [it is the factor $R(t)$ which enables us to normalise $\mathrm{d}\sigma^2$ in this way].

Once we know that the 3–dimensional space with metric $\mathrm{d}\sigma^2$ has constant curvature and is isotropic (has no preferred directions), this uniquely determines the metric.

This is beyond the scope of the course, so we shall simplify matters by assuming that $\mathrm{d}\sigma^2$ is spherically symmetric, and has the form

$$\mathrm{d}\sigma^2 = \mathrm{d}r^2 + f(r)^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2) \ , \tag{344}$$

for some function $f(r)$. We know in principle how to calculate the Ricci scalar $R^{(3)}$, which for a 3 dimensional manifold is proportional to the Gaussian curvature, $K$, with the answer

$$3K = -\left(\frac{2ff'' - 1 + (f')^2}{f^2}\right) \tag{345}$$

$$\text{Putting } K = \begin{cases} 1 \\ 0 \\ -1 \end{cases} \text{gives the solutions} \begin{cases} f = \sin r \ , \\ f = r \ , \\ f = \sinh r \ . \end{cases} \tag{346}$$

With a metric of the form (344), $r$ measures radial distance. However, as you know, we can also make $r$ measure the area of a sphere of coordinate $r$, or play another role, and so it is sometimes useful to consider the alternative, equivalent, metrics

$$\mathrm{d}\sigma^2 = \frac{\mathrm{d}\rho^2}{(1 - k\,\rho^2)} + \rho^2(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2) \ , \tag{347}$$

$$\mathrm{d}\sigma^2 = \frac{\mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2)}{(1 + \frac{1}{4}k\,r^2)^2} = \frac{\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2}{(1 + \frac{1}{4}k\,r^2)^2} \ . \tag{348}$$

The form (347) is arrayed so that the area of a sphere of coordinate $\rho$ is $4\pi\rho^2$. The form (348) is *conformally equivalent* to the flat 3-dimensional metric. The full space-time metric

$$\mathrm{d}s^2 = \mathrm{d}t^2 - R^2(t)\mathrm{d}\sigma^2 \ , \tag{349}$$

with $\mathrm{d}\sigma^2$ of type (347) or (348) was independently discovered by several people and so can be called, variously, a "Friedmann-Lemaître-Robertson-Walker"metric or by any subsetm such as a "Robertson-Walker" or "Friedmann-Walker" metric.

We shall eventually want to find dynamical equations for $R(t)$, but before we do that it will be useful to consider what sort of universes we are describing with metrics (349), and what physical meaning $k$ and $R(t)$ have.

## 6.1 The physical meaning of $k$

We so far haven't discussed the physical meaning of $k$ as it affects the spatial metric in its various forms:

$$\mathrm{d}\sigma^2 = \begin{cases} \mathrm{d}r^2 + \sin^2 r\,(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2) & k=1 \\ \mathrm{d}r^2 + r^2\,(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2) & k=0 \\ \mathrm{d}r^2 + \sinh^2 r\,(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2) & k=-1 \end{cases}, \tag{257}$$

$$\mathrm{d}\sigma^2 = \frac{\mathrm{d}\rho^2}{1-k\,\rho^2} + \rho^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2)\,, \tag{258}$$

$$\mathrm{d}\sigma^2 = \frac{\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2}{(1+k\,r^2/4)^2}\,. \tag{259}$$

For this purpose it is best to consider form (344) in which $r$ is a direct measure of radial distance: the proper distance along the radial path between two points at $r=0$ and $r=r$ in metric (344) is

$$\int \mathrm{d}\sigma = \int_0^r \mathrm{d}r = r\,, \qquad \text{since } \mathrm{d}\theta = \mathrm{d}\phi = 0 \text{ along radial paths}\,.$$

So, using this radial coordinate, we can consider the sphere if all points radial distance $r$ from the origin, and ask what is the volume of space contained in this sphere?

It is outside the scope of this course, but the volume of a region of space with metric $g_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j$ is

$$\int \mathrm{d}x^1\,\mathrm{d}x^2\,\mathrm{d}x^3\,\sqrt{\det(g_{ij})}\,,$$

where the factor $\sqrt{\det(g_{ij})}$ can be seen as the Jacobian for the transformation to coordinates in locally flat coordinates. For the metric (344),

$$g_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & f(r)^2 & 0 \\ 0 & 0 & f(r)^2\,\sin^2\theta \end{pmatrix}\,, \qquad \sqrt{\det(g_{ij})} = f(r)^2\,\sin\theta\,,$$

and the volume of the sphere of radius $r$ is

$$\begin{aligned} \mathrm{Vol}(r) &= \int_{r=0}^{r}\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi} f(r)^2\,\sin\theta\,\mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}\phi \\ &= \begin{cases} \pi\,(\,2r - \sin(2r)\,) = \frac{4}{3}\pi r^3\left(1 - \frac{r^2}{5}\ldots\right) & ,\ k=1 \\ \frac{4}{3}\pi r^3 & ,\ k=0 \\ \pi\,(\sinh(2r) - 2r\,) = \frac{4}{3}\pi r^3\left(1 + \frac{r^2}{5}\ldots\right) & ,\ k=-1 \end{cases}\,. \end{aligned} \tag{350}$$

These are shown in figure 45

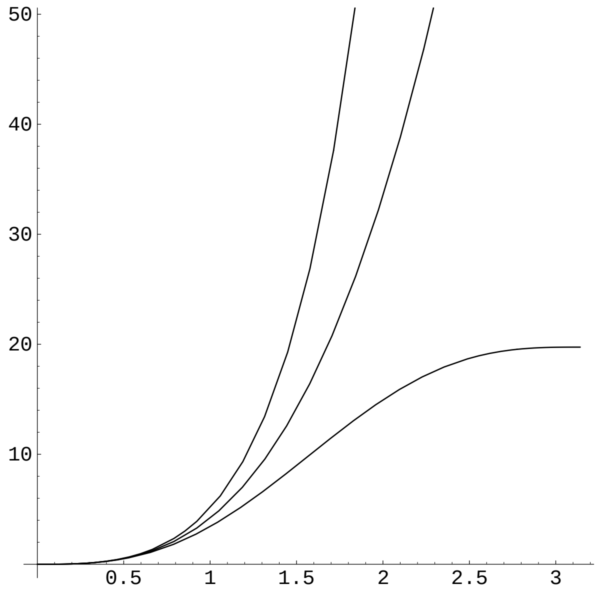



Figure 45: The volume of a sphere of radius $r$ in the three metrics (344).

The spaces with $k = 0$ and $k = -1$ both have infinite volume – $\mathrm{Vol}(r) \to \infty$ tends to infinity as $r \to \infty$ – and the resulting spacetimes are called 'open'. The space with $k = 0$ is just the usual flat Euclidean 3–space, but the space with $k = -1$ is 'larger' in that the sphere of radius $r$ has larger volume than the corresponding sphere in Euclidean 3–space.

The space with $k = 1$ in fact has finite volume, since $r$ has a maximal value of $\pi$ at which the area of the sphere at radial value $r = \pi$ is in fact zero. This means that the whole "sphere" at $r = \pi$ is in fact a single point - just like the whole "circle" at latitude $\theta = \pi$ is a single point in two-dimensions. This space is in fact a 3–sphere, and the resulting spacetime is called 'closed'.

It may indeed be possible to measure the curvature of the universe directly by counting the number of galaxies in a sphere of a given radius.

## 6.2 The physical meaning of $R(t)$

We have already described $R(t)$ as a 'scale factor' - and we can see that this is indeed what it is when we try to find the spatial distance between 'fixed' observers.

Consider two observers, one at $r = 0$, and one at $r = r_0$, in coordinates (348). The spatial distance between them is given by

$$\int |\mathrm{d}s| = R(t) \int_0^{r_0} \frac{\mathrm{d}r}{1 + k\, r^2/4} = R(t)\, f(r_0)\ , \tag{351}$$

where $f(r_0)$ is the integral $\int_0^{r_0} \mathrm{d}r/(1 + k\, r^2/4)$, whose exact form we do not need here.

So, the distance between two points at radial coordinates $r = 0$ and $r = r_0$ is

$$d(t) = R(t)\, f(r_0)\ .$$

If $R(t)$ is increasing with time, then the distance between these points is also increasing, and the universe appears to be expanding; conversely, if $R(t)$ is decreasing, then the distance between these points is decreasing and the universe appears to be contracting.

There is also another property, which is that the relative speed at which points appear to be moving away or towards each other is proportional to their separation,

$$\dot{d}(t) = \dot{R}(t)\, f(r_0) = \dot{R}(t) \left(\frac{d(t)}{R(t)}\right) = \left(\frac{\dot{R}(t)}{R(t)}\right) d(t)\ , \tag{352}$$

where the proportionality factor $\dot{R}/R$ has a special name,

$$\frac{\dot{R}(t)}{R(t)} = H(t)\ , \tag{353}$$

Hubble's constant, and eqn. (352) is known as Hubble's law. Cosmology really took off as a subject when the relative motions of nearby galaxies were measured and seen to obey this law.

However, as we mentioned before, we cannot actually determine the relative velocities of galaxies *now*, i.e. at the same cosmological time, since we only have information about them at the time the light that we see left them. this means that rather than deducing $R(t)$ as it is now, when we measure the speeds of far off galaxies, we instead are able to deduce the whole past history of the scale factor $R(t)$.

It turns out that it is possible to measure the ratios of the scale factor now $R(t_0)$ to the scale factor $R(t_1)$ at the time light was emitted from a distant galaxy at some earlier time $t_1$ directly. The reason is that the ratio of these two scale factors shows up directly in the *cosmological red-shift* of the light. This effect is not the same as the gravitational red-shift in static or stationary spacetimes, and we work out the formula for it now.

## 6.3 The Cosmological redshift

Although the cosmological redshift has a different origin to the gravitational redshift in a static spacetime, we work it out using the same ideas; that is, we consider two successive wave-crests of a photon emitted from a distant galaxy and find the difference in proper time between their arrival at us.

Assuming (as is the case) that purely radial lightlike geodesics are possible in Friedmann-Walker spacetimes, we can work put the equations of these geodesics from the fundamental property of lightlike paths that

$$\mathrm{d}s^2 = 0\ .$$

If the path is purely radial, then $\mathrm{d}\theta = \mathrm{d}\phi = 0$, so that we have (in coordinates (348),

$$\mathrm{d}s^2 = \mathrm{d}t^2 - \frac{R(t)^2}{(1 + k\,r^2/4)^2}\,\mathrm{d}r^2 = 0\ , \quad \frac{\mathrm{d}t}{R(t)} = -\frac{\mathrm{d}r}{1 + k\,r^2/4}\ , \tag{354}$$

where as usual the '-' sign is due to the fact that the photons are moving along incoming radial geodesics, paths of decreasing $r$. This means that the time of emission $t_1$ and reception $t_0$ of the photon are given implicitly in terms of the $r$–coordinate by

$$\int_{t_1}^{t_0} \frac{\mathrm{d}t}{R(t)} = -\int_{r_0}^{0} \frac{\mathrm{d}r}{1 + k\,r^2/4} = f(r_0)\ , \tag{355}$$

where $f(r_0)$ is the same function as before.

Suppose the next wave-crest of the photon is emitted at time $t_1 + \delta t_1$, then since $t$ is the proper time of stationary observers, $\delta_1 = 1/\nu_1$, where $\nu_1$ is the frequency of the photon as emitted. This wave-crest will then arrive at us at time $t_0 + \delta t_0$, where again $\delta t_0 = 1/\nu_0$, where $\nu_0$ is the frequency of the photon as we observe it. Since this second photon also leaves $r_0$ and arrives at $r = 0$, its trajectory must also obey (355), that is

$$\int_{t_1+\delta t_1}^{t_0+\delta t_0} \frac{\mathrm{d}t}{R(t)} = f(r_0) \; . \tag{356}$$

But if $\delta t_0$ and $\delta t_1$ are small compared (as they will be on cosmological time-scales), we have

$$\begin{aligned} f(r_0) &= \int_{t_1+\delta t_1}^{t_0+\delta t_0} \frac{\mathrm{d}t}{R(t)} = (\int_{t_1}^{t_0} + \int_{t_0}^{t_0+\delta t_0} - \int_{t_1}^{t_1+\delta t_1})\frac{\mathrm{d}t}{R(t)} \sim (\int_{t_1}^{t_0}\frac{\mathrm{d}t}{R(t)}) + \frac{\delta t_0}{R(t_0)} - \frac{\delta t_1}{R(t_1)} \\ &= f(r_0) + \frac{\delta t_0}{R(t_0)} - \frac{\delta t_1}{R(t_1)} \; , \end{aligned} \tag{357}$$

and hence

$$\frac{\delta t_0}{R(t_0)} \sim \frac{\delta t_1}{R(t_1)} \; , \quad \delta t_0 \sim \left(\frac{R(t_0)}{R(t_1)}\right) \delta t_1 \; , \quad \nu_0 \sim \left(\frac{R(t_1)}{R(t_0)}\right) \nu_1 \; . \tag{358}$$

Thus the change in frequency is directly given by the ratio of the scale factors then and now.

The standard cosmologist's definition of redshift is in fact in terms of the wavelength of the photon, and is

$$z = \frac{\lambda_0}{\lambda_1} - 1 = \frac{c/\nu_0}{c/\nu_1} - 1 = \frac{\nu_1}{\nu_0} - 1 = \frac{R(t_0)}{R(t_1)} - 1 \; .$$

This means that the redshift is a direct measure of the change in separation of galaxies during the time it has taken the photon to reach us. If a galaxy is said to be at redshift 5, say, this means that it is now 6 times further away from us than the time at which the photons we are receiving were emitted.

It is important to note that the redshift does not give any direct information on the distance of the source, nor does it have to be a faithful indicator of the distance – sources at different distances can have the same or similar redshifts. The redshift only depends on the ratio of the scale factor at the time of emission and now, so that if the redshift stayed more or less constant for a period, then all the photons emitted during that period would reach us with the same or similar redshifts. Similarly, if the scale factor ever decreased then increased, there would also be sources at very different distances which would appear to have the same redshift (See figure 46)

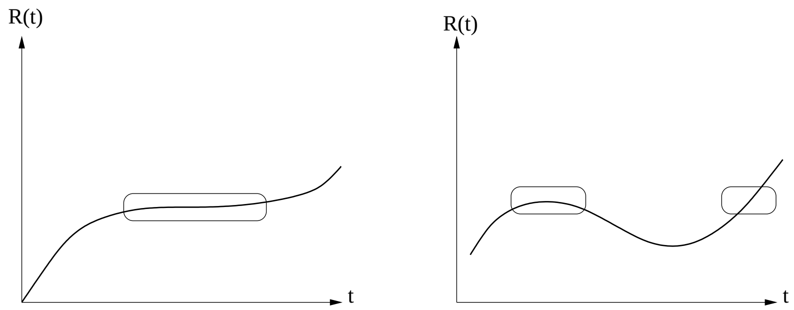



Figure 46:
Light emitted at the same values of $R(t)$ arrives with the same cosmological redshift.

## 6.4 Dynamical equations for $R(t)$

Having investigated the kinematics of Robertson-Walker spacetimes, we now have to find an equation that will give us $R(t)$, and tell us which value of $k$ describes our universe. The equations which govern our spacetime are (of course) Einstein's equations. In the presence of matter, and with a Cosmological constant, these equations are

$$G_{\mu\nu} - \Lambda\, g_{\mu\nu} = (8\pi G) T_{\mu\nu} \tag{359}$$

where $G_{\mu\nu} = R_{\mu\nu} - \frac{R}{2} g_{\mu\nu}$ is the Einstein tensor, and $T_{\mu\nu}$ is the energy-momentum tensor of the matter. $G$ is Newton's constant. Remember that it is a consequence of (359) that

$$\nabla^\mu T_{\mu\nu} = 0\ .$$

The equations (there are 16 eqns in (359), since $\mu$ and $\nu$ can each take 4 values) will be differential equations for $R(t)$ that will depend upon the parameter $k$ (which takes values $0, \pm 1$) and on the type of matter in the universe. Again, you know how to work out $G_{\mu\nu}$ for the metric (349). The equations for $R(t)$ clearly do not depend on which form of the metric we take, so it turns out to be easiest to use (348) for $\mathrm{d}\sigma^2$. It turns out that the only non-zero components are the diagonal components, and that of these four, only two are independent:

$$G_{tt} \;=\; 3\frac{\dot{R}^2}{R^2} \;+\; 3\frac{k}{R^2}\ , \tag{360}$$

$$\frac{G_{xx}}{g_{xx}} = \frac{G_{yy}}{g_{yy}} = \frac{G_{zz}}{g_{zz}} \;=\; \frac{2\ddot{R}}{R} \;+\; \frac{\dot{R}^2}{R^2} \;+\; \frac{k}{R^2}\ . \tag{361}$$

We now have to investigate the right-hand side of eqn (359).

We have briefly discussed the energy-momentum tensor before, and it turns out that for the sorts of matter we shall consider, we can take

$$T_{\mu\nu} = (p+\rho) U_\mu U_\nu - p\, g_{\mu\nu}\ , \tag{362}$$

where $\rho$ is the density, $p$ the pressure and $U^\mu$ a 4–velocity. The relation between $\rho$ and $p$ is known as the equation of state of the matter in question, and we shall consider only two special sorts: dust and radiation.

- Dust: This is an idealisation of a system of particles which are assumed to have a smooth density $\rho$ but to have no internal forces, and hence zero pressure. For us 'dust' will mean an energy-momentum tensor of the form (362), but with

$$\boxed{p = 0} \tag{363}$$

- Radiation: This can again be characterised by an energy-momentum tensor of the form (362), but with

$$\boxed{p = \frac{\rho}{3}} \tag{364}$$

In general, $U^\mu$ can take on any value, but in our cosmological situation, the cosmological principle implies that $U^i$ must be zero, since it is a 3-vector, and would define a preferred direction at each point if it were non-zero. Furthermore, since $U^\mu$ is a 4-velocity,

$$U^\mu U_\mu = g_{\mu\nu} U^\mu U^\nu = 1 \; . \tag{365}$$

If $U^\mu = (U^0, 0, 0, 0)$, then eqn (365) becomes $g_{00}\,(U^0)^2 = 1$. For the metrics we are considering, (349), $g_{00} = 1$ and hence $U^\mu = (1, 0, 0, 0)$, and $U_\mu = g_{\mu\nu}U^\nu = (g_{00}, 0, 0, 0) = (1, 0, 0, 0)$. Putting these values into (362) we find that $T_{\mu\nu}$ also has only two independent components:

$$T_{00} \;=\; \rho \; , \tag{366}$$

$$\frac{T_{xx}}{g_{xx}} = \frac{T_{yy}}{g_{yy}} = \frac{T_{zz}}{g_{zz}} \;=\; -p \; . \tag{367}$$

(All other components of $T_{\mu\nu}$ are zero).

Putting eqns (360), (361), (366) and (367) together, we find Einstein's eqns are

$$\frac{3\dot{R}^2}{R^2} + \frac{3k}{R^2} - \Lambda \;=\; (8\pi G)\rho \; , \tag{368}$$

$$\frac{2\ddot{R}}{R} + \frac{\dot{R}^2}{R^2} + \frac{k}{R^2} - \Lambda \;=\; -(8\pi G)p \; . \tag{369}$$

It is useful to rewrite eqn. (368) in the form

$$\dot{R}^2 = \frac{8\pi G}{3}\,\rho\, R^2 + \frac{\Lambda R^2}{3} - k \; . \tag{370}$$

In this form, it is known as Friedmann's equation. If we could only find $\rho$ as a function of $R$, we could then plot $\dot{R}^2$ against $R$, and so find the qualitative behaviour of $R(t)$ without further ado. This is what we shall do later on in the course.

It is also useful to know the following two equations can be deduced from (368) and (369):

$$\dot{\rho} + (\rho + p)\frac{3\dot{R}}{R} \;=\; 0 \; , \tag{371}$$

$$\frac{\ddot{R}}{R} \;=\; \frac{\Lambda}{3} - \frac{(4\pi G)}{3}(\rho + 3p) \; . \tag{372}$$

To derive (371), we see that we need to find $\dot{\rho}$. The only place this can come from is differentiating (368) by $t$, to get

$$\dot{\rho} = \frac{3}{8\pi\, G}\left(\frac{\dot{R}\,\ddot{R}}{R^2} - \frac{2\dot{R}^2}{R^3} - \frac{2k\,\dot{R}}{R^3}\right) = \frac{1}{8\,\pi\, G}\left(3\frac{\dot{R}}{R}\right)\left(\frac{2\ddot{R}}{R} - \frac{2\dot{R}}{R^2} - \frac{2k}{R^2}\right) \; . \tag{373}$$

However, by direct substitution from eqns. (368) and (369), we also have

$$(p + \rho)\,\frac{3\,\dot{R}}{R} = -\frac{1}{8\,\pi\, G}\left(\frac{3\dot{R}}{R}\right)\left(\frac{2\ddot{R}}{R} - \frac{2\dot{R}}{R^2} - \frac{2k}{R^2}\right) \; , \tag{374}$$

so that eqns. (368) and (369) together imply

$$\dot{\rho} + (p + \rho)\,\frac{3\dot{R}}{R} = 0 \; . \tag{375}$$

This equation can be deduced directly from the conservation of the energy momentum tensor:

$$\nabla^{\mu} T_{\mu\nu} = 0 \ . \tag{376}$$

We have already seen in exercise sheet 3 that (376) for an energy momentum tensor of type (362) leads to the equation

$$\frac{\mathrm{D}\rho}{\mathrm{d}t} + (p+\rho)\,\nabla_{\mu} U^{\mu} \ , \tag{377}$$

which already looks very much like (375). In exercise sheet 6 it is shown that they are indeed the same for the Robertson-Walker metric.

Equation (371) is known as the continuity equation, and (372) as the quasi-Newtonian equation. They are especially useful in two ways. Firstly, eqn. (372) finally allows us to see why Einstein introduced the cosmological constant.

### 6.4.1 Einstein's reason for introducing $\Lambda$

Let us suppose for the moment that $\Lambda = 0$ (as it was in Einstein's original equations) then equation (372) implies

$$\frac{\ddot{R}}{R} = -\frac{(4\pi G)}{3}(\rho + 3p) \ .$$

So, if we want to have a static universe with non-zero matter content, then $p > 0, \rho > 0$, then it is a consequence that $\ddot{R} < 0$ (if $\Lambda = 0$). This bothered Einstein a lot when he first wrote down his equations since in the original form the $\Lambda$–term was missing, and he wanted to find a *static* universe with matter in it, since that was how the universe was believed to be.

Substituting $\dot{R} = \ddot{R} = 0$ into the full eqns. (368), (369) we find

$$\frac{3k}{R^2} - \Lambda = (8\pi G)\rho \ , \tag{378}$$

$$\frac{k}{R^2} - \Lambda = -(8\pi G)p \ . \tag{379}$$

From these we deduce that

$$\frac{k}{R^2} = 4\pi G(\rho + p) \ , \quad \Lambda = 4\pi G(\rho + 3p) \ . \tag{380}$$

Secondly, we note that equation (371) can be directly integrated in many cases of interest. Firstly, there are two alternative ways of writing equation (371):

$$\frac{\mathrm{d}}{\mathrm{d}t}(\rho R^3) = -p\frac{\mathrm{d}}{\mathrm{d}t}R^3 \;, \tag{381}$$

and (more useful for us)

$$\frac{\mathrm{d}}{\mathrm{d}R}(\rho R^3) = -3pR^2 \;. \tag{382}$$

This is useful because one often has a formula for $p$ just in terms of $\rho$, something of the form $p = f(\rho)$. Such as equation is called an "equation of state". We have exactly this situation for the two cases we are interested in, "dust", $p = 0$, (363) and "radiation", $p = \rho/3$, (364).

Knowing these equations of state, we can integrate eqn. (382)

$$\text{Dust} \quad : \quad \frac{\mathrm{d}}{\mathrm{d}R}(\rho R^3) = 0 \qquad \Rightarrow \quad \rho = \frac{\rho_0}{R^3} \;, \quad , \tag{383}$$

$$\text{Radiation} \quad : \quad \frac{\mathrm{d}}{\mathrm{d}R}(\rho R^3) = -\rho R^2 \quad \Rightarrow \quad \rho = \frac{\rho_0}{R^4} \;. \;. \tag{384}$$

Now that we have both $\rho$ and $p$ as functions of $R$, we have both $\ddot{R}$ and $\dot{R}$ as functions of $R$:

$$\text{Dust} \quad \begin{cases} \ddot{R} = \frac{1}{3}\Lambda R - \frac{4\pi G}{3}\frac{\rho_0}{R^2} \\ \dot{R}^2 = \frac{1}{3}\Lambda R^2 - k + \frac{8\pi G}{3}\frac{\rho_0}{R} \end{cases} \tag{385}$$

$$\text{Radiation} \quad \begin{cases} \ddot{R} = \frac{1}{3}\Lambda R - \frac{8\pi G}{3}\frac{\rho_0}{R^3} \\ \dot{R}^2 = \frac{1}{3}\Lambda R^2 - k + \frac{8\pi G}{3}\frac{\rho_0}{R^2} \end{cases} \tag{386}$$

It is quite possible to find $R(t)$ from these equations by integration, but it is much easier to find the qualitative form of the evolution of the universe by examining the equations themselves.

One of the simplest approaches is to plot $\dot{R}^2$ against $R$, and consider what this implies for the qualitative form of $R(t)$.

In each of these two cases, $\dot{R}^2$ is given by the Friedmann eqn as

$$\begin{aligned} \dot{R}^2 &= \frac{1}{3}\Lambda R^2 - k + \frac{8\pi G}{3}\rho R^2 \\ &= \frac{1}{3}\Lambda R^2 - k + \frac{C}{R^\alpha} \qquad \alpha = \begin{cases} 1\;, & \text{Dust,} \\ 2\;, & \text{Radiation.} \end{cases} \end{aligned} \tag{387}$$

The first thing to note is that qualitatively there is little difference between dust and radiation[13]

- For small $R$, $\dot{R}^2$ is dominated by the density term $C/R^\alpha$ and $|\dot{R}| \to \infty$ as $R \to 0$, so this is a period of rapid expansion or contraction.

  Given

  $$\dot{R}^2 \sim \frac{C}{R^\alpha} \quad \Rightarrow \quad \dot{R} \sim \pm\frac{\sqrt{C}}{R^{\alpha/2}} \;,$$

[13] In exercise sheet 6 this is taken further and it is shown that for a large range of equations of state there is little qualitative difference in the behaviour of $R(t)$.

we can solve this equation easily to find (assuming that $\dot{R} > 0$)

$$R \sim \text{constant}\, |t|^{2/(\alpha+2)} \,.$$

For dust, $\alpha = 1$ and $R \sim |t|^{2/3}$; for radiation $\alpha = 2$ and $R \sim |t|^{1/2}$. In both these cases $R$ will expand from zero to finite size in finite time, or collapse from finite size to zero size in finite size.

- For large $R$, the behaviour depends principally on $\Lambda$, and if $\Lambda = 0$ then on $k$.

We now consider in more detail the two special classes of solutions which have been regarded as most relevant to physics: these are the solutions with $\Lambda = 0$ (regarded as important until recently) and those with $k = 0$ (nowadays regarded as more likely to be the case). We finish with sketches (without discussion) of all the possible qualitative forms of the evolution of the universe.

We start with the detailed discussion of the $k = 0$ cases.

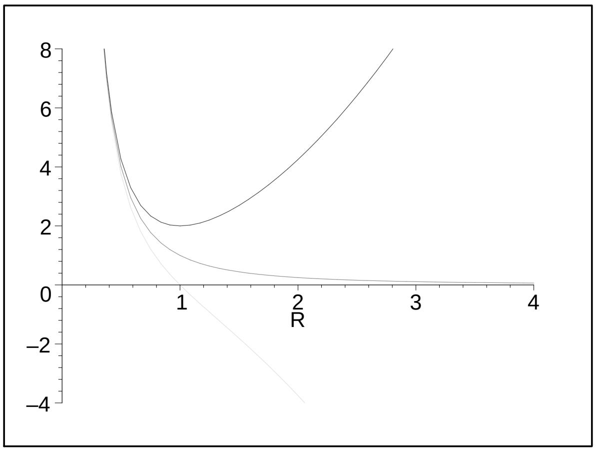
## 6.5 Cosmological solutions with $k = 0$

Putting $k = 0$ in the Friedmann equation, we have

$$\dot{R}^2 = \frac{\Lambda}{3} R^2 + \frac{C}{R^\alpha} \,. \tag{388}$$

- For small $R$, $\dot{R}^2$ is large and positive.
- For $R$ large, $\dot{R}^2$ is large and positive for $\Lambda$ positive, small and positive for $\Lambda = 0$, and negative (unphysical) for $\Lambda$ negative

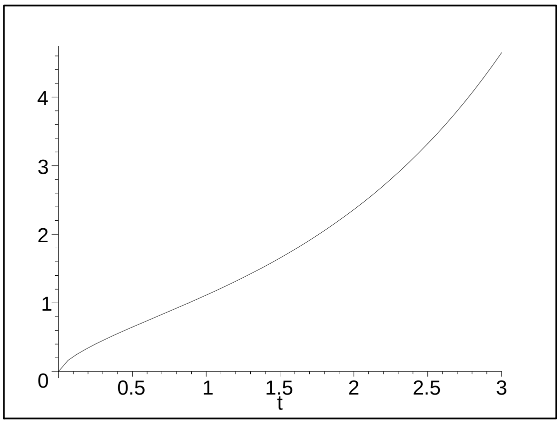
These features are sketched below:




Given these simple sketches, we can work out the qualitative form of the evolution of the universe.

A For $\Lambda > 0$, $\dot{R}^2$ is never negative, and so $\dot{R}$ must be always positive or always negative.

• For $\dot{R} > 0$, $R$ starts off small with a rapid expansion, then the rate of expansion decreases to a minimum rate of expansion before the rate increases again

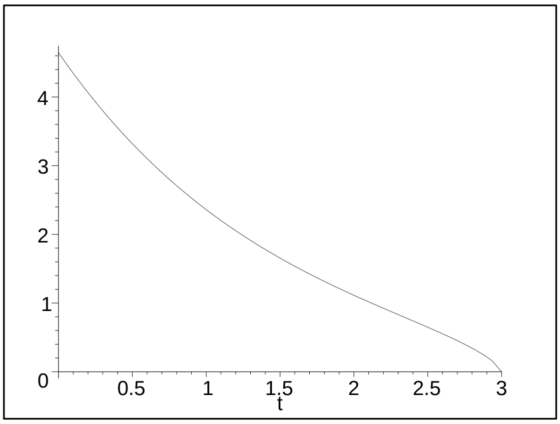



To be honest, we need to think a bit more carefully to see that the universe can expand from zero in a finite time; for small $R$, $\dot{R}^2 \sim C/R^\alpha$, so the time to increase from scale factor $R_0$ to $R_1$ is

$$\int_{R_0}^{R_1} \mathrm{d}R/\dot{R} \sim \int_{R_0}^{R_1} R^{\alpha/2}\mathrm{d}R/\sqrt{C} = \frac{(2+\alpha)}{\alpha\sqrt{C}}[R_1^{(2+\alpha)/2} - R_0^{(2+\alpha)/2}]$$

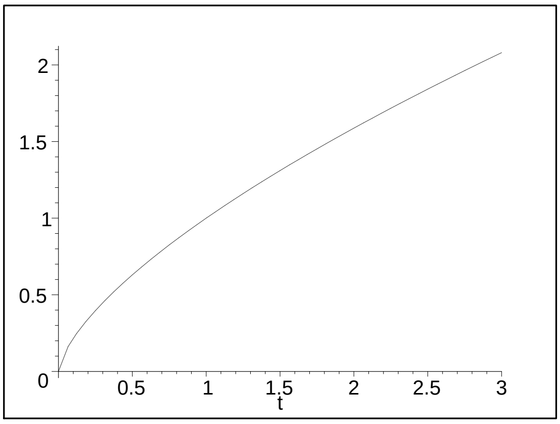
For the two cases we consider, $\alpha = 1$ and $\alpha = 2$, this gives a finite time to increase from zero size to finite size.

• For $\dot{R} < 0$, the evolution is the opposite: $R$ starts off large and decreasing fast, before the rate of collapse decreases to a minimum rate before increasing again until the universe collapses to zero size.

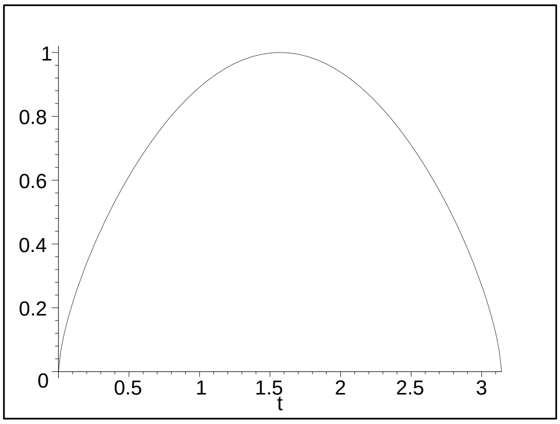



(again, in principle we need to check that the universe will reach zero size in finite time, but this is he case as we showed above).

B For $\Lambda = 0$, $\dot{R}^2$ is again always positive so $\dot{R}$ cannot change sign. We consider $\dot{R} > 0$ (the other case is the time reverse of this). The universe starts off at zero size, expands rapidly, and then continues expanding but at an ever decreasing rate




C For $\Lambda < 0$, there is a critical value of $R = R_c$ at which $\dot{R}^2 = 0$. It turns out that if $k = 0$, then $\ddot{R} < 0$ at this value of $R$, and so it can't correspond to a static solution. Instead, if $R$ is initially increasing, it slows until it reaches $R_c$ and then $R$ starts decreasing again. This corresponds to a universe that initially expands then contracts back to zero size in finite time:




We can also find the exact solutions without too much difficulty. We treat the cases of dust and radiation separately.

• The case of dust, $\alpha = 1$. Let us at first assume that $\Lambda > 0$ and $\dot{R} > 0$, then we have

$$\begin{aligned} \dot{R}^2 &= \frac{\Lambda}{3} R^2 + \frac{C}{R} = \frac{\Lambda/3 R^3 + C}{R} \\ \Rightarrow \quad \int \mathrm{d}t &= \int \frac{\mathrm{d}R}{\dot{R}} = (\Lambda/3)^{-1/2} \int \frac{R^{1/2} \mathrm{d}R}{\sqrt{R^3 + a^2}} \end{aligned} \tag{389}$$

where $a^2 = 3C/\Lambda$. Substituting $r = R^{3/2}$ and $r = a \sinh\theta$ in turn, we get

$$\sqrt{\Lambda/3}\, t = \int \frac{R^{1/2} \mathrm{d}R}{\sqrt{R^3 + a^2}} = \frac{2}{3} \int \frac{\mathrm{d}r}{\sqrt{r^2 + a^2}} = \frac{2}{3}\theta \tag{390}$$

Hence

$$R_{\Lambda>0} = r^{2/3} = (a \sinh\theta)^{2/3} = \left( \sqrt{\frac{3C}{\Lambda}} \sinh(\frac{3}{2}\sqrt{\Lambda/3} t) \right)^{2/3} = \left(\tfrac{3C}{\Lambda}\right)^{1/3} \sinh\left( \frac{3}{2}\sqrt{\Lambda/3} t \right)^{2/3} . \tag{391}$$

This is fine for $\Lambda > 0$. To get the solution $\Lambda = 0$, we can simply take this limit to get

$$R_{\Lambda=0} = \lim_{\Lambda\to 0}\left(\sqrt{\tfrac{3C}{\Lambda}}\,\sinh(\tfrac{3}{2}\sqrt{\Lambda/3}t)\right)^{2/3} = (\frac{3}{2}\sqrt{C})^{2/3}\,t^{2/3}\ . \tag{392}$$

For $\Lambda < 0$, we put $\sqrt{\Lambda} = i\sqrt{-\Lambda}$ and get

$$R_{\Lambda<0} = \left(\tfrac{3C}{-\Lambda}\right)^{1/3}\ \sin\left(\frac{3}{2}\sqrt{-\Lambda/3}\ t\right)^{2/3}\ . \tag{393}$$

• The case of radiation, $\alpha = 2$. We repeat the same steps as above: we assume that $\Lambda > 0$ and $\dot{R} > 0$, so

$$\begin{aligned}\dot{R}^2 &= \frac{\Lambda}{3}R^2 + \frac{C}{R^2} = \frac{\Lambda/3R^4 + C}{R^2}\\ \Rightarrow\quad \int \mathrm{d}t &= \int \frac{\mathrm{d}R}{\dot{R}} = (\Lambda/3)^{-1/2}\int \frac{R\mathrm{d}R}{\sqrt{R^4+a^2}}\end{aligned} \tag{394}$$

where $a^2 = C/\Lambda$. Substituting $r = R^2$ and $r = a\sinh\theta$ in turn, we get

$$\sqrt{\Lambda/3}\,t = \int \frac{R\mathrm{d}R}{\sqrt{R^4+a^2}} = \frac{1}{2}\int\frac{\mathrm{d}r}{\sqrt{r^2+a^2}} = \frac{1}{2}\,\theta \tag{395}$$

Hence

$$R_{\Lambda>0} = r^{1/2} = (a\sinh\theta)^{1/2} = \left(\sqrt{\frac{3C}{\Lambda}}\ \sinh(2\sqrt{\Lambda/3}t)\right)^{1/2} = \left(\tfrac{3C}{\Lambda}\right)^{1/4}\ \sinh\left(2\sqrt{\Lambda/3}t\right)^{1/2}\ . \tag{396}$$

To get the solution $\Lambda = 0$, we can simply take this limit to get

$$R_{\Lambda=0} = \lim_{\Lambda\to 0}\left(\sqrt{\tfrac{3C}{\Lambda}}\,\sinh(2\sqrt{\Lambda/3}t)\right)^{1/2} = (2\sqrt{C})^{1/2}\,t^{1/2}\ . \tag{397}$$

For $\Lambda < 0$, we put $\sqrt{\Lambda} = i\sqrt{-\Lambda}$ and get

$$R_{\Lambda<0} = \left(\tfrac{3C}{-\Lambda}\right)^{1/4}\ \sin\left(2\sqrt{-\Lambda/3}\ t\right)^{1/2}\ . \tag{398}$$

## 6.6 Cosmological solutions with $\Lambda = 0$

The first thing to note is that we can re-scale the solution and we only need distinguish the three cases $k = 1,\ k = 0$ and $k = -1$

- If $k = 1$, there is a maximum value of $R$ for which $\dot{R}^2$ is positive or zero, and this means that the initial expansion is followed by a pt $R_{\text{max}}$ at which $\dot{R}$ is zero, and then a subsequent contraction.
- If $k = 0$ or $-1$, the universe continues to expand, but at different rates. $\dot{R} \to 1$ for $k = -1$, but $\dot{R} \to 0$ for $k = 0$, as $R \to \infty$.

We plot these qualitative features in figures 47 and 48:

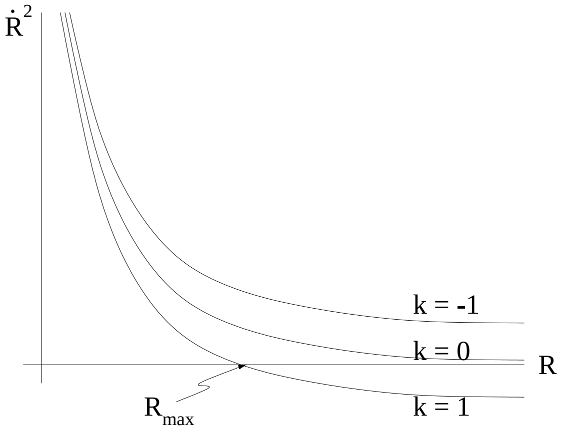



Figure 47:
The qualitative behaviour of $\dot{R}^2$ vs. $R$ for $k = 0, \pm 1$.

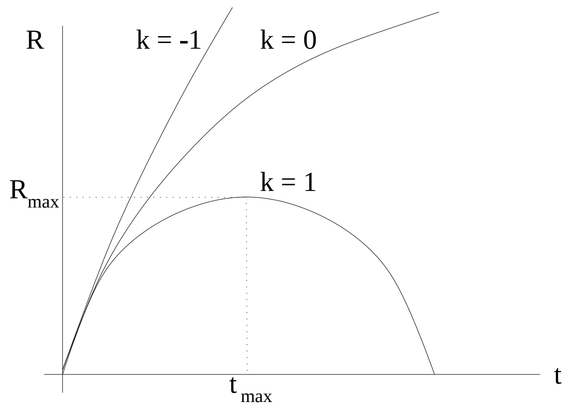



Figure 48:
The qualitative behaviour of $R$ vs. $t$ for $k = 0, \pm 1$.

We can also find the asymptotic behaviour of $R(t)$ as $R \to 0$, since we can neglect $k$ in the Friedmann eqn in this case. Neglecting $k$, we have

$$\dot{R}^2 = \frac{C}{R^\alpha} \qquad \text{where } \alpha = \begin{cases} 1 & \text{'Dust'} \\ 2 & \text{'Radiation'} \end{cases}$$

$$\Rightarrow \quad \dot{R} = \sqrt{C} R^{-\alpha/2}$$

$$\Rightarrow \quad \int R^{\alpha/2}\, \mathrm{d}R = \sqrt{C} \int \mathrm{d}t$$

$$\Rightarrow \quad \frac{2}{2+\alpha} R^{\frac{\alpha}{2}+1} = \sqrt{C}\, t$$

$$\Rightarrow \quad R = \left[ \left( \frac{\alpha+2}{2} \right) \sqrt{C} t \right]^{2/(2+\alpha)} = \begin{cases} \left( \sqrt{\frac{9C}{4}}\, t \right)^{2/3} & \text{Dust, } \alpha = 1 \\ \left( \sqrt{\frac{9C}{3}}\, t \right)^{1/2} & \text{Radiation, } \alpha = 2 \end{cases} \tag{399}$$

So, for all $k$,

$$\begin{aligned} R &\sim t^{2/3} \quad \text{for dust, small t} \\ R &\sim t^{1/2} \quad \text{for radiation, small t} \end{aligned}$$

Furthermore, equations (399) are the exact solutions for $k = 0$ for all $t$.

The exact solutions for $k = \pm 1$ and $\alpha = 1, 2$ are also known, but only in parametric form - that is to say, one can write

$$R = R(\psi) \ , \quad t = t(\psi)$$

such that for all $\psi$,

$$\dot{R}^2 = \frac{R'^2}{t'^2} = \frac{C}{R^\alpha} - k \ .$$

These solutions are:

| | $k = -1$ | $k = 0$ | $k = 1$ |
|---|---|---|---|
| dust | $R = \frac{C}{2}(\cosh\psi - 1)$<br>$t = \frac{C}{2}(\sinh\psi - \psi)$ | $R = \frac{C}{2}\frac{\psi^2}{2}$<br>$t = \frac{C}{2}\frac{\psi^3}{6}$ | $R = \frac{C}{2}(1 - \cos\psi)$<br>$t = \frac{C}{2}(\psi - \sin\psi)$ |
| radiation | $R = \sqrt{C}\sinh\psi$<br>$t = \sqrt{C}(\cosh\psi - 1)$ | $R = \sqrt{C}\,\psi$<br>$t = \sqrt{C}\,\frac{\psi^2}{2}$ | $R = \sqrt{C}\sin\psi$<br>$t = \sqrt{C}(1 - \cos\psi)$ |

## 6.7 The derivation of the equations for a dust–filled universe by Newtonian methods

The full field equations of general relativity for an isotropic homogeneous universe are (368) and (369),

$$-\frac{3\dot{R}^2}{R^2} - \frac{3k}{R^2} + \Lambda = -(8\pi G)\,\rho\;,$$
$$-\frac{2\ddot{R}}{R} - \frac{\dot{R}^2}{R^2} - \frac{k}{R^2} + \Lambda = \;(8\pi G)\,p\;.$$

In the case $p = \Lambda = 0$, these equations are equivalent to the two equations ($k$ being a constant of integration)

$$\dot{\rho} + 3\rho\,\frac{\dot{R}}{R} = 0\;, \tag{400}$$
$$\ddot{R} = -\frac{4\pi G}{3}\,R\,\rho\;. \tag{401}$$

It turns out that these two equations can be derived by Newtonian arguments as follows:

1. Since Einstein's equations reduce to Newton's equations for weak fields, and since the scale factor $R$ is directly related to the distance between any two dust particles, then provided we restrict to a small enough region and to small enough $R$, the equation for $R$ (being the separation of two dust particles very close together) should be given by Newton's laws.

2. Consider a dust particle and a sphere of radius $R$ surrounding it at time $t_0$. Now consider the dust particles in this sphere of radius $R$ at time $t_0$. As time evolves, the distance of the dust particles in this sphere from our original particle (which we can say is stationary at the origin) will change; we denote this distance by $R(t)$.

3. Since no matter crosses this sphere of dust particles, the total mass $M$ contained in the volume $(4\pi R^3/3)$ inside that sphere stays constant, and hence the density $\rho$ is given by

   $$\rho = \frac{\text{Mass}}{\text{Volume}} = \frac{M}{4\pi R^3/3}\;, \tag{402}$$

   for some constant $M$. This is the solution of equation (400).

4. Now we have to work out the equation for $R$. Consider a dust particle at distance $R$ – then it is a classical result of Newtonian gravity that a spherically symmetric mass distribution outside a sphere of radius $R$ exerts no net force inside. Hence the only gravitational force on the particle is due to the mass inside the sphere

5. The mass inside the sphere is given by equation (402), so that the gravitational acceleration on the particle at distance $R$ due to the mass inside $R$ is

   $$\ddot{R} = -\frac{G \times \text{Mass}}{R^2} = -\frac{4\pi G}{3}\,\rho\,R\;. \tag{403}$$

   This is equation (401)

## 6.8 The critical density

Since the sign of $k$ is so important for the future fate of the universe, a lot of effort has been put into trying to measure it.

This is often expressed in terms of $\Omega$, which is defined through

$$k/(3H^2) = \left( \left( \frac{8\pi G}{3H^2} \right) \rho + \frac{\Lambda}{3H^2} \right) - 1 = \Omega - 1 \ , \tag{404}$$

$$\Omega = \Omega_m + \Omega_\Lambda \ , \quad \Omega_m = \frac{\rho}{\rho_{\text{crit}}} \ , \ \ \Omega_\Lambda = \frac{\Lambda}{3H^2} \ , \tag{405}$$

where $H$ is Hubble's 'constant'.

The 'critical density' is the value of $\rho$ required such that $\Omega_m = 1$, i.e. that the value $k = 0$ is reached through matter energy density only:

$$\rho_{\text{crit}} = \frac{3}{8\pi G} \left( \frac{\dot{R}^2}{R^2} \right) = \frac{3}{8\pi G} H^2 \ .$$

Current experiments lead to $\Omega \simeq 1$, but $.2 \lesssim \Omega_m \lesssim .5$, $.5 \lesssim \Omega_\Lambda \lesssim .8$, but these results are very sensitive to both observational error and to the models which lie behind the analysis of the data, and so could well change in the next few years.

This is a subject of ongoing debate.

## 6.9 Exercises

**6.1 The Continuity equation.**

This question shows how the continuity equations can be derived directly by substituting the form of the energy momentum-tensor of a fluid,

$$T_{\mu\nu} = (p+\rho)U_\mu U_\nu - p\, g_{\mu\nu} \ , \tag{406}$$

into the equations for the conservation of the energy-momentum tensor:

$$\nabla_\mu T^{\mu\nu} = 0 \ . \tag{407}$$

We have already shown in a previous exercise that substituting the form (406) into eqn. (407) leads to:

$$0 = U^\mu \nabla_\mu \rho \ + \ (\nabla_\mu U^\mu)\,(p+\rho) = \nabla_\mu(\rho\, U^\mu) \ + \ p\,(\nabla_\mu U^\mu) \ . \tag{408}$$

- Using the formula true for any vector $V^\mu$ and metric $g_{\mu\nu}$

$$\nabla_\mu V^\mu = \frac{1}{\sqrt{|\det g_{\mu\nu}|}}\,\partial_\mu \left(\sqrt{|\det g_{\mu\nu}|}\ V^\mu\right) \ ,$$

in the special case of the 4–vector $U^\mu = (1,0,0,0)$ and the Robertson-Walker metric

$$\mathrm{d}s^2 = \mathrm{d}t^2 - R(t)^2 \frac{\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2}{(1+k\,r^2/4)^2} \ , \tag{409}$$

show that (408) leads to the continuity equation

$$\dot\rho + 3\frac{\dot R}{R}(p+\rho) = 0 \ , \tag{410}$$

**6.2 The Continuity equation for a general equation of state.**

**a)** Suppose that there existed an exotic form of matter for which the equation of state was

$$p = \alpha\,\rho \ , \tag{411}$$

for some constant $\alpha$. Find the relation for $p$ as a function of $R$ by solving the continuity equation in the form

$$\frac{\mathrm{d}\rho}{\mathrm{d}R} = -3\,\frac{(p+\rho)}{R} \ . \tag{412}$$

Check your solution for the two cases $\alpha = 0$ (dust) and $\alpha = 1/3$ (radiation) for which

$$\rho_{\mathrm{dust}}(R) = \frac{A}{R^3} \ , \qquad \rho_{\mathrm{radiation}}(R) = \frac{B}{R^4} \ . \tag{413}$$

**b)** Which values of $\alpha$ would leave the quanlitative behaviour of the time-evolution of the universe unchanged? Which values of $\alpha$ would alter the qualitative behaviour of the time-evolution of the universe? What would be the time evolution in this case?
[ hint - plot $\dot{R}^2$ against $R$ ]

### 6.3 A strange property of the Robertson Walker universe with $k = 1$

Consider a Robertson-Walker universe with $k = 1$. It is possible in principle for light emitted from a distant star to reach us from two diametrically opposite directions.

Supposing that $R = ct$, show that the redshifts $z_1$ and $z_2$ of the light arriving at us at any given time satisfy

$$(1 + z_1)(1 + z_2) = \exp(2\pi) \; .$$

### 6.4 The solutions for $R(t)$ with $\Lambda = 0$.

The exact solutions for $R(t)$ with $\Lambda = 0$ in the cases $k = 0, \pm 1$ for dust ($p = 0$) and radiation ($p = \frac{1}{3}\rho$) in parametric form are

| | $k = -1$ | $k = 0$ | $k = 1$ |
|---|---|---|---|
| dust | $R = \frac{\kappa}{6}(\cosh\psi - 1)$<br>$t = \frac{\kappa}{6}(\sinh\psi - \psi)$ | $R = \frac{\kappa}{6}\frac{\psi^2}{2}$<br>$t = \frac{\kappa}{6}\frac{\psi^3}{6}$ | $R = \frac{\kappa}{6}(1 - \cos\psi)$<br>$t = \frac{\kappa}{6}(\psi - \sin\psi)$ |
| radiation | $R = \sqrt{\frac{\kappa}{3}}\sinh\psi$<br>$t = \sqrt{\frac{\kappa}{3}}(\cosh\psi - 1)$ | $R = \sqrt{\frac{\kappa}{3}}\psi$<br>$t = \sqrt{\frac{\kappa}{3}}\frac{\psi^2}{2}$ | $R = \sqrt{\frac{\kappa}{3}}\sin\psi$<br>$t = \sqrt{\frac{\kappa}{3}}(1 - \cos\psi)$ |

**a)** Check that the parametric forms for $k = 0$ give the exact solutions

$$R_{\text{dust}}(t) = \left(\sqrt{\frac{3\kappa}{4}}t\right)^{2/3} \quad , \quad R_{\text{radiation}}(t) = \left(\sqrt{\frac{4\kappa}{3}}t\right)^{1/2} \quad .$$

**b)** Check that the exact parametric solutions for $k = \pm 1$ have the correct asymptotic behaviour as $R \to 0$.

[hint – expand the parametric solutions for small $\psi$]

**c)** Check that substituting the correct form of the matter density $\rho$ from (413), the parametric solutions for $k = \pm 1$ satisfy the Friedmann equation (with $\Lambda = 0$) for a suitable value of $\rho$

$$\dot{R}^2 = \frac{8\pi G}{3}\rho R^2 - k \; .$$

### 6.5 Question 3, 1998 exam

The metric of a static spacetime with coordinates $x^\mu$ may be written in the form

$$ds^2 = g_{\mu\nu}\, dx^\mu\, dx^\nu = g_{00}\, dt^2 \ + \ g_{ij}\, dx^i dx^j \ , \tag{1}$$

where $t = x^0$ and where $g_{\mu\nu}$ is independent of $t$.

(a) From the definition $g^{\mu\nu}\, g_{\nu\rho} = \delta^\mu_\rho$, find $g^{00}$ and $g^{0i}$ in terms of $g_{\mu\nu}$.

(b) Show that the Christoffel symbols $\Gamma^0_{00}$, $\Gamma^0_{0i} = \Gamma^0_{i0}$, $\Gamma^0_{ij}$ and $\Gamma^i_{00}$ satisfy

$$\Gamma^0_{00} = 0 \ , \quad \Gamma^0_{0i} = \frac{1}{2\, g_{00}} \frac{\partial}{\partial x^i} g_{00} \ , \quad \Gamma^0_{ij} = 0 \ , \quad \Gamma^i_{00} = -\frac{1}{2}\, g^{ij} \frac{\partial}{\partial x^j} g_{00} \ .$$

In the coordinates $(t, x^i)$, the vector field $\xi^\mu$ has components $(1, 0, 0, 0)$, that is

$$\xi^0 = 1 \ , \quad \xi^i = 0 \ . \tag{2}$$

(c) Find the components of $\xi_\mu = g_{\mu\nu}\, \xi^\nu$.

A vector $\zeta^\mu$ is said to be a "Killing vector" for a metric $g_{\mu\nu}$ if

$$\nabla_\mu\, \xi_\nu \ + \ \nabla_\nu\, \xi_\mu = 0 \ ,$$

where $\nabla_\mu\, \xi_\nu = \partial \xi_\nu / \partial x^\mu - \Gamma^\rho_{\mu\nu}\, \xi_\rho$.

(d) Show that the vector field $\xi^\mu$ defined by (2) is a Killing vector for the metric (1).

### 6.6 The Einstein tensor for a Friedmann-Robertson-Walker spacetime

Consider the FRW metric in the form

$$\mathrm{d}s^2 = \mathrm{d}t^2 - \frac{R^2}{f^2}(\mathrm{d}x^2 + \mathrm{d}y^s + \mathrm{d}z^2) \ , \quad f = (1 + \frac{kr^2}{4}) \ .$$

(a) Find all the Christoffel symbols.

[ Answers: The non-zero Christoffel symbols are

$$\Gamma^t_{xx} = R\,\dot{R}\,/\,f^2 \ , \quad \Gamma^x_{xt} = \dot{R}\,/\,R \ ,$$

$$-\Gamma^x_{xx} = \Gamma^x_{yy} = \Gamma^x_{zz} = k\,x\,/(2\,f) \ , \quad \Gamma^x_{xy} = -k\,y\,/(2\,f) \ , \quad \Gamma^x_{xz} = -k\,z\,/(2\,f)$$

and those related by symmetry ]

(b) Find the components of the Ricci tensor $R_{tt}$, $R_{tx}$, $R_{xx}$ and $R_{xy}$.

[ Answer:

$$R_{tt} = -3\,\ddot{R}\,/\,R \ , \quad R_{tx} = 0 \ , \quad R_{xx} = (2\dot{R}^2 + R\ddot{R} + 2k)/f^2 \ , \quad R_{xy} = 0 \ .$$

(c) Hence show that

$$G_{tt} = \frac{3\dot{R}^2}{R^2} + \frac{3k}{R^2} \ , \quad \frac{G_{xx}}{g_{xx}} = \frac{2\ddot{R}}{R} + \frac{\dot{R}^2}{R^2} + \frac{k}{R^2} \ .$$

## Contents