

Spacetime Geometry and General Relativity (CM334A)

Neil Lambert
Department of Mathematics
King's College London
Strand
London WC2R 2LS, U. K.
Email: `neil.lambert@kcl.ac.uk`

Contents

1	Introduction	3
2	Special Relativity	4
3	Elementary geometry	10
3.1	\mathbb{R}^2 the hard way	10
3.2	S^2	13
4	Manifolds and Tensors	14
4.1	Manifolds	14
4.2	Tensors	15
4.3	Covariant Derivatives	17
4.4	Geodesics	20
4.5	Causal Curves Again	23
4.6	Integration	24
4.7	Riemann Normal Coordinates	24
4.8	Curvature	26
5	General Relativity	28

6	Schwarzschild Solution	29
6.1	Geodesics	31
6.2	Newtonian Limit	33
6.3	Perihilion Shift of Mercury	33
6.4	Bending of Light	35
6.5	Gravitational Redshift and Time Dilation	37
6.6	Schwarzschild Solution as a Black Hole	38
7	Cosmology: FRW	42

1 Introduction

This course is meant as introduction to what is widely considered to be the most beautiful and imaginative physical theory ever devised: General Relativity.¹ It is well-known as the brain-child of Einstein who developed it most or less single-handedly in the years leading up to its publication in 1915. It seems that everyone knows something about it. Perhaps most impressively the theory was developed without any experimental input (although it immediately explained at least one known observational anomaly). Rather Einstein made ‘thought experiments’² which led him to the notion of the equivalence principle and then the hypothesis that spacetime is curved. From here one simply adapts Newton’s famous $F = ma$ law to predict Newton’s gravitational law $F = -G_N m_1 m_2 / r^2$ along with corrections. In this way the motion of falling apples, astronauts, moons, planets and stars are simply tracing the natural ‘straight lines’ of a curved spacetime. Furthermore it opened up the entire Universe and its history to study in the form of Cosmology which is now a highly developed scientific field.

The plan of the course is as follows. We will first review Special Relativity. It is assumed that you have a reasonable knowledge of this as well as tensors such as those encountered in Electromagnetism (*e.g.* $F_{\mu\nu}$). We will review (or perhaps introduce) covariant index notation and spacetime. Next we will discuss Einstein’s famous thought experiment and follow his logic that leads to the notion that, in the presence of gravity, spacetime is curved. We will then spend some time developing the mathematics needed to understand curved spaces. These are known as manifolds in the mathematical literature although we will try to avoid using the abstract mathematical machinery. We will then have to spend some getting used to tensors in curved spacetime. This includes the all important notions of covariant derivative and curvature.

We can then write down Einstein’s theory. For the rest of the course we will examine two classic solutions. The first is the Schwarzschild solution that models the curvature about a spherical mass. Here we can derive Newton’s gravitational law as the leading order effect. However we will see that there are corrections which lead to new predictions. The first is the so called perihelion shift. This is where the planets depart slightly from a pure elliptical orbit about the sun. This explained observational anomaly of Mercury that was known already since the 1800’s (although a triumph in retrospect at the time it was not clear that there wasn’t another explanation, such as a new planet). Secondly we will see that light is bent as it passes by the sun. This was a signature prediction of General Relativity and was confirmed in 1919 thereby establishing the theory and confirming the picture of spacetime as curved. The second solution that we will discuss is the Freedman-Robertson-Walker (FRW) metric which describes the cosmological structure of the entire Universe, predicting that it started with a Big Bang.

That’s quite a lot for one idea!

¹WARNING: this course can make you want to do theoretical physics for the rest of your life.

²Unfortunately this has inspired generations of crackpots ever since.

2 Special Relativity

Let us start by a review of Special Relativity. To do physics one introduces a coordinate system typically x, y, z and t . Using such a coordinate system one can measure positions and velocities, make predictions and test them. That is physics. Here t of course represents time and x, y, z space. Before Einstein it was thought that time was absolute. And space was absolute. That meant that everyone surely agreed on the passage of time. And space was, in some sense, a solid object that we all play around in. Of course the coordinate systems that people use need not agree. However there are well defined rules that tell you how to convert quantities in one coordinate system (sometimes called a frame) to those of another. But time was always just time and measured in seconds, minutes, hours, days,... and we surely all agree on the absolute passage of time no?...

A fundamental idea of special relativity is that the law of physics do not depend on any particular choice of frame. This is so much a part of modern physics that its hard to think otherwise. It is formulated in the so-called principle of relativity. To quote Einstein himself:

- *Principle of Relativity:* If a system of coordinates K is chosen so that, in relation to it, physical laws hold good in their simplest form, the same laws hold good in relation to any other system of coordinates K' moving in uniform translation relatively to K .

Here moving in uniform translation relatively to K means that it moves with a constant velocity with respect to K . In the modern parlance we'd say that the laws of physics are covariant with respect to coordinate transformations.

One important feature of the principle of relativity is that there is no preferred frame that the universe uses. Although it is now almost forgotten the prevailing view in the 19th century was that there must be some medium throughout space and time, called the ether, where electromagnetic waves 'wave' in. This then defines a preferred reference frame. However it is well-known that the famous Michelson-Morley experiment failed to detect any motion of the earth through the ether. Einstein's theory simply did away with the ether.

Apart from this the Principle of Relativity alone is not so new and the older theory theories of Newton and others satisfy a form of relativity known as Galilean relativity. In Galilean relativity if the coordinates of K are (x, y, z, t) and those of K' (x', y', z', t') and K' is moving along the x -axis with speed v then the relation is

$$\begin{aligned}t' &= t \\x' &= x + vt \\y' &= y \\z' &= z\end{aligned}\tag{2.1}$$

There are also other possible transformations, for example the spatial coordinates (x', y', z') could be related to those of K by a rotation. But $t' = t$ assuming you measure time in seconds in both frames!

Note that with this change of coordinates we can compute the ‘addition’ rule of velocities. Suppose in frame K a particle is moving with speed u along the x-axis. If at $t_1 = 0$ it is at $x_1 = 0$ then at time t_2 it will be at $x_2 = ut_2$. In frame K' we find that it is still at $x'_1 = 0$ when $t'_1 = 0$ but now at time $t'_2 = t_2$ it is at $x'_2 = ut_2 + vt_2$. Thus in the frame K' one observes the speed

$$u' = x'_2/t'_2 = u + v$$

which of course is the familiar rule for addition of velocities.

The second idea is more radical and defines Special Relativity:

- *Principle of a constant speed of light:* The speed of light as measured in any two frames that are move with constant velocity respect to each other is the same universal constant called c

What made Einstein think of this? Well the answer is that Maxwell’s equations of electromagnetism (developed while Maxwell was at King’s) have a factor of c^{-1} on the right-hand-side. Thus they don’t make any sense in a frame where $c = 0$. Thus they can’t be Galilean invariant.

Clearly this violates the usual rules of changing frames where one would simply adds the velocities. This leads to a new set of rules about how to transform between K and K'. In fact Maxwell’s equations were known to be invariant under a set of coordinate transformations known Lorentz transformations which predates Einstein. However Einstein saw that Maxwell’s equations and Lorentz transformations were fundamental (thus in a sense King’s College is the birth place of Relativity). In particular for a frame K' moving along the x-axis of a frame K the Lorentz transformations are

$$\begin{aligned} x' &= \gamma(x + vt) \\ t' &= \gamma(t + vx/c^2) \\ y' &= y \quad z' = z \end{aligned} \tag{2.2}$$

where $\gamma = 1/\sqrt{1 - v^2/c^2}$.

Problem: Show that the addition law for velocities is

$$u' = \frac{u + v}{1 + uv/c^2}$$

Indeed one sees from this that $c' = c$ and also $u' < c$ provided that $u, v < c$. Thus one can never go faster than the speed of light by switching from one frame to another (assuming no frame is moving faster than the speed of light). This transformation also leads to the well-known effects of time-dilation and length contraction. We will not review these calculations here as they appear in any course on Special Relativity.

Soon afterwards a deeper geometrical structure was shown by Minkowski to underlie Special Relativity. In particular one sees from the Lorentz transformations that $t' \neq t$

and furthermore t' is mixed into t and x . Thus time is no longer separate and distinct from space. Indeed given the principle of relativity one cannot even pick frame with a preferred or absolute notion of time.

Rather Minkowski introduced spacetime which consists of all four coordinates together: (t, x, y, z) . It is helpful (on dimensional grounds) to introduce the coordinates

$$x^0 = ct$$

along with $x^1 = x$, $x^2 = y$ and $x^3 = z$. Now three-dimensional Euclidean space can be thought of as \mathbf{R}^3 where the rule for calculating lengths is

$$\Delta l^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2$$

It is important to note that length is a 'scalar' quantity, meaning that in any admissible frame all observers will agree on the length of an object. Minkowski generalized this to spacetime with

$$\Delta s^2 = -c^2(\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 = \eta_{\mu\nu}\Delta x^\mu\Delta x^\nu$$

where

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

The key idea now is that Δs is a scalar quantity that all observers will agree on. Note the minus sign. Although time can be, in some sense, be rotated into space, it is also clearly different.

An important part of Minkowski spacetime is that $(\Delta s)^2$ can be negative, positive or zero. Vectors with such lengths are called timelike, spacelike or null respectively. In particular the set of vectors with null length defines a surface known as the lightcone centred at the origin. Thus spacetime is split into two distinct areas that are separated by the lightcone.

The point is that signals can only travel at or below the speed of light and thus must be timelike. This means that an event located at the origin can only be seen in the $t > 0$ timelike (interior of the light cone) region. And conversely only events in the $t < 0$ timelike (interior of the lightcone) can be seen by an observer at the origin. The rest of Minkowski space (the spacelike region outside the lightcone) cannot communicate with the origin. It is said to not be in causal contact with the origin whereas the interior region is in causal contact.

In ordinary three-dimensional space observers do not agree on the length of an object along a given axis, such as the x-axis, because they disagree on what they've called the x-axis. For example if you take a rod and rotate it you can change its 'length' along any given axis that is, for example defined by the walls of the room you are in. The overall length, as measured by Δl is of course invariant but the projection of the length onto a particular axis can be changed by a choice of frame.

This all happens in Minkowski space too. The only difference is now time is another coordinate that can be ‘rotated’ in to the spatial directions. Thus the projection of the length of a rod onto a spatial or temporal axis can change. This is the origin of length contraction and time dilation. However there are invariant notions of length defined by Δs .

From this perspective the Lorentz transformations arise from requiring that Δs is the same in all frames. Thus if $x'^{\mu} = \Lambda^{\mu}_{\nu} x^{\nu}$ then we must have that

$$\begin{aligned}
 (\Delta s')^2 &= \eta_{\mu\nu} \Delta x'^{\mu} \Delta x'^{\nu} \\
 &= \eta_{\mu\nu} \Lambda^{\mu}_{\sigma} \Lambda^{\nu}_{\rho} \Delta x^{\sigma} \Delta x^{\rho} \\
 &= (\Delta s)^2 \\
 &= \eta_{\sigma\rho} \Delta x^{\sigma} \Delta x^{\rho}
 \end{aligned}
 \tag{2.3}$$

Here we have used Einstein’s summation convention whereby a repeat upstairs and downstairs index is summed over. Thus we must have, since Δx^{μ} is arbitrary,

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

If we think in term of matrices then this equation is simply $\eta = \Lambda^T \eta \Lambda$ where we have used matrix multiplication.

More generally a tensor, such as the electromagnetic field strength, transforms in the same, covariant, way: $F'_{\lambda\sigma} = F_{\mu\nu} \Lambda^{\mu}_{\lambda} \Lambda^{\nu}_{\sigma}$. Thus in a sense the rule Special Relativity is that the metric tensor is invariant under Lorentz transformations (however other tensors are not).

Problem: Consider a transformation of the form

$$\begin{aligned}
 x'^0 &= ax^0 + bx^1 \\
 x'^1 &= cx^0 + dx^1 \\
 x'^2 &= x^2 \quad x'^3 = x^3
 \end{aligned}
 \tag{2.4}$$

Show that the condition (2) implies that we can parameterize (for simplicity assume that $a, d > 0$)

$$a = d = \cosh \beta \quad b = c = \sinh \beta$$

for some parameter β . Using this rederive the Lorentz transformations by writing $\tanh \beta = v/c$.

This was the situation about just before 1915. However it has the problem in that it cannot describe acceleration and therefore, according to Newtons laws, any force. Although it will work in situations where the acceleration is small and also for electromagnetism which is relativistic. Thus Einstein was led to a more general principle of relatively.

The new principle is motivated by the following thought experiment (slightly modernised): You’ve seen pictures of the astronauts in the space shuttle. They are weightless

but why? The naive answer is that it is because they are in outer space away from the pull of earth's gravity. But then if this is the case, why do they say in orbit about the earth? This is clearly not the case. Rather what happens is that in Newton's laws one has

$$F = ma = -\frac{G_N M m}{r^2}$$

where M is the mass of the earth, G_N is Newton's constant, r the distance from the centre of the earth and m the mass of the astronaut. But m is also the mass of anything else in the space shuttle. The point is that m drops out from this equation and one simply has

$$a = -\frac{G_N M}{r^2}$$

for all objects, astronauts or otherwise. Thus, in free fall, they all follow the same path, *i.e.* there is a universal path valid for all observers in free fall. This of course is essentially just the famous leaning tower of Pisa experiment of Galileo, revisited many years later. Einstein's genius (among other things) was to notice that this is a rather strange feature that points to a deeper more fundamental interpretation.

There is a well-known issue here that I am not going to give much time to. It could be argued that the m that appears in $F = ma$ is different from the m that appears in $F = -G_N M m / r^2$. The former is known as the inertial mass m_i whereas the latter is called the gravitational mass m_g . It should be clear from our discussion that we are heavily relying on the assumption that

$$m_i = m_g$$

for all objects. All I can say is: "yes we are". But all evidence, to a high degree of accuracy suggests that this is true. So let us say no more about this.

Conversely if we sit in a car and it accelerates we feel a force, just like the force of gravity, only it acts to push us backwards, rather than down. The great realisation of Einstein was that a single observer cannot perform any local experiment which can tell if s/he is in free fall in a gravitational field or if there is no gravitational force at all. Conversely they also cannot tell, using local experiments, the differences between acceleration and a gravitational field.

The important concept here is locally. Obviously a freely falling observer can tell that they are in the gravitational field of the earth if they can see the whole earth, along with their orbit. Certainly they will know if they ever hit the earth. The point is that the laws of physics should be local, *i.e.* they should only depend on the properties of spacetime and fields as evaluated at each single point independently. This leads to the 'happiest thought of Einstein's life', the famous principle of equivalence

- *Principle of Equivalence:* The laws of physics cannot distinguish between motion in a gravitational field and acceleration

Thus we aim to reconcile the following observation. A particle which is freely falling in a gravitational field is physically equivalent to a particle which feels no force. And furthermore the paths followed by all free-falling objects are the same. Now by Newton's

law no force should mean that the particle does not accelerate, that it moves in a ‘straight line’. The key idea here is to realise that spacetime must therefore be curved and that ‘straight lines’ are not actually straight in the familiar sense of the word. The space shuttle moves in orbit around the earth because locally, that is from one instant to the next, it is following a straight path - the path of shortest length - in a spacetime that is curved. Just as an airplane travels in a great circle that passes over Greenland whenever it flies from Washington DC to London.

Mathematically we can impose these ideas by noting that a curved space is described by the concept of a manifold. Roughly speaking a manifold is a space that in a neighbourhood of each point looks like \mathbb{R}^n for some n . Here we have $n = 4$. For infinitesimal variations dx^μ the proper distance is

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu \quad (2.5)$$

where $g_{\mu\nu}$ is called the metric and locally determines the geometry of spacetime by determining the lengths and angles in an infinitesimal neighbourhood of each point. By definition $g_{\mu\nu}$ is symmetric: $g_{\mu\nu} = g_{\nu\mu}$. We will review manifolds and their tensors in more detail soon.

The general principle of relativity states that the laws of physics are invariant under an arbitrary - but invertible - coordinate transformation (not just linear)

$$x^\mu \longrightarrow x'^\mu = x'^\mu(x^\nu) \quad (2.6)$$

under which we have that ds^2 is invariant. The same calculation as above leads to

$$ds'^2 = g'_{\mu\nu} dx'^\mu dx'^\nu \quad (2.7)$$

$$= g'_{\mu\nu} \frac{\partial x'^\mu}{\partial x^\rho} dx^\rho \frac{\partial x'^\nu}{\partial x^\sigma} dx^\sigma \quad (2.8)$$

$$= g_{\rho\sigma} dx^\rho dx^\sigma \quad (2.9)$$

Note that the transformation needs to be invertible so that the Jacobian

$$\Lambda^\mu{}_\rho = \frac{\partial x'^\mu}{\partial x^\rho} \quad (2.10)$$

is an invertible 4 matrix whose inverse is

$$\Lambda_\sigma{}^\nu = \frac{\partial x^\nu}{\partial x'^\sigma} \quad (2.11)$$

since

$$\delta_\lambda^\mu = \frac{\partial x'^\mu}{\partial x^\rho} \frac{\partial x^\rho}{\partial x'^\lambda} \quad \text{and} \quad \delta_\lambda^\mu = \frac{\partial x'^\rho}{\partial x^\lambda} \frac{\partial x^\mu}{\partial x'^\rho} \quad (2.12)$$

Such a change of variables is called a diffeomorphism. We now see that the invariance of the infinitesimal proper distance implies that

$$g'_{\rho\sigma} = \frac{\partial x^\mu}{\partial x'^\rho} \frac{\partial x^\nu}{\partial x'^\sigma} g_{\mu\nu} \quad (2.13)$$

This is the defining property of a tensor field and we will discuss these in more detail soon.

Thus we see that we have generalised the linear transformation property $x'^{\mu} = \Lambda^{\mu}_{\nu} x^{\nu}$ of Special Relativity to arbitrary, but invertible, transformations by including a general metric $g_{\mu\nu}(x)$ rather than a fixed one $\eta_{\mu\nu}$ (this is essentially the same as gauging in a gauge theory). The theory of General Relativity treats the metric $g_{\mu\nu}$ as a dynamical object and its evolution is obtained from Einstein's equation.

3 Elementary geometry

3.1 \mathbb{R}^2 the hard way

To understand these ideas we need to get used to metrics and curved spaces. We will start with an example that we all know, where there is a non-constant metric, although the space is not curved: polar coordinates for \mathbb{R}^2 . The Euclidean rule for lengths is

$$ds^2 = dx^2 + dy^2$$

So here the metric is (using matrix notation)

$$g = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

However one frequently wants to use so-called polar coordinates:

$$x = r \cos \theta \quad y = r \sin \theta$$

What is the metric in these coordinates? Well we can easily compute that

$$dx = dr \cos \theta - r \sin \theta d\theta \quad dy = dr \sin \theta + r \cos \theta d\theta$$

and thus we see that

$$\begin{aligned} ds^2 &= (dr \cos \theta - r \sin \theta d\theta)^2 + (dr \sin \theta + r \cos \theta d\theta)^2 \\ &= dr^2 + r^2 d\theta^2 . \end{aligned} \tag{3.1}$$

We now have (again using matrix notation)

$$g = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}$$

and thus has non-constant components. But we are clearly still just taking about Euclidean \mathbb{R}^2 !

These coordinates are good for somethings, but not everything. Indeed they actually suffer a pathology at $r = 0$. Here θ is not defined. In particular the inverse map is

$$r = \sqrt{x^2 + y^2} \quad \theta = \arctan(y/x)$$

but y/x is ambiguous at $r = 0$ *i.e.* $x = y = 0$ (you might object to $x = 0$ but it is clear that $\theta = \pm\pi/2$ there, depending on the sign of y). Thus this coordinate system is not ‘global’ meaning that it does not cover all of the space one is interested in. In particular it doesn’t cover $r = 0$. However this is sufficiently simple that one can work around it and pretend as if all is okay.

In physics we are often interested in two types of computation. The first is to solve differential equations such as

$$\nabla^2\Phi = J$$

where Φ is some field and J is some source. For example in electrostatics Maxwell’s equation for the time-component of the vector potential is $\nabla^2 A_0 = \rho$, where ρ is the charge density. In Cartesian coordinates we have (in \mathbb{R}^2)

$$\nabla^2\Phi = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) \Phi$$

let us see what this is in polar coordinates. We first note that, by the chain rule,³

$$\begin{aligned} \frac{\partial}{\partial x} &= \frac{\partial r}{\partial x} \frac{\partial}{\partial r} + \frac{\partial \theta}{\partial x} \frac{\partial}{\partial \theta} \\ &= \frac{x}{r} \frac{\partial}{\partial r} - \frac{y}{x^2} \frac{1}{1 + y^2/x^2} \frac{\partial}{\partial \theta} \\ &= \cos\theta \frac{\partial}{\partial r} - \frac{1}{r} \sin\theta \frac{\partial}{\partial \theta} \end{aligned} \tag{3.2}$$

Similarly

$$\frac{\partial}{\partial y} = \sin\theta \frac{\partial}{\partial r} + \frac{1}{r} \cos\theta \frac{\partial}{\partial \theta}$$

Next we find

$$\begin{aligned} \frac{\partial^2}{\partial x^2} &= \left(\cos\theta \frac{\partial}{\partial r} - \frac{1}{r} \sin\theta \frac{\partial}{\partial \theta} \right) \left(\cos\theta \frac{\partial}{\partial r} - \frac{1}{r} \sin\theta \frac{\partial}{\partial \theta} \right) \\ &= \cos^2\theta \frac{\partial^2}{\partial r^2} + \frac{1}{r^2} \sin^2\theta \frac{\partial^2}{\partial \theta^2} + \frac{2}{r^2} \sin\theta \cos\theta \frac{\partial}{\partial \theta} + \frac{1}{r} \sin^2\theta \frac{\partial}{\partial r} \end{aligned} \tag{3.3}$$

and similarly

$$\begin{aligned} \frac{\partial^2}{\partial y^2} &= \left(\sin\theta \frac{\partial}{\partial r} + \frac{1}{r} \cos\theta \frac{\partial}{\partial \theta} \right) \left(\sin\theta \frac{\partial}{\partial r} + \frac{1}{r} \cos\theta \frac{\partial}{\partial \theta} \right) \\ &= \sin^2\theta \frac{\partial^2}{\partial r^2} + \frac{1}{r^2} \cos^2\theta \frac{\partial^2}{\partial \theta^2} - \frac{2}{r^2} \sin\theta \cos\theta \frac{\partial}{\partial \theta} + \frac{1}{r} \cos^2\theta \frac{\partial}{\partial r} \end{aligned} \tag{3.4}$$

Thus we find

$$\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2}$$

³As well as the fact that $(\arctan z)' = 1/(1 + z^2)$.

which is often rewritten as

$$\nabla^2 = \frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2}$$

Hopefully you've seen that (or maybe the 3D version) before.

The second thing one needs to do is compute integrals of some quantity over some region of space. Suppose that one has some function $f(x, y)$ that we want to integrate over some region:

$$\int f = \int \int dx dy f(x, y)$$

But let us now do it in polar coordinates. We have already computed dx and dy in terms of dr and $d\theta$. From calculus you know that to change the integration measure you must introduce the jacobian factor

$$\int f = \int \int dr d\theta \left| \det \left(\frac{\partial(x, y)}{\partial(r, \theta)} \right) \right| f(r, \theta)$$

Here $\partial(x, y)/\partial(r, \theta)$ is the matrix formed by taking all derivatives:

$$\begin{aligned} \begin{pmatrix} \frac{\partial(x, y)}{\partial(r, \theta)} \end{pmatrix} &= \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix} \\ &= \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix} \end{aligned} \tag{3.5}$$

Thus we see that

$$\det \left(\frac{\partial(x, y)}{\partial(r, \theta)} \right) = r$$

and so

$$\int f = \int \int dr d\theta r f(r, \theta)$$

The thing to notice here is that $r = \sqrt{\det g}$ (using a matrix notation for the metric).

Why am I bothering to go through this, what's it got to do with curved space? Well although this space is not curved, in these coordinates the Laplacian and volume form have non-trivial form due to the non-constant metric. We will recover these forms later from so-called covariant derivatives and more general considerations of Riemannian geometry.

Problem: Consider spherical coordinates on \mathbb{R}^3 :

$$x = r \sin \theta \cos \varphi \quad y = r \sin \theta \sin \varphi \quad z = r \cos \theta$$

Following the same steps show that

$$\nabla^2 = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \varphi^2}$$

and

$$\int f = \int \int \int dr d\theta d\varphi \sqrt{\det g} f(r, \theta, \varphi)$$

3.2 S^2

Let us next consider something slightly less trivial that is curved: S^2 . To construct S^2 of radius R we start in \mathbb{R}^3 and impose the constraint

$$x^2 + y^2 + z^2 = R^2$$

We can solve this by taking

$$x = R \sin \theta \cos \varphi \quad y = R \sin \theta \sin \varphi \quad z = R \cos \theta$$

Problem: Show that the entire space is covered by taking $\theta \in [0, \pi]$ and $\varphi \in [0, 2\pi)$. Show that the coordinates break down at $\theta = 0, \pi$.

We start with the normal Euclidean metric on \mathbb{R}^3 :

$$ds^2 = dx^2 + dy^2 + dz^2$$

We now substitute in

$$dx = R \cos \theta \cos \varphi d\theta - \sin \theta \sin \varphi d\varphi \quad dy = R \cos \theta \sin \varphi d\theta + \sin \theta \cos \varphi d\varphi \quad dz = -R \sin \theta d\theta$$

to find (note first that the cross $d\theta d\varphi$ terms cancel)

$$\begin{aligned} ds^2 &= R^2 (\cos^2 \theta d\theta^2 + \sin^2 \theta d\varphi^2 + \sin^2 \theta d\theta^2) \\ &= R^2 (d\theta^2 + \sin^2 \theta d\varphi^2) \end{aligned} \tag{3.6}$$

This looks similar to polar coordinates on \mathbb{R}^2 with $r = \theta$ except that the metric coefficient of $d\varphi^2$ is $\sin^2 r$ and not r^2 . Thus in effect there are two origins: one at $\theta = 0$ and the other at $\theta = \pi$. Both of these are so-called coordinate singularities where φ is not well-defined.

Another way to think of this is that at a fixed value of θ there is a circle defined by the $\varphi \in [0, 2\pi)$ coordinate with radius $R \sin \theta$. At $\theta = 0$ this circle has shrunk to zero size. As we increase θ the circle grows until it reaches at maximum size at $\theta = \pi/2$ and then it shrinks again until it disappears at $\theta = \pi$.

Let us see now why airplanes take the paths they do when flying. Suppose we consider a flight path defined by a curve $\theta = f(\varphi)$. The infinitesimal length of a segment of this curve is

$$ds = R \sqrt{f'^2 + \sin^2 f} d\varphi$$

and hence the total length is

$$l = R \int \sqrt{f'^2 + \sin^2 f} d\varphi$$

We wish to minimize this integral. To do this we can simply use the Euler-Lagrange equation which gives:

$$-\frac{d}{d\varphi} \left(\frac{f'}{\sqrt{f'^2 + \sin^2 f}} \right) + \frac{\sin f \cos f}{\sqrt{f'^2 + \sin^2 f}} = 0$$

This is clearly a rather tough looking equation. However we can find at least one solution by setting f to a constant. This implies $f = 0, \pi/2, \pi$. The first and last case mean that there is no curve at all: the plane is stuck at the north or south pole (in fact the coordinates break down there and also the denominator vanishes so we should ignore it). The middle case implies that it circles the equator.

In fact, with a little thought, this is enough. Since the sphere has an $SO(3)$ symmetry we can always pick our coordinates so that Washington DC and London both lie on the ‘equator’ defined by $\theta = \pi/2$.⁴ We see that the shortest path is then a so-called ‘great circle’ between two points. That is a circle whose circumference is R . That is indeed the path that planes take (when air traffic control lets them and without volcanoes). We will return to this example later when we have some more experience.

4 Manifolds and Tensors

Warning: This is a physicists version of a deep and beautiful mathematical subject. You won’t need to know more in the course but if you’d like to know more you can take Manifolds (CM437Z/CMMS18). No apologies will be made here for brutalizing this subject, I will just say a few words in passing before we simply go ahead and develop tensor calculus without knowing what we are really doing.

4.1 Manifolds

We have studied some elementary examples of differential geometry. However in the S^2 case we thought of it as embedded inside \mathbb{R}^3 . It then inherited its geometry from the Euclidean metric of \mathbb{R}^3 . This is said to be extrinsic geometry. However one need not think this way. One can simply throw away the embedding and think of S^2 as its own object with its own intrinsic geometry.

An n -dimensional manifold is a space that locally looks like \mathbb{R}^n .

Formally the definition involves taking an open cover of a topological space \mathcal{M} , that is a set of pairs (U_i, ϕ_i) where U_i is an open set of \mathcal{M} and $\phi_i : U_i \rightarrow \mathbb{R}^n$ is a homeomorphism onto its image, *i.e.* it is continuous, invertible (when restricted to its image in \mathbb{R}^n) and its inverse is continuous. These are subject to two key constraints:

i) $\mathcal{M} = \cup_i U_i$

⁴This of course is not the standard equator that is used in geography. It may seem politically motivated but we simply want to find the simplest mathematical solution.

ii) If $U_i \cap U_j \neq \emptyset$ then $\phi_j \circ \phi_i^{-1} : \phi_i(U_i \cap U_j) \subset \mathbb{R}^n \rightarrow \phi_j(U_i \cap U_j) \subset \mathbb{R}^n$ is differentiable (for our purpose we assume that all the partial derivatives exist to all orders).

What does this mean? Each point $p \in \mathcal{M}$ is contained in an open set $U_i \subset \mathcal{M}$ called a neighbourhood. The map ϕ_i then provides coordinates for the point p and all the other points in that neighbourhood:

$$\phi_i(p) = (x^1(p), x^2(p), x^3(p), \dots, x^n(p)) \in \mathbb{R}^n \quad (4.1)$$

The key point of manifolds is that there can be many possible coordinate systems for the same point and its neighbourhood, corresponding to different choices of ϕ_i . Furthermore a particular coordinate system does not have to (and in general won't) cover the whole manifold. The second point guarantees that if two coordinate systems overlap then the transformation between one and the other is smooth.

The classic example of a manifold is the surface of a sphere, such as the earth.

Common coordinates are longitude and latitude. However these don't cover the whole space as the north and south poles do not have a well defined longitude.

We are in addition interested in Riemannian (or technically pseudo-Riemannian) manifold which means that we also have a metric $g_{\mu\nu}$ this is an invertable matrix located at each point which determines lengths and angles of vector fields at that point, viz:

$$\|V(x)\|^2 = g_{\mu\nu}(x)V^\mu(x)V^\nu(x) \quad (4.2)$$

Thus if we think of $V^\mu(x)$ as the infinitesimal variation of a curve $V^\mu = dx^\mu$ then we recover the definition above for the length of an infinitesimally small curve passing through the point x^μ

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu \quad (4.3)$$

4.2 Tensors

We now need to develop the rules that the 'good' (covariant) objects will obey when we pass from one coordinate system to another. Such objects are called tensors. A tensor is 'something that transforms like a vector'. Indeed it is just a generalisation of a vector field. You have probably already encountered these in a course on Electromagnetism ($F_{\mu\nu}$ is a tensor) but in the special case that the transformation between coordinates is linear (*i.e.* Lorentz transformations).

We saw earlier that for the notion of the proper distance to be invariant under coordinate transformations, *i.e.* diffeomorphisms, the metric had to transform as

$$g'_{\rho\sigma} = \frac{\partial x^\mu}{\partial x'^\rho} \frac{\partial x^\nu}{\partial x'^\sigma} g_{\mu\nu} \quad (4.4)$$

This is an example of a (0, 2)-tensor. We could also consider the inverse metric, that is the object $g^{\rho\sigma}$ which is the matrix inverse of $g_{\mu\nu}$:

$$g_{\mu\nu}g^{\nu\sigma} = \delta_\mu^\sigma \quad (4.5)$$

Problem: Show that

$$g^{\nu\sigma} = \frac{\partial x'^{\nu}}{\partial x^{\mu}} \frac{\partial x'^{\sigma}}{\partial x^{\lambda}} g^{\mu\lambda} \quad (4.6)$$

Thus we can define a (p, q) -tensor, or rank (p, q) tensor, on a manifold to be an object $T^{\mu_1\mu_2\mu_3\cdots\mu_p}_{\nu_1\nu_2\nu_3\cdots\nu_q}$ with p upstairs indices and q downstairs indices that transforms under a diffeomorphism $x^{\mu} \rightarrow x'^{\mu}(x^{\nu})$ as

$$\begin{aligned} T'(x')^{\mu_1\mu_2\mu_3\cdots\mu_p}_{\nu_1\nu_2\nu_3\cdots\nu_q} \\ = \left(\frac{\partial x'^{\mu_1}}{\partial x^{\rho_1}} \frac{\partial x'^{\mu_2}}{\partial x^{\rho_2}} \frac{\partial x'^{\mu_3}}{\partial x^{\rho_3}} \cdots \frac{\partial x'^{\mu_p}}{\partial x^{\rho_p}} \right) \left(\frac{\partial x^{\lambda_1}}{\partial x'^{\nu_1}} \frac{\partial x^{\lambda_2}}{\partial x'^{\nu_2}} \frac{\partial x^{\lambda_3}}{\partial x'^{\nu_3}} \cdots \frac{\partial x^{\lambda_q}}{\partial x'^{\nu_q}} \right) T(x)^{\rho_1\rho_2\rho_3\cdots\rho_p}_{\lambda_1\lambda_2\lambda_3\cdots\lambda_q} \end{aligned} \quad (4.7)$$

N.B.: Note the positions of the primed and unprimed coordinates! So the inverse metric is an example of a $(2, 0)$ -tensor.

A tensor field is simply a tensor which is defined at each point on the manifold.

Thus a scalar is a $(0, 0)$ -tensor;

$$\phi'(x') = \phi(x) \quad (4.8)$$

a vector is a $(1, 0)$ -tensor;

$$V'^{\mu}(x') = \frac{\partial x'^{\mu}}{\partial x^{\nu}} V^{\nu}(x) \quad (4.9)$$

and a covector is a $(0, 1)$ -tensor;

$$A'_{\mu}(x') = \frac{\partial x^{\nu}}{\partial x'^{\mu}} A_{\nu}(x) \quad (4.10)$$

In older books the upstairs and downstairs indices are referred to as contravariant and covariant respectively.

Given two tensors we can obtain a new one in various ways. If they have the same rank then any linear combination of them is also a tensor.

In addition a (p, q) -tensor can be multiplied by an (r, s) -tensor to produce a $(p + r, q + s)$ -tensor.

Finally a (p, q) -tensor $T(x)^{\rho_1\rho_2\rho_3\cdots\rho_p}_{\lambda_1\lambda_2\lambda_3\cdots\lambda_q}$ with $p, q \geq 1$ can be contracted to form a $(p - 1, q - 1)$ -tensor:

$$T(x)^{\rho_2\rho_3\cdots\rho_p}_{\lambda_2\lambda_3\cdots\lambda_q} = T(x)^{\mu\rho_2\rho_3\cdots\rho_p}_{\mu\lambda_2\lambda_3\cdots\lambda_q} \quad (4.11)$$

Clearly this can be done in pq ways depending on which pair of indices we sum over

Also since we have a metric $g_{\mu\nu}$ and its inverse $g^{\mu\nu}$ we can lower and raise indices on a tensor (this doesn't really create a new tensor so it keeps the same symbol). For example if V^{μ} is a vector then

$$V_{\mu} = g_{\mu\nu} V^{\nu} \quad (4.12)$$

is a covector.

Problem: What are the mistakes in the following equation:

$$A_{\mu\nu}{}^{\nu\lambda} B_{\rho\nu\pi\rho}{}^{\mu\sigma} - 34 C_{\mu\rho\pi}{}^{\lambda\sigma} = D_{\rho\pi}{}^{\sigma} \quad (4.13)$$

We can also take the symmetric and anti-symmetric parts of a tensor. Consider a $(0, q)$ -tensor $T_{\mu_1\mu_2\mu_3\dots\mu_q}$ then we have

$$\begin{aligned} T_{(\mu_1\mu_2\mu_3\dots\mu_q)} &= \frac{1}{q!} (T_{\mu_1\mu_2\mu_3\dots\mu_q} + T_{\mu_2\mu_1\mu_3\dots\mu_q} + \dots) \\ T_{[\mu_1\mu_2\mu_3\dots\mu_q]} &= \frac{1}{q!} (T_{\mu_1\mu_2\mu_3\dots\mu_q} - T_{\mu_2\mu_1\mu_3\dots\mu_q} + \dots) \end{aligned} \quad (4.14)$$

where the sum is over all permutations of the indices and in the second line the plus (minus) sign occurs for even (odd) permutations.

4.3 Covariant Derivatives

Having introduced tensors we can consider their derivatives. However the partial derivative of a tensor is not a tensor. To see this we can consider a vector field:

$$\begin{aligned} \frac{\partial}{\partial x'^{\nu}} V'^{\mu} &= \frac{\partial x^{\lambda}}{\partial x'^{\nu}} \frac{\partial}{\partial x^{\lambda}} \left(\frac{\partial x'^{\mu}}{\partial x^{\rho}} V^{\rho} \right) \\ &= \frac{\partial x^{\lambda}}{\partial x'^{\nu}} \frac{\partial x'^{\mu}}{\partial x^{\rho}} \frac{\partial}{\partial x^{\lambda}} V^{\rho} + \frac{\partial x^{\lambda}}{\partial x'^{\nu}} \frac{\partial^2 x'^{\mu}}{\partial x^{\rho} \partial x^{\lambda}} V^{\rho} \end{aligned} \quad (4.15)$$

The first term is fine but the second one isn't. Although we note that the derivative of a scalar is a vector, *i.e.* a $(0, 1)$ -tensor. To correct for this we must introduce the notion of a covariant derivative which respect the tensorial property.

The solution to this is well-known in physics. We introduce a so-called connection which modifies the derivative into a so-called covariant derivative and transforms in such a way that the covariant derivative of a tensor is again a tensor. Thus we introduce $\Gamma_{\lambda\rho}^{\mu}$ - called a connection - and define

$$D_{\nu} V^{\mu} = \partial_{\nu} V^{\mu} + \Gamma_{\nu\rho}^{\mu} V^{\rho} \quad (4.16)$$

We then require that under a diffeomorphism $\Gamma_{\lambda\rho}^{\mu}$ transforms as

$$\Gamma'^{\lambda}_{\mu\nu} = \frac{\partial x^{\rho}}{\partial x'^{\mu}} \frac{\partial x^{\sigma}}{\partial x'^{\nu}} \frac{\partial x'^{\lambda}}{\partial x^{\tau}} \Gamma^{\tau}_{\rho\sigma} - \frac{\partial x^{\sigma}}{\partial x'^{\mu}} \frac{\partial x^{\rho}}{\partial x'^{\nu}} \frac{\partial^2 x'^{\lambda}}{\partial x^{\sigma} \partial x^{\rho}} \quad (4.17)$$

We then see that

$$\begin{aligned}
D'_\nu V'^\mu &= \frac{\partial x^\lambda}{\partial x'^\nu} \frac{\partial x'^\mu}{\partial x^\rho} \frac{\partial}{\partial x^\lambda} V^\rho + \frac{\partial x^\lambda}{\partial x'^\nu} \frac{\partial^2 x'^\mu}{\partial x^\rho \partial x^\lambda} V^\rho \\
&\quad + \left(\frac{\partial x^\lambda}{\partial x'^\nu} \frac{\partial x^\sigma}{\partial x'^\mu} \frac{\partial x'^\mu}{\partial x^\rho} \Gamma_{\lambda\sigma}^\rho - \frac{\partial x^\sigma}{\partial x'^\mu} \frac{\partial x^\rho}{\partial x'^\nu} \frac{\partial^2 x'^\lambda}{\partial x^\sigma \partial x^\rho} \right) \frac{\partial x'^\mu}{\partial x^\pi} V^\pi \\
&= \frac{\partial x'^\mu}{\partial x^\rho} \frac{\partial x^\lambda}{\partial x'^\nu} D_\lambda V^\rho + \frac{\partial x^\lambda}{\partial x'^\nu} \frac{\partial^2 x'^\mu}{\partial x^\rho \partial x^\lambda} V^\rho - \delta_\pi^\sigma \frac{\partial x^\rho}{\partial x'^\nu} \frac{\partial^2 x'^\lambda}{\partial x^\sigma \partial x^\rho} V^\pi \tag{4.18}
\end{aligned}$$

$$= \frac{\partial x'^\mu}{\partial x^\rho} \frac{\partial x^\lambda}{\partial x'^\nu} D_\lambda V^\rho \tag{4.19}$$

Problem: Show that we can also write

$$\Gamma_{\mu\nu}^{\lambda\sigma} = \frac{\partial x^\rho}{\partial x'^\mu} \frac{\partial x^\sigma}{\partial x'^\nu} \frac{\partial x'^\lambda}{\partial x^\tau} \Gamma_{\rho\sigma}^\tau + \frac{\partial x'^\lambda}{\partial x^\sigma} \frac{\partial^2 x^\sigma}{\partial x'^\nu \partial x'^\mu}$$

(HINT: consider the chain rule on the second term of the original expression for $\Gamma_{\mu\nu}^{\lambda\sigma}$ and note the identity $\partial M^{-1}M = -M^{-1}\partial M$ for matrices)

For covectors we define

$$D_\lambda V_\mu = \partial_\lambda V_\mu - \Gamma_{\lambda\mu}^\rho V_\rho \tag{4.20}$$

This ensures that the scalar obtained by contracting V^μ and U_μ satisfies

$$D_\lambda(V^\mu U_\mu) = D_\lambda V^\mu U_\mu + V^\mu D_\lambda U_\mu = \partial_\lambda(V^\mu U_\mu) \tag{4.21}$$

as we expect for scalars.

The covariant derivative is then defined on a general (p, q) -tensor to be

$$\begin{aligned}
D_\mu T^{\rho_1 \rho_2 \rho_3 \dots \rho_p}{}_{\lambda_1 \lambda_2 \lambda_3 \dots \lambda_q} &= \partial_\mu T^{\rho_1 \rho_2 \rho_3 \dots \rho_p}{}_{\lambda_1 \lambda_2 \lambda_3 \dots \lambda_q} \\
&\quad + \Gamma_{\mu\nu}^{\rho_1} T^{\nu \rho_2 \rho_3 \dots \rho_p}{}_{\lambda_1 \lambda_2 \lambda_3 \dots \lambda_q} + \dots \\
&\quad - \Gamma_{\mu\lambda_1}^\nu T^{\rho_1 \rho_2 \rho_3 \dots \rho_p}{}_{\nu \lambda_2 \lambda_3 \dots \lambda_q} - \dots \tag{4.22}
\end{aligned}$$

where each index gets contracted with $\Gamma_{\nu\lambda}^\mu$

Problem: Convince yourself that the covariant derivative of a (p, q) -tensor is a $(p, q+1)$ -tensor.

Note that the anti-symmetric part of a connection $\Gamma_{[\mu\nu]}^\lambda$ is a $(1, 2)$ -tensor (why?). This is called the torsion and it is usually set to zero. In addition the difference between any two connections is a $(1, 2)$ -tensor.

To determine the connection $\Gamma_{\mu\nu}^\lambda$ we impose another condition, namely that the metric is covariantly constant, $D_\lambda g_{\mu\nu} = 0$. This is called the Levi-Civita connection. It is the unique connection which annihilates the metric and is torsion free.

To determine it (and show that it is unique) we consider the following

$$\begin{aligned}
D_\lambda g_{\mu\nu} &= \partial_\lambda g_{\mu\nu} - \Gamma_{\lambda\mu}^\rho g_{\rho\nu} - \Gamma_{\lambda\nu}^\rho g_{\mu\rho} = 0 \\
D_\mu g_{\nu\lambda} &= \partial_\mu g_{\nu\lambda} - \Gamma_{\mu\nu}^\rho g_{\rho\lambda} - \Gamma_{\mu\lambda}^\rho g_{\nu\rho} = 0 \\
D_\nu g_{\lambda\mu} &= \partial_\nu g_{\lambda\mu} - \Gamma_{\nu\lambda}^\rho g_{\rho\mu} - \Gamma_{\nu\mu}^\rho g_{\lambda\rho} = 0
\end{aligned} \tag{4.23}$$

Next we take the sum of the 2nd and 3rd equation minus the 1st (and use the fact that $\Gamma_{\mu\nu}^\lambda = \Gamma_{\nu\mu}^\lambda$):

$$0 = \partial_\nu g_{\lambda\mu} + \partial_\mu g_{\nu\lambda} - \partial_\lambda g_{\mu\nu} - 2\Gamma_{\mu\nu}^\rho g_{\rho\lambda} \tag{4.24}$$

Thus we find that the Levi-Civita is

$$\Gamma_{\mu\nu}^\lambda = \frac{1}{2} g^{\lambda\rho} (\partial_\mu g_{\rho\nu} + \partial_\nu g_{\rho\mu} - \partial_\rho g_{\mu\nu}) \tag{4.25}$$

Thus we see that this is indeed symmetric in its lower two indices. It is easy to verify that

$$\begin{aligned}
D_\lambda g_{\mu\nu} &= \partial_\lambda g_{\mu\nu} - \Gamma_{\lambda\mu}^\rho g_{\rho\nu} - \Gamma_{\lambda\nu}^\rho g_{\mu\rho} \\
&= \partial_\lambda g_{\mu\nu} - \frac{1}{2} (\partial_\mu g_{\lambda\nu} + \partial_\lambda g_{\mu\nu} - \partial_\nu g_{\lambda\mu}) - \frac{1}{2} (\partial_\nu g_{\lambda\mu} + \partial_\lambda g_{\mu\nu} - \partial_\mu g_{\lambda\nu}) \\
&= 0
\end{aligned} \tag{4.26}$$

The last thing to do is prove that $\Gamma_{\mu\nu}^\lambda$ indeed transforms as it should. To do this we note that, from (4.24), we have, by construction,

$$0 = \partial'_\nu g'_{\lambda\mu} + \partial'_\mu g'_{\nu\lambda} - \partial'_\lambda g'_{\mu\nu} - 2\Gamma'_{\mu\nu}{}^\rho g'_{\rho\lambda} \tag{4.27}$$

in the new coordinates. Now we know that

$$g'_{\lambda\mu} = \frac{\partial x^\tau}{\partial x'^\lambda} \frac{\partial x^\sigma}{\partial x'^\mu} g_{\tau\sigma} \quad \partial'_\nu = \frac{\partial x^\rho}{\partial x'^\nu} \frac{\partial}{\partial x^\rho}$$

We just need to substitute this into (4.27) and determine $\Gamma'_{\mu\nu}{}^\lambda$.

Problem: Show that $\Gamma_{\mu\nu}^\lambda$ indeed transforms as a connection (HINT: first consider the case where the matrices $\partial x^\mu / \partial x'^\nu$ are constant).

Problem: Compute the components of $\Gamma_{\mu\nu}^\lambda$ for S^2 .

Problem: Consider polar coordinates of \mathbb{R}^2 . Compute $g^{\mu\nu} D_\mu D_\nu f$ for an arbitrary function $f(r, \theta)$ and show that this agrees with $\nabla^2 f$ computed above.

4.4 Geodesics

We wish to find paths which minimize their proper length, what we can think of as the analogues of straight lines. If $X^\mu(\tau)$ is a path in spacetime where τ parameterizes the curve and runs from $\tau = a$ to $\tau = b$ we need to minimize the functional

$$l = \int_a^b ds = \int_a^b \sqrt{|g_{\mu\nu} dx^\mu dx^\nu|} = \int_a^b \sqrt{|g_{\mu\nu} \dot{X}^\mu \dot{X}^\nu|} d\tau$$

We do not worry about boundary terms, as we wish to find a local condition on the curve $X^\mu(\tau)$. This is simply a variational problem that you should have encountered in classical mechanics where the equations of motion are determined by extremizing the Lagrangian $\mathcal{L} = \sqrt{|g_{\mu\nu} \dot{X}^\mu \dot{X}^\nu|}$.

This turns out to have a technically subtle because it is so non-linear. I don't want to bother you with the full details (except as a problem!) so a quick way to avoid this is note that we can just as well use the 'Lagrangian' \mathcal{L}^2 and hence extremize

$$“l^2” = \int g_{\nu\lambda} \dot{X}^\nu \dot{X}^\lambda d\tau$$

The Euler-Lagrange equations give

$$-2 \frac{d}{d\tau} (g_{\mu\nu} \dot{X}^\nu) + \frac{\partial g_{\lambda\nu}}{\partial x^\mu} \dot{X}^\lambda \dot{X}^\nu = 0 \quad (4.28)$$

Expanding things out a bit we find

$$\begin{aligned} 0 &= -2g_{\mu\nu} \ddot{X}^\nu - \left(2 \frac{\partial g_{\mu\nu}}{\partial x^\lambda} - \frac{\partial g_{\lambda\nu}}{\partial x^\mu} \right) \dot{X}^\lambda \dot{X}^\nu \\ &= -2 \left(g_{\mu\nu} \ddot{X}^\nu + \frac{1}{2} \frac{\partial g_{\mu\nu}}{\partial x^\lambda} + \frac{1}{2} \frac{\partial g_{\mu\lambda}}{\partial x^\nu} - \frac{1}{2} \frac{\partial g_{\lambda\nu}}{\partial x^\mu} \right) \dot{X}^\lambda \dot{X}^\nu \\ &= -2g_{\mu\nu} \left(\ddot{X}^\mu + \Gamma_{\lambda\nu}^\mu \dot{X}^\lambda \dot{X}^\nu \right) \end{aligned} \quad (4.29)$$

This is equivalent to

$$\ddot{X}^\mu + \Gamma_{\lambda\nu}^\mu \dot{X}^\lambda \dot{X}^\nu = 0 \quad (4.30)$$

A geodesic that satisfies (4.30) is said to be affinely parameterized and in what follows we will always assume this to be the case.

The technical point that we skipped is that one can find geodesics which don't satisfy this. If we compute the Euler-Lagrange equations that arise from minimizing l then there will be factors of $\sqrt{|g_{\mu\nu} \dot{X}^\mu \dot{X}^\nu|}$ in the denominator. We can ignore these if they are constant. And this is just what happens:

Theorem: Along an affinely parameterized geodesic $g_{\mu\nu} \dot{X}^\mu \dot{X}^\nu$ is constant

Proof: We simply differentiate

$$\begin{aligned}
\frac{d}{d\tau} (g_{\mu\nu} \dot{X}^\mu \dot{X}^\nu) &= \partial_\lambda g_{\mu\nu} \dot{X}^\lambda \dot{X}^\mu \dot{X}^\nu + 2g_{\mu\nu} \dot{X}^\mu \ddot{X}^\nu \\
&= \partial_\lambda g_{\mu\nu} \dot{X}^\lambda \dot{X}^\mu \dot{X}^\nu - 2g_{\mu\nu} \dot{X}^\mu \Gamma_{\lambda\rho}^\nu \dot{X}^\lambda \dot{X}^\rho \\
&= (\partial_\lambda g_{\mu\nu} - \partial_\nu g_{\mu\lambda} - \partial_\lambda g_{\mu\nu} + \partial_\mu g_{\lambda\nu}) \dot{X}^\lambda \dot{X}^\mu \dot{X}^\nu \\
&= 0
\end{aligned} \tag{4.31}$$

Thus $ds = \text{const.} \times d\tau$ along and affinely parameterised geodesic and hence we can think of s and τ as the same, up to an overall constant.

In the original calculation we see that l does not depend on the choice of the parameterization $X^\mu(\tau)$. This is because we can redefine the variable $\tau \rightarrow \tau(\tau')$. As such we see that

$$\dot{X}^\mu = \frac{d\tau'}{d\tau} \frac{dX^\mu}{d\tau'} \quad d\tau = \frac{d\tau}{d\tau'} d\tau'$$

and hence

$$\begin{aligned}
l &= \int_a^b \sqrt{|g_{\mu\nu} \dot{X}^\mu \dot{X}^\nu|} d\tau \\
&= \int_{\tau^{-1}(a)}^{\tau^{-1}(b)} \sqrt{\left| g_{\mu\nu} \frac{d\tau'}{d\tau} \frac{dX^\mu}{d\tau'} \frac{d\tau'}{d\tau} \frac{dX^\nu}{d\tau'} \right|} \frac{d\tau}{d\tau'} d\tau' \\
&= \int_{\tau^{-1}(a)}^{\tau^{-1}(b)} \sqrt{\left| g_{\mu\nu} \frac{dX^\mu}{d\tau'} \frac{dX^\nu}{d\tau'} \right|} d\tau'
\end{aligned} \tag{4.32}$$

where the factors of $d\tau/d\tau'$ have canceled. The new curve, viewed as a function of τ' won't satisfy (4.30). However one can show that there is always a choice of τ so that the geodesic does satisfy (4.30).

Problem: Show this.

Note that this does not work in the expression for “ l^2 ”. “ l^2 ” only works if we use affine parameterization.

Let us return to our sphere example again. Here we have

$$l = R \int \sqrt{\dot{\theta}^2 + \sin^2 \theta \dot{\varphi}^2} d\tau$$

Before we imagined that $\theta = f(\varphi)$ which in this language means that we choose $\tau = \varphi$. This is possible since we have reparameterization invariance, however other choices, such as an affine parameter, might be better. Let us exam things in the new way using

$$\text{“}l^2\text{”} = R^2 \int \dot{\theta}^2 + \sin^2 \theta \dot{\varphi}^2 d\tau \tag{4.33}$$

Since we don't have reparameterization invariance (we must use an affine parameter) we can't choose τ how we like. We must proceed keeping τ general.

This is just an Euler-Lagrange system for two fields θ and φ . However let us note a simplification. The ‘‘Lagrangian’’ $\mathcal{L} = R^2(\dot{\theta}^2 + \sin^2 \theta \dot{\varphi}^2)$ depends on $\dot{\varphi}$ but not directly on φ (there is a symmetry $\varphi \rightarrow \varphi + \text{const}$). This means that there is a conserved quantity:

$$L = \frac{\partial \mathcal{L}}{\partial \dot{\varphi}} = R^2 \sin^2 \theta \dot{\varphi} \quad (4.34)$$

it’s called L because it is related the angular momentum of the motion. The equation of motion now reduces to the equation of motion for θ :

$$\ddot{\theta} - \sin \theta \cos \theta \dot{\varphi}^2 = \ddot{\theta} - \frac{L^2 \cos \theta}{R^4 \sin^3 \theta} = 0$$

Maybe this doesn’t seem much better than what we had before. There is still the constant solution at $\theta = \pi/2$ but now $\varphi = L\tau/R^4$.

One trick is step back a bit and notice that this is the equation of motion for a particle with position θ moving in a potential $V = L^2/R^4 \sin^2 \theta$. Indeed multiplying by $\dot{\theta}$ and integrating once leads to

$$\frac{1}{2} \dot{\theta}^2 + \frac{L^2}{2R^4 \sin^2 \theta} = E$$

where E is a constant (the total energy). Alternatively we can just substitute (4.34) into (4.33). So even if we can’t solve the equation we can get a good picture of the behaviour of the solutions by thinking of a particle moving in the potential $L^2/R^4 \sin^2 \theta$.

Actually one can solve this equation by substituting $\theta = f(y)$ for some y so that $\dot{\theta} = f'y$. The equation can now be written as

$$\frac{1}{2} f'^2 \dot{y}^2 + \frac{L^2}{2R^4 \sin^2 f} = E$$

If we choose $f' = L/R^2 \sin f$ then we get

$$\dot{y}^2 + 1 = \frac{2ER^4}{L^2} \sin^2 f$$

Now the solution to $\sin f f' = L/R^2$ is $-\cos f = Ly/R^2$ and hence $\sin^2 f = 1 - L^2 y^2/R^4$. In this way we find

$$\dot{y}^2 + 2Ey^2 = \frac{2ER^4}{L^2} - 1$$

Since the left hand side is positive definite we require that $E > L/2R^2$ and the solutions are then simply

$$y = \left(\frac{R^4}{L^2} - \frac{1}{2E} \right) \cos \left(\sqrt{2E}(\tau - \tau_0) \right)$$

Finally we put the pieces back together to find

$$\theta = \arccos \left(\cos \theta_0 \cos \left(\sqrt{\frac{2EL}{R^2}}(\tau - \tau_0) \right) \right)$$

where $\cos \theta_0 = L/2ER^2 - R^2/L$. We can also compute φ by noting that

$$\dot{\varphi} = \frac{L}{R^4 \sin^2 \theta} = \frac{L}{R^4} \frac{1}{1 - \cos^2 \theta_0 \cos^2 \left(\sqrt{\frac{2EL}{R^2}} (\tau - \tau_0) \right)}$$

This can be integrated to give

$$\varphi = \varphi_0 + \sqrt{\frac{L}{2ER^6}} \csc \theta_0 \operatorname{arccot} \left(\sin \theta_0 \cot \left(\sqrt{\frac{2EL}{R^2}} (\tau - \tau_0) \right) \right)$$

4.5 Causal Curves Again

We have been a bit cavalier with the expression $\sqrt{|-g_{\mu\nu} dx^\mu dx^\nu|}$. Since $g_{\mu\nu}$ is a symmetric matrix it has real eigenvalues. Furthermore in flat spacetime (in the usual coordinates) we have

$$g_{\mu\nu} = \eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (4.35)$$

Note that since $g_{\mu\nu}$ must be invertible its eigenvalues can never pass through zero. Thus $g_{\mu\nu}$ always has one negative eigenvalue and three positive ones. The eigenvector associated to the negative eigenvalue is ‘time’. Thus vectors can have a length-squared which is positive, negative or zero. These are called space-like, time-like or null respectively. This is the same as in Special relativity.

If $X^\mu(\tau)$ is a curve, *i.e.* a map from some interval in \mathbb{R} to \mathcal{M} , then we can construct the ‘tangent’ vector at a point $X^\mu(\tau_0)$ along the curve to be

$$T^\mu = \left. \frac{dX^\mu}{d\tau} \right|_{\tau=\tau_0} \quad (4.36)$$

The length of this vector is determined by the metric

$$\|T\|^2 = g_{\mu\nu}(X(\tau_0)) \left. \frac{dX^\mu}{d\tau} \right|_{\tau=\tau_0} \left. \frac{dX^\nu}{d\tau} \right|_{\tau=\tau_0} \quad (4.37)$$

An important consequence of this is that $\|T\|^2$ can be positive, negative or zero. Indeed for a static curve, where only the time coordinate is changing, $\|T\|^2 < 0$. For light rays we have $\|T\|^2 = 0$. Finally curves for which every point is at the same ‘time’ have $\|T\|^2 > 0$. Similarly curves are called time-like, null and space-like respectively if their tangent vectors are everywhere time-like, null or spacelike respectively.

The familiar statement of Special Relativity that nothing can travel faster than the speed of light is the statement that physical observers always follow time-like curves, in particular time-like geodesics, that is curves for which (at all points)

$$g_{\mu\nu} \dot{X}^\mu \dot{X}^\nu < 0 \quad (4.38)$$

where the derivative is with respect to the coordinate along the particles world-line. Similarly light travels along null curves:

$$g_{\mu\nu}\dot{X}^\mu\dot{X}^\nu = 0 \quad (4.39)$$

at each point.

Two events, *i.e.* two points in spacetime, are said to be causally related if there is a time-like or null curve that passes through them. In which case the earlier one (as defined by the coordinate of the worldline) can influence the later one. If no such curve exists then the two points are said to be spacelike separated and an observer at one cannot know anything about the events at the other.

4.6 Integration

Let us look at how to compute the integral of some scalar quantity f over spacetime (or a part of spacetime). We need the integral to be well defined, *i.e.* independent of coordinate transformations. We first note that, under a diffeomorphism, $x^\mu \rightarrow x'^\mu(x)$

$$d^4x' = \det\left(\frac{\partial x'}{\partial x}\right) d^4x \quad (4.40)$$

This transformation can be cancelled by noting that

$$\begin{aligned} \sqrt{-\det(g'_{\mu\nu})} &= \sqrt{-\det\left(\frac{\partial x^\lambda}{\partial x'^\mu}\frac{\partial x^\rho}{\partial x'^\nu}g_{\lambda\rho}\right)} \\ &= \sqrt{\det\left(\frac{\partial x^\lambda}{\partial x'^\mu}\right)}\sqrt{\det\left(\frac{\partial x^\rho}{\partial x'^\nu}\right)}\sqrt{-\det(g_{\lambda\rho})} \\ &= \det\left(\frac{\partial x}{\partial x'}\right)\sqrt{-\det(g)} \end{aligned} \quad (4.41)$$

Here we have neglected indices and viewed the various two-index expressions as matrices. Thus we have that

$$d^4x'\sqrt{-\det(g')} = d^4x\sqrt{-\det(g)} \quad (4.42)$$

is a coordinate independent measure and so the a coordinate system invariant expression for the integral of f is

$$\int f = \int d^4x\sqrt{-\det(g)}f \quad (4.43)$$

4.7 Riemann Normal Coordinates

There is a local coordinate choice that we can use to set $\Gamma_{\mu\nu}^\lambda = 0$ at any given point p . Note that this is not possible for a tensor which vanishes at p in all coordinates systems if it vanishes in one. From $D_\lambda g_{\mu\nu} = 0$ this is equivalent to $\partial_\lambda g_{\mu\nu} = 0$ at p . If we suppose

that the coordinates of p correspond to $x^\mu = 0$ then this means that the metric has the form

$$g_{\mu\nu}(x) = g_{\mu\nu}(0) + \mathcal{O}(x^2)$$

To arrange this we start with a general metric

$$g_{\mu\nu}(x) = g_{\mu\nu}(0) + G_{\mu\nu\lambda}x^\lambda + \mathcal{O}(x^2)$$

and consider a coordinate transformation

$$x'^\mu = x^\mu + C_{\nu\lambda}^\mu x^\nu x^\lambda + \mathcal{O}(x^2)$$

Note that without loss of generality we have that $C_{\nu\lambda}^\mu$ is symmetric in ν, λ . Thus it has the same symmetries and index structure as $\Gamma_{\nu\lambda}^\mu$ so one may hope to use it to, locally at p , set $\Gamma_{\nu\lambda}^\mu = 0$. Indeed we compute

$$\frac{\partial x'^\mu}{\partial x^\nu} = \delta_\nu^\mu + 2C_{\nu\lambda}^\mu x^\lambda \quad \frac{\partial^2 x'^\sigma}{\partial x^\nu \partial x^\mu} = 2C_{\mu\nu}^\sigma$$

We also need $\partial x^\mu / \partial x'^\nu$ but all we need is the very lowest order in x^μ and this is simply obtained as the inverse to $\partial x'^\mu / \partial x^\nu$:

$$\frac{\partial x^\mu}{\partial x'^\nu} = \delta_\nu^\mu + \mathcal{O}(x)$$

Next we recall that

$$\Gamma_{\mu\nu}^{\lambda} = \frac{\partial x^\rho}{\partial x'^\mu} \frac{\partial x^\sigma}{\partial x'^\nu} \frac{\partial x'^\lambda}{\partial x^\tau} \Gamma_{\rho\sigma}^\tau - \frac{\partial x^\sigma}{\partial x'^\mu} \frac{\partial x^\rho}{\partial x'^\nu} \frac{\partial^2 x'^\lambda}{\partial x^\sigma \partial x^\rho} \quad (4.44)$$

We want to show that if $\Gamma_{\mu\nu}^\lambda = \Gamma_{\mu\nu}^\lambda(0) + \mathcal{O}(x)$ we can chose $C_{\mu\nu}^\lambda$ so that $\Gamma_{\mu\nu}^\lambda = 0 + \mathcal{O}(x)$. This leads to the condition

$$0 + \mathcal{O}(x) = \Gamma_{\mu\nu}^\lambda(0) + \mathcal{O}(x) - 2C_{\mu\nu}^\lambda + \mathcal{O}(x)$$

and hence we simply take $C_{\mu\nu}^\lambda = \frac{1}{2}\Gamma_{\mu\nu}^\lambda(0)$.

Such a coordinate system is called the Riemann normal coordinate system about p .

Note that if we consider geodesics that start at p then they take the simple form

$$X^\mu(\tau) = T^\mu \tau + \mathcal{O}(\tau^3)$$

for a constant vector T^μ . Thus the simple flows along the coordinates are geodesics to lowest order in τ

4.8 Curvature

Partial derivatives commute: $[\partial_\mu, \partial_\nu] = \partial_\mu \partial_\nu - \partial_\nu \partial_\mu = 0$. However this is not the case with covariant derivatives. Indeed

$$\begin{aligned}
[D_\mu, D_\nu]V_\lambda &= D_\mu (\partial_\nu V_\lambda - \Gamma_{\nu\lambda}^\rho V_\rho) - (\mu \leftrightarrow \nu) \\
&= \partial_\mu (\partial_\nu V_\lambda - \Gamma_{\nu\lambda}^\rho V_\rho) - \Gamma_{\mu\nu}^\sigma (\partial_\sigma V_\lambda - \Gamma_{\sigma\lambda}^\rho V_\rho) \\
&\quad - \Gamma_{\mu\lambda}^\sigma (\partial_\nu V_\sigma - \Gamma_{\nu\sigma}^\rho V_\rho) - (\mu \leftrightarrow \nu) \\
&= -\partial_\mu (\Gamma_{\nu\lambda}^\rho V_\rho) - \Gamma_{\mu\lambda}^\sigma (\partial_\nu V_\sigma - \Gamma_{\nu\sigma}^\rho V_\rho) - (\mu \leftrightarrow \nu) \\
&= -\partial_\mu \Gamma_{\nu\lambda}^\rho V_\rho + \Gamma_{\mu\lambda}^\sigma \Gamma_{\nu\sigma}^\rho V_\rho - (\mu \leftrightarrow \nu) \\
&= R_{\mu\nu\lambda}{}^\rho V_\rho
\end{aligned} \tag{4.45}$$

where

$$R_{\mu\nu\lambda}{}^\rho = -\partial_\mu \Gamma_{\nu\lambda}^\rho + \partial_\nu \Gamma_{\mu\lambda}^\rho - \Gamma_{\nu\lambda}^\sigma \Gamma_{\mu\sigma}^\rho + \Gamma_{\mu\lambda}^\sigma \Gamma_{\nu\sigma}^\rho \tag{4.46}$$

is the Riemann curvature.

N.B.: Different books have different conventions for $R_{\mu\nu\lambda}{}^\rho$.

For higher tensors one finds that

$$\begin{aligned}
[D_\mu, D_\nu]T^{\rho_1\rho_2\rho_3\dots\rho_p}{}_{\lambda_1\lambda_2\lambda_3\dots\lambda_q} &= R_{\mu\nu\lambda_1}{}^\pi T^{\rho_1\rho_2\rho_3\dots\rho_p}{}_{\pi\lambda_2\lambda_3\dots\lambda_q} + \dots \\
&\quad - R_{\mu\nu\pi}{}^{\rho_1} T^{\pi\rho_2\rho_3\dots\rho_p}{}_{\lambda_1\lambda_2\lambda_3\dots\lambda_q} - \dots
\end{aligned} \tag{4.47}$$

$R_{\mu\nu\lambda}{}^\rho$ is a tensor. To see this we note that $[D_\mu, D_\nu]V_\lambda$ is a $(0, 3)$ -tensor for any V_λ . Thus under a diffeomorphism

$$R'{}_{\mu\nu\lambda}{}^\rho V'_\rho = \frac{\partial x^\sigma}{\partial x'^\mu} \frac{\partial x^\tau}{\partial x'^\nu} \frac{\partial x^\pi}{\partial x'^\lambda} R_{\sigma\tau\pi}{}^\rho V_\rho \tag{4.48}$$

Now since

$$V_\rho = \frac{\partial x'^\theta}{\partial x^\rho} V'_\theta \tag{4.49}$$

we see that

$$R'{}_{\mu\nu\lambda}{}^\rho V'_\rho = \frac{\partial x^\sigma}{\partial x'^\mu} \frac{\partial x^\tau}{\partial x'^\nu} \frac{\partial x^\pi}{\partial x'^\lambda} R_{\sigma\tau\pi}{}^\rho \frac{\partial x'^\theta}{\partial x^\rho} V'_\pi \tag{4.50}$$

Since V'_ρ is arbitrary we deduce that $R_{\sigma\tau\pi}{}^\rho$ is a $(1, 3)$ -tensor;

$$R'{}_{\mu\nu\lambda}{}^\rho = \frac{\partial x^\sigma}{\partial x'^\mu} \frac{\partial x^\tau}{\partial x'^\nu} \frac{\partial x^\pi}{\partial x'^\lambda} \frac{\partial x'^\theta}{\partial x^\rho} R_{\sigma\tau\pi}{}^\theta \tag{4.51}$$

It has some identities:

$$R_{(\mu\nu)\lambda}{}^\rho = 0, \quad R_{[\mu\nu\lambda]}{}^\rho = 0, \quad R_{\mu\nu\lambda\rho} = R_{\lambda\rho\mu\nu}, \quad D_{[\tau} R_{\mu\nu]\lambda}{}^\rho = 0 \tag{4.52}$$

The first identity is obvious.

To establish the other identities use Riemann Normal coordinates to find a system of coordinates such that $\Gamma_{\mu\nu}^\lambda = 0$ at some point p (but not everywhere). Thus at this point p

$$R_{\mu\nu\lambda}{}^\rho = -\partial_\mu\Gamma_{\nu\lambda}^\rho + \partial_\nu\Gamma_{\mu\lambda}^\rho \quad (4.53)$$

and so

$$\begin{aligned} R_{\mu\nu\lambda\rho} &= -g_{\rho\tau}\partial_\mu\Gamma_{\nu\lambda}^\tau + g_{\rho\tau}\partial_\nu\Gamma_{\mu\lambda}^\tau \\ &= \frac{1}{2}(-\partial_\mu\partial_\lambda g_{\rho\nu} - \partial_\mu\partial_\nu g_{\rho\lambda} + \partial_\mu\partial_\rho g_{\nu\lambda} + \partial_\nu\partial_\lambda g_{\rho\mu} + \partial_\nu\partial_\mu g_{\rho\lambda} - \partial_\nu\partial_\rho g_{\mu\lambda}) \\ &= \frac{1}{2}(-\partial_\mu\partial_\lambda g_{\rho\nu} + \partial_\mu\partial_\rho g_{\nu\lambda} + \partial_\nu\partial_\lambda g_{\rho\mu} - \partial_\nu\partial_\rho g_{\mu\lambda}) \end{aligned} \quad (4.54)$$

Here we can simply check that, at p , $R_{[\mu\nu\lambda]\rho} = 0$ and $R_{\mu\nu\lambda\rho} - R_{\lambda\rho\mu\nu} = 0$. Furthermore since these are tensors if they vanishes at p in one coordinate system then it vanishes in all. Finally there was nothing special about the point p . Therefore the identities everywhere.

Problem: Prove the final identity $D_{[\tau}R_{\mu\nu]\lambda}{}^\rho = 0$.

From the Riemann tensor we can construct the Ricci tensor by contraction

$$R_{\mu\nu} = R_{\mu\rho\nu}{}^\rho \quad (4.55)$$

and it follows that $R_{\mu\nu} = R_{\nu\mu}$. And lastly there is the Ricci scalar which requires us to contract using the metric

$$R = g^{\mu\nu}R_{\mu\nu} \quad (4.56)$$

Problem: Show that

$$D^\mu \left(R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R \right) = 0 \quad (4.57)$$

Problem: Compute $R_{\mu\nu\lambda}{}^\rho$, $R_{\mu\nu}$ and R for S^2 . Recall that previously you should have computed

$$\Gamma_{\varphi\varphi}^\theta = -\sin\theta\cos\theta \quad \Gamma_{\theta\varphi}^\varphi = \Gamma_{\varphi\theta}^\varphi = \cot\theta \quad (4.58)$$

(HINT: The only independent component is $R_{\theta\varphi\theta}{}^\varphi$.)

Problem: Show that in two dimensions one always has

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 0$$

5 General Relativity

We are finally in a position to write down Einstein's equation that determines the dynamics of the metric field $g_{\mu\nu}$ and examine some physical consequences.

Einstein's idea was that matter causes spacetime to become curved so that geodesics in the presence of large masses can explain the motion of bodies in a gravitational field. The next important step is to postulate an equation for the metric in the presence of matter (or energy since they are interchangeable in relativity). In addition since gravity is universal the coupling of geometry to matter should only depend on the mass and energy present and not what kind of matter it is.

Thus we need to look for an equation of the form

$$\text{Geometry} = \text{Matter} \tag{5.1}$$

The bulk properties of matter are described by the energy-momentum tensor $T_{\mu\nu}$. Furthermore we want an equation that is second order in derivatives of the metric tensor (since we want to mimic a Newtonian style force law). Another hint comes from the fact that in flat space the energy-momentum tensor is conserved $\partial^\mu T_{\mu\nu} = 0$. Since this is not covariant we postulate that the general expression is $D^\mu T_{\mu\nu} = 0$ Given the identity

$$D^\mu \left(R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R \right) = 0 \tag{5.2}$$

an obvious choice is

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + \Lambda g_{\mu\nu} = \kappa^2 T_{\mu\nu} \tag{5.3}$$

where Λ and κ are constants. We won't fix κ in this course. For the curious it can be computed by comparing to the Newtonian Theory and one finds $\kappa^2 = 8\pi G_N/c^4$ where G_N is Newton's constant.

Here we have included an additional term which is obviously covariantly conserved: $D^\mu g_{\mu\nu} = 0$. There is a long story about Λ - the so-called cosmological constant. It is in fact the biggest mystery in the exact sciences. The problem is that we don't know why the cosmological constant is so small, by a factor of 10^{-120} from what one would expect. Furthermore recent observations strongly imply that it is just slightly greater than zero.

Nowadays one usually absorbs the cosmological constant term into the energy momentum tensor $T_{\mu\nu} \rightarrow T_{\mu\nu} + \Lambda \kappa^{-2} g_{\mu\nu}$ where it is identified with the energy density of the vacuum. We will do the same. Thus we take Einstein's equation to be

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \kappa^2 T_{\mu\nu} \tag{5.4}$$

It is often said that this equations says "matter tells space how to bend and space tells matter how to move". It turns out that this guess is at least mathematically good: it is not overdetermined, *i.e.* as a set of differential equations it is well posed with a suitable set of initial conditions. This would not be the case if we hadn't chosen the left hand side to be covariantly conserved.

Problem: Show that Einstein's equation is equivalent to

$$R_{\mu\nu} = \kappa^2 T_{\mu\nu} - \frac{1}{2} \kappa^2 g_{\mu\nu} T \quad (5.5)$$

6 Schwarzschild Solution

Let us look for a time-independent and spherically symmetric solution to Einstein's equation with $T_{\mu\nu} = 0$. We take the ansatz:

$$ds^2 = -e^{2A(r)} c^2 dt^2 + e^{2B(r)} dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2) \quad (6.1)$$

Note that we could have allowed for terms of the form $g_{tr}(r) dt dr$. This would still be time-independent and spherically symmetric but it turns out we can ignore it and this will save us a lot of work. Technically having a time-independent metric is called stationary and a stationary metric which has no $dt dx$ terms (for any spatial coordinate x) is called static. The difference between the two is like the difference between a lake and a river: both are constant in time yet a river is somehow moving and transferring momentum whereas the lake is truly static.

So we have

$$g_{\mu\nu} = \begin{pmatrix} -c^2 e^{2A} & 0 & 0 & 0 \\ 0 & e^{2B} & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2 \sin^2 \theta \end{pmatrix}, \quad g^{\mu\nu} = \begin{pmatrix} -c^{-2} e^{-2A} & 0 & 0 & 0 \\ 0 & e^{-2B} & 0 & 0 \\ 0 & 0 & \frac{1}{r^2} & 0 \\ 0 & 0 & 0 & \frac{1}{r^2 \sin^2 \theta} \end{pmatrix} \quad (6.2)$$

We start by calculating the non-vanishing Christoffel coefficients. There are two (equivalent) ways to do this. The first is to use the formula (4.25). The other is to consider the equations for geodesics using the action “ l^2 ” introduced above, from the Euler-Lagrange equations one can read off the components of $\Gamma_{\mu\nu}^\lambda$ from the terms in the X^λ equation that are first order in derivatives. For the metric we have we find the “ l^2 ” Lagrangian is

$$\mathcal{L} = e^{2A(r)} c^2 \dot{t}^2 - e^{2B(r)} \dot{r}^2 - r^2 \dot{\theta}^2 - r^2 \sin^2 \theta \dot{\varphi}^2$$

and thus the Euler-Lagrange equations are (suitably normalized so that the coefficient of the second order term is one):

$$\begin{aligned} \ddot{t} + 2A' \dot{r} \dot{t} &= 0 \\ \ddot{r} + B' \dot{r} \dot{r} - r e^{-2B} \dot{\theta} \dot{\theta} - r \sin^2 \theta e^{-2B} \dot{\varphi} \dot{\varphi} + A' e^{2(A-B)} c^2 \dot{r} \dot{t} &= 0 \\ \ddot{\theta} + \frac{2}{r} \dot{r} \dot{\theta} - \sin \theta \cos \theta \dot{\varphi} \dot{\varphi} &= 0 \\ \ddot{\varphi} + \frac{2}{r} \dot{r} \dot{\varphi} + \cot \theta \dot{\theta} \dot{\varphi} &= 0 \end{aligned} \quad (6.3)$$

This leads to (note the factors of 1/2 in $\Gamma_{\mu\nu}^\lambda$ when $\mu \neq \nu$).

$$\Gamma_{r\varphi}^\varphi = \frac{1}{r}$$

$$\begin{aligned}
\Gamma_{\theta\varphi}^{\varphi} &= \Gamma_{\varphi\theta}^{\varphi} = \cot \theta \\
\Gamma_{\varphi\varphi}^{\theta} &= -\sin \theta \cos \theta \\
\Gamma_{r\theta}^{\theta} &= \Gamma_{\theta r}^{\theta} = \frac{1}{r} \\
\Gamma_{\varphi\varphi}^r &= -r \sin^2 \theta e^{-2B} \\
\Gamma_{\theta\theta}^r &= -r e^{-2B} \\
\Gamma_{rr}^r &= B' \\
\Gamma_{tt}^r &= A' e^{2(A-B)} c^2 \\
\Gamma_{rt}^t &= \Gamma_{tr}^t = A'
\end{aligned} \tag{6.4}$$

Next we calculate $R_{\mu\nu}$ from

$$R_{\mu\nu} = R_{\mu\rho\nu}{}^{\rho} = -\partial_{\mu}\Gamma_{\nu\rho}^{\rho} + \partial_{\rho}\Gamma_{\mu\nu}^{\rho} - \Gamma_{\rho\nu}^{\sigma}\Gamma_{\mu\sigma}^{\rho} + \Gamma_{\mu\nu}^{\sigma}\Gamma_{\sigma\rho}^{\rho} \tag{6.5}$$

The non-vanishing components are

$$\begin{aligned}
R_{\varphi\varphi} &= \left((rB' - rA' - 1)e^{-2B} + 1 \right) \sin^2 \theta \\
R_{\theta\theta} &= (rB' - rA' - 1)e^{-2B} + 1 \\
R_{rr} &= -A'' - A'^2 + A'B' + \frac{2}{r}B' \\
R_{tt} &= \left(A'' + A'^2 - A'B' + \frac{2}{r}A' \right) e^{2(A-B)} c^2
\end{aligned} \tag{6.6}$$

Next we see that $R_{rr} = 0$ and $R_{tt} = 0$ imply $B' = -A'$. Hence $B = a - A$ for some constant a . In fact we can remove this constant a by rescaling t (which is equivalent to shifting A by a constant). This leaves us with two equations

$$\begin{aligned}
A'' + 2A'^2 + \frac{2}{r}A' &= 0 \\
-(2rA' + 1)e^{2A} + 1 &= 0
\end{aligned} \tag{6.7}$$

We can solve the second one by writing it as

$$r(e^{2A})' + e^{2A} - 1 = 0 \tag{6.8}$$

so if $e^{2A} = 1 + X$ then $rX' = -X$ hence we find

$$e^{2A} = 1 - \frac{R}{r} \tag{6.9}$$

for an arbitrary R . Lastly we must check that the second order equation is satisfied. Again to see this note that it can be written as

$$(e^{2A})'' + \frac{2}{r}(e^{2A})' = 0 \tag{6.10}$$

which is just the harmonic equation for e^{2A} in three dimensions. Thus our final answer is

$$ds^2 = - \left(1 - \frac{R}{r}\right) c^2 dt^2 + \frac{dr^2}{1 - \frac{R}{r}} + r^2(d\theta^2 + \sin^2 \theta d\varphi^2) \quad (6.11)$$

This seems like a miracle since we solved two equations for a single function A . However the second order equation was guaranteed to be solved if the first order one is. This is a consequence of the Bianchi identity (why?) and ensures that the Einstein equations are well posed.

The Schwarzschild metric is in fact it is unique, given our assumptions:

Theorem: (Birkhoff) The Schwarzschild metric is the unique, up to diffeomorphisms, stationary and spherically symmetric solution to the vacuum Einstein equations.

We will not prove this theorem here, it simply follows from the Einstein equations. It is an elementary version of a ‘no-hair’ theorem. The surprising thing about it is that the solution contains no information about what makes up the matter. The result remains true if more complicated matter terms are added to the Lagrangian.

It seems as if there is a problem at $r = R$ and also at $r = 0$. Certainly we can’t use the metric at $r = R$. If you calculate some curvature invariant, such as $R_{\mu\nu\lambda}{}^\rho R^{\mu\nu\lambda}{}_\rho$ then there is no apparent problem at $r = 2R$ but there is a divergence at $r = 0$. For $r < R$ we can still use the Schwarzschild solution. But note that the role of ‘time’ is played by r . In particular, for $r < R$, the metric is no longer time-independent. To understand more we must look at geodesics. We will do this later.

For now it is important to note that we do not have to consider the Schwarzschild metric to be valid everywhere. For example it is also the unique solution outside a static spherically symmetry distribution of matter, whose total mass is M . In particular it describes the spacetime geometry outside a star such as the sun or a planet such as the earth (ignoring their rotation).

Problem: Using the expressions for $R_{\mu\nu}$ and setting $B = -A$ but not imposing any condition on A , show that the components of $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R$ are

$$\begin{aligned} G_{\varphi\varphi} &= r^2 \sin^2 \theta e^{2A} \left(A'' + 2A'^2 + \frac{2}{r}A' \right) \\ G_{\theta\theta} &= r^2 e^{2A} \left(A'' + 2A'^2 + \frac{2}{r}A' \right) \\ G_{rr} &= -e^{-2A} \left(\frac{2}{r}A' e^{2A} + \frac{1}{r^2}e^{2A} - \frac{1}{r^2} \right) \\ G_{tt} &= -c^2 e^{2A} \left(\frac{2}{r}A' e^{2A} + \frac{1}{r^2}e^{2A} - \frac{1}{r^2} \right) \end{aligned} \quad (6.12)$$

6.1 Geodesics

Now we need to physically examine this solution. We start by looking at Geodesics, which, according to Einstein, are the paths of free-falling observers. As we have discussed

we can obtain these from the modify action

$$\begin{aligned} S &= - \int d\tau g_{\mu\nu} \dot{X}^\mu \dot{X}^\nu \\ &= \int d\tau c^2 \left(1 - \frac{R}{r}\right) \dot{t}^2 - \left(1 - \frac{R}{r}\right)^{-1} \dot{r}^2 - r^2 \dot{\theta}^2 - r^2 \sin^2 \theta \dot{\varphi}^2 \end{aligned} \quad (6.13)$$

If we think of this as a Lagrangian for four fields $t(\tau), r(\tau), \theta(\tau)$ and $\varphi(\tau)$, then t and θ do not have and ‘potential’ terms. Thus it follows that

$$E = c^2 \left(1 - \frac{R}{r}\right) \dot{t} \quad (6.14)$$

and

$$L = r^2 \sin^2 \theta \dot{\varphi} \quad (6.15)$$

are constant along any geodesic.

We still have the θ equation. We note here that $\theta = \pi/2$ is guaranteed to be a solution to the Euler-Lagrange equations (why?). Thus we can simply take $\theta = \pi/2$. In fact one can argue that this is sufficient. Just as in the case of Newtonian physics in a spherically symmetric potential, the conservation of angular momentum implies that the motion takes place in a plane. This turns out to be true here too so we simply take that plane to be $\theta = \pi/2$.

We have not identified what τ is but from our discussion of affinely parameterized geodesic we saw that $ds \propto d\tau$. The exact constant of proportionality doesn’t matter so there are three distinct cases: $ds^2 < 0$, $ds^2 = 0$ and $ds^2 > 0$ corresponding to timelike, null or spacelike geodesics. In this case we find

$$\epsilon = - \left(\frac{ds}{d\tau}\right)^2 = \left(1 - \frac{R}{r}\right)^{-1} \frac{E^2}{c^2} - \left(1 - \frac{R}{r}\right)^{-1} \dot{r}^2 - \frac{L^2}{r^2} \quad (6.16)$$

where $\epsilon = 1$ for timelike, $\epsilon = 0$ for null and $\epsilon = -1$ for spacelike geodesics. A little rearranging leads us to

$$\frac{1}{2} \dot{r}^2 + \frac{1}{2} \left(\epsilon + \frac{L^2}{r^2}\right) \left(1 - \frac{R}{r}\right) = \frac{1}{2c^2} E^2 \quad (6.17)$$

This is the equation for a particle with position r in a potential

$$\begin{aligned} V &= \frac{1}{2} \left(\epsilon + \frac{L^2}{r^2}\right) \left(1 - \frac{R}{r}\right) \\ &= \frac{1}{2} \epsilon - \frac{\epsilon R}{2r} + \frac{L^2}{2r^2} - \frac{RL^2}{2r^3} \end{aligned} \quad (6.18)$$

The $1/r$ and $1/r^2$ terms are the usual Newtonian potential if we take $\epsilon = 1$ for a timelike falling observer (the first ϵ term is an irrelevant constant), but we see that there are relativistic corrections due to the additional $1/r^3$ term.

6.2 Newtonian Limit

Let us look at radial ($L = 0$) timelike ($\epsilon = 1$) geodesics at large $r \gg R$. This corresponds to a massive object falling straight down to earth or what ever the source of the curvature is. In this case the leading order equations are

$$E = c^2 \dot{t} \quad \frac{1}{2} \dot{r}^2 - \frac{R}{2r} = \frac{E^2}{2c^2} - 1$$

Now for a slowly moving observer (slow compared to c) it is sensible to take $ct = \tau$, so that the curve is parameterized by time, or what an observer at infinity would call time. Note the factor of c . This is needed on dimensional grounds as we choose τ to have the same dimensions as the proper length s and radius r (we did this when we set $\epsilon = 1$). Thus $E = c$. Differentiating the second equation we recover Newton's law:

$$\ddot{r} = \frac{1}{c^2} \frac{d^2 r}{dt^2} = -\frac{R}{2r^2}$$

provided that we identify $R = 2G_N M/c^2$.

With this in mind we have fixed the Schwarzschild metric to the more familiar form:

$$ds^2 = - \left(1 - \frac{2G_N M/c^2}{r} \right) c^2 dt^2 + \frac{dr^2}{1 - \frac{2G_N M/c^2}{r}} + r^2 (d\theta^2 + \sin^2 \theta d\phi^2) \quad (6.19)$$

But in addition we have recovered Newton's gravitation law $F = -G_N Mm/r^2$ as the leading order effect of the curvature of spacetime.

6.3 Perihelion Shift of Mercury

The higher order corrections to Newton's law that lead to two classic predictions of General Relativity which we now discuss in detail

The first effect is that planets no longer move in elliptical orbits about the sun. This deviation is extremely small for most planets but it was already observed prior to Einstein that the closest planet to the sun, namely Mercury, does not follow an exactly elliptical path. Rather the 'ellipse' slowly rotates. This is called the perihelion shift of Mercury.

To start with lets recover the elliptical motion in the Newtonian limit. Here we look for geodesics with $L \neq 0$. Now it is more helpful to think of r as a function of φ . Thus we have

$$\dot{r} = \frac{d\varphi}{d\tau} \frac{dr}{d\varphi} = Lr'/r^2$$

where $r' = dr/d\varphi$. Let us introduce $u = 1/r$ so that $\dot{r} = -Lu'$. The equation for \dot{r} becomes

$$\frac{1}{2} L^2 u'^2 - \frac{G_N M}{c^2} u + \frac{1}{2} L^2 u^2 - \frac{G_N M L^2}{c^2} u^3 = \frac{1}{2c^2} E^2 - \frac{1}{2} \quad (6.20)$$

At large orbits u is small so we can neglect the u^3 term. If we write $u = G_N M / c^2 L^2 + \tilde{u}$ and drop the u^3 term then we find

$$\frac{1}{2} L^2 \tilde{u}'^2 + \frac{1}{2} L^2 \tilde{u}^2 = \frac{1}{2c^2} E^2 - \frac{1}{2} - \frac{1}{2} \frac{G_N^2 M^2}{c^4 L^2}$$

To clean this up we write $E^2 = c^2 + G_N^2 M^2 / c^2 L^2 + G_N^2 M^2 E_0^2 / c^2 L^2$ and obtain

$$\tilde{u}'^2 + \tilde{u}^2 = \frac{G_N^2 M^2 E_0^2}{c^4 L^4}$$

which has the solutions (putting back the original variables)

$$u = \frac{1}{r} = \frac{G_N M}{c^2 L^2} (1 + E_0 \cos(\varphi - \varphi_0)) \quad (6.21)$$

This is the classic Kepler-Newtonian orbit with eccentricity E_0 . It is an ellipse if $E_0 < 1$ (a circle if $E_0 = 0$), parabola if $E_0 = 1$ or a hyperbola if $E_0 > 1$

Problem: Show that this indeed describes elliptical, parabolic or hyperbolic orbits depending on E_0 .

Next we need to consider the effect of including the u^3 term. Let us start by taking the derivative of (6.20) so that we find

$$L^2 u'' - G_N M / c^2 + L^2 u - 3 \frac{G_N M L^2}{c^2} u^2 = 0$$

Now we expand $u = u_0 + \delta u$ where u_0 is the solution we found in (6.21). Expanding this all out gives

$$L^2 u_0'' - \frac{G_N M}{c^2} + L^2 u_0 - \frac{3G_N M L^2}{c^2} u_0^2 + L^2 \delta u'' + L^2 \delta u - \frac{6G_N M L^2}{c^2} u_0 \delta u - \frac{3G_N M L^2}{c^2} (\delta u)^2 = 0$$

The first three terms vanish since u_0 is a solution of the Newtonian problem. We also drop the $u_0 \delta u$ and $(\delta u)^2$ terms since we assume that both u_0 and δu are small but also that $\delta u \ll u_0$. Thus the leading order correction comes from u_0^2 rather than $u_0 \delta u$. The $(\delta u)^2$ term gives an even smaller correction. Thus we need to solve

$$\delta u'' + \delta u = \frac{3G_N M}{c^2} u_0^2$$

If we substitute in the exact solution for u_0 (setting $\varphi_0 = 0$ for simplicity) we find

$$\delta u'' + \delta u = \frac{3G_N^3 M^3}{c^6 L^4} \left(1 + 2E_0 \cos \varphi + \frac{E_0^2}{2} (\cos(2\varphi) + 1) \right)$$

The solution to this is

$$\delta u = \frac{3G_N^3 M^3}{c^2 L^4} \left(\left(1 + \frac{1}{2} E_0^2 \right) + E_0 \varphi \sin \varphi - \frac{1}{6} E_0^2 \cos(2\varphi) \right)$$

Ugh. However the interesting thing is the $\varphi \sin \varphi$ term that grows without bound. This arises because the $\cos \varphi$ term is acting as a source for the harmonic oscillator but at the resonant frequency. Thus ultimately this term will grow large and invalidate the assumption that δu is a small perturbation. Certainly it will be the dominant correction. From the point of view of the full solution nothing too dramatic can happen. The interpretation of this term is to observe that

$$\frac{G_N M E_0}{c^2 L^2} \cos \left(\left(1 - \frac{6G_N^2 M^2}{c^4 L^2} \right) \varphi \right) = \frac{G_N M E_0}{c^2 L^2} \cos \varphi + \frac{3G_N^3 M^3 E_0}{c^6 L^4} \varphi \sin \varphi + \dots$$

Thus this term can be interpreted as changing the periodicity of the solution. In particular we find

$$u = \frac{1}{r} = \frac{G_N M}{c^2 L^2} \left(1 + E_0 \cos \left(\left(1 + \frac{3G_N^2 M^2}{c^4 L^2} \right) \varphi \right) \right) + \dots$$

where the ellipsis denotes smaller terms.

This gives the perihilion shift of Mercury (it's too small for other planets to observe). We see that the minimum distance of Mercury to the sun occurs at

$$\cos \left(\left(1 + \frac{3G_N^2 M^2}{c^4 L^2} \right) \varphi \right) = 1$$

We have normalized things so that at $\varphi = 0$ Mercury is at its closest point to the sun. But the next time this happens is

$$\varphi = 2\pi \left(1 + \frac{3G_N^2 M^2}{c^4 L^2} \right)^{-1} = 2\pi - \frac{6\pi G_N^2 M^2}{c^4 L^2} + \dots$$

Thus the orbit is 'rotating' forwards.

6.4 Bending of Light

The second prediction of General Relativity beyond the Newtonian theory is that a light ray passing close by the sun will bend. It is not clear in the Newtonian theory how to calculate the bending of light since it has no mass. You might try to interpret it as a particle and assign some effective mass for it due to its energy, *i.e.* by including effects of Special Relativity. However the amount predicted is only half that observed. But in General Relativity it is clear what to do: we look for geodesics with $\epsilon = 0$ but $L \neq 0$. So we just follow the steps that we did for the perihilion shift of Mercury. The equation is now

$$\frac{1}{2} \dot{r}^2 + \frac{L^2}{2r^2} - \frac{G_N M L^2}{c^2 r^3} = \frac{1}{2c^2} E^2 \quad (6.22)$$

which in terms of u is

$$\frac{1}{2} L^2 u'^2 + \frac{1}{2} L^2 u^2 - \frac{G_N M L^2}{c^2} u^3 = \frac{1}{2c^2} E^2 \quad (6.23)$$

Again we start by neglecting the u^3 term. This time we take for our solution (let us set, with out loss of generality, $\varphi_0 = 0$)

$$u = \frac{1}{r} = \frac{E}{cL} \sin \varphi \quad (6.24)$$

This represents a light ray that comes in from infinity at $\varphi = 0$ and then out again to infinity at $\varphi = \pi$. In particular if we write $x = r \cos \varphi$ and $y = r \sin \varphi$ then the solution is just

$$y = \frac{cL}{E}$$

Thus the light ray is undeflected and its closest distance to the run is $r_{min} = cL/E$.

Let us consider the effect of the u^3 term that comes from General Relativity. Again writing $u = u_0 + \delta u$, where u_0 is the solution (6.24), and differentiating we obtain the equation

$$L^2(u_0'' + \delta u'') + L^2(u_0 + \delta u) - \frac{3G_N M L^2}{c^2} (u_0 + \delta u)^2 = 0$$

Once again we only need the term that is linear in δu (again the terms linear in u_0 vanish since u_0 is a solution). This gives

$$\delta u'' + \delta u = \frac{3G_N M}{c^2} u_0^2 + \dots$$

Substituting in for u_0 we find

$$\delta u'' + \delta u = \frac{3G_N M E^2}{2c^4 L^2} (1 - \cos(2\varphi))$$

so the solution is

$$\delta u = \frac{3G_N M E^2}{2c^4 L^2} + \frac{G_N M E^2}{2c^4 L^2} \cos(2\varphi)$$

or

$$\begin{aligned} \frac{1}{r} &= \frac{E}{cL} \sin \varphi + \frac{3G_N M E^2}{2c^4 L^2} + \frac{G_N M E^2}{2c^4 L^2} \cos(2\varphi) \\ &= \frac{E}{cL} \sin \varphi + \frac{3G_N M E^2}{2c^4 L^2} + \frac{G_N M E^2}{2c^4 L^2} (1 - 2 \sin^2 \varphi) \end{aligned} \quad (6.25)$$

Next we need to study this solution in the asymptotic regions where $r \rightarrow \infty$. This gives a quadratic equation for $\sin \varphi$:

$$\sin^2 \varphi - \frac{c^3 L}{G_N M E} \sin \varphi - 2 = 0$$

The solutions are

$$\begin{aligned} \sin \varphi &= \frac{c^3 L}{2G_N M E} \pm \frac{1}{2} \sqrt{\frac{c^6 L^2}{G_N^2 M^2 E^2} + 8} \\ &= \frac{c^3 L}{2G_N M E} \left(1 \pm 1 \pm \frac{1}{2} \frac{8G_N^2 M^2 E^2}{c^6 L^2} + \dots \right) \end{aligned} \quad (6.26)$$

Now the effect needs to be small as $G_N M \rightarrow 0$ and thus only the lower sign solution makes sense (*i.e.* has $|\sin \varphi| \leq 1$):

$$\sin \varphi = -\frac{2G_N M E}{c^3 L}$$

Expanding for small φ we find the ray comes in at

$$\varphi_1 = -\frac{2G_N M E}{c^3 L} + \dots$$

and leaves at

$$\varphi_2 = \pi + \frac{2G_N M E}{c^3 L} + \dots$$

Corresponding to a deflection angle of

$$\Delta\varphi = \frac{4G_N M E}{c^3 L} + \dots$$

Note that E can be related to the smallest value of r that the light ray takes. In the unperturbed solution we see that the minimum value of r is $r_{min} = cL/E$. To leading order this is unchanged in the perturbed solution so that

$$\Delta\varphi = \frac{4G_N M}{c^2 r_{min}} + \dots$$

6.5 Gravitational Redshift and Time Dilation

There are other important effects of General Relativity. Just as in Special Relativity spacetime has the ‘ability’ to cause time dilation and Doppler-like red-shifts. Let us look at a two observers. For simplicity we assume that they are both at a constant value of r . Each observer sits at a constant values of r, θ and φ . To them time is ‘proper time’ in the sense that $(\Delta s)^2 = -c^2 \Delta t_i^2$ and hence

$$\Delta t_i = \sqrt{1 - \frac{2G_N M / c^2}{r_i}} \Delta t$$

for $i = 1, 2$. This leads to

$$\frac{\Delta t_1}{\sqrt{1 - \frac{2G_N M / c^2}{r_1}}} = \frac{\Delta t_2}{\sqrt{1 - \frac{2G_N M / c^2}{r_2}}}$$

More generally we have (as long as g_{00} is time-independent)

$$\frac{\Delta t_1}{\Delta t_2} = \sqrt{\frac{g_{00}(x_1)}{g_{00}(x_2)}}.$$

A standard notion of a clock is given by the (inverse) frequency of light: $\Delta t = 1/\nu$. Suppose that an observer at $r = r_1$ emits a light ray that escapes to some larger value of $r_2 > r_1$. If the frequency of emitted light is ν_1 this corresponds to taking $\Delta t_1 = 1/\nu_1$. However at infinity an observer will measure:

$$\nu_2 = \sqrt{\frac{1 - \frac{2G_N M/c^2}{r_1}}{1 - \frac{2G_N M/c^2}{r_2}}} \nu_1$$

Thus $\nu_2 < \nu_1$. This is known as a red-shift since light will sign as more red (lower frequency) at infinity than it is as seen from where it was emitted. In the limit that $r_2 \rightarrow \infty$ we get

$$\nu_\infty = \sqrt{1 - \frac{2G_N M/c^2}{r_1}} \nu_1$$

and in particular $\nu_\infty < \nu_1$ if $2G_N M/c^2 < r_1 < \infty$. Thus the light seen at infinity has a lower frequency, ie it has been red-shifted.

6.6 Schwarzschild Solution as a Black Hole

So far we have been treating the Schwarzschild solution at distances $r > 2G_N M/c^2$. However it is most famous for it's behaviour at $r < 2G_N M/c^2$. Let us look at radial geodesics, these correspond to in-falling observers traveling at constant values of θ and φ . Recall that our effective dynamical system is

$$\frac{1}{2} \dot{r}^2 + \frac{1}{2} \epsilon - \frac{G_N M \epsilon}{r} = \frac{1}{2c^2} E^2 \quad (6.27)$$

There is nothing special from this equation about $r = 2G_N M/c^2$. Indeed only $r = 0$ seems singular (there is an infinite potential there). It follows from general considerations that one will pass through $r = 2G_N M/c^2$ in a finite affine time (that is a finite value of τ). In addition one will hit $r = 0$ too. However these geodesics are parameterized by the proper time τ , which is the time that the in-falling observer feels. Let us consider what an observer at a safe distance from $r = 2G_N M/c^2$ sees.

Thus we want to consider dr/dt as opposed to $dr/d\tau$. We simply note that

$$\begin{aligned} \frac{dr}{d\tau} &= \frac{dt}{d\tau} \frac{dr}{dt} \\ &= \dot{t} \frac{dr}{dt} \\ &= \frac{E}{c^2} \left(1 - \frac{2G_N M/c^2}{r}\right)^{-1} \frac{dr}{dt} \end{aligned} \quad (6.28)$$

From here we see that the geodesic equation is

$$\frac{1}{2c^4} E^2 \left(\frac{dr}{dt}\right)^2 = \left(1 - \frac{2G_N M/c^2}{r}\right)^2 \left(\frac{1}{2c^2} E^2 - \frac{1}{2} \epsilon \left(1 - \frac{2G_N M/c^2}{r}\right)\right) \quad (6.29)$$

Near $r \rightarrow \infty$ this modification does not do much as $t \sim E\tau$. However as we approach $r = 2G_N M/c^2$ we see that

$$\begin{aligned} \frac{1}{2c^2} E^2 \left(\frac{d\delta r}{dt} \right)^2 &= \frac{1}{2} E^2 \left(1 - \frac{2G_N M/c^2}{2G_N M/c^2 + \delta r} \right)^2 + \dots \\ &= \frac{1}{2} \frac{E^2}{4G_N^2 M^2/c^4} (\delta r)^2 + \dots \end{aligned} \tag{6.30}$$

where $r = 2G_N M/c^2 + \delta r$. Thus we see that

$$\frac{d\delta r}{dt} = -\frac{\delta r}{2G_N M/c^2} + \dots \tag{6.31}$$

and hence, near $r = 2G_N M/c^2$ we have

$$\delta r = e^{-\frac{t-t_0}{2G_N M/c^2}} \tag{6.32}$$

This shows that r never reaches $r = 2G_N M/c^2$ for any finite value of t . Thus an observer at infinity, for whom t is the time variable, will never see an in-falling observer reach $r = 2G_N M/c^2$. Whereas we saw that the in-falling observer will pass smoothly through $r = 2G_N M/c^2$ in a finite proper time.

This is gravitational red-shifting gone mad. If we fix one observer at infinity, $r_1 = \infty$ then our red-shifting formula states that, for every one second in our frame the in-falling observe has

$$\Delta t_2 = \frac{1}{1 - \frac{2G_N M/c^2}{r_2}} \rightarrow \infty$$

as $r_2 \rightarrow 2G_N M/c^2$. Thus to an observer at infinity the clocks on the in-falling observe slow down to zero as they get closer and closer to $r = 2G_N M/c^2$. They appear to become frozen and live for ever.

Thus to an outside observer the region $r \leq 2G_N M/c^2$ is causally disconnected, they cannot send in any probe, say a light beam or an astronaut, which will be able to go into this region and return. The surface $r = 2G_N M/c^2$ is called the event horizon because observers outside the horizon will never be able to probe what is beyond $r = 2G_N M/c^2$, whereas a freely falling observer will smoothly pass through in a finite time. Of course it also can be shown that, as is well known, no signal inside $r = 2G_N M/c^2$ can reach the outside. This is the classic example of a black hole.

Clearly we need to understand the spacetime near $r = 2G_N M/c^2$ better. In particular the coordinates are breaking down but this does not mean that spacetime is breaking down. This could be, and in fact is, just a coordinate singularity.

To find a new set of coordinates let us consider a coordinate that is well-suited to the in-falling observer. In particular consider such an observer who is lightlike, so that $ds^2 = 0$

$$0 = - \left(1 - \frac{2G_N M/c^2}{r} \right) c^2 dt^2 + \frac{dr^2}{1 - \frac{2G_N M/c^2}{r}}$$

This gives

$$\pm cdt = \frac{dr}{\left(1 - \frac{2G_N M/c^2}{r}\right)^2} = dr + \frac{2G_N M/c^2}{r - 2G_N M/c^2} dr$$

We can integrate both sides of this to find

$$\pm ct = r + \frac{2G_N M}{c^2} \ln \left(\frac{c^2}{2G_N M} r - 1 \right)$$

Here we see that $t \rightarrow \infty$ as $r \rightarrow 2G_N M/c^2$, just as we argued above for the timelike case. Therefore we can choose to use the new coordinate

$$u = ct + r + \frac{2G_N M}{c^2} \ln \left(\frac{c^2}{2G_N M} r - 1 \right)$$

so that $du = 0$ along an in-falling null geodesic.

By construction we have

$$du = cdt + \frac{dr}{1 - \frac{2G_N M/c^2}{r}}$$

so we can rewrite the Schwarzschild metric as

$$\begin{aligned} ds^2 &= - \left(1 - \frac{2G_N M/c^2}{r}\right) \left(du - \frac{dr}{1 - \frac{2G_N M/c^2}{r}}\right)^2 + \frac{dr^2}{1 - \frac{2G_N M/c^2}{r}} + r^2(d\theta^2 + \sin^2 \theta d\varphi^2) \\ &= - \left(1 - \frac{2G_N M/c^2}{r}\right) du^2 + 2dudr + r^2(d\theta^2 + \sin^2 \theta d\varphi^2) \end{aligned} \quad (6.33)$$

These are called in-falling Eddington-Finkelstein coordinates. Now look: there is nothing singular about $r = 2G_N M/c^2$. While it is true that $g_{uu} = 0$ there the metric is still invertible due to the $drdu$ term. However $r = 0$ is singular. Indeed one can show that $R_{\mu\nu\lambda}{}^\rho R^{\mu\nu\lambda}{}_\rho$ diverges at $r = 0$.

Problem: Compute the inverse metric and show it exists at $r = 2G_N M/c^2$.

Problem: Show that it is impossible to find a coordinate transformation that makes $R_{\mu\nu\lambda}{}^\rho R^{\mu\nu\lambda}{}_\rho$ well behaved at $r = 0$.

We can get an intuitive notion for what is happening by observing that the original coordinates r, t are valid at $r > 2G_N M/c^2$ and also at $r < 2G_N M/c^2$. The first region we have studied a length in the previous sections. Consider the second region. Here the metric is better written as

$$ds^2 = - \frac{dr^2}{\frac{2G_N M/c^2}{r} - 1} + \left(\frac{2G_N M/c^2}{r} - 1 \right) c^2 dt^2 - r^2(d\theta^2 + \sin^2 \theta d\varphi^2)$$

so that here r is time and t is space. The metric is now wildly time dependent. Thus the roles of space and time have interchanged. As we have seen they have done so smoothly. Looking again at the Eddington-Finkelstein coordinates we see that at large r we have

$$ds^2 = -du^2 + 2dudr + r^2(d\theta^2 + \sin^2\theta d\varphi^2)$$

If we write $r = (v + u)/2$ then we have

$$ds^2 = dudv + r^2(d\theta^2 + \sin^2\theta d\varphi^2) \quad (6.34)$$

These are the normal coordinates for Minkowski space where u, v describe the lightcone about $t = r = 0$. That is the light cone consists of the points $u = \text{constant}$ or $v = \text{constant}$.

Problem: Show that in flat space the with $u = ct + r$, $v = -ct + r$ the metric is (6.34) and that the light cone through $t = r = 0$ is parameterized by $u = 0$ or $v = 0$.

As we fall into $r = 2G_N M/c^2$ the light cone defined by u, v or r, t ‘tips’. Eventually we reach $r = 2G_N M/c^2$ where the Eddington-Finkelstein metric is

$$ds^2 = 2dudr + r^2(d\theta^2 + \sin^2\theta d\varphi^2)$$

Thus the light cones have tipped so much that u and r are both lightlike (u is lightlike everywhere). Once in the region $r < 2G_N M/c^2$ we again have lightcones but they have tipped so much that the singularity at $r = 0$ (which should now be thought of as a time in the future) now lies inside the light cone of an in-falling observer. This means that they are doomed to hit $r = 0$. It has become an event in their future, that exists throughout space, and it therefore unavoidable.

The surface defined by $r = 2G_N M/c^2$ is called an event horizon and the radius $2G_N M/c^2$ is called the Schwarzschild radius. We have seen that lightrays and observers who fall in through $r = 2G_N M/c^2$ cannot escape back to $r > 2G_N M/c^2$ and in fact are doomed to hit $r = 0$ where there is a singularity. This is known as a Black Hole.

Does this mean that there is a Black Hole at the centre of our sun? No. The point is that the Schwarzschild metric is the solution in the absence of matter (*i.e.* with $T_{\mu\nu} = 0$). A star, like our sun, has matter in the region $r < R_s$ where R_s is the radius of the sun. The Schwarzschild solution is then valid for $r > R_s$.

A different, non-singular, solution is required for $r < R_s$ corresponding to the fact that $T_{00} \neq 0$ there. However if there is a body that is so dense that its radius is less than its Schwarzschild radius then it will be a black hole as the Schwarzschild solution is valid for values of r that include $2G_N M/c^2$.

For example the sun has $R_s = 7 \times 10^7$ m and $2G_N M/c^2 = 3 \times 10^3$ m. However there is a black hole at the centre of our (and most) galaxies.

7 Cosmology: FRW

The idea now is to model the cosmic history of our universe. In particular we assume that the spacetime has spacelike hyper-surfaces with the maximum symmetry. Therefore the 3-dimensional spatial cross sections are either hyperbolic space H^3 , flat space \mathbf{R}^3 or spheres S^3 . Let us endow this space with local coordinates x^i and metric γ_{ij} . Without loss of generality we can take the four-dimensional metric to be of the form

$$ds^2 = -c^2 dt^2 + a^2(t) \gamma_{ij} dx^i dx^j \quad (7.1)$$

Thus the only parameter is the scale factor $a(t)$ which gives the physical size of the spatial hypersurfaces.

We also consider a non-vanishing energy-momentum tensor. To be consistent with the symmetries we assume that

$$T_{\mu\nu} = \begin{pmatrix} \rho & 0 \\ 0 & p a^2 \gamma_{ij} \end{pmatrix} \quad (7.2)$$

i.e. we assume that the only non-vanishing components consist of an energy density ρ as well as an isotropic pressure p_i . Note that we do not assume that ρ and p are constant. Thus the general assumption of the model is that, on a large cosmological scale, the universe is isotropic and homogeneous. The former means that there is no preferred direction whereas the latter means that there are no preferred points. These seem like very reasonable assumptions. Certainly as far as we can tell, on the largest scales that we can observe, the universe consists of an even but sparse distribution of galaxies and hence can be thought of as homogeneous. As far as we can tell the universe is also isotropic however since we can only look out from where we are it could be the universe is not isotropic, so that it has some kind of centre, in which case we must be relatively near the centre. However such an earth-centric view has been out of fashion in cosmology for hundreds of years, *i.e.* since Copernicus.

This means that we choose the metric γ_{ij} to be one of three possibilities: \mathbf{R}^3 (a 3D plane), S^3 (a 3D sphere) or H^3 (3D hyperbolic space). These are the three manifolds that are isotropic and homogenous, meaning that there are no preferred points and no preferred directions. Since we can absorb a constant scale factor these metric take the form

$$ds_3^2 = \gamma_{ij} dx^i dx^j = \frac{dr^2}{1 - kr^2} + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2)$$

With $k = -1, 0, 1$. Clearly for $k = 0$ we have \mathbf{R}^3 in spherical coordinates:

$$ds_3^2(k = 0) = dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\varphi^2$$

For $k = 1$ we write $r = \sin \psi$ and find

$$ds_3^2(k = 1) = d\psi^2 + \sin^2 \psi d\theta^2 + \sin^2 \psi \sin^2 \theta d\varphi^2$$

which is just the metric on S^3 . For $k = -1$ we write $r = \sinh \rho$ in which case we have

$$ds_3^2(k = -1) = d\rho^2 + \sinh^2 \rho d\theta^2 + \sinh^2 \rho \sin^2 \theta d\varphi^2$$

This is the metric on hyperbolic space. Unfortunately we don't have time to go into the details here. In all these cases one finds the the Ricci tensor r_{ij} constructed from γ_{ij} satisfies

$$r_{ij} = 2k\gamma_{ij}$$

Experiments indicate that, in our universe, $k = 0$ so if you'd like you can simply take this case and use $\gamma_{ij} = \delta_{ij}$ in Cartesian coordinates.

We need to solve the Einstein equation

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \kappa^2 T_{\mu\nu}$$

To begin with we must calculate the Levi-Civita connection coefficients

$$\begin{aligned}\Gamma_{ij}^t &= \frac{1}{2} \left(-\frac{1}{c^2} \right) (\partial_i g_{tj} + \partial_j g_{ti} - \partial_t g_{ij}) = \frac{1}{c^2} a \dot{a} \gamma_{ij} \\ \Gamma_{tj}^i &= \frac{1}{2} \frac{1}{a^2} \gamma^{il} (\partial_t g_{lj} + \partial_j g_{lt} - \partial_l g_{tj}) = \frac{\dot{a}}{a} \delta_j^i \\ \Gamma_{jk}^i &= \frac{1}{2} \frac{1}{a^2} \gamma^{il} (\partial_j g_{lk} + \partial_k g_{lj} - \partial_l g_{jk}) = \gamma_{jk}^i\end{aligned}\tag{7.3}$$

where γ_{jk}^i are the Levi-Civita connection coefficients of the spatial metric γ_{ij} . From these we can calculate the Ricci tensor

$$R_{\mu\nu} = R_{\mu\lambda\nu}{}^\lambda = -\partial_\mu \Gamma_{\nu\lambda}^\lambda + \partial_\lambda \Gamma_{\mu\nu}^\lambda - \Gamma_{\mu\lambda}^\sigma \Gamma_{\nu\sigma}^\lambda + \Gamma_{\mu\nu}^\sigma \Gamma_{\sigma\lambda}^\lambda\tag{7.4}$$

We find the non-zero components are

$$\begin{aligned}R_{tt} &= -\partial_t \left(3 \frac{\dot{a}}{a} \right) - 3 \frac{\dot{a}^2}{a^2} \\ &= -3 \frac{\ddot{a}}{a} \\ R_{ij} &= r_{ij} + \frac{1}{c^2} \partial_t (a \dot{a} \gamma_{ij}) - \frac{2}{c^2} \dot{a}^2 \gamma_{ij} + \frac{3}{c^2} \dot{a}^2 \gamma_{ij} \\ &= r_{ij} + \frac{1}{c^2} a \ddot{a} \gamma_{ij} + \frac{2}{c^2} \dot{a}^2 \gamma_{ij}\end{aligned}\tag{7.5}$$

$$\tag{7.6}$$

where r_{ij} is the Ricci tensor for the spatial manifold. Since it is maximally symmetric we have that

$$r_{ij} = 2k\gamma_{ij}\tag{7.7}$$

for some $k = 0, -1, 1$. Continuing we see that the Ricci scalar is

$$\begin{aligned} R &= -c^{-2}R_{tt} + a^{-1}\gamma^{ij}R_{ij} \\ &= \frac{6}{c^2}\frac{\ddot{a}}{a} + 6\frac{k}{a^2} + \frac{6}{c^2}\frac{\dot{a}^2}{a^2} \end{aligned} \quad (7.8)$$

Putting these together we find that the Einstein equation is

$$\begin{aligned} R_{tt} + \frac{c^2}{2}R &= 3\frac{\dot{a}^2}{a^2} + 3\frac{kc^2}{a^2} \\ &= \kappa^2 T_{tt} \\ R_{ij} - \frac{1}{2}a^2\gamma_{ij}R &= \left(-\frac{2}{c^2}a\ddot{a} - \frac{1}{c^2}\dot{a}^2 - k\right)\gamma_{ij} \\ &= \kappa^2 T_{ij} \end{aligned} \quad (7.9)$$

Thus we find the equations

$$\begin{aligned} 3\frac{\dot{a}^2}{a^2} + 3\frac{kc^2}{a^2} &= \kappa^2\rho \\ 2\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2} + \frac{kc^2}{a^2} &= -\kappa^2 c^2 p \end{aligned} \quad (7.10)$$

The first equation is known as the Friedman equation. Note that it is first order in time derivatives. This is a consequence of the Bianchi identity. Quite often the second equation is rewritten, using the Friedman equation, so that

$$\begin{aligned} \frac{\dot{a}^2}{a^2} + \frac{kc^2}{a^2} &= \frac{1}{3}\kappa^2\rho \\ \frac{\ddot{a}}{a} &= -\frac{1}{6}\kappa^2(3pc^2 + \rho) \end{aligned} \quad (7.11)$$

One famous feature of these equations is that we can only find a static universe if $p = \rho = 0$ (in which case we have Minkowski space). Thus the universe is evolving. Indeed not long after General Relativity was invented Hubble observed the expansion of the universe (which at that time was not a prediction of General Relativity).

Often one introduces the Hubble ‘constant’ $H = \dot{a}/a$ however, except for the case of exponential expansion, H is not constant (although its time variation is over cosmic scales).

Note that at late times, meaning large a , one can drop the k term from the equations.

In addition the matter ‘equation of state’ is often taken to be $p = w\rho/c^2$ where w is a constant. For any known type of matter one has that $w \geq -1$ and so this is generally assumed to be the case. In this case one can solve the equations. Let us assume for simplicity that $k = 0$ and try

$$a = a_1 t^\gamma \quad (7.12)$$

for some constant γ . From the Friedman equation we see that

$$\frac{1}{3}\kappa^2\rho = \gamma^2 t^{-2} \quad (7.13)$$

and hence substitution into the remaining equation gives

$$\begin{aligned}\gamma(\gamma - 1)t^{-2} &= -\frac{1}{6}\kappa^2(3w + 1)\rho \\ &= -\frac{3w + 1}{2}\gamma^2t^{-2}\end{aligned}\tag{7.14}$$

This implies that

$$\frac{3w + 3}{2}\gamma^2 = \gamma\tag{7.15}$$

and hence we find

$$a(t) = a_1t^{\frac{2}{3w+3}}\tag{7.16}$$

where $a_1 = a(1)$. In the limit that $w = -1$ the solution becomes exponential:

$$a = a_1e^{\sqrt{\frac{\kappa^2\rho}{3}}t}$$

This spacetime is known as de Sitter space.

Notice that in the $w > -1$ case there is a singularity at some early time where $a = 0$. This is a genuine curvature singularity. Thus the present observation of the expansion of the universe points to an initial ‘Big Bang’ singularity. Note that the Universe isn’t expanding into anything and indeed for $k \neq 1$ the ‘size’ of the spatial cross sections is infinite.

Problem: Find a change of variables that maps the metric

$$ds^2 = -c^2dt^2 + a^2(t)\gamma_{ij}dx^i dx^j$$

to so-called conformal time:

$$ds^2 = a^2(\eta)(-d\eta^2 + \gamma_{ij}dx^i dx^j)$$

What is η exactly for the case that $a(t) = a_1t^\gamma$?

Another concept that arises is that of a particle horizon. Consider a massless particle moving such that, along Σ , $\gamma_{ij}dx^i dx^j = dr^2$ for some variable r . It follows a null geodesic

$$0 = -c^2dt^2 + a^2dr^2\tag{7.17}$$

Thus in the time after $t = t_0$ the particle will travel a distance

$$R = c \int_{t_0}^t \frac{dt'}{a(t')}\tag{7.18}$$

The point is that this can be bounded. For example if the Hubble parameter $H = \dot{a}/a$ is constant, *i.e.* $a = a_0e^{Ht}$ where $a_0 = a(0)$, then

$$R = -ca_0^{-1}H^{-1}e^{-Ht}\Big|_{t_0}^t = ca_0^{-1}H^{-1}(e^{-Ht_0} - e^{-Ht})\tag{7.19}$$

Thus if we send the particle out today so that $t_0 = 0$, even if we wait until $t \rightarrow \infty$ we see that R will be bounded, *i.e.* the particle can only make it out a finite distance due to the expansion of the universe.

Another point to notice is that at different periods in the expansion different fields may have dominated, with different values of w , thereby leading to different expansion rates (but always positive). In particular initially the Universe was hot and the matter in the form of atoms had not yet formed and hence the main contribution was largely photons and radiation. The energy momentum tensor of the electromagnetic field is traceless $g^{\mu\nu}T_{\mu\nu} = 0$ and hence $w = 1/3$. In this ‘radiation’ era the Universe expanded as $a \propto t^{1/2}$.

Problem: Show that $g^{\mu\nu}T_{\mu\nu} = 0$ and $w = 1/3$ for the electromagnetic field. (HINT: you may recall that $T_{\mu\nu} = F_{\mu\lambda}F_{\nu}^{\lambda} - \frac{1}{4}g_{\mu\nu}F^2$).

After the Universe cooled down enough matter did form and the dominant source of the energy momentum tensor was from a pressureless gas of matter. Here $w = 0$ and hence $a \sim t^{2/3}$. This is known as the matter dominated era.

We now think there are two other era’s in the universe’s expansion. Both of these correspond to $w = -1$ (*i.e.* $p = -\rho$) and exponential expansion. In this case we see that

$$T_{\mu\nu} = \begin{pmatrix} \rho & 0 \\ 0 & -\rho a^2/c^2 \gamma_{ij} \end{pmatrix} = -\rho/c^2 g_{\mu\nu}$$

This type of energy momentum tensor is caused by a non-vanishing, vacuum energy ρ . Such a contribution to Einstein’s equation is called a cosmological constant. it corresponds to adding in a term to the Einstein equation

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + \Lambda g_{\mu\nu} = \kappa^2 T_{\mu\nu}$$

However this is now normal viewed as coming from a vacuum energy term $Vg_{\mu\nu}$ in $T_{\mu\nu}$.

What are the two era’s? One is now. Observation now strongly supports the notion that we are undergoing small but exponential expansion. The associated cosmological constant Λ is small but, in time, the exponential growth in a has come to dominate.

The other era was in the early universe, before the radiation dominated era. This is called inflation and the cosmological constant or vacuum energy in that case was large. This era is still somewhat speculative but it has become the dominant idea in cosmology and has begun to pass observational tests. However the original motivation for inflation was to solve some problems with the standard expansion scenario. In particular there are two key problems (there are others):

- Why is the universe in thermal equilibrium across the whole sky? We have measure the temperature of space and it is the same everywhere up to fluctuations of the order of 10^{-5} . Under the power-law expansions of the matter and radiation era’s different parts of the sky have not been in causal contact with each other. So it is rather striking that their temperature is the same to within a factor of 10^{-5} .

- Why is the universe so close to $k = 0$. The universe that we observe today is very flat. Although expansion clearly flattens space (the effect of k drops as a^{-2}) the power law expansions of the matter and radiation era's are not enough to account for this unless we assume that the initial universe was very very flat.

To solve these problems one postulates a period of vast expansion. In particular one requires a to have grown by a factor of e^{60} , *i.e.*

$$a(\text{after}) = e^{60}a(\text{before})$$

So the most natural thing to imagine is the exponential expansion associated to a cosmological constant. The number e^{60} is referred to as 60 e-foldings. Therefore the part of the universe that we see all came from an extremely small, quantum-sized, portion of the universe immediately after the big bang. Indeed the temperature fluctuations that we see, and which led to structure formation, are just quantum fluctuations of the initial space.

Appendix: Comparison with Gerard Watt's Course

In recent years this course has been given by Dr Watts. Although the content of his course and this one are essentially the same there are some conventions and notations that differ. Thus in order to read his notes and practice his exams perhaps it is helpful to list here some of the differences between his course and this one. With these comments in mind you should be able to translate all of the questions in his exams into a form that you can try to solve.

- He uses the $(+, -, -, -)$ signature for the metric, *i.e.* Minkowski space has the metric $ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2$.
- He discusses the Newtonian force law more generally (*i.e.* not just about a spherically symmetric source):

$$F_i = -m\partial_i\varphi \quad \nabla^2\varphi = 4\pi G_N\rho(x)$$

where $\rho(x)$ is the mass distribution. The $F = -G_N Mm/r^2$ law that we used simply corresponds to a point source at the origin: $\rho = M\delta(x)$.

- He calls a (p, q) tensor a $\begin{bmatrix} p \\ q \end{bmatrix}$ tensor
- When discussing geodesics he introduces the notion of covariant derivative along the curve:

$$\frac{DU^\lambda}{D\tau} \equiv \dot{X}^\mu D_\mu U^\lambda = \dot{X}^\mu (\partial_\mu U^\lambda + \Gamma_{\mu\nu}^\lambda U^\nu)$$

where U^λ is an arbitrary vector and $\dot{X}^\mu = dX^\mu/d\tau$ is the tangent to the curve $X^\mu(\tau)$.

Alternatively we can think of this as

$$\frac{DU^\lambda}{D\tau} = \dot{U}^\lambda + \Gamma_{\mu\nu}^\lambda \dot{X}^\mu U^\nu$$

because $\dot{U}^\lambda = \dot{X}^\mu \partial_\mu U^\lambda$ by the chain rule.

A vector U^λ is said to be parallel transported along a curve $X^\mu(\tau)$ if $DU^\lambda/D\tau = 0$. An affine geodesic is a curve whose tangent vector is parallel transported along itself: $D\dot{X}^\lambda/D\tau = 0$.

- In cosmology he writes out the cosmological constant term explicitly so that

$$T_{\mu\nu} = \Lambda g_{\mu\nu} + \begin{pmatrix} \rho & 0 \\ 0 & pa^2\gamma_{ij} \end{pmatrix}$$

whereas we absorb Λ into ρ and p and only consider the second term (note that $-c^{-2}\rho = p = \Lambda$ in (7.2) is just a cosmological constant).

References

- [1] J. Hartle, *Gravity*, Addison Wesley, 2003.
- [2] R. Wald, *General Relativity*, University of Chicago Press, 1984.