Gravitation och Kosmologi
Lecture Notes

Joseph A. Minahan c

Uppsala, 2002-2012

Chapter 0

Overview

This course is an introduction to Einstein's theory of general relativity. It is assumed that you are already familiar with special relativity, which is relevant when particles are traveling near to the speed of light[1]. General relativity takes the concepts from special relativity but then generalizes them so that one may also include gravity. Hence, in general relativity we will see many notions already introduced in special relativity, such as 4-vectors, tensors, proper times *etc.* But we will also learn some new concepts such as metrics, local inertial frames, parallel transport and curvature. Eventually we will put together these new pieces to construct the Einstein equations.

Once we have the equations, we will look for solutions. When Einstein first formulated his equations, he assumed that exact solutions would never be found. However, within six months a 42 year old German physicist named Karl Schwarzchild did find an exact solution while serving in the German army. Unfortunately for Schwarzchild he died the next year at the Russian front (although from disease). At first, Schwarzchild's solution was considered a curiosity, for it had many strange properties. But it was later realized that it was the solution for a black hole. We will spend some time in this class studying Schwarzchild's solution.

It is also possible to find another class of exact solutions to Einstein's equations. These are the solutions that describe a uniform expanding universe, which happily, is very close to how our own universe has evolved through time. We will spend a significant part of time on these solutions, which makes up the "Kosmologi" part of the course.

0.1 A note on units

The speed of light is a natural constant that appears throughout relativity, so physicists often use units to simplify this constant. In this course we will adopt this convention and choose units where c, the speed of light, is $c = 1$. So for example, if we use one meter as our unit of length, then the unit of time is also a meter. In other words, this is the amount of time that it takes for light to travel one meter. As you can see, the speed

[1] A fairly complete set of notes on special relativity can be found at http://www.teorfys.uu.se/people/minahan/Courses/SR/notes.pdf

is $c = \frac{1\,\text{meter}}{1\,\text{meter}} = 1$. If we measure masses using kilograms, then energy and momentum are also measured in kilograms. Velocities are dimensionless.

The big advantage of using these units, is that it is no longer necessary to write down factors of c. Note that we can always go back to the more conventional units at the end of any computation. So for example, suppose you are computing a time, T and your answer is 100 meters. The conversion rate is $1\,\text{second} = 3\times10^8\,\text{meters}$. The time in seconds is then $T = 100\,\text{meters}/(3\times10^8\,\text{meters/second}) \approx 3.3\times10^{-7}\,\text{seconds}$.

Contents

Chapter 1

Review of Special Relativity

Before turning to general relativity we will first review some aspects of special relativity where gravity is absent.

1.1 The basics

In special relativity we learned that observers in different reference frames can have different measurements of physical quantities. For us, a reference frame **S** will be a set of coordinates in 4 space-time dimensions (t, x, y, z), but in order to uniformize the coordinates we will instead write this as

$$(x^0, x^1, x^2, x^3)\,, \tag{1.1.1}$$

where $x^0 = ct = t$ is the time coordinate. Notice that we have replaced c with 1. A point in space-time is called an *event* and the displacement between two events we will write as $\overrightarrow{\Delta x}$ where $\overrightarrow{\Delta x}$ is a *4-vector* with components Δx^μ where $\mu = 0 \ldots 3$. An observer at rest in a different reference frame $\mathbf{S}'$ will have a different set of coordinates

$$(x^{0'}, x^{1'}, x^{2'}, x^{3'})\,, \tag{1.1.2}$$

and will measure different components for the displacement 4-vector, $\Delta x^{\mu'}$. We are assuming that both observers are referring to the same two events so the displacement 4-vector is the same in both frames. What differs is how the vector is written in the components of the respective frames. To perhaps make this clearer, consider the two dimensional vector $\vec{A}$ in figure 1.1. Two sets of coordinates are given (x, y) and (x', y') and the primed coordinates are related to the unprimed coordinates by a rotation of angle θ. In the unprimed coordinates the components of $\vec{A}$ are $(a, 0)$, while in the primed coordinates they are given by $(\cos\theta\, a, -\sin\theta\, a)$, where a is the length of $\vec{A}$. But the vector itself has not changed.

In special relativity we are usually concerned with a particular type of reference frame called an *inertial frame*. An inertial frame has the properties:

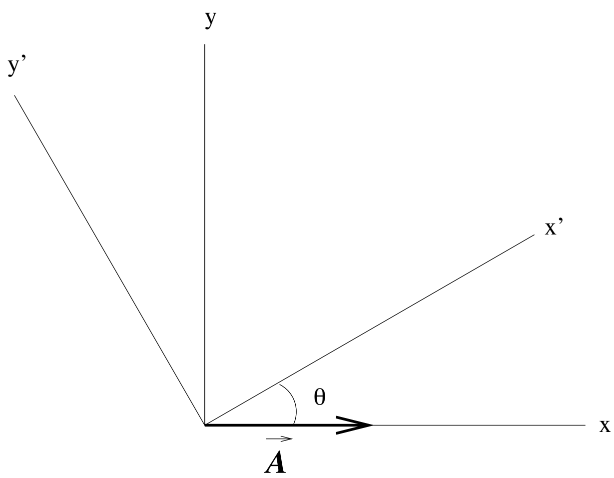

Figure 1.1: Vector $\vec{A}$ in terms of two different coordinate systems.

1. *There is a universal time coordinate that can be synchronized everywhere in the inertial frame.* This means is that at every spatial point in the reference frame we can place a clock and that all the clocks agree with each other.

2. *The spatial components are Euclidean.* That is, the spatial components satisfy all axioms of Euclidean geometry.

3. *A body with no forces acting on it will travel at constant velocity according to the clocks and measuring sticks in the inertial frame.*

An important question is how a measurement differs for two observers who are at rest in different inertial frames, which we call **S** and **S**′. To answer this question we turn to Einstein's two postulates for special relativity,

1. The laws of physics are identical in any inertial frame.

2. The speed of light in a vacuum, c, is the same in any inertial frame.

The first postulate tells us that the relation between coordinates in **S** and **S**′ is linear. To see this, suppose that there is a body with no forces on it and so moves at constant velocity in both frames. The body moves on a space-time trajectory called a world line. In figure 1.2 we show the world-line for this body on a *space-time diagram.* Since there are no forces on the body, the world-line is a straight line.

To measure a velocity we observe the displacement between any two events on the world-line, say for instance **P** and **Q**. The velocity components in **S** are $v^i = \frac{\Delta x^i}{\Delta x^0}$ where $i = 1, 2, 3$ (we will often use latin indices to signify spatial components) while those in **S**′ are $v^{i'} = \frac{\Delta x^{i'}}{\Delta x^{0'}}$. The velocity components for either frame are the same no matter which two events we choose. In order for this to be true for any choice of the two events, we must have

$$\Delta x^{\mu'} = \Lambda^{\mu'}{}_{\nu}\, \Delta x^{\nu}\,, \tag{1.1.3}$$

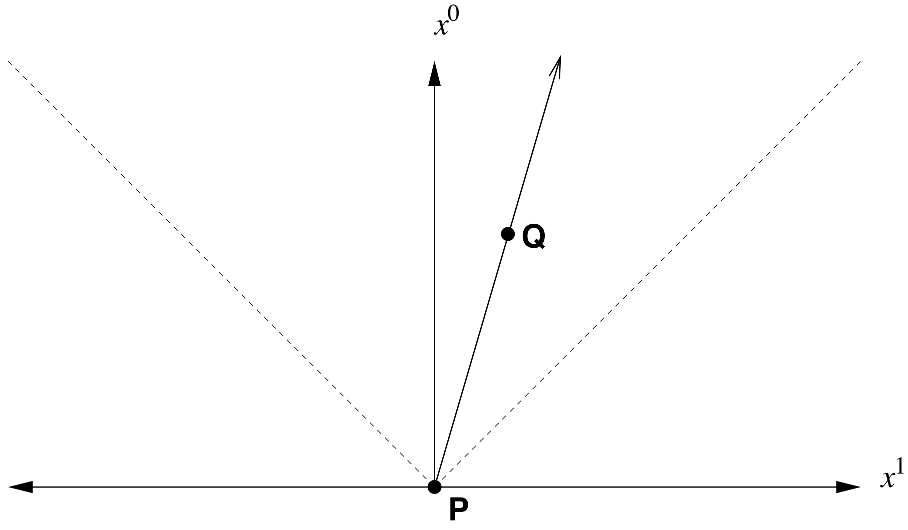

Figure 1.2: A space-time diagram showing the world-line for a body passing throught the space-time points **P** at the origin and **Q**. The dashed lines are light-like trajectories.

where $\Lambda^{\mu'}{}_{\nu}$ are constants in x^μ and we use the repeated index convention where an identical index in an up and down position is summed over. Such repeated indices are sometimes called *dummy indices*, while an unrepeated index is called a *free index*. The $\Lambda^{\mu'}{}_{\nu}$ are components of the Lorentz transformation matrix relating frame **S** to frame **S′**. The different frames can be related to each other through *boosts*, or by rotations in space, or combinations of the two. For a boost in the x-direction, we say that **S′** is moving with velocity $v\,\hat{x}$ with respect to (wrt) **S**. Einstein's second postulate tells us that if two observers were to measure the speed of a light ray, they would both find the universal constant c, which in our units is 1. In order for this to be true, the Lorentz transformation matrix for the boost must be

$$\Lambda = \begin{pmatrix} \gamma & -v\,\gamma & 0 & 0 \\ -v\,\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad \text{where } \gamma = \frac{1}{\sqrt{1-v^2}}. \tag{1.1.4}$$

The inverse Lorentz transformation is

$$\Delta x^\mu = (\Lambda^{-1})^\mu{}_{\nu'}\,\Delta x^{\nu'}\,, \tag{1.1.5}$$

where for a boost in the $\hat{x}$ direction is

$$\Lambda^{-1} = \begin{pmatrix} \gamma & +v\,\gamma & 0 & 0 \\ +v\,\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{1.1.6}$$

It is obvious from the placement of the primed and unprimed indices that $(\Lambda^{-1})^\mu{}_{\nu'}$ is the inverse, so we drop the exponent "-1" and simply write it as $\Lambda^\mu{}_{\nu'}$.

A boost in the y direction with velocity v would look like

$$\Lambda_{y-\mathrm{boost}} = \begin{pmatrix} \gamma & 0 & -v\,\gamma & 0 \\ 0 & 1 & 0 & 0 \\ -v\,\gamma & 0 & \gamma & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{1.1.7}$$

which is a different Lorentz transformation. A Lorentz transformations that is a rotation of angle ϕ in the $x-y$ plane leaves the x^0 and x^3 coordinates alone and only mixes the x^1 and x^2 coordinates. The matrix for this Lorentz transformation is given by

$$\Lambda_{\mathrm{rot}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi & 0 \\ 0 & \sin\phi & \cos\phi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{1.1.8}$$

There is a way of reformulating boosts so that they look like the rotations in (1.1.8). We can define the *rapidity*, ξ, as $\cosh\xi = \gamma$ and $\sinh\xi = v\,\gamma$. Notice that the identity $\cosh^2\xi - \sinh^2\xi = 1$ is automatically satisfied. Then the boost matrix Λ in (1.1.4) becomes

$$\Lambda = \begin{pmatrix} \cosh\xi & -\sinh\xi & 0 & 0 \\ -\sinh\xi & \cosh\xi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{1.1.9}$$

with an obvious similarity to the transformation in (1.1.8).

Instead of finite displacements $\overrightarrow{\Delta x}$ we will have occasion to consider infinitesimal displacements $\overrightarrow{d\,x}$. In fact, once we consider gravity we will be forced to do so. The infinitesimal displacement has the same Lorentz transformation properties,

$$d\,x^{\mu'} = \Lambda^{\mu'}{}_{\nu}\, d\,x^{\nu}\,. \tag{1.1.10}$$

By the rules of differential calculus with several variables, we know that when we change the variables the relation between the differentials is given by

$$d\,x^{\mu'} = \frac{\partial\,x^{\mu'}}{\partial\,x^{\nu}}\, d\,x^{\nu}\,, \tag{1.1.11}$$

hence we identify $\Lambda^{\mu'}{}_{\nu} = \frac{\partial\,x^{\mu'}}{\partial\,x^{\nu}}$. Notice that (1.1.11) applies for any two reference frames; neither $\mathbf{S}$ nor $\mathbf{S'}$ has to be an inertial frame in order for this to be true. However, (1.1.10) is true only if $\mathbf{S}$ and $\mathbf{S'}$ are both inertial frames.

For a general 4-vector $\vec{A}$, the Lorentz transformation properties for the components are exactly the same

$$A^{\mu'} = \Lambda^{\mu'}{}_{\nu}\, A^{\nu}\,. \tag{1.1.12}$$

A 4-vector with these transformation properties is called *contravariant*.

1.2 Lorentz invariants and the metric in Minkowski space

We are often interested in quantities that are the same as measured by any observer, no matter what the reference frame. If we are talking about inertial frames then we call such quantities *Lorentz invariants*, but if we are considering noninertial frames, then we simply call these *invariants*. One of the most important invariants is the *length squared* Δs^2. For an inertial frame **S** this is given as

$$\Delta s^2 \equiv -(\Delta x^0)^2 + (\Delta x^1)^2 + (\Delta x^2)^2 + (\Delta x^3)^2 \equiv \overrightarrow{\Delta x} \cdot \overrightarrow{\Delta x}\,, \tag{1.2.1}$$

or in differential form,

$$d\,s^2 = -(d\,x^0)^2 + (d\,x^1)^2 + (d\,x^2)^2 + (d\,x^3)^2 \equiv \overrightarrow{d\,x} \cdot \overrightarrow{d\,x}\,. \tag{1.2.2}$$

As you can see (1.2.1) looks like the length squared for a Euclidean space, except for the minus sign in front of $(\Delta x^0)^2$. This is clearly invariant under rotations in the spatial directions. It is also invariant under boosts, as is most easy to check using the form of the transformations in (1.1.9).

We can also define $\Delta s^2 \equiv -\Delta\tau^2$, where τ is called the *proper time*. The proper time is the time elapsed in the rest frame of the body whose time we are considering. In the rest frame, the spatial components of the displacement are $\Delta x^i = 0$, so clearly $\Delta x^0 = \Delta\tau$.

We can write (1.2.1) and (1.2.2) as

$$\begin{aligned} \Delta\,s^2 &= \eta_{\mu\nu}\,\Delta x^\mu\,\Delta x^\nu \\ d\,s^2 &= \eta_{\mu\nu}\,d\,x^\mu\,d\,x^\nu\,, \end{aligned}$$

where we again use the repeated index notation and where

$$\eta_{\mu\nu} = \mathrm{diag}(-1, +1, +1, +1)\,. \tag{1.2.3}$$

$\eta_{\mu\nu}$ is called the *metric* for *flat Minkowski space*. We will explain what we mean by flat in a later lecture. A more general metric for a noninertial frame, that is one that is not flat Minkowski space, will be written as $g_{\mu\nu}$ and the length squared as

$$ds^2 = g_{\mu\nu}\,dx^\mu dx^\nu\,. \tag{1.2.4}$$

It is not uncommon to refer to ds^2 as the metric since it contains the metric components. Usually when we are speaking of noninertial frames we will need to consider the infinitesimal displacements. Notice that since $dx^\mu dx^\nu = dx^\nu dx^\mu$, only the symmetric part of $g_{\mu\nu}$ will contribute, so without any loss of generality we can let $g_{\mu\nu} = g_{\nu\mu}$.

As an example, the invariant length squared for flat Euclidean space in three dimensions can be written as

$$ds^2 = (d\,x^1)^2 + (d\,x^2)^2 + (d\,x^3)^2 = g_{ij}\,d\,x^i\,dx^j\,, \qquad g = \mathrm{diag}(+1, +1, +1)\,. \tag{1.2.5}$$

(It is convential to use latin letters i, j, k *etc.* for spatial indices). However, if we use cylindrical coordinates (r, ϕ, z) where $x^1 = r\,\cos\phi$, $x^2 = r\,\sin\phi$, $x^3 = z$, then

$$dx^1 = \cos\phi\, dr - r\sin\phi\, d\phi\,, \qquad dx^2 = \sin\phi\, dr + r\cos\phi\, d\phi \qquad dx^3 = dz. \tag{1.2.6}$$

Hence the invariant length squared is

$$ds^2 = dr^2 + r^2 d\phi^2 + dz^2\,, \tag{1.2.7}$$

and the various components of the metric are

$$g_{rr} = 1 \qquad g_{\phi\phi} = r^2 \qquad g_{zz} = 1\,. \tag{1.2.8}$$

It is possible to have off-diagonal terms, such as $g_{r\phi}$, but in this case they are zero. We can also see for the cylindrical coordinates that it was necessary to restrict to infinitesimal displacements, since the metric is not constant but depends on the coordinate r.

1.3 Relativistic physics

Throughout this section we assume that $\mathbf{S}'$ is moving with velocity $v\hat{x}$ *wrt* $\mathbf{S}$.

1.3.1 More on space-time diagrams

Let us return to the space-time diagram in figure 1.2. The diagram shows two trajectories (dashed lines) emanating from $\mathbf{P}$ at 45 degree angles. On these trajectories, we have that $x^1 = \pm x^0$. Hence, if these are trajectories of particles we see that they have velocity $u = x^1/x^0 = \pm 1$. Therefore, these particles are traveling at the speed of light and we call the trajectories *light-like* trajectories. They are also called *null* trajectories since for a displacement along this trajectory, $(\Delta s)^2 = (\Delta x^1)^2 - (\Delta x^0)^2 = 0$.

The x^0 axis is itself the trajectory for a body at rest in frame $\mathbf{S}$. Hence this is called a *time-like* trajectory. A time-like trajectory has $(\Delta s)^2 < 0$. The x^1 axis is called a *space-like* trajectory. These trajectories have $(\Delta s)^2 > 0$. Trajectories for bodies are also known as *world-lines*.

We can also include the axes for a different frame on our space-time diagram. Notice that for frame $\mathbf{S}$ the x^1 axis is defined by the line $x^0 = 0$, while the x^0 axis is defined by the line $x^1 = 0$. Hence, we can draw axes for a different frame $\mathbf{S}'$ by finding the lines $x^{0'} = 0$ and $x^{1'} = 0$. Let us assume that the origin is the same for both frames. Then the line $x^{0'} = 0$ is $\gamma x^0 - v\,\gamma x^1 = 0$. Hence, this is the line $x^0 = v\,x^1$. Assuming that $v < 1$, then the slope of this line is less than 1. Likewise, the line $x^{1'} = 0$ is $\gamma x^1 - v\,\gamma x^0 = 0$ which leads to the line $x^0 = \frac{1}{v}\,x^1$ which has a slope greater than one. The space-time diagram showing both sets of axes is shown in figure 1.3. Notice that as v gets closer to c, the slopes of the $x^{0'}$ axis and the $x^{1'}$ axis get closer to one. In other words they approach the null trajectory from opposite sides. If $\mathbf{S}'$ is the rest frame of a body, then the $x^{0'}$ axis is the body's trajectory, assuming that the trajectory goes through the origin at $\mathbf{P}$.

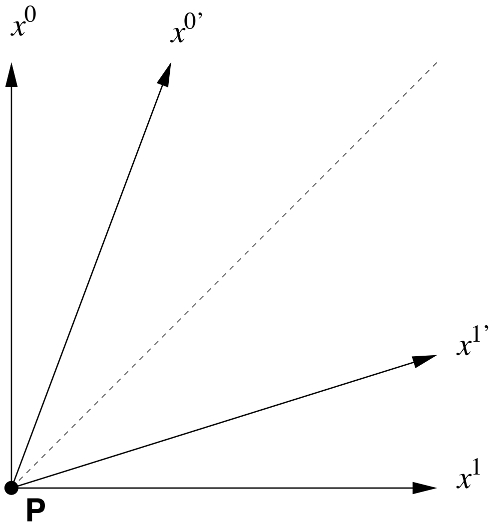

Figure 1.3: A space-time diagram showing the axes for reference frames **S** and **S′**.

1.3.2 The relativity of simultaneity and causality

Figure 1.4 shows the space-time diagram with events **R** and **Q** included. Notice that event **R** occurs at the same time as event **P** according to an observer in **S** since both events sit on the x^1 axis which has $x^0 = 0$. But an observer in **S′** would see something different. According to this observer, the event **R** happened *before* **P** since **R** is below the $x^{1'}$ axis and hence occurred for some time when $t' < 0$. On the other hand, event **Q** is simultaneous with **P** according to the **S′** observer, but occurs after **P** according to the **S** observer. Thus we see that the notion of simultaneous events is a relative concept.

We can now ask if event **P** can *cause* event **Q**. We can say that this causality would occur if a signal can emanate from **P** and travel to **Q**, thus causing it. Let us suppose that the signal is some body whose world-line is the $x^{1'}$ axis. One consequence of this is that the body's speed is greater than the speed of light. One can already see that trouble occurs if we try to boost to a frame moving faster than the speed of light, because γ will be imaginary. But there is a more fundamental problem. According to the third reference frame in figure 1.4, **S″**, event **Q** happened *before* event **P**. Hence, there is no way that **P** could cause **Q**. Hence, we must conclude that there is no way to send a signal along $x^{1'}$. In fact, we cannot send a signal along *any* space-like trajectory, only along time-like or null trajectories. From this we also must conclude that no physical particle can have a world-line along a space-like trajectory. Looking again at figure 1.4, we see that event **W** is connected to **P** by a time-like trajectory. Hence, **P** can cause **W**. When this occurs for two events we say that they are *causally connected*. It is also true that **W** and **P** are not simultaneous in *any* inertial frame.

Figure 1.5 is known as a light-cone diagram. Any event in the shaded region below **P**, including the boundary can cause **P**, because there is a time-like or null world-line that connects this event to **P**. This region is called the *past light-cone*. Similarly, any event in shaded region *above* **P** can be caused by **P** because there is a time-like or null world-line connecting **P** to the event.

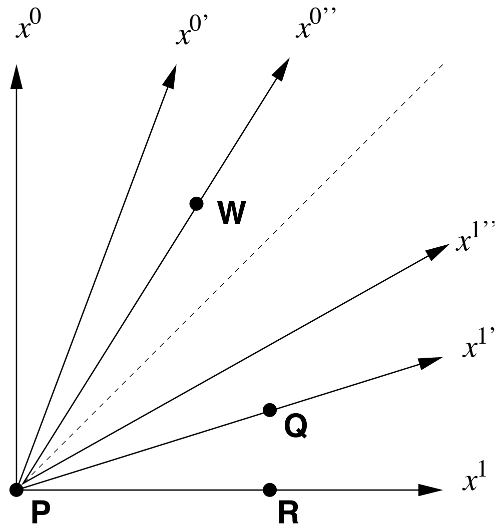

Figure 1.4: Event **R** is simultaneous with **P** according to an observer in **S** but not according to an observer in **S**$'$. Similarly, event **Q** is simultaneous with event **P** according to an observer in **S**$'$, but not according to an observer in **S**. Event **W** is causally connected to **P**.

1.3.3 Length contraction

Suppose we have a bar of length L that is stationary in **S**$'$ and is aligned along the x axis. What length would an observer in **S** measure?

In problems of this type, the key to solving this question is properly setting up the equations. In this case one should ask "how would an observer in **S** measure the length of the bar?" A reasonable thing to do is to measure the positions of the front and back of the bar simultaneously, that is find their positions at the same time t, and then measure the displacement, $\Delta x \equiv \Delta x^1$. Since the observer is making a simultaneous measurement according to his clocks, we have that $\Delta t \equiv \Delta x^0 = 0$. We also know that the displacement in **S**$'$ is $\Delta x' = L$ since the bar is stationary in this frame.

We now use the relations in (1.1.5) and (1.1.6) to write

$$\begin{aligned} \Delta t = 0 &= \gamma\,(\Delta t' + v\,\Delta x') \qquad \Rightarrow \qquad \Delta t' = -v\,\Delta x' \\ \Delta x &= \gamma\,(\Delta x' + v\Delta t') = \gamma\left(1 - v^2\right)\Delta x' = \frac{L}{\gamma}\,. \end{aligned} \tag{1.3.1}$$

Hence the observer in **S** measures a contracted length since $\gamma > 1$.

This is particularly clear if we look at the corresponding space-time diagram shown in figure 1.6. In this diagram we have shown the trajectories for the front and back of the bar. We are still assuming that **S**$'$ is the rest-frame. The length of the bar in the rest-frame is the length between the intersection points of the trajectories with the $x^{1'}$ axis, while the length in the **S** frame is the length between the intersection points on the x^1 axis. Clearly the distances are not the same and in fact the distance is longer on the $x^{1'}$ axis.

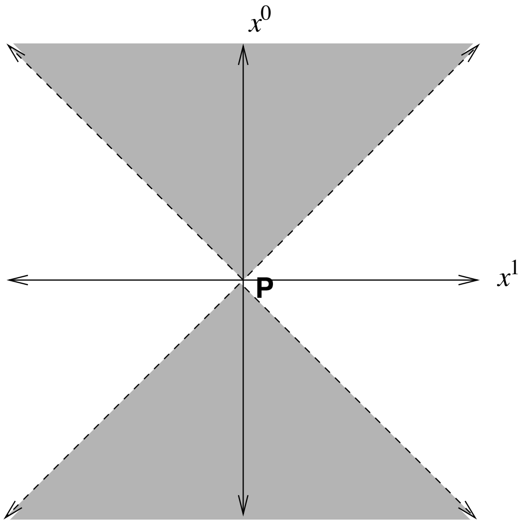

Figure 1.5: Light-cone diagram for event **P**. Those events in the past light-cone can cause **P** and those events in the future light-cone can be caused by **P**.

1.3.4 Time dilation

Another interesting phenomenon is *time dilation*. Suppose we have a clock at a fixed point in $\mathbf{S}'$ and it measures a time interval $\Delta t'$. What is the time interval measured by an observer in $\mathbf{S}$? Note that $\Delta t' = \Delta\tau$ is the elapsed proper time on the clock.

Again, the key to solving this question is setting up the problem correctly. Since the clock is stationary in $\mathbf{S}'$ we have that there is no spatial displacement in this frame, thus $\Delta x' = 0$. Then we just use (1.1.3) and (1.1.4) to obtain

$$\Delta t = \gamma \Delta t' \,. \tag{1.3.2}$$

Since $\gamma > 1$, the observer in $\mathbf{S}$ measures a longer elapsed time than the proper time of the clock. Therefore, the clock in $\mathbf{S}'$ seems to be running slow according the observer's clocks in $\mathbf{S}$.

1.3.5 The twin paradox

In our discussion of time dilation, you might have noticed the possibility of a contradiction. We argued that an observer in $\mathbf{S}$ would measure the stationary clocks in $\mathbf{S}'$ to be running slow. But an observer in $\mathbf{S}'$ would also think the clocks in $\mathbf{S}$ are running slow. We can then spin this into a paradox as follows:

Suppose there are two clocks **A** and **B** and the world-lines of the clocks both pass through the space-time point **P**. We assume that **A** is at rest in $\mathbf{S}$ and **B** is at rest in $\mathbf{S}'$. **B** then travels to a space-time point **Q** at which point it accelerates into a new frame with velocity $-v$ *wrt* $\mathbf{S}$ and travels back to meet up with **A** again at the space-time point **R**. Which clock has more elapsed time? The paradox is that naive thinking would say that they both think the other clock is slower, since at all times one clock was moving with velocity $\pm v$ *wrt* the other clock. But that can't be right. But a little bit of more

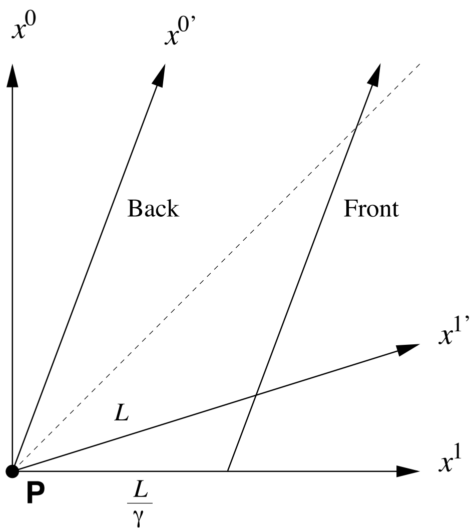

Figure 1.6: Space-time diagram for a bar stationary in $\mathbf{S}'$. The world-lines for the front and back of the bar are shown.

careful thinking shows that the problem is not entirely symmetric, since **B** undergoes an instantaneous acceleration halfway through its journey, while **A** does not. This is especially clear if we look at the space-time diagram in figure 1.7, where obviously the world-line for **A** looks different than the world-line for **B**.

We can then compare the times of the clocks by measuring the invariant lengths along the trajectories. From the diagram, it is clear that **B**'s invariant length is twice that in going from **P** to **Q**. Letting Δt_A be the elapsed time along **A**'s trajectory between **P** and **R**, then the elapsed time in going on **B**'s trajectory is

$$\Delta t_B = 2\sqrt{\left(\frac{\Delta t_A}{2}\right)^2 - v^2\left(\frac{\Delta t_A}{2}\right)^2} = \frac{\Delta t_A}{\gamma}\,. \tag{1.3.3}$$

Hence, **B**'s clock has less elapsed time and both sides would agree. Thus, there is no paradox.

Notice that while **B**'s trajectory looks longer in the diagram, the elapsed time is shorter. This is because of the relative minus sign that appears in the invariant in (1.2.1).

We can also obtain our result another way. Suppose that the distance that **B** travels away from **A** is L according to an observer in **S**. Then **A** would think the total time for **B**'s journey is $\Delta t_A = 2vL$. But **B** would see this length contracted to L/γ, so according to his clock the time for the journey is $\Delta t_B = 2vL/\gamma = \Delta t_A/\gamma$.

1.3.6 Velocity transformations

Suppose that a body is moving with 3-velocity $\vec{u}\,'$ in inertial frame $\mathbf{S}'$. What is the 3-velocity $\vec{u}$ measured by an observer in **S**? To measure a velocity, one measures the spatial displacement and divides by the elapsed time. Thus, the velocity components in

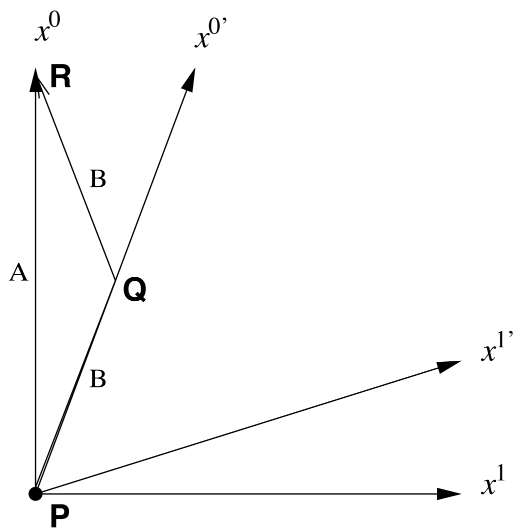

Figure 1.7: Space-time diagram for two clocks **A** and **B**. **B** has an instaneous acceleration at point **Q**.

$\mathbf{S}'$ are given by

$$u_1' = \frac{\Delta x^{1'}}{\Delta x^{0'}}\,, \qquad u_2' = \frac{\Delta x^{2'}}{\Delta x^{0'}}\,, \qquad u_3' = \frac{\Delta x^{1'}}{\Delta x^{0'}}\,, \tag{1.3.4}$$

while the components in **S** are

$$u_1 = \frac{\Delta x^1}{\Delta x^0}\,, \qquad u_2 = \frac{\Delta x^2}{\Delta x^0}\,, \qquad u_3 = \frac{\Delta x^1}{\Delta x^0}\,, \tag{1.3.5}$$

We again make use of the Lorentz transformations to write the velocity components in (1.3.5) as

$$\begin{aligned} u_1 &= \frac{\Delta x^{1'} + v\Delta x^{0'}}{\Delta x^{0'} + v\,\Delta x^{1'}} = \frac{u_1' + v}{1 + u_1' v}\,, \\ u_2 &= \frac{\Delta x^{2'}}{\gamma(\Delta x^{0'} + v\,\Delta x^{1'})} = \frac{u_2'}{\gamma(1 + u_1' v)}\,, \\ u_3 &= \frac{\Delta x^{3'}}{\gamma(\Delta x^{0'} + v\,\Delta x^{1'})} = \frac{u_3'}{\gamma(1 + u_1' v)}\,, \end{aligned} \tag{1.3.6}$$

From (1.3.6) one can show that if $|\vec{u}\,'| < c$ and $v < c$, then $|\vec{u}| < c$. To see this, consider the combination

$$1 - (\vec{u}\,')^2 = \frac{(\Delta x^{0'})^2 - (\Delta x^{1'})^2 - ((\Delta x^{2'})^2 - ((\Delta x^{3'})^2}{(\Delta x^{0'})^2} > 0\,. \tag{1.3.7}$$

The numerator is the invariant $-\Delta s^2$ and since $(\Delta x^{0'})^2 > 0$, the numerator must also be greater than zero. Therefore

$$1 - (\vec{u})^2 = \frac{(\Delta x^{0'})^2 - (\Delta x^{1'})^2 - ((\Delta x^{2'})^2 - ((\Delta x^{3'})^2}{(\Delta x^{0'})^2} > 0\,. \tag{1.3.8}$$

1.3.7 Doppler shifts

Suppose we have a light source with wavelength $\lambda = 1/\nu$ whose rest-frame is $\mathbf{S}'$ and is shining light at an observer in $\mathbf{S}$. We can think of the light source as a clock which is sending a signal at a regular time interval, $\Delta t' = 1/\nu$. In other words, this is the time for one wavelength of light to be emitted. The signal of course travels at the speed of light. This clock is time dilated in $\mathbf{S}$ to $\Delta t = \gamma \Delta t'$. But the question is more involved than just time dilation. The light being emitted is being observed by an observer at a fixed position in $\mathbf{S}$. We are *not* comparing the clock in $\mathbf{S}'$ with a series of clocks it passes in $\mathbf{S}$ as the light source moves along.

Let us say that at $t = 0$ the light source is at $x = 0$ moving with velocity v in the x direction. The observer is fixed at $x = 0$. If the light source sends a signal at $t = 0$ then the observer receives it instantaneously because the signal has zero distance to travel. The source then sends another signal at time $t = \Delta t$. But the light source is moving away from the observer and this signal is sent from position $x = v\Delta t$, and so the observer does not receive it immediately, but at a later time

$$t = \Delta T = \Delta t + v\Delta t = \gamma\, \Delta t'\,(1 + v) = \Delta t' \sqrt{\frac{1+v}{1-v}}\,. \tag{1.3.9}$$

Hence the time it takes the observer to see one wavelength is ΔT and so the light frequency as seen by the observer is

$$\nu_{\rm obs} = \frac{1}{\Delta T} = \nu \sqrt{\frac{1-v}{1+v}}\,, \tag{1.3.10}$$

and the wavelength is

$$\lambda_{\rm obs} = \frac{1}{\nu_{\rm obs}} = \lambda \sqrt{\frac{1+v}{1-v}}\,. \tag{1.3.11}$$

Since the wavelength increases, we call this a *red-shift*. To find the shift if the source is moving toward the observer, all we need to do is replace v with $-v$, hence we have

$$\lambda_{\rm obs} = \lambda \sqrt{\frac{1-v}{1+v}}\,. \tag{1.3.12}$$

Since the wavelength is smaller, we call this a *blue-shift*.

It is instructive to look at the space-time diagram for this process. Figure 1.8 shows the world-line of the source as the $x^{0'}$ axis. Included are the world-lines for light signals sent a time $\Delta t'$ apart according to the source's clock and sent back toward the observer at $x = 0$. The diagram shows the difference between Δt and ΔT.

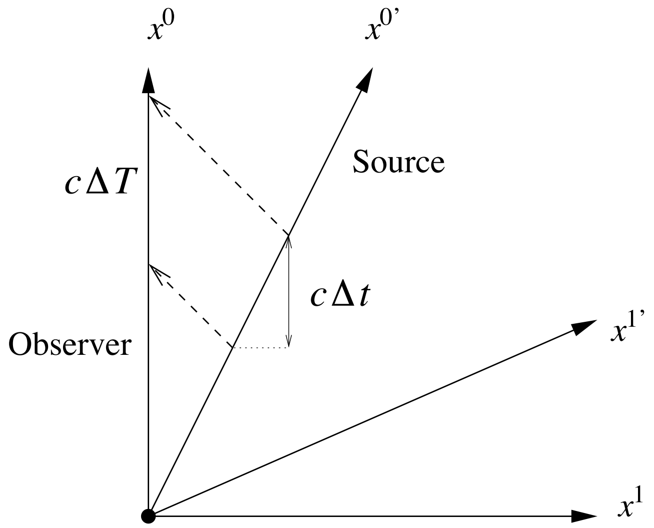

Figure 1.8: A space-time diagram for a light source emitting light toward an observer. The source's world-line is the $x^{0'}$ axis, while the observer's world-line is the x^0 axis.

Chapter 2

The equivalence principle

In this chapter we discuss Einstein's equivalence principle in the weak form that equates a constant gravitational field with a uniformly accelerating reference frame.

2.1 Noninertial frames: the accelerating frame

2.1.1 Acceleration

Let us now consider the case of acceleration and how an observer in a frame different from the accelerating object would see this. The *proper acceleration*, $\vec{\alpha}$ is defined as the 3-acceleration in the rest frame of the accelerated body. Now since this rest frame is accelerating with respect to the observer's rest frame, which we are assuming is an inertial frame, it means that the body's rest frame is *not* an inertial frame. However, at any time there is an inertial frame where the body is at rest in. We call this inertial frame its *instantaneous rest frame*. This can also be called the *momentary rest frame*. At a later time, the instantaneous inertial frame is different from the one it is in now.

To avoid too many complications, let us simplify the problem a bit and assume that the velocities and accelerations are in the $x^1 \equiv x$ direction only, so $\vec{\alpha} = \alpha\hat{x}$. We also use t for x^0. We then let $\mathbf{S}$ be the frame of the observer and $\mathbf{S}'(t)$ be the instantaneous rest frame at time t. If $\vec{u} = u\,\hat{x}$ is the velocity of the body as seen by the observer, then $v = u$ at t. In $\mathbf{S}'(t)$ we have that $u' = 0$. At a slightly later time $t + dt$, where dt is assumed to be an infinitesimal displacement, we have that u' has an infinitesimal change, du', since $\mathbf{S}'(t)$ is no longer the instantaneous rest frame, $\mathbf{S}'(t + dt)$ is. But from the definition of the proper acceleration, we have that

$$du' = \alpha\, dt' \,. \tag{2.1.1}$$

Since $\mathbf{S}'(t)$ is the instantaneous rest frame, dt is related to dt' by time dilation:

$$dt = \gamma_u\, dt' \,, \qquad \gamma_u = \frac{1}{\sqrt{1-u^2}} \,. \tag{2.1.2}$$

The change in velocity du as seen by the observer in $\mathbf{S}$ is then found using the velocity transformations in (1.3.6) of chapter 1. Given the velocity in $\mathbf{S}'(t)$ is du' and the velocity in **S** is $u + du$ then the transformation formula leads to

$$du = \frac{du' + u}{1 + u\,du'} - u \approx (du' + u)\,(1 - u\,du') - u \approx du'\left(1 - u^2\right), \tag{2.1.3}$$

where we used that $|u| < c$ and $|du'| << 1$ to justify the approximations.

The acceleration, a, as measured by the observer in **S** is then

$$a = \frac{du}{dt} = \frac{du'}{dt'}\left(1 - u^2\right)^{3/2} = \frac{\alpha}{{\gamma_u}^3}\,. \tag{2.1.4}$$

Notice that a is approaching 0 as u approaches c, which is reasonable since we should not be able to accelerate the body through the speed of light.

Let us take the further special case that the proper acceleration α is constant. We then observe that

$$\frac{d}{dt}(\gamma_u\,u) = \gamma_u\frac{du}{dt} + u\frac{d}{dt}\frac{1}{\sqrt{1-u^2}} = \left(\gamma_u + u^2{\gamma_u}^3\right)\frac{du}{dt} = {\gamma_u}^3\,\frac{du}{dt}\,. \tag{2.1.5}$$

Hence, we can write (2.1.4) as

$$\frac{d}{dt}(\gamma_u\,u) = \alpha \tag{2.1.6}$$

which has the solution

$$\gamma_u\,u = \alpha t + u_0 \tag{2.1.7}$$

where u_0 is a constant. If we take the initial condition that $u = 0$ at $t = 0$, then $u_0 = 0$. Squaring both sides of the above equation gives

$$\frac{u^2}{1-u^2} = \alpha^2 t^2\,, \tag{2.1.8}$$

which has the solution

$$u = \frac{\alpha t}{\sqrt{1+\alpha^2 t^2}}\,. \tag{2.1.9}$$

Now using $u = \frac{dx}{dt}$ we get

$$dx = \frac{\alpha t\,dt}{\sqrt{1+\alpha^2 t^2}}\,, \tag{2.1.10}$$

which can be integrated to give

$$x = \frac{1}{\alpha}\sqrt{1+\alpha^2 t^2} + x_0, \tag{2.1.11}$$

where x_0 is a constant. Since the constant is just a shift in x we drop it. Hence, this solution can be written as

$$x^2 - t^2 = \frac{1}{\alpha^2}\,, \tag{2.1.12}$$

which is the equation for a hyperbola. This solution is graphed in figure 2.1, where the horizontal axis is the x axis and the vertical axis is the t axis. The hyperbola crosses the x axis at $x = \alpha^{-1}$. Notice that as t becomes large, the hyperbola approaches the light-like trajectory which is the dashed line in the plot. Interestingly, if at $t = 0$ a light ray is emitted at $x = 0$, we see that it will never catch up with the accelerating body since the dashed line never intersects the hyperbola. By the way, notice that α has units of inverse length as is clear from (2.1.12). Hence αx and αt are dimensionless.

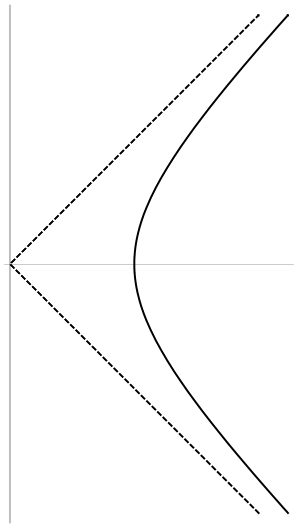

Figure 2.1: Graph of the hyperbolic trajectory in (2.1.12). The horizontal axis is the x coordinate and the vertical axis is the $x^0 = t$ coordinate. The dashed lines are light-like lines that define the limits of the hyperbola. The hyperbola intersects the x-axis at $x = 1/\alpha$.

It is of interest to find the proper time τ of the accelerating body in terms of the time t. By time dilation we can relate $d\tau$ to dt by

$$d\tau = \frac{dt}{\gamma(t)} = dt\sqrt{1-\left(\frac{dx}{dt}\right)^2} = dt\sqrt{1-\frac{(\alpha t)^2}{(\alpha t)^2+1}} = \frac{dt}{\sqrt{(\alpha t)^2+1}}\,. \tag{2.1.13}$$

We can integrate the first and the last expressions to give

$$\begin{aligned}\int_0^\tau d\tau &= \int_0^t \frac{dt}{\sqrt{(\alpha t)^2+1}}\\ \tau &= \frac{1}{\alpha}\operatorname{arcsinh}(\alpha\tau) = \frac{1}{\alpha}\operatorname{arctanh}\left(\frac{\alpha t}{\sqrt{(\alpha t)^2+1}}\right) = \frac{1}{\alpha}\operatorname{arctanh}\left(\frac{t}{x}\right)\end{aligned} \tag{2.1.14}$$

The arctanh is the inverse of the tanh function, where $\tanh(\xi)$ is given by

$$\tanh(\xi) = \frac{e^{\xi}-e^{-\xi}}{e^{\xi}+e^{-\xi}}\,. \tag{2.1.15}$$

It then follows that

$$\frac{t}{x} = \tanh(\alpha\tau)\,, \tag{2.1.16}$$

and so using (2.1.15)

$$\tau = \frac{1}{2\alpha}\log\left(\frac{x+t}{x-t}\right). \tag{2.1.17}$$

As $t \to \infty$ we see by Taylor expanding (2.1.12) that $x \approx t + \frac{1}{2\alpha^2 t}$ in which case

$$\tau \to \frac{1}{2\alpha}\log\left(4\alpha^2 t^2\right) = \frac{1}{\alpha}\log(2\alpha t)\,. \tag{2.1.18}$$

Hence as $t \to \infty$, τ also diverges, but only logarithmically.

2.1.2 The accelerating frame

Let us now assume that we have many bodies with different but fixed proper accelerations. Let us further assume that these bodies are separated from each other in the x direction as in figure 2.2, such that their intersection on the x axis is their proper acceleration. The difference between the different curves is just a different choice of α in (2.1.12). Let us now choose a reference frame where all of these accelerating bodies are at rest. We will call this the *accelerating frame*. We can then use $\bar{x} \equiv \alpha^{-1}$ as a spatial coordinate for this frame, where the world-lines for the accelerating bodies correspond to curves of constant $\bar{x}$.

For $\bar{t}$, the time coordinate in the accelerating frame, we will use the proper time for one particular accelerating body, say one with constant acceleration α_0, which is thus fixed at spatial coordinate $\bar{x} = 1/\alpha_0$. Hence we see from (2.1.14) that the lines of constant $\bar{t}$ are those that keep the ratio t/x fixed, in other words these are straight lines emanating from the origin. These are shown by the dashed red lines in the figure, with the thicker lines showing the limiting cases. The values of $\bar{t}$ range from $\bar{t} = -\infty$ where $t/x = -1$ to $\bar{t} = \infty$ where $t/x = +1$. To summarize, the coordinates in the accelerating frame with respect to the coordinates in the original inertial frame are given by

$$\begin{aligned}
\bar{t} &= \frac{1}{\alpha_0}\,\text{arctanh}\left(\frac{t}{x}\right) \\
\bar{x} &= \text{sgn}(x)\sqrt{x^2 - t^2} \\
\bar{y} &= y \qquad \bar{z} = z\,.
\end{aligned} \tag{2.1.19}$$

These coordinates are called *Rindler coordinates* and the accelerating frame is also called *Rindler space*.

Let us now suppose that there are a series of light sources fixed at different positions in $\bar{x}$ and with frequency ν. We further assume that there is an observer at $\bar{x} = 1/\alpha_0$, and so has proper time $\bar{t}$. This observer can measure the clock rates at the other spatial points by measuring the frequency of the observed light emanating from those points. We want to show that this corresponds to the proper time at the position of the light source. Without any loss of generality we focus on one source located at $\bar{x} = \bar{x}_1$ which emits a light ray at $\bar{t} = 0$. This corresponds to $t = 0$, $x = \bar{x}_1$ and so the momentary inertial frame when the light is emitted is $\mathbf{S}$. If $\bar{x}_1 < 1/\alpha_0$ then the world-line of the light

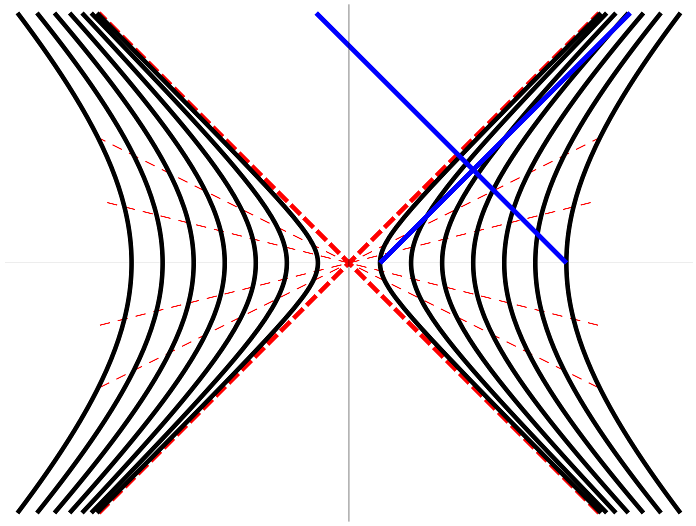

Figure 2.2: The world-lines for several accelerating bodies. Each curve corresponds to a constant value of the spatial coordinate $\bar{x}$ in the accelerating frame. The curves on the left hand side correspond to bodies with negative constant accelerations. Along the x axis $\bar{x} = x$. The dashed red lines are the lines of constant $\bar{t}$. The blue lines are forward and backward world-lines for light rays emitted at $\bar{t} = 0$

seen by the observer at $\bar{x} = 1/\alpha_0$ is parameterized by $x = t + \bar{x}_1$, while if $\bar{x} > 1/\alpha_0$ then the world-line of the light is $x = -t + \bar{x}_1$ (see figure 2.2). For $\bar{x} < 1/\alpha_0$ the light ray intersects the observer's world-line when

$$(t + \bar{x}_1)^2 = t^2 + \alpha_0^{-2} \tag{2.1.20}$$

which we can solve for but don't actually need. At this time t the velocity of the accelerating body *wrt* **S** is

$$v = \frac{dx}{dt} = \frac{t}{x} = \frac{\alpha_0 t}{\sqrt{(\alpha_0 t)^2 + 1}}\,, \tag{2.1.21}$$

Therefore the light is redshifted by the factor (see the Doppler shift section of the first chapter)

$$\sqrt{\frac{1-v}{1+v}} = \sqrt{(\alpha_0 t)^2 + 1} - \alpha_0 t = \alpha_0 x - \alpha_0 t = \alpha_0 \bar{x}_1\,, \tag{2.1.22}$$

where we inserted the expression for the light's world-line in the last step. Hence, the observer thinks the clock at $\bar{x}_1$ is slow by a factor of $\alpha_0 \bar{x}_1$. We can repeat the steps for $\bar{x} > 0$, or we can just continue the result in (2.1.22) to $\bar{x}_1 > 1/\alpha_0$. In any case, if we had directly computed the proper time at $\bar{x}_1$ we would have found

$$\tau_{\bar{x}_1} = \bar{x}_1 \text{arctanh}(t/x) = \alpha_0 \bar{x}_1 \bar{t}\,, \tag{2.1.23}$$

which is consistent with the redshift result.

We now want to use this information to construct ds^2 for these coordinates. From (2.1.22) we know that the proper time for a body at fixed $\bar{x}$ will come with a factor of $\alpha_0\bar{x}$ as compared to $\bar{t}$. Hence, we expect ds^2 to have the form

$$ds^2 = -\alpha_0^2\,\bar{x}^2 d\bar{t}^2 + f^2(\alpha_0\bar{x})d\bar{x}^2 + d\bar{y}^2 + d\bar{z}^2\,, \tag{2.1.24}$$

where $f(\alpha_0\bar{x})$ is a yet to be determined function. A straightforward calculation (which is left as an exercise) using the expressions for the Rindler coordinates and comparing to the metric in terms of t and x,

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2\,, \tag{2.1.25}$$

shows that $f(\alpha_0\bar{x}) = 1$.

For later purposes it will be convenient to do one further change of variables. To this end, we define a new coordinate $\tilde{x}$ such that

$$1 + 2\alpha_0\tilde{x} = (\alpha_0\bar{x})^2\,. \tag{2.1.26}$$

It then follows that the differentials are related by

$$2\alpha_0 d\tilde{x} = 2\alpha_0^2\,\bar{x}\,d\bar{x}\,. \tag{2.1.27}$$

Letting $\tilde{t} = \bar{t}$, $\tilde{y} = \bar{y}$, $\tilde{z} = \bar{z}$ and substituting into ds^2 we then find

$$ds^2 = -(1 + 2\alpha_0\tilde{x})d\tilde{t}^2 + \frac{d\tilde{x}^2}{1 + 2\alpha_0\tilde{x}} + d\tilde{y}^2 + d\tilde{z}^2\,. \tag{2.1.28}$$

Actually this transformation is only good if $\tilde{x} > -1/(2\alpha_0)$. If $\tilde{x} < -1/(2\alpha_0)$ then $t > x$ and so the argument of the arctanh in (2.1.19) is greater than 1, in which case the arctanh is complex. Hence for $t > x$ we choose $\tilde{t} = \frac{1}{\alpha_0}\mathrm{arctanh}(x/t)$ which differs from $\bar{t}$ in (2.1.19) by the imaginary constant $\frac{1}{2\alpha_0}\log(-1)$, as you can see by looking at (2.1.17).

Let us now make some observations of the metrics in (2.1.28) and (2.1.24):

1. The metric in (2.1.28) looks singular when $\tilde{x} = -1/(2\alpha_0)$. This however is a fake. We know that nothing untoward happens for this space-time since by a change of coordinates we can go back to the (t, x) basis where the metric is that of flat Minkowski space which has no singularities. The singularity that we are seeing here is is known as a *coordinate singularity*. Later on we will see more serious singularities, as well as find out how we can distinguish these real singularities from the coordinate singularities.

2. Even more peculiar is that if we go to region where $\tilde{x} < -1/(2\alpha_0)$ then the $\tilde{t}$ coordinate becomes space-like and the $\tilde{x}$ coordinate becomes time-like, since the overall sign changes in the metric.

3. As a light source is moved closer to $\tilde{x} = -1/(2\alpha_0)$ it becomes more and more redshifted until it reaches the *horizon* where it's frequency drops to zero. What happens as the source goes beyond $\tilde{x} = -1/(2\alpha_0)$? Then the observer at $\tilde{x} = 0$ will never see it; the source has gone beyond the horizon where the light is trapped inside.

4. The metric in (2.1.24) looks like the metric for cylindrical coordinates discussed in the last lecture. In that case, $r = \sqrt{x^2 + y^2}$ and $\phi = \arctan(y/x)$. If you compare to the Rindler coordinates you can see why the metrics look similar. In fact, if you replace t with iy, $\bar{x}$ with r and $\bar{t}$ with $i\phi$ then they are the same. This transformation is called a *rotation to Euclidean coordinates*.

2.2 The weak equivalence principle

Suppose we have an observer at rest at $\tilde{x} = 0$ in the accelerated frame using the metric in (2.1.28). Let there also be a body that is moving with velociy $v\hat{y}$ in the frame **S** and that intersects the observer's world line at $t = \tilde{t} = 0$. What does the motion of the body look like to the observer? We can convert the coordinates to determine the motion in the accelerated frame. In particular, the world-line in **S** is parameterized by $x = 1/\alpha_0$, $y = vt$. Therefore, in the accelerated coordinates we can parameterize $\tilde{x}$ in terms of $\tilde{y}$ as

$$\tilde{x} = -\frac{\alpha_0}{2v^2}\,\tilde{y}^2\,. \tag{2.2.1}$$

In other words, the trajectory as seen by the observer in the accelerated coordinates is a downward falling parabolic arc, where downward means the negative $\tilde{x}$ direction. This is exactly what an observer would see if they were near the surface of the earth and were watching a free-falling body. The observer thinks the body has a downward acceleration of α_0, so if $\alpha_0 = g$ where g is the acceleration of gravity at the earth's surface, then to the observer this would look exactly the same as if he or she were at rest near the earth's surface.

Furthermore, it did not matter what the mass of the body is, or what its composition is. Given its velocity in the inertial frame, it would have the same trajectory as observed by the fixed observer in the accelerated frame. We know that gravity has this same property, namely that the trajectory of any falling body is the same (ignoring of course wind resistance).

We can now turn this argument around. Suppose that we had an observer sitting at the surface of the earth. Because of the gravity, this is *not* an inertial frame. However, a body that is free-falling toward the earth *is* in an inertial frame. We can now make the statement that a body free-falling in a uniform gravitational field is in an inertial frame. Furthermore the observer who is at a fixed position in the gravitational field is in an accelerated frame. This statement is known as the *principle of equivalence*. Actually, this is the weak version of the equivalence. The strong version states that all laws of physics are the same in the free-falling frame as in special relativity.

We should explain what we mean by a uniform gravitational field. It means that the gravitational acceleration in $\tilde{x}$ with respect to the *inertial* time t is constant. It is not constant with respect to the proper time of a fixed observer in the accelerating frame. Note that in terms of the time $\tilde{t}$, the position $\tilde{x}$ is given by

$$\tilde{x} = -\frac{1}{2}\alpha_0\,t^2 = -\frac{1}{2\alpha_0}\tanh^2(\alpha_0\,\tilde{t})\,, \tag{2.2.2}$$

and so the acceleration with respect to t is $-\alpha_0$ while the acceleration with respect to $\tilde{t}$ is

$$\frac{d^2\tilde{x}}{d\tilde{t}^2} = -\alpha_0 \left(3 - 2\mathrm{sech}^2(\alpha_0\,\tilde{t})\right)\mathrm{sech}^2(\alpha_0\,\tilde{t})\,. \tag{2.2.3}$$

For $\alpha_0\,\tilde{t} << 1$ we can approximate (2.2.3) as

$$\frac{d^2\tilde{x}}{d\tilde{t}^2} \approx -\alpha_0\,, \tag{2.2.4}$$

which is the constant free-falling form, but for $\alpha_0\,\tilde{t} >> 1$ the acceleration falls off exponentially as $\tilde{x}$ approaches the horizon at $\tilde{x} = -\frac{1}{2\alpha_0}$.

Actually, since the acceleration of gravity toward the earth depends on the distance to the earth's center, a free falling body is not exactly in an inertial frame. In fact, whenever a gravitating body is present an inertial frame does not exist. Instead, we can say that the free-falling body is in a *local inertial frame*. This basically means that if we only look at very short distances and times, it looks the same as an inertial frame, but as the distances get larger it is no longer inertial. A frame that is an inertial frame everywhere is sometimes called a *global inertial frame*.

The equivalence principle also tells us what will happen to a light ray that is directed upward at the earth's surface. Since it is equivalent to being in an accelerated frame (at least locally), then we know it will be red-shifted. This was confirmed by Pound and Rebka in 1960. One of the exercises concerns this effect.

One final comment. The metric in (2.1.28) has the form

$$ds^2 = -(1+2\Phi(\tilde{x}))d\tilde{t}^{\,2} + \frac{d\tilde{x}^{\,2}}{1+2\Phi(\tilde{x})} + d\tilde{y}^2 + d\tilde{z}^2\,. \tag{2.2.5}$$

Note that $\Phi(\tilde{x}) = \alpha_0\tilde{x}$ is the gravitational potential for a constant gravitational field with acceleration α_0. We will see later on that metrics in the presence of gravitational fields will have this general form even for non-constant gravitational fields.

Chapter 3

Tensors and basis vectors

3.1 One-forms and tensors

Let $\vec{A}$ be a general 4-vector, with the displacement vector being one example of this. We define a *one-form* as a *linear map* of all 4-vectors to the real numbers, the result of which is an invariant. The word *map* is just a fancy way of saying it takes a vector as an input and spits out a real number as an output. We will write this as $\Phi(\vec{A})$. Since the map is linear, we have that

$$\Phi(c_1\,\vec{A}_1 + c_2\vec{A}_2) = c_1\,\Phi(\vec{A}_1) + c_2\,\Phi(\vec{A}_2)\,, \tag{3.1.1}$$

where c_1 and c_2 are arbitrary real numbers and $\vec{A}_1$ and $\vec{A}_2$ are any two 4-vectors. Hence $\Phi(\vec{A})$ must have the form

$$\Phi(\vec{A}) = \Phi_\mu A^\mu \tag{3.1.2}$$

where Φ_μ are the components of the one-form. The index is down so that we know to sum over it. In order for this to be invariant the components must have special transformation properties. Given the transformation properties for the components of $\vec{A}$ in (1.1.12) of chapter 1, the components of Φ transform under the inverse Lorentz transformation

$$\Phi_{\mu'} = \Lambda^{\nu}{}_{\mu'}\Phi_\nu\,. \tag{3.1.3}$$

Since Φ has 4 components it is often called a *covariant 4-vector*. In fact, the down index on the component tells us that it is covariant.

A nice way to illustrate a one-form is shown in figure 3.1 for vectors in two spatial dimensions. The dashed lines represent the one-form Φ. The map gives the number of times the vector crosses the lines, so from the figure we see that $\Phi(\vec{A}) = 2$ and $\Phi(\vec{B}) = 6$. Notice that these numbers are the same whether we use frame $\mathbf{S}$ or $\mathbf{S}'$, even though the vectors have different components in the two frames. Since $A^{1'} = \cos\theta A^1 + \sin\theta A^2$ and $A^{2'} = -\sin\theta A^1 + \cos\theta A^2$, we have that $\Phi_{1'} = \cos\theta\Phi_1 - \sin\theta\Phi_2$, $\Phi_{1'} = \sin\theta\Phi_1 + \cos\theta\Phi_2$. Notice decreasing the spacing between the dashed lines increases the value of the map.

We can generalize the one-forms to other maps called $\binom{0}{n}$ *tensors*. These are linear maps of n vectors to the real numbers, which we write as $T(\vec{A}_1, \vec{A}_2, \ldots \vec{A}_n)$. The map is

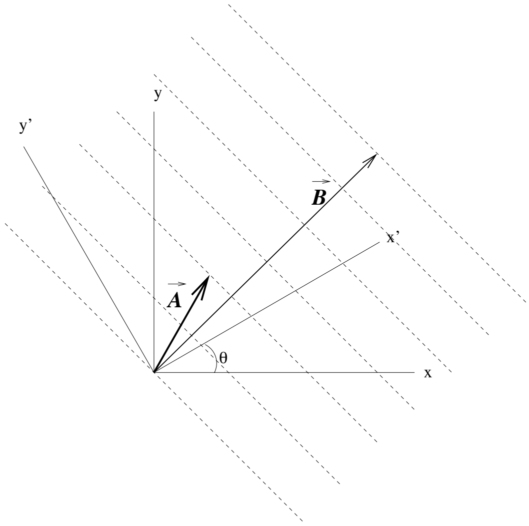

Figure 3.1: The one-form Φ is represented by the dashed lines. The map gives the number of times a vector crosses a dashed line.

assumed to be linear for every vector input, in other words

$$\begin{aligned}
T(a\,\vec{A}_1 + b\,\vec{B}_1, \vec{A}_2, \ldots \vec{A}_n) &= a\,T(\vec{A}_1, \vec{A}_2, \ldots \vec{A}_n) + b\,T(\vec{B}_1, \vec{A}_2, \ldots \vec{A}_n) \\
T(\vec{A}_1, a\,\vec{A}_2 + b\,\vec{B}_2, \ldots \vec{A}_n) &= a\,T(\vec{A}_1, \vec{A}_2, \ldots \vec{A}_n) + b\,T(\vec{A}_1, \vec{B}_2, \ldots \vec{A}_n) \\
&\vdots \\
T(\vec{A}_1, \vec{A}_2, \ldots, \vec{a}\,\vec{A}_n + b\,\vec{B}_n) &= a\,T(\vec{A}_1, \vec{A}_2, \ldots \vec{A}_n) + b\,T(\vec{A}_1, \vec{A}_2, \ldots \vec{B}_n)\,. \qquad (3.1.4)
\end{aligned}$$

In order for this to be true we must have

$$T(\vec{A}_1, \vec{A}_2, \ldots \vec{A}_n) = T_{\mu_1\mu_2\ldots\mu_n} A^{\mu_1} A^{\mu_2} \ldots A^{\mu_n}\,, \qquad (3.1.5)$$

where the $T_{\mu_1\mu_2\ldots\mu_n}$ are the components of the map. In order for this to be Lorentz invariant we also require that each index transform accordingly under the inverse Lorentz transformation:

$$T_{\mu'_1\mu'_2\ldots\mu'_n} = \Lambda^{\nu_1}{}_{\mu'_1} \Lambda^{\nu_2}{}_{\mu'_2} \ldots \Lambda^{\nu_n}{}_{\mu'_n} T_{\nu_1\nu_2\ldots\nu_n}\,. \qquad (3.1.6)$$

Notice that we have already seen a $\binom{0}{2}$ tensor, namely the η-tensor with components $\eta_{\mu\nu}$. In other words, we have the linear map

$$\eta(\vec{A}_1, \vec{A}_2) = \eta_{\mu\nu} A_1^\mu A_2^\nu = \vec{A}_1 \cdot \vec{A}_2 \qquad (3.1.7)$$

that results in a Lorentz invariant. As an exercise, show that under a Lorentz transformation

$$\eta_{\mu'_1\mu'_2} = \Lambda^{\nu_1}{}_{\mu'_1} \Lambda^{\nu_2}{}_{\mu'_2} \eta_{\nu_1\nu_2} \qquad (3.1.8)$$

has the same form, diag$(-1, +1, +1, +1)$, in $\mathbf{S}'$ as it does in $\mathbf{S}$.

3.2 Lowering indices, the inverse metric and raising indices.

Notice that in (3.1.7) we can group the expression as

$$\eta(\vec{A},\vec{B}) = (\eta_{\mu\nu}A^\mu)B^\nu\,, \tag{3.2.1}$$

where the part inside the parentheses has one free index ν which is down. Hence we can write this as $A_\nu \equiv \eta_{\mu\nu}A^\mu$ where we know that A_ν must transform as a covariant vector. Thus it follows that with the η-tensor we can *lower an index* on a vector. If we can lower the index we should be able to reverse the process and raise it again. This requires the *inverse metric* $\eta^{\mu\nu}$ where $\eta^{\mu\nu}\eta_{\nu\lambda} = \delta^\mu{}_\lambda$, where $\delta^\mu{}_\lambda$ is the Kronecker δ which equals 1 if $\mu = \lambda$ and is 0 otherwise. So starting with A_ν we have

$$\eta^{\mu\nu}A_\nu = \eta^{\mu\nu}(\eta_{\nu\lambda}A^\lambda) = \delta^\mu{}_\lambda\, A^\lambda = A^\mu\,. \tag{3.2.2}$$

Hence, with an inverse metric we can *raise the index*. In fact, we could have started with a one-form with components Φ_μ and raised its index, $\Phi^\nu \equiv \eta^{\mu\nu}\Phi_\mu$. For a general $\binom{0}{n}$ tensor, we can raise one or more down indices using an inverse metric for each index. Hence after raising m of the indices for a $\binom{0}{m+n}$ tensor we can consider an $\binom{m}{n}$ tensor whose transformation properties are

$$T^{\mu'_1\mu'_2\ldots\mu'_m}_{\nu'_1\nu'_2\ldots\nu'_n} = \Lambda^{\mu'_1}{}_{\mu_1}\Lambda^{\mu'_2}{}_{\mu_2}\ldots\Lambda^{\mu'_m}{}_{\mu_m}\Lambda^{\nu_1}{}_{\nu'_1}\Lambda^{\nu_2}{}_{\nu'_2}\ldots\Lambda^{\nu_n}{}_{\nu'_n}T^{\mu_1\mu_2\ldots\mu_m}_{\nu_1\nu_2\ldots\nu_n}\,, \tag{3.2.3}$$

3.3 Scalar functions and derivatives

A *scalar* function $\phi(x^\lambda)$ is a function of the coordinates x^λ that is invariant under Lorentz transformations (don't confuse the index on x^λ in the function with a free index). In other words,

$$\phi'(x^{\lambda'}) = \phi(x^\lambda) \tag{3.3.1}$$

where the different coordinates are related by Lorentz transformations. Notice that the form of the function could change under the transformation, but the new function evaluated at the new coordinates is the same as the old function evaluated at the old coordinates.

Now suppose we take a derivative on this function with respect to one of the coordinates, $\partial_\mu\phi(x^\lambda) \equiv \frac{\partial}{\partial x^\mu}\phi(x^\lambda) \equiv \phi_{,\mu}(x^\lambda)$. Then in $\mathbf{S'}$ this is

$$\partial_{\mu'}\phi'(x^{\lambda'}) \equiv \frac{\partial}{\partial x^{\mu'}}\phi'(x^{\lambda'}) = \frac{\partial x^\nu}{\partial x^{\mu'}}\partial_\nu\phi(x^\lambda) = \Lambda^\nu{}_{\mu'}\partial_\nu\phi(x^\lambda)\,, \tag{3.3.2}$$

In other words $\phi_{,\mu}$ transforms as a covariant vector.

We can generalize this to derivatives acting on vectors and tensors, but we will postpone this discussion until later.

3.4 Tensors for noninertial frames

Lorentz transformations are linear transformations between different inertial frames. As was argued in the first chapter, the components of the Lorentz transformation matrix between two inertial frames $\mathbf{S}$ and $\mathbf{S}'$ are given by

$$\Lambda^{\mu'}{}_{\nu} = \frac{\partial x^{\mu'}}{\partial x^{\nu}}\,. \tag{3.4.1}$$

So we could have replaced $\Lambda^{\mu'}{}_{\nu}$ with $\frac{\partial x^{\mu'}}{\partial x^{\nu}}$ in (3.2.3). However, once in this form there is no reason why either of the frames needs to be an inertial frame. Hence, we can have tensors in any frame, not just inertial frames which transform between any two frames as

$$T^{\mu'_1\mu'_2\dots\mu'_m}_{\nu'_1\nu'_2\dots\nu'_n} = \frac{\partial x^{\mu'_1}}{\partial x^{\mu_1}}\frac{\partial x^{\mu'_2}}{\partial x^{\mu_2}}\cdots\frac{\partial x^{\mu'_m}}{\partial x^{\mu_m}}\frac{\partial x^{\nu_1}}{\partial x^{\nu'_1}}\frac{\partial x^{\nu_2}}{\partial x^{\nu'_2}}\cdots\frac{\partial x^{\nu_n}}{\partial x^{\nu'_n}}T^{\mu_1\mu_2\dots\mu_m}_{\nu_1\nu_2\dots\nu_n}\,, \tag{3.4.2}$$

If all components of a tensor are zero in one frame, then they are zero in any frame as is obvious from (3.4.2). This is a useful property that we will exploit when deriving Einstein's equations

In the general case we can also raise indices with the inverse metric, $g^{\mu\nu}$, which satisfies

$$g^{\mu\nu}g_{\nu\lambda} = \delta^{\mu}{}_{\lambda}\,. \tag{3.4.3}$$

Since

$$ds^2 = g_{\mu\nu}\,dx^{\mu}\,dx^{\nu}\,, \tag{3.4.4}$$

is an invariant, it is clear that

$$g_{\mu'\nu'} = \frac{\partial x^{\mu}}{\partial x^{\mu'}}\frac{\partial x^{\nu}}{\partial x^{\nu'}}\,g_{\mu\nu}\,. \tag{3.4.5}$$

Hence, if A^{μ} are the components of a contravariant vector transforming as

$$A^{\mu'} = \frac{\partial x^{\mu'}}{\partial x^{\mu}}\,A^{\mu}\,, \tag{3.4.6}$$

then $A_{\nu} = g_{\mu\nu}A^{\nu}$ transforms covariantly:

$$A_{\mu'} = g_{\mu'\nu'}A^{\nu'} = \frac{\partial x^{\mu}}{\partial x^{\mu'}}\frac{\partial x^{\nu}}{\partial x^{\nu'}}\frac{\partial x^{\mu'}}{\partial x^{\mu}}\,g_{\mu\nu}\,A^{\mu} = \frac{\partial x^{\nu}}{\partial x^{\nu'}}A_{\nu}\,. \tag{3.4.7}$$

3.5 Basis vectors

A general 4-vector $\vec{A}$ can be expressed as a linear combination of *basis vectors* $\vec{e}_{\mu}$, as

$$\vec{A} = \vec{e}_{\mu}\,A^{\mu}\,. \tag{3.5.1}$$

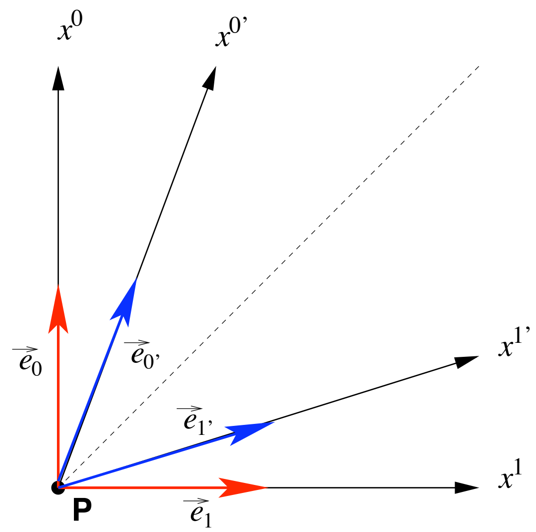

Figure 3.2: Basis vectors for the inertial frame **S** (red) and **S′** (blue).

The basis vectors are also 4-vectors, where the coordinate index μ is a label for the basis vector. Hence for each value of μ there is a different basis vector. The basis vectors satisfy the orthonormality relations

$$\vec{e}_\mu \cdot \vec{e}_\nu = \eta_{\mu\nu}\,, \tag{3.5.2}$$

for an inertial frame and

$$\vec{e}_\mu \cdot \vec{e}_\nu = g_{\mu\nu}\,, \tag{3.5.3}$$

for a more general frame. The basis vectors for different frames are different 4-vectors. For example consider two inertial frames **S** and **S′** with basis vectors $\vec{e}_\mu$ and $\vec{e}_{\mu'}$. The 4-vectors $\vec{e}_0$ and $\vec{e}_1$ point in different directions in space-time than $\vec{e}_{0'}$ and $\vec{e}_{1'}$, as you can see in figure 3.2. Hence they are different 4-vectors. It might seem that the basis vectors just add a trivial complication, but they turn out to be quite useful when we are considering spaces that are not flat Minkowski.

If we are discussing an inertial frame, then the basis vectors are constant in all space. However, for a general frame the basis vectors may not be constant. Hence if we take a derivative with respect to the coordinates on a basis vector, $\partial_\mu \vec{e}_\nu$, in general this will be nonzero. The derivative of 4-vector is also a 4-vector, and since the basis vectors form a complete set, it must be true that the derivative has the form

$$\partial_\mu \vec{e}_\nu = \Gamma^\lambda_{\mu\nu}\,\vec{e}_\lambda\,, \tag{3.5.4}$$

where the repeated λ index is summed over. The objects $\Gamma^\lambda_{\mu\nu}$ are called *Christoffel symbols.*

As a simple example, let us consider the basis vectors for the cylindrical coordinates with metric

$$ds^2 = dr^2 + r^2\,d\phi^2 + dz^2\,. \tag{3.5.5}$$

The basis vectors we label by $\vec{e}_r$, $\vec{e}_\phi$ and $\vec{e}_z$. In figure 3.3 we show $\vec{e}_r$ and $\vec{e}_\phi$ at two different values of ϕ in the r,ϕ plane. Clearly, as we move from one point to another, the basis vectors have changed. Since $\vec{e}_\phi\cdot\vec{e}_\phi = r^2$ and $\vec{e}_r\cdot\vec{e}_r = 1$, we can see from the figure that

$$\partial_\phi\vec{e}_\phi = -r\vec{e}_r\,,\qquad \partial_\phi\vec{e}_r = \frac{1}{r}\vec{e}_\phi\,. \tag{3.5.6}$$

We also have that as we change r

$$\partial_r\vec{e}_\phi = \frac{1}{r}\vec{e}_\phi\qquad \partial_r\vec{e}_r = 0\,. \tag{3.5.7}$$

Therefore, the nonzero Christoffel symbols are

$$\Gamma^r_{\phi\phi} = -r\,,\qquad \Gamma^\phi_{\phi r} = \frac{1}{r}\,,\qquad \Gamma^\phi_{r\phi} = \frac{1}{r}\,. \tag{3.5.8}$$

Notice that $\Gamma^\phi_{\phi r} = \Gamma^\phi_{r\phi}$. In fact for all spaces that that will concern us in this course the Christoffel symbols will satisfy $\Gamma^\lambda_{\mu\nu} = \Gamma^\lambda_{\nu\mu}$. Such spaces are called *torsion free*.

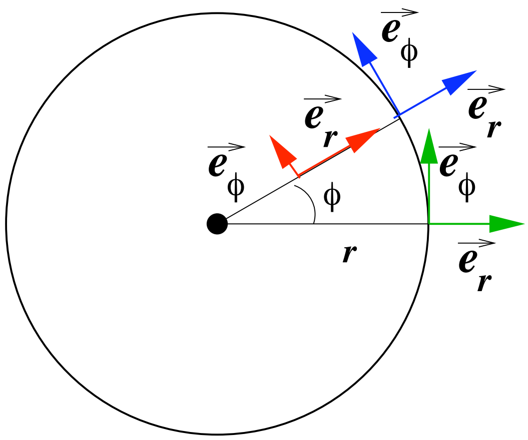

Figure 3.3: The basis vectors $\vec{e}_r$ and $\vec{e}_\phi$ at two different values of ϕ and fixed r (green, blue) and the basis vectors at two different values of r and fixed ϕ (blue, red). For decreasing r the length of $\vec{e}_\phi$ is proportionally shorter but $\vec{e}_r$ does not change.

3.6 The velocity vector

We have previously considered velocity 3-vectors in space, but we would like to also have a velocity 4-vector, $\vec{U}$ that transforms as a contravariant vector. Suppose then we have a body moving through space-time. If we look at an infinitesimal change in its space-time position then we can construct the 4-vector $\overrightarrow{dx}$. If we then divide this by a differential of the proper time, $d\tau$, where recall $d\tau^2 = -\overrightarrow{dx}\cdot\overrightarrow{dx}$, then we can define $\vec{U}$ as

$$\vec{U} = \frac{\overrightarrow{dx}}{d\tau}\,. \tag{3.6.1}$$

Clearly $\vec{U}$ satisfies $\vec{U}\cdot\vec{U} = -1$.

Let us suppose that $\mathbf{S}'(\tau)$ is the momentary rest-frame of the moving body at proper time τ. Recall that the momentary rest-frame is an inertial frame, or at least a local inertial frame. In the rest frame the spatial components of $\overrightarrow{dx}$ are zero, while $d\tau = dx^{0'}$. Hence, $U^{0'} = 1$ and the spatial components satisfy $U^{i'} = 0$. So it follows that $\vec{U} = \vec{e}_{0'}(\tau)$, the basis vector for the time component in the momentary rest frame. In figure 3.4 we show the world-line of a body moving through space-time with nonconstant velocity. In the momentary rest frame the trajectory points along the time direction of the frame. Hence $\vec{U}$ is the tangent vector of the world-line.

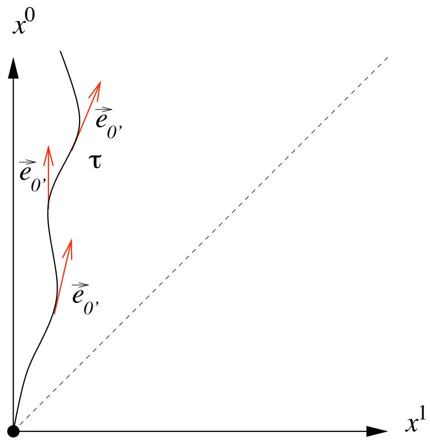

Figure 3.4: A world-line with nonconstant velocity parameterized by the proper time τ. The tangent vectors (red) are $\vec{e}_{0'}(\tau)$, the basis vector along the time direction for the momentary rest frame.

3.7 The covariant derivative

Let us return to the case of derivatives on vectors and more general tensors. As was emphasized in the first lecture, a 4-vector $\vec{A}$ is invariant under a coordinate transformation, although the components of $\vec{A}$ can change as the coordinates are changed. Since $\vec{A}$ is invariant, $\vec{e}_\mu$ must tranform as

$$\vec{e}_{\mu'} = \frac{\partial x^\mu}{\partial x^{\mu'}}\,\vec{e}_{\mu'}\,. \tag{3.7.1}$$

If we then consider the derivative $\partial_\lambda\vec{A}$, this transforms as

$$\partial_{\lambda'}\vec{A} = \frac{\partial x^\lambda}{\partial x^{\lambda'}}\partial_\lambda\vec{A}\,. \tag{3.7.2}$$

But we also have that

$$\partial_\lambda\vec{A} = \partial_\lambda(\vec{e}_\mu A^\mu) = \vec{e}_\mu\partial_\lambda A^\mu + \Gamma^\nu_{\lambda\mu}\vec{e}_\nu A^\mu\,. \tag{3.7.3}$$

Relabeling the indices we can rewrite this as

$$\partial_\lambda \vec{A} = \vec{e}_\mu(\partial_\lambda A^\mu + \Gamma^\mu_{\lambda\nu} A^\nu)\,. \tag{3.7.4}$$

Defining the *covariant derivative* as

$$\nabla_\lambda A^\mu \equiv \partial_\lambda A^\mu + \Gamma^\mu_{\lambda\nu} A^\nu \equiv A^\mu{}_{;\lambda} \tag{3.7.5}$$

we see from (3.7.1) and (3.7.2) that it transforms as the components of a $\binom{1}{1}$ tensor. For the inertial frame the covariant derivative is the same as the ordinary derivative. It is also clear that $\Gamma^\mu_{\lambda\nu}$ by itself is not a tensor, only when it is combined with a derivative term. Actually, we can see this another way as well. We know that if we have an inertial frame $\Gamma^\mu_{\lambda\nu} = 0$, but if we choose a different set of coordinates then in general at least some of the $\Gamma^\mu_{\lambda\nu}$ are nonzero. But if all components of a tensor are zero in one frame then they are zero in all frames. Clearly $\Gamma^\mu_{\lambda\nu}$ does not satisfy this property.

Since a derivative acting on a scalar ϕ is a tensor, we define its covariant derivative as $\nabla_\lambda\phi = \partial_\lambda\phi$. If we next consider a derivative acting on the scalar made from a one-form acting on a vector $\Phi(\vec{A}) = \Phi_\mu A^\mu$ and we assume that the covariant derivative satisfies the Leibniz rule, then we have

$$\begin{aligned}\nabla_\lambda(\Phi_\mu A^\mu) &= (\nabla_\lambda\Phi_\mu)\,A^\mu + \Phi_\mu\,\nabla_\lambda A^\mu \\ &= (\partial_\lambda\Phi_\mu)\,A^\mu + \Phi_\mu\,\partial_\lambda A^\mu\,.\end{aligned} \tag{3.7.6}$$

It thus follows from (3.7.5) that

$$\nabla_\lambda\Phi_\mu = \partial_\lambda\Phi_\mu - \Gamma^\nu_{\lambda\mu}\Phi_\nu \equiv \Phi_{\mu;\lambda}\,. \tag{3.7.7}$$

For the covariant derivative of a general tensor one should add a Christoffel symbol with a $+$ sign for every up index and a Christoffel symbol with a $-$ sign for every down index:

$$\nabla_\lambda\, T^{\mu_1\ldots\mu_m}_{\nu_1\ldots\nu_n} = \partial_\lambda\, T^{\mu_1\ldots\mu_m}_{\nu_1\ldots\nu_n} + \sum_{j=1}^{m}\Gamma^{\mu_j}_{\lambda\sigma}T^{\mu_1\ldots\sigma\ldots\mu_m}_{\nu_1\ldots\nu_n} - \sum_{j=1}^{n}\Gamma^{\sigma}_{\lambda\nu_j}T^{\mu_1\ldots\mu_m}_{\nu_1\ldots\sigma\ldots\nu_n} \equiv T^{\mu_1\ldots\mu_m}_{\nu_1\ldots\nu_n;\lambda}\,. \tag{3.7.8}$$

We close this lecture by finding the expression for the Christoffel symbols in terms of the metric $g_{\mu\nu}$. From their definitions we have that

$$\begin{aligned}\partial_\lambda(\vec{e}_\mu\cdot\vec{e}_\nu) &= \partial_\lambda\, g_{\mu\nu} \\ &= (\Gamma^\sigma_{\lambda\mu}\vec{e}_\sigma)\cdot\vec{e}_\nu + \vec{e}_\mu\cdot(\Gamma^\sigma_{\lambda\nu}\vec{e}_\sigma) = \Gamma^\sigma_{\lambda\mu}g_{\sigma\nu} + \Gamma^\sigma_{\lambda\nu}g_{\mu\sigma}\,.\end{aligned} \tag{3.7.9}$$

If we then rotate indices and take the following combination of terms we have

$$\begin{aligned}\partial_\lambda\, g_{\mu\nu} + \partial_\mu\, g_{\lambda\nu} - \partial_\nu\, g_{\mu\lambda} &= \Gamma^\sigma_{\lambda\mu}g_{\sigma\nu} + \Gamma^\sigma_{\lambda\nu}g_{\mu\sigma} + \Gamma^\sigma_{\mu\lambda}g_{\sigma\nu} + \Gamma^\sigma_{\mu\nu}g_{\lambda\sigma} - \Gamma^\sigma_{\nu\lambda}g_{\sigma\mu} - \Gamma^\sigma_{\nu\mu}g_{\lambda\sigma} \\ &= 2\Gamma^\sigma_{\lambda\mu}g_{\sigma\nu}\,,\end{aligned} \tag{3.7.10}$$

where we used the symmetry properties of $\Gamma^\sigma_{\lambda\mu}$ and $g_{\sigma\nu}$. Hence we find using the inverse metric that

$$\Gamma^\sigma_{\lambda\mu} = \tfrac{1}{2}g^{\sigma\nu}\left(\partial_\lambda\, g_{\mu\nu} + \partial_\mu\, g_{\lambda\nu} - \partial_\nu\, g_{\mu\lambda}\right)\,. \tag{3.7.11}$$

As an example, let's again consider the cylindrical coordinates. The components of the inverse metric are $g^{rr} = 1$, $g^{\phi\phi} = 1/r^2$ and $g^{zz} = 1$. Let us consider $\Gamma^{\phi}_{r\phi}$. According to (3.7.11) this is

$$\begin{aligned}
\Gamma^{\phi}_{r\phi} &= \tfrac{1}{2}\, g^{\mu\phi}\left(\partial_r g_{\mu\phi} + \partial_\phi g_{\mu r} - \partial_\nu g_{r\phi}\right) \\
&= \frac{1}{2}\frac{1}{r^2}\left(\partial_r r^2 + 0 + 0\right) = \frac{1}{r}\,,
\end{aligned} \tag{3.7.12}$$

which agrees with our result from before. In doing this calculation we used the fact that $g_{\mu\nu}$ is diagonal so there is only a contribution if $\mu = \phi$. The computation of the other components is left as an exercise.

Chapter 4

Equations of motion in curved space

In this chapter we compute the trajectories for freely falling bodies and light rays in gravitational fields.

4.1 Free-falling and geodesics

In an inertial frame, a body with no forces acting on it travels with constant velocity. It follows that its world-line is a straight line. We saw that when we go to the accelerating frame the trajectory is a parabola, precisely what we would expect for free-falling in a constant gravitational field. In this chapter we will study the trajectories of bodies in space-times where a global inertial frame does not exist. This happens when nonconstant gravitational fields are present.

The metric itself is found by solving Einstein's equations, which will be discussed in later lectures. For the present we just assume there is a generic metric $g_{\mu\nu}$ that affects the motion of a particle. If the particle itself has a mass, then it too should be a gravitational source and hence change the metric. However, if the mass is small enough, then the effect on the metric is practically negligible and can be ignored. In this case we say the particle is a *test particle* and that it does not *backreact* on the metric.

4.1.1 The metric is covariantly constant

Before getting into the main parts of this section, we begin with two remarks about covariant derivatives. We first observe that $\nabla_\lambda g_{\mu\nu} = 0$, or in words, "the metric is covariantly constant". To show this, consider the derivative

$$\partial_\lambda g_{\mu\nu} = \partial_\lambda(\vec{e}_\mu \cdot \vec{e}_\nu) = \Gamma^\sigma_{\lambda\mu}\vec{e}_\sigma \cdot \vec{e}_\nu + \Gamma^\sigma_{\lambda\nu}\vec{e}_\mu \cdot \vec{e}_\sigma = \Gamma^\sigma_{\lambda\mu} g_{\sigma\nu} + \Gamma^\sigma_{\lambda\nu} g_{\mu\sigma}\,. \tag{4.1.1}$$

Thus,

$$\partial_\lambda g_{\mu\nu} - \Gamma^\sigma_{\lambda\mu} g_{\sigma\nu} - \Gamma^\sigma_{\lambda\nu} g_{\mu\sigma} = \nabla_\lambda g_{\mu\nu} = 0\,. \tag{4.1.2}$$

Secondly, since the metric is covariantly constant, we can raise and lower indices through a covariant derivative. In other words

$$\begin{aligned}
\nabla_\lambda A_\mu &= \nabla_\lambda(g_{\mu\nu}A^\nu) = g_{\mu\nu}\nabla_\lambda A^\nu \\
\nabla_\lambda A^\mu &= \nabla_\lambda(g^{\mu\nu}A_\nu) = g^{\mu\nu}\nabla_\lambda A_\nu \\
\nabla_\mu A^\mu &= \nabla_\mu(g^{\mu\nu}A_\nu) = g^{\mu\nu}\nabla_\mu A_\nu = \nabla^\nu A_\nu \,,
\end{aligned} \tag{4.1.3}$$

et cetera.

4.1.2 The local inertial frame

We mentioned in a previous lecture that while a global inertial frame might not exist, there exists a local inertial frame at each space-time point. Let us show that this is true at any space-time point, which witihout any loss of generality we can set to be $x^\mu = 0$. If the frame is locally inertial then the metric $g_{\mu\nu}(x)$ satisfies $g_{\mu\nu}(0) = \eta_{\mu\nu}$ and the Christoffel symbols evaluated at $x^\mu = 0$ are $\Gamma^\mu_{\nu\lambda}(0) = 0$ ($g_{\mu\nu}(x)$ and $\Gamma^\mu_{\nu\lambda}(x)$ are assumed to depend on all 4 position components, but we abbreviate the notation and drop the index inside the parentheses).

Let us first start with a frame that is not locally inertial at $x^\mu = 0$, but instead has an arbitrary $g_{\mu\nu}(0)$ and $\Gamma^\mu_{\nu\lambda}(0)$. We then look for transformations that can take us to the local inertial frame. Let us assume that we make a small change in our coordinate system and write the new variables $x^{\mu'}$ in terms of the old variables x^μ as

$$x^{\mu'} = x^\mu + \xi^\mu(x)\,, \tag{4.1.4}$$

where ξ^μ(x) are the components of a spatially dependent 4-vector which we are assuming to be small, at least near $x^\mu = 0$. Let us investigate how this changes the metric. Consider the invariant

$$ds^2 = g_{\mu'\nu'}(x')dx^{\mu'}dx^{\nu'} = g_{\mu\nu}(x)dx^\mu dx^\nu\,. \tag{4.1.5}$$

The metric in the new coordinates is a different tensor function than the metric in the old coordinates. Assuming that $\xi^\mu(x)$ is small, we can then approximate the new metric as

$$g_{\mu'\nu'}(x') \approx g_{\mu\nu}(x) + \xi^\lambda(x)\partial_\lambda g_{\mu\nu}(x) + \delta g_{\mu\nu}(x)\,, \tag{4.1.6}$$

where the second term comes from Taylor expanding the function about the old coordinates and the third term is the change in the function. If we expand the invariant in (4.1.5) to first order we then find

$$\begin{aligned}
g_{\mu'\nu'}(x')dx^{\mu'}dx^{\nu'} \approx\; & g_{\mu\nu}(x)dx^\mu dx^\nu + g_{\mu\nu}(x)\partial_\lambda\xi^\mu(x)dx^\lambda dx^\nu + g_{\mu\nu}(x)\partial_\lambda\xi^\nu(x)dx^\mu dx^\lambda \\
& + \xi^\lambda(x)\partial_\lambda g_{\mu\nu}(x)dx^\mu dx^\nu + \delta g_{\mu\nu}(x)dx^\mu dx^\nu\,.
\end{aligned} \tag{4.1.7}$$

Hence it follows after changing some indices that

$$\begin{aligned}
\delta g_{\mu\nu} &= -g_{\lambda\nu}\partial_\mu\xi^\lambda - g_{\mu\lambda}\partial_\nu\xi^\lambda - \xi^\lambda\partial_\lambda g_{\mu\nu}\\
&= -\partial_\mu\xi_\nu - \partial_\nu\xi_\mu + \xi^\lambda\Big(\partial_\mu g_{\lambda\nu} + \partial_\nu g_{\mu\lambda} - \partial_\lambda g_{\mu\nu}\Big)\\
&= -\partial_\mu\xi_\nu - \partial_\nu\xi_\mu + \xi^\lambda\Big(\Gamma^\sigma_{\mu\lambda}g_{\sigma\nu} + \Gamma^\sigma_{\mu\nu}g_{\lambda\sigma} + \Gamma^\sigma_{\nu\mu}g_{\sigma\lambda} + \Gamma^\sigma_{\nu\lambda}g_{\mu\sigma} - \Gamma^\sigma_{\lambda\mu}g_{\sigma\nu} - \Gamma^\sigma_{\lambda\nu}g_{\mu\sigma}\Big)\\
&= -\partial_\mu\xi_\nu - \partial_\nu\xi_\mu + 2\xi_\sigma\Gamma^\sigma_{\nu\mu}\\
&= -\nabla_\mu\xi_\nu - \nabla_\nu\xi_\nu\,. \qquad (4.1.8)
\end{aligned}$$

In going from the first line to the second we brought $g_{\lambda\nu}$ and $g_{\mu\lambda}$ inside the derivatives so that we could lower the index on ξ^λ. The third line follows from the second because $g_{\mu\nu}$ is covariantly constant. Finally the last line follows from the fourth because of the form of covariant derivatives acting on covariant vectors. Notice that the change in $g_{\mu\nu}$ is also a symmetric $\binom{0}{2}$ tensor. The transformations in (4.1.4) and (4.1.8) are called *diffeomorphisms*. They are just changes in the coordinate systems and do not change the physics. These are the analogs of gauge transformations in electrodynamics, where a change in the gauge field A_μ by a derivative $\partial_\mu\phi$ does not change the electric and magnetic fields.

Let us now assume that $\xi^\mu(0) = 0$ so that the origin of the first set of coordinates stays the origin for the second set. Then the expansion of $\xi_\mu(x)$ about $x^\mu = 0$ has the form

$$\xi_\mu(x) = \epsilon_{\mu\nu}\,x^\nu + \tfrac{1}{2}\,b_{\mu\nu\lambda}\,x^\nu x^\lambda + \ldots \qquad (4.1.9)$$

where $\epsilon_{\mu\nu}$ and $b_{\mu\nu\lambda}$ are constants in x. Because of the symmetry in (4.1.9) we can assume that $b_{\mu\nu\lambda}$ is symmetric in the last two indices. Hence we have that

$$\delta g_{\mu\nu}(0) = -\nabla_\mu\xi_\nu(0) - \nabla_\nu\xi_\nu(0) = -\partial_\mu\xi_\nu(0) - \partial_\nu\xi_\mu(0) = -\epsilon_{\mu\nu} - \epsilon_{\nu\mu}\,, \qquad (4.1.10)$$

where we used that $\xi^\mu(0) = 0$ so that the covariant derivatives become ordinary derivatives. Hence, by choosing $\epsilon_{\mu\nu}$ appropriately we can shift the metric closer to $\eta_{\mu\nu}$. Actually, in order for the expansion in (4.1.7) to be valid we need that $\partial_\lambda\xi^\mu$ is small, at least near $x^\mu = 0$. This means that the $\epsilon_{\mu\nu}$ are small. However, if $g_{\mu\nu}(0)$ starts far from $\eta_{\mu\nu}$, we can shift to $\eta_{\mu\nu}$ in small steps, where we define new coordinates for each step, and eventually reach the desired form of the metric.

Notice that only the symmetric part of $\epsilon_{\mu\nu}$ contributes to $\delta g_{\mu\nu}$. To understand why the antisymmetric part does not shift the metric, let us suppose we have the transformation in (4.1.9) but with $\epsilon_{\mu\nu}$ small and antisymmetric. Then to leading order near $x^\mu = 0$ we have that

$$x_{\mu'} = x_\mu + \epsilon_{\mu\nu}x^\nu\,. \qquad (4.1.11)$$

Now consider an infinitesimal Lorentz boost in the $\hat{x}$ direction, where we use the rapidity variable $\gamma = \cosh\zeta \approx 1$, $v\gamma = \sinh\zeta \approx \zeta$, then

$$\begin{aligned}
x_{0'} &= \eta_{0'0'}\Lambda^{0'}{}_\mu x^\mu \approx x_0 - \zeta x^1\\
x_{1'} &= \eta_{1'1'}\Lambda^{1'}{}_\mu x^\mu \approx x_1 + \zeta\,x^0\,. \qquad (4.1.12)
\end{aligned}$$

Comparing this to (4.1.11) we see that $\epsilon_{01} = -\epsilon_{10} = -\zeta$. Hence the antisymmetric $\epsilon_{\mu\nu}$ correspond to local Lorentz transformations which we previously knew did not change the metric.

Once we have set $g_{\mu\nu}(0) = \eta_{\mu\nu}$ we can then deal with the Christoffel symbols. Now we assume that

$$\xi_\mu(x) = \tfrac{1}{2}\, b_{\mu\nu\lambda}\, x^\nu x^\lambda + \dots \tag{4.1.13}$$

so that we leave the origin and the metric at the origin unchanged. Notice that for arbitrary $b_{\mu\nu\lambda}$ the expansion in (4.1.7) is valid so long as x^μ is small enough. With this form, the change to the Christoffel symbols at the origin is

$$\begin{aligned}
\delta\Gamma^\mu_{\nu\lambda}(0) &= \frac{1}{2}\eta^{\mu\sigma}\left(\partial_\lambda \delta g_{\nu\sigma} + \partial_\nu \delta g_{\sigma\lambda} - \partial_\sigma \delta g_{\nu\lambda}\right) \\
&= -\frac{1}{2}\eta^{\mu\sigma}\left(b_{\sigma\nu\lambda} + b_{\nu\sigma\lambda} + b_{\lambda\sigma\nu} + b_{\sigma\lambda\nu} - b_{\lambda\nu\sigma} - b_{\nu\lambda\sigma}\right) \\
&= -\eta^{\mu\sigma}\, b_{\sigma\nu\lambda}
\end{aligned} \tag{4.1.14}$$

Hence, by choosing $b_{\sigma\nu\lambda} = \eta_{\mu\sigma}\,\Gamma^\mu_{\nu\lambda}(0)$, $\delta\Gamma^\mu_{\nu\lambda}(0)$ will cancel off the Christoffel symbols at the origin, leaving us with the local inertial frame.

4.1.3 Free-falling

Now that we have established the existence of a local inertial frame, let us find the trajectories for free-falling in the noninertial frame. The key point is that free-falling is the same as constant velocity in the local inertial frame. To this end, let us consider the 4-velocity $\overrightarrow{U} = \frac{\overrightarrow{dx}}{d\tau}$, where τ is the proper time of the falling body. This is an invariant since we are considering the full vector and not just its components. We can then make an acceleration 4-vector $\overrightarrow{A} = \frac{d}{d\tau}\overrightarrow{U}$ by taking another τ derivative. Since this is the 4-acceleration for the free-falling body, we know that in the local inertial frame it is zero. But then it must be zero in all frames. Now write the vector $\overrightarrow{dx}$ in components, $\overrightarrow{dx} = \vec{e}_\mu dx^\mu$. If we take a τ derivative on this we find

$$\begin{aligned}
\frac{d}{d\tau}\frac{\overrightarrow{dx}}{d\tau} &= \frac{d}{d\tau}\left(\vec{e}_\mu\, \dot{x}^\mu\right) \\
&= \vec{e}_\mu \ddot{x}^\mu + \left(\frac{d}{d\tau}\vec{e}_\mu\right)\dot{x}^\mu = \vec{e}_\mu \ddot{x}^\mu + \left(\dot{x}^\nu \partial_\nu \vec{e}_\mu\right)\dot{x}^\mu \\
&= \vec{e}_\mu\left(\ddot{x}^\mu + \Gamma^\mu_{\nu\lambda}\dot{x}^\nu\dot{x}^\lambda\right) = 0\,,
\end{aligned} \tag{4.1.15}$$

where $\dot{x}^\mu \equiv \frac{d}{d\tau}x^\mu$. We used here that $\vec{e}_\mu$ does not explicitly depend on τ, it only depends on τ indirectly through the coordinates x^σ. Since the $\vec{e}_\mu$ are a complete set of basis vectors it must be true that all components in the last line are zero. Therefore, we are led to the equations of motion for free-fall,

$$\ddot{x}^\mu + \Gamma^\mu_{\nu\lambda}\dot{x}^\nu\dot{x}^\lambda = 0 \tag{4.1.16}$$

4.1.4 Geodesics

Equation (4.1.16) is called the *geodesic equation*. Why do we call it this? Instead of a space-time, let us consider a Euclidean space[1] with some metric and let us suppose that we want to find the shortest path between two points in this space. For example, suppose our space is the earth's surface and the two points are Logan airport in Boston and Ciampano airport in Rome. Boston and Rome are practically at the same latitude (42 degrees North). Yet when an airplane flies from Boston to Rome it does not go directly east, but starts out heading northeast. Eventually the flight starts turning toward the south and is heading in a southeasterly direction by the time it approaches Ciampano. Why does it take this route? Because it is the shortest! The route it takes is called a *geodesic*, or sometimes a *great circle*.

Let us see how this is related to (4.1.16). Let the metric on the earth's surface be $ds^2 = g_{ij}dx^i dx^j$ and suppose we have a continuous path between two points A and B. We parameterize the path by a variable λ, where $\lambda = 0$ at A and $\lambda = 1$ at B, although we could have chosen a different initial or final value. We write the coordinates on the path as $x^i(\lambda)$. To find the length L between A and B we then integrate ds along the path, since ds is giving us the infinitesimal length of the path between λ and $\lambda + d\lambda$. Hence we have

$$L = \int_A^B ds = \int_A^B \sqrt{g_{ij}dx^i dx^j} = \int_0^1 \sqrt{g_{ij}\dot{x}^i\dot{x}^j}\, d\lambda\,, \tag{4.1.17}$$

where $\dot{x}^i \equiv \frac{dx^i}{d\lambda}$. Let us assume that $x^i(\lambda)$ is the path that minimizes L. Then if we add a small correction $\delta x^i(\lambda)$ to the path, then to leading order δL is zero, since L is the minimum. Hence we have

$$0 = \int_0^1 g^{-1/2}\left(\frac{1}{2}\partial_k g_{ij}\dot{x}^i\dot{x}^j\delta x^k + g_{ij}\dot{x}^i\frac{d}{d\lambda}\delta x^j\right)d\lambda\,, \tag{4.1.18}$$

where $g \equiv g_{ij}\dot{x}^i\dot{x}^j$. Integrating the second term by parts we find

$$0 = \int_0^1 g^{-1/2}\left(\frac{1}{2}\partial_k g_{ij}\dot{x}^i\dot{x}^j - g_{ik}\ddot{x}^i - \partial_j g_{ik}\dot{x}^i\dot{x}^j + \frac{1}{2g}g_{ik}\dot{x}^i\frac{d}{d\lambda}g\right)\delta x^k d\lambda\,, \tag{4.1.19}$$

where there is no contribution from boundary terms since $\delta x^i(0) = \delta x^i(1) = 0$. Since (4.1.19) is 0 for any δx^i, the term inside the parentheses is 0. Let us first assume that $\frac{d}{d\lambda}g = 0$, showing afterward that this is consistent. Then we have

$$\begin{aligned} 0 &= g_{ik}\ddot{x}^i + \partial_j g_{ik}\dot{x}^i\dot{x}^j - \frac{1}{2}\partial_k g_{ij}\dot{x}^i\dot{x}^j \\ &= g_{ik}\ddot{x}^i + \frac{1}{2}\left(\partial_j g_{ik}\dot{x}^i\dot{x}^j + \partial_i g_{jk}\dot{x}^i\dot{x}^j - \partial_k g_{ij}\dot{x}^i\dot{x}^j\right). \end{aligned} \tag{4.1.20}$$

[1] The convention is to use the phrase "Euclidean space" when are all components are spatial. It does not need to be a flat space.

Multiplying through by a metric factor g^{kl} we get,

$$\begin{aligned} 0 &= g^{kl}\left(g_{ik}\ddot{x}^i + \frac{1}{2}\left(\partial_j g_{il}\dot{x}^i\dot{x}^j + \partial_i g_{jk}\dot{x}^i\dot{x}^j - \partial_k g_{ij}\dot{x}^i\dot{x}^j\right)\right) \\ &= \ddot{x}^l + \frac{g^{kl}}{2}\left(\partial_j g_{il}\dot{x}^i\dot{x}^j + \partial_i g_{jk}\dot{x}^i\dot{x}^j - \partial_k g_{ij}\dot{x}^i\dot{x}^j\right) = \ddot{x}^l + \Gamma^l_{ij}\dot{x}^i\dot{x}^j\,, \end{aligned} \tag{4.1.21}$$

matching the form of the equation in (4.1.16). Hence, the free fall equation in space-time is the analog of the shortest path between two points on the earth's surface. Actually, nothing in this derivation is specific to the surface of the earth. These arguments would have worked for any Euclidean space. Let us now show that $\frac{d}{d\lambda}g = 0$ is consistent. To this end we have

$$\begin{aligned} \frac{d}{d\lambda}g &= 2g_{ij}\ddot{x}^i\dot{x}^j + \partial_j g_{ik}\dot{x}^i\dot{x}^k\dot{x}^j \\ &= 2g_{ik}\ddot{x}^i\dot{x}^k + \partial_j g_{ik}\dot{x}^i\dot{x}^k\dot{x}^j + \partial_i g_{jk}\dot{x}^i\dot{x}^k\dot{x}^j - \partial_k g_{ij}\dot{x}^i\dot{x}^k\dot{x}^j \\ &= 2\left[g_{ik}\ddot{x}^i + \frac{1}{2}\left(\partial_j g_{ik}\dot{x}^i\dot{x}^j + \partial_i g_{jk}\dot{x}^i\dot{x}^j - \partial_k g_{ij}\dot{x}^i\dot{x}^j\right)\right]\dot{x}^k\,. \end{aligned} \tag{4.1.22}$$

But the term inside the square brackets is the same term as the last line of (4.1.20), which is zero. Hence $\frac{d}{d\lambda}g = 0$ is consistent.

4.1.5 The particle action

We have just shown that by minimizing the length of a path we get the geodesic equation. We know that in particle mechanics that by minimizing an action we should get the equations of motion. The equations of motion are the geodesic equations, so the action should be closely related to (4.1.17). For the particle motion, the particle is on a time-like trajectory, while the measured length is space-like. Hence, instead of ds we should use $d\tau$ and instead of g we should use $-g = -g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu$, where $\dot{x}^\mu = \frac{d}{d\tau}x^\mu$. Finally, the action should have dimension of (energy)$\times$(time). In our units where $c = 1$, mass has the same units as energy, so to get an action with appropriate dimension we will multiply by the particle's rest mass m. Hence, the particle action is

$$S = m\int_{\tau_i}^{\tau_f}\sqrt{-g}\,d\tau\,, \tag{4.1.23}$$

where τ_i and τ_f are the initial and final proper times for the particle's space-time trajectory. By the same arguments as the last section, minimizing S will lead to the equations of motion in (4.1.16).

4.1.6 An example

Suppose we have a metric in polar coordinates that looks like[2]

$$ds^2 = -\frac{r^2}{R^2}dt^2 + \frac{R^2}{r^2}dr^2 + R^2(d\theta^2 + \sin^2\theta\,d\phi^2) = -d\tau^2 \tag{4.1.24}$$

[2]This is the metric for the space-time $AdS_2 \times S^2$. We will learn the origin of this name later in the course.

where R is a constant. Let us also suppose that a body that is free-falling radially, such that $\dot\theta = \dot\phi = 0$, in which case we can ignore the θ and ϕ part of the metric and work with the two dimensional space-time parameterized by (t, r).

Since the metric is diagonal, the components of the inverse metric are also diagonal and have the form $g^{aa} = 1/g_{aa}$, where the repeated indices are *not* summed over. The Christoffel symbols are also easier to compute with a diagonal metric and have the form

$$\Gamma^a_{bb} = -\tfrac{1}{2}g^{aa}\partial_a g_{bb}\,, \qquad \Gamma^b_{ab} = \tfrac{1}{2}g^{bb}\partial_a g_{bb}\,, \qquad \Gamma^a_{aa} = \tfrac{1}{2}g^{aa}\partial_a g_{aa}\,, \qquad \Gamma^a_{bc} = 0\,, \tag{4.1.25}$$

where a, b and c are assumed to be different indices and repeated indices are not summed over. Hence, for the metric in (4.1.24) the nonzero Christoffel symbols involving r and t are

$$\begin{aligned}
\Gamma^r_{rr} &= \frac{1}{2}g^{rr}\partial_r g_{rr} = \frac{1}{2}\frac{r^2}{R^2}\partial_r\frac{R^2}{r^2} = -\frac{1}{r} \\
\Gamma^r_{tt} &= -\frac{1}{2}g^{rr}\partial_r g_{tt} = -\frac{1}{2}\frac{r^2}{R^2}\partial_r\left(-\frac{r^2}{R^2}\right) = \frac{r^3}{R^4} \\
\Gamma^t_{tr} &= \frac{1}{2}g^{tt}\partial_r g_{tt} = \frac{1}{2}\left(-\frac{R^2}{r^2}\right)\partial_r(-\frac{r^2}{R^2}) = \frac{1}{r}
\end{aligned} \tag{4.1.26}$$

Hence we find

$$\begin{aligned}
\ddot r - \frac{1}{r}\dot r^2 + \frac{r^3}{R^4}\dot t^2 &= 0 \\
\ddot t + 2\frac{1}{r}\dot r\dot t &= 0\,.
\end{aligned} \tag{4.1.27}$$

At first glance (4.1.27) looks very complicated since we have two coupled nonlinear second order differential equations. However, if we multiply the last equation by r^2, we can rewrite this as

$$\frac{d}{d\tau}(r^2\dot t) = 0\,, \tag{4.1.28}$$

which has the solution

$$r^2\dot t = C \tag{4.1.29}$$

where C is an integration constant. Hence, we can then write the first equation as

$$\ddot r - \frac{1}{r}\dot r^2 + \frac{C^2}{rR^4} = 0\,, \tag{4.1.30}$$

which after multiplying by $\dot r/r^2$ becomes

$$\frac{\dot r\ddot r}{r^2} - \frac{\dot r^3}{r^3} + \frac{C^2\dot r}{r^3R^4} = \frac{d}{d\tau}\left(\frac{1}{2}\frac{\dot r^2}{r^2} - \frac{1}{2}\frac{C^2}{r^2R^4}\right) = 0\,. \tag{4.1.31}$$

It then follows that

$$\dot r^2 = \frac{C^2}{R^4} - \bar C r^2\,, \tag{4.1.32}$$

where $\bar{C}$ is another integration constant.

However, there is a faster way of deriving (4.1.32) which in addition shows that $\bar{C}$ must be fixed to a particular value. If we take the metric in (4.1.24) and divide by $d\tau^2$, mulitply by r^2/R^2, set $\dot{\phi} = \dot{\theta} = 0$ and use (4.1.29) we get the relation

$$\left(\frac{dr}{d\tau}\right)^2 = \frac{r^4}{R^4}\left(\frac{dt}{d\tau}\right)^2 - \frac{r^2}{R^2}\left(\frac{d\tau}{d\tau}\right)^2 = \frac{C^2}{R^4} - \frac{r^2}{R^2}\,, \tag{4.1.33}$$

which is (4.1.32) with $\bar{C} = 1/R^2$.

Equation (4.1.33) is solved by quadrature,

$$\tau = R\int_{r_i}^{r} \frac{dr}{\sqrt{\frac{C^2}{R^2} - r^2}}\,, \tag{4.1.34}$$

where r_i is the initial value of r. In this case the solution is

$$\tau = R\left(\arcsin\left(\frac{R}{C}\,r\right) - \arcsin\left(\frac{R}{C}\,r_i\right)\right), \tag{4.1.35}$$

which we can invert to

$$r = \frac{C}{R}\sin\left(\frac{1}{R}(\tau + \tau_0)\right), \tag{4.1.36}$$

where τ_0 is chosen to give r_i at $\tau = 0$. Notice that there is a maximum value for r, $r_{\mathrm{max}} = \frac{C}{R}$ where $\dot{r} = 0$, hence the body can never escape to infinity for any finite value of r_i and C.

We can now go back and solve for t, which has the form

$$\begin{aligned} t &= \int_0^\tau \frac{C d\tau}{r^2(\tau)} = \frac{R^2}{C}\int_0^\tau \frac{d\tau}{\sin^2(R^{-1}(\tau + \tau_0))} \\ &= \frac{R^3}{C}\left(\cot\left(\frac{1}{R}\,\tau_0\right) - \cot\left(\frac{1}{R}\,(\tau + \tau_0)\right)\right) \\ &= R\sqrt{\frac{R^2}{r_i^2} - \frac{R^4}{C^2}} - R\sqrt{\frac{R^2}{r^2} - \frac{R^4}{C^2}}\,. \end{aligned} \tag{4.1.37}$$

Assuming that $\dot{r}$ starts out positive, then we first use the positive branch of the square root to evaluate t. Once r reaches r_{max} we switch to the negative branch so that $\dot{t}$ is always positive. Notice that it takes an infinite amount of coordinate time t to reach $r = 0$.

4.1.7 Light-like trajectories

The geodesic equations are also valid for light rays, although in this case we need to replace τ with a variable λ that can not be interpreted as the proper time, but is still a parameter that marks the points on the space-time trajectory. However, it is actually easier to just use the metric, setting $ds^2 = 0$ on the trajectory to make it light-like. Returning to the example in the previous section, we find for the metric in (4.1.24) that

$$dt = \pm \frac{R^2}{r^2}\, dr\,, \tag{4.1.38}$$

assuming that the motion is along fixed θ and ϕ. We then integrate (4.1.38) to find

$$t = \pm \int_{r_i}^{r} \frac{R^2 dr}{r^2} = \pm\left(\frac{R^2}{r_i} - \frac{R^2}{r}\right), \tag{4.1.39}$$

where we have $+$ for an outgoing trajectory and $-$ for an ingoing trajectory. Notice that for the outgoing trajectory the light reaches $r = \infty$ in a finite amount of coordinate time, $t = R^2/r_i$. For the ingoing trajectory it takes an infinite amount of time to reach $r = 0$. One further observation is that if we take the limit $C \to \infty$ in (4.1.37) then the relation between t and r approaches (4.1.39).

4.2 Nonrelativistic motion in weak gravitational fields

In a previous chapter we suggested that the g_{00} component of the metric was related to the gravitational potential. Let us now explicitly show this by considering nonrelativistic motion for a test particle of rest mass m and show that the geodesic equations are consistent with this. We assume that the metric has the form

$$ds^2 = -(1 + 2\Phi(\underset{\rightarrow}{x}))dt^2 + d^2x^i\,, \tag{4.2.1}$$

where the repeated i index is summed over and $\underset{\rightarrow}{x}$ symbolizes the spatial 3-vector. $\Phi(\underset{\rightarrow}{x})$ only depends on the spatial coordinates and is assumed to be small, that is $|\Phi(\underset{\rightarrow}{x})| << 1$. It is possible to include small corrections to the spatial components of the metric, but they will be unimportant for free-fall in the nonrelativistic limit (In a subsequent lecture we will see that we need to include such corrections when considering space-time curvature).

For nonrelativistic motion, we can approximate the proper time to be $\tau = t$ so that $\dot t \approx 1$ and $\dot x^i \approx \frac{d}{dt}x^i$ and $|\dot x^i| << 1$. From the geodesic equations and our approximations we have

$$m\,\frac{d^2}{dt^2}\,x^i \approx m\,\ddot x^i = -m\,\Gamma^i_{\mu\nu}\dot x^\mu \dot x^\nu \approx -m\,\Gamma^i_{00}\,. \tag{4.2.2}$$

The relevant Christoffel symbol is easily computed, where we find

$$\Gamma^i_{00} = -\frac{1}{2}g^{ij}\partial_j g_{00} = \partial_i \Phi(\underset{\rightarrow}{x})\,, \tag{4.2.3}$$

where we used that $g^{ij} = \delta^{ij}$. Hence the equation of motion is

$$m\,\frac{d^2}{dt^2}\,\underset{\rightarrow}{x} = -m\,\nabla\Phi(\underset{\rightarrow}{x})\,. \tag{4.2.4}$$

If we identify $\Phi(\underrightarrow{x})$ with the gravitational potential then $-m\,\nabla\Phi(\underrightarrow{x})$ is the gravitational force and so we find that (4.2.4) is Newton's force equation for motion in a gravitational field.

For example, suppose we consider the gravitational potential outside a spherically symmetric potential of mass M. In this case Φ is given by

$$\Phi(r) = -\frac{MG}{r}\,, \tag{4.2.5}$$

where G is Newton's gravitational constant, $G \approx 6.674 \times 10^{-11}\,\mathrm{m}^3\,\mathrm{kg}^{-1}\mathrm{s}^{-2} \approx 7.42 \times 10^{-28}\,\mathrm{m\,kg}^{-1}$ where the latter value is given in units where $c = 1$. If this is to be a weak potential with $|\Phi(r)| << 1$, then $r >> MG$. For a solar mass where $M \approx 2 \times 10^{30}$ kg, then $r >> 1.5 \times 10^3$ m. The radius of the sun is about 7×10^8 m, so even at this distance the potential is sufficiently weak.

4.3 Isometries and Killing vectors

4.3.1 Existence of Killing vectors

It might seem miraculous that we are able to solve for the equations in (4.1.27). It turns out that there is an underlying reason why we could do it — there exists constants of the motion. The idea should be familiar to you from ordinary mechanics, where conservation of energy can be used to convert a second order differential equation to a first-order equation. Motion in three spatial dimensions can also be solved if there is another conserved quantity. For example, a central force conserves angular momentum. With the angular momentum conserved, we can convert a differential equation in terms of time derivatives on r, θ and ϕ to one where only r appears.

For motion in curved space there are conserved quantities if the metric has *isometries*. An isometry is coordinate transformation that leaves the metric unchanged. Hence if our coordinate transformation is $x^\mu \to x^\mu + \xi^\mu(x)$, then this is an isometry if

$$\nabla_\mu \xi_\nu + \nabla_\nu \xi_\mu = 0\,. \tag{4.3.1}$$

Notice that because this is a tensor equation, if it is true for one frame then it is true for all frames. A vector $\vec{\xi}$ that satisfies (4.3.1) is called a *Killing vector*.

Now let us see how the existence of a Killing vector leads to a conserved quantity. Consider the combination

$$\vec{\xi}\cdot\vec{U} = \vec{\xi}\cdot\frac{d}{d\tau}\overrightarrow{dx} = g_{\mu\nu}\,\xi^\mu\dot{x}^\nu\,, \tag{4.3.2}$$

where $\vec{U}$ is the velocity 4-vector, and take a derivative on this combination with respect to τ. This gives

$$
\begin{aligned}
\frac{d}{d\tau}\left(\vec{\xi}\cdot\frac{d}{d\tau}\overrightarrow{dx}\right) &= \partial_\lambda g_{\mu\nu}\,\xi^\mu\dot{x}^\lambda\dot{x}^\nu + g_{\mu\nu}\,\partial_\lambda\xi^\mu\dot{x}^\lambda\dot{x}^\nu + g_{\mu\nu}\,\xi^\mu\ddot{x}^\nu \\
&= \partial_\lambda g_{\mu\nu}\,\xi^\mu\dot{x}^\lambda\dot{x}^\nu + g_{\mu\nu}\,\partial_\lambda\xi^\mu\dot{x}^\lambda\dot{x}^\nu - g_{\mu\nu}\,\xi^\mu\Gamma^\nu_{\lambda\sigma}\dot{x}^\lambda\dot{x}^\sigma \\
&= \partial_\lambda g_{\mu\nu}\,\xi^\mu\dot{x}^\lambda\dot{x}^\nu + g_{\mu\nu}\,\partial_\lambda\xi^\mu\dot{x}^\lambda\dot{x}^\nu - \frac{1}{2}\xi^\mu\left(\partial_\lambda g_{\mu\sigma} + \partial_\sigma g_{\lambda\mu} - \partial_\mu g_{\lambda\sigma}\right)\dot{x}^\lambda\dot{x}^\sigma \\
&= \frac{1}{2}\left(g_{\lambda\nu}\partial_\mu\xi^\lambda + g_{\mu\lambda}\partial_\nu\xi^\lambda + \xi^\lambda\partial_\lambda g_{\mu\nu}\right)\dot{x}^\mu\dot{x}^\nu\,, \qquad (4.3.3)
\end{aligned}
$$

where in going from the first to the second line we used the geodesic equation (4.1.16), and from the second to the third we wrote the Christoffel symbols in terms of the metric. The fourth line follows from the third after a relabeling of indices and a cancellation of terms. If we now consult the first line of (4.1.8), we see that the term inside the parentheses in the last line of (4.3.3) is $-\delta g_{\mu\nu}$. Since $\vec{\xi}$ is a Killing vector, $\delta g_{\mu\nu} = 0$ and so $\vec{\xi}\cdot\frac{d}{d\tau}\overrightarrow{dx}$ is constant in τ.

4.3.2 The example again

Let us see how this works with our example in section 4.1.6. If we let $\xi^t \equiv \xi^0 = 1$ and set all spatial components to zero, then

$$
\nabla_\mu\xi_\nu + \nabla_\nu\xi_\mu = g_{\nu\lambda}\nabla_\mu\xi^\lambda + g_{\lambda\mu}\nabla_\nu\xi^\lambda = g_{\nu\lambda}\Gamma^\lambda_{\mu\sigma}\xi^\sigma + g_{\lambda\mu}\Gamma^\lambda_{\nu\sigma}\xi^\sigma\,. \qquad (4.3.4)
$$

Most combinations of μ and ν obviously give 0. If μ is r and ν is t, then we have

$$
\nabla_r\xi_t + \nabla_t\xi_r = g_{tt}\Gamma^t_{rt}\xi^t + g_{rr}\Gamma^r_{tt}\xi^t = -\frac{r^2}{R^2}\frac{1}{r} + \frac{R^2}{r^2}\frac{r^3}{R^4} = 0\,. \qquad (4.3.5)
$$

Hence $\vec{\xi}$ is a Killing vector. Actually, there is an easier way to see that $\vec{\xi}$ is a Killing vector just by inspection of the metric. The only time dependence in ds^2 is in the dt^2 term. Hence, if we shift t by a constant the metric is unchanged. Therefore $\vec{\xi}$ is a Killing vector.

The constant of the motion that we extract from $\vec{\xi}$ is

$$
\vec{\xi}\cdot\frac{d}{d\tau}\overrightarrow{dx} = g_{\mu\nu}\,\xi^\mu\dot{x}^\nu = -\frac{r^2}{R^2}\dot{t}\,, \qquad (4.3.6)
$$

which up to a constant factor is the expression in (4.1.29).

Since $\vec{\xi}\cdot\vec{\xi} = g_{\mu\nu}\,\xi^\mu\xi^\nu = -\frac{r^2}{R^2} < 0$, we call $\vec{\xi}$ a *time-like Killing vector*. If we inspect the metric in (4.1.24), we see that the only ϕ dependence is in $d\phi^2$, hence the metric is invariant under a constant shift in ϕ and so $\vec{\zeta}$ is a Killing vector that has $\zeta^\phi = 1$ and all other components zero. This Killing vector is space like and the conserved quantity that we make from it is

$$
\vec{\zeta}\cdot\dot{\vec{x}} = g_{\mu\nu}\zeta^\mu\dot{x}^\nu = R^2\sin^2\theta\,\dot{\phi}\,. \qquad (4.3.7)
$$

4.3.3 The conserved quantities

Physically, how should one understand the conserved quantities? For this, let us consider motion of a body with rest mass m in an inertial frame. In this case any constant vector $\vec{\xi}$ is a Killing vector and $\vec{\xi}\cdot\vec{U} = \eta_{\mu\nu}\xi^\mu U^\nu$ is a constant of the motion. If we take $\xi^0 = 1$ and all other components zero, and further multiply by a factor of $-m$ we have

$$-m\,\vec{\xi}\cdot\vec{U} = m\,U^0\,, \tag{4.3.8}$$

which is the energy of the particle. In the previous example which does not have a global inertial frame, the conserved quantity derived from the time-like Killing vector is then related to the energy of the particle. However, since gravity is built into the metric, the energy that is conserved is the combined kinetic plus gravitational potential energy.

The space-like Killing vectors in the inertial frame lead to other conserved quantities which are the components of the linear momentum. If we had used polar coordinates with metric

$$ds^2 = -dt^2 + dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)\,, \tag{4.3.9}$$

then we can have the space-like Killing vector with $\zeta^\phi = 1$ and all other components zero, which leads to the conserved quantity $r^2\dot{\phi}$. Multiplying by m, this is the angular momentum. Hence Killing vectors along angular directions lead to conserved angular momentum.

Chapter 5

Curvature and the Riemann tensor

In this chapter we give a mathematical description of curvature. This culminates in the construction of a tensor called the Riemann tensor.

5.1 Parallel transport and curvature

In the last chapter we discussed how to determine the motion of massive bodies or light given a metric. The metric itself is determined by solving Einstein's equations which encodes the gravitational interactions. The source of the gravitational fields can be massive bodies, radiation or even something called *dark energy*. These gravitational sources end up curving the space, which changes the metric and hence effects the motion of test bodies in the gravitational fields.

It thus behooves us to learn how space-time gets curved by the gravitational sources. In order to do this we require a precise definition of curvature which will enter into Einstein's equations, which roughly speaking will be of the form

$$\text{“curvature=gravitational sources”}\,.$$

Obviously, we will need to make this more precise.

5.1.1 Parallel transport

Suppose we have a vector at a particular point in space-time and we want to move it to another point in a series of infinitesimal steps, such that the vector at the beginning of the step is parallel to the vector at the end of the step. Such a process is called “parallel transport”. If the space-time is flat, then the result of the transport is independent of the path taken to get from the initial point to the final point. However, if the space is curved, then the result will be path dependent.

Figure 5.1 shows parallel transport in a two dimensional flat space along two different paths. As you can see, the arrow stays parallel to the preceding arrow in the transport. You can also see that at the end of the transport, the arrow points in the same direction for both paths. Another way to look at this transport as parallel transporting the vector around a loop, where the transport goes out along one path and them comes back to the

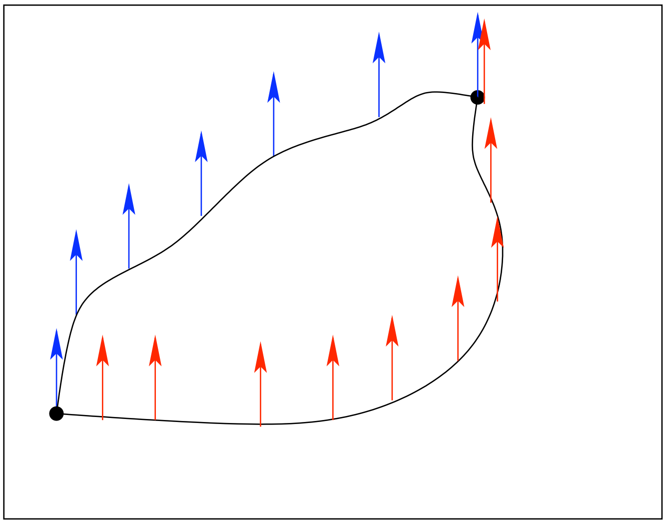

Figure 5.1: Parallel transport along two different paths in a flat two dimensional space. Alternatively, this is parallel transport around a loop with the red arrows for the transport going out (say) and the blue arrows for the transport going in.

original point along the other path. In this case the vector that returns is pointing in the same direction as the starting vector.

Next consider transport along the surface of a two-dimensional sphere, such as the surface of the earth. Figure 5.2 shows parallel transport of a vector that starts out pointing north on a point on the earth's equator. It is then parallel transported 90 degrees east along the equator, after which it is still pointing north. Then we transport the vector northward all the way to the North pole. Finally, we transport southward toward the origin of the transport. As you can see, the vector is not the original vector, but has been rotated by 90 degrees and is now pointing west. This change in the vector under the parallel transport is due to the earth's curvature. Curvature is one of those things where you know it and when you see it, but parallel transport will give us a way of defining curvature precisely. In particular, it will give us a way of defining the local curvature, that is the curvature at a particular point in space-time. The earth has roughly the same curvature everywhere on its surface, but we will consider many space-times where the curvature is position dependent.

To proceed, suppose we have a vector $\vec{A}$, which we write in terms of basis vectors as $\vec{e}_\mu A^\mu$. We want to parallel transport the vector along some path parameterized by a variable λ. If the vector is parallel transported along the path then it does not vary along the path, so $\frac{\partial}{\partial\lambda}\vec{A} = 0$. Since we can parallel transport along any path we choose, it might just seem that one way to insure parallel transport is for $\vec{A}$ to have constant components everywhere in space-time. But the components depend on the basis vectors and the basis vectors themselves could change as we move along the path. Hence, the more general case might be that $\vec{A}$ is made up of components where each component is a single valued function on space-time, although not necessarily a constant function. But even this is wrong as can be seen by our example of parallel transport along the earth's surface, where the vector is clearly not single valued under parallel transport around the

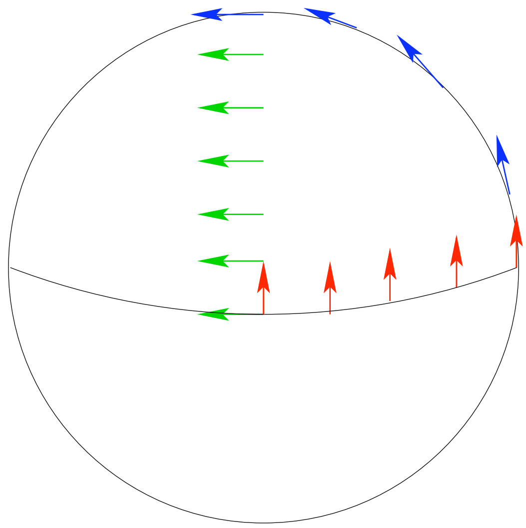

Figure 5.2: Parallel transport on a two dimensional sphere. The path first goes east (red), and then north to the North pole (blue) and then south back to the original point on the equator (green). After the transport he vector has turned by 90 degrees.

loop (the original vector is different from the final vector even though both are at the same point.).

To sort this out, let us say we have some vector $\vec{A}_{12}(x)$ and some coordinate system such that the components $A^\mu_{12}(x)$ *are* single valued functions in space-time (as in the last hand-out, the functions depend on all components of the coordinates x^μ, but we drop the index inside the functions). Let $\vec{A} = \vec{A}_{12}(x)$ for some particular point with coordinates x^μ and let us parallel transport $\vec{A}$ a short distance away to the point with coordinates $x^\mu + \delta\, x^\mu_1$ and let us further assume that this parallel transported vector *is* $\vec{A}_{12}(x + \delta\, x_1)$. Under parallel transport we have

$$\delta x^\sigma_1 \partial_\sigma \vec{A}_{12}(x) = 0 = \delta x^\sigma_1 \vec{e}_\mu \big(\partial_\sigma A^\mu_{12}(x) + \Gamma^\mu_{\sigma\nu}(x) A^\nu_{12}(x)\big) = \delta x^\sigma_1 \vec{e}_\mu \, \nabla_\sigma \, A^\mu(x)\,, \qquad (5.1.1)$$

that is, the covariant derivative of A^μ_{12} is zero along the parallel transport direction. Then the components of the parallel transported vector, $A^\mu_{12}(x + \delta\, x_1)$, satisfy to leading order

$$A^\mu_{12}(x + \delta\, x_1) \approx A^\mu_{12}(x) + \delta x^\sigma_1 \partial_\sigma A^\mu_{12}(x) = A^\mu_{12}(x) - \delta x^\sigma_1 \Gamma^\mu_{\sigma\nu}(x) A^\nu_{12}(x) = A^\mu - \delta x^\sigma_1 \Gamma^\mu_{\sigma\nu}(x) A^\nu\,. \qquad (5.1.2)$$

Now suppose we take the parallel transported vector at $x^\mu + \delta\, x^\mu_1$ and parallel transport it a short distance away to $x^\mu + \delta\, x^\mu_1 + \delta\, x^\mu_2$ where we assume that $\delta\, \vec{x}_2$ is not parallel to $\delta\, \vec{x}_1$. Let us further assume that $\vec{A}_{12}(x + \delta\, x_1 + \delta\, x_2)$ is the parallel transported vector.

Since we got to this point by parallel transport, we have that

$$\begin{aligned}
\vec{A}_{12}(x+\delta\, x_1+\delta\, x_2) &\approx A^\mu_{12}(x+\delta\, x_1)+\delta x_2^\sigma\partial_\sigma A^\mu_{12}(x+\delta\, x_1)\\
&= A^\mu_{12}(x+\delta\, x_1)-\delta x_2^\sigma\Gamma^\mu_{\sigma\nu}(x+\delta\, x_1)A^\nu_{12}(x+\delta\, x_1)\\
&\approx A^\mu-\delta x_1^\sigma\Gamma^\mu_{\sigma\nu}\,A^\nu-\delta x_2^\sigma\Gamma^\mu_{\sigma\nu}\,A^\nu\\
&\qquad -\delta x_2^\sigma\delta x_1^\rho\,\partial_\rho\Gamma^\mu_{\sigma\nu}\,A^\nu+\delta x_2^\sigma\delta x_1^\rho\,\Gamma^\mu_{\sigma\nu}\Gamma^\nu_{\rho\kappa}\,A^\kappa\\
&= A^\mu-\delta x_1^\sigma\Gamma^\mu_{\sigma\nu}\,A^\nu-\delta x_2^\sigma\Gamma^\mu_{\sigma\nu}\,A^\nu\\
&\qquad -\delta x_2^\sigma\delta x_1^\rho\left(\partial_\rho\Gamma^\mu_{\sigma\nu}-\Gamma^\mu_{\kappa\sigma}\Gamma^\kappa_{\nu\rho}\right)A^\nu
\end{aligned}\tag{5.1.3}$$

where the Christoffel symbols with no explicit argument are evaluated at x^μ. In going to the last line we exchanged the ν and the κ dummy indices and used the symmetry properties of the Christoffel symbols.

Now suppose that we had parallel transported in the opposite order, that is first we parallel transport along $\delta\, x_2^\mu$ and then $\delta\, x_1^\mu$. Then we would have defined a *different* single valued vector function $\vec{A}_{21}(x)$ such that $\vec{A}_{21}(x+\delta\, x_2)$ is the vector after parallel transporting from x^μ to $x^\mu+\delta\, x_2$ and $\vec{A}_{21}(x+\delta\, x_2+\delta\, x_1)$ is the vector after parallel transporting to $x^\mu+\delta\, x_2+\delta\, x_1$. Exchanging 1 and 2 in (5.1.3) we get

$$\begin{aligned}
\vec{A}_{21}(x+\delta\, x_2+\delta\, x_1) &= A^\mu-\delta x_2^\sigma\Gamma^\mu_{\sigma\nu}\,A^\nu-\delta x_1^\sigma\Gamma^\mu_{\sigma\nu}\,A^\nu\\
&\qquad -\delta x_1^\sigma\delta x_2^\rho\left(\partial_\rho\Gamma^\mu_{\sigma\nu}-\Gamma^\mu_{\kappa\sigma}\Gamma^\kappa_{\nu\rho}\right)A^\nu\\
&= A^\mu-\delta x_2^\sigma\Gamma^\mu_{\sigma\nu}\,A^\nu-\delta x_1^\sigma\Gamma^\mu_{\sigma\nu}\,A^\nu\\
&\qquad -\delta x_2^\sigma\delta x_1^\rho\left(\partial_\sigma\Gamma^\mu_{\rho\nu}-\Gamma^\mu_{\kappa\rho}\Gamma^\kappa_{\nu\sigma}\right)A^\nu\,,
\end{aligned}\tag{5.1.4}$$

where we exchanged the σ and ρ indices in the last line. If we now consider the parallel transport around the vanishingly small loop where first we transport in the $\delta\,\vec{x}_1$ direction, then $\delta\,\vec{x}_2$, then $-\delta\,\vec{x}_1$ and finally $-\delta\,\vec{x}_2$, then the change in the components of the vector δA^μ are given by the difference between the two vector functions A^μ_{12} and A^μ_{21} at $x^\mu+\delta\, x_2^\mu+\delta\, x_1^\mu$. This is illustrated in figure 5.3. Thus, the resulting change in the vector from parallel transport around the loop is

$$\begin{aligned}
\delta A^\mu &= A^\mu_{12}(x+\delta\, x_1+\delta\, x_2)-A^\mu_{21}(x+\delta\, x_2+\delta\, x_1)\\
&= \left(\partial_\sigma\Gamma^\mu_{\rho\nu}-\partial_\rho\Gamma^\mu_{\sigma\nu}+\Gamma^\mu_{\kappa\sigma}\Gamma^\kappa_{\nu\rho}-\Gamma^\mu_{\kappa\rho}\Gamma^\kappa_{\nu\sigma}\right)A^\nu\,\delta x_2^\sigma\delta x_1^\rho\\
&\equiv R^\mu{}_{\nu\sigma\rho}\,A^\nu\,\delta x_2^\sigma\delta x_1^\rho\,.
\end{aligned}\tag{5.1.5}$$

$R^\mu{}_{\nu\sigma\rho}$ is called the *Riemann curvature tensor*, or just Riemann tensor, and as it names implies it is a tensor. To show this, consider the commutator of covariant derivatives $[\nabla_\sigma,\nabla_\rho]$ acting on A^μ, where we find

$$\begin{aligned}
[\nabla_\sigma,\nabla_\rho]A^\mu &\equiv (\nabla_\sigma\nabla_\rho-\nabla_\rho\nabla_\sigma)A^\mu\\
&= \partial_\sigma(\nabla_\rho A^\mu)-\cancel{\Gamma^\nu_{\rho\sigma}\nabla_\nu A^\mu}+\Gamma^\mu_{\nu\sigma}\nabla_\rho A^\nu-\partial_\rho(\nabla_\sigma A^\mu)+\cancel{\Gamma^\nu_{\sigma\rho}\nabla_\nu A^\mu}-\Gamma^\mu_{\nu\rho}\nabla_\sigma A^\nu\\
&= \cancel{\partial_\sigma\partial_\rho A^\mu}+\partial_\sigma(\Gamma^\mu_{\nu\rho}A^\nu)+\Gamma^\mu_{\nu\sigma}\partial_\rho A^\nu+\Gamma^\mu_{\nu\sigma}\Gamma^\nu_{\kappa\rho}A^\kappa\\
&\qquad -\cancel{\partial_\rho\partial_\sigma A^\mu}-\partial_\rho(\Gamma^\mu_{\nu\sigma}A^\nu)-\Gamma^\mu_{\nu\rho}\partial_\sigma A^\nu-\Gamma^\mu_{\nu\rho}\Gamma^\nu_{\kappa\sigma}A^\kappa\\
&= \left(\partial_\sigma\Gamma^\mu_{\nu\rho}-\partial_\rho\Gamma^\mu_{\nu\sigma}+\Gamma^\mu_{\kappa\sigma}\Gamma^\kappa_{\nu\rho}-\Gamma^\mu_{\kappa\sigma}\Gamma^\kappa_{\nu\rho}\right)A^\nu\\
&= R^\mu{}_{\nu\sigma\rho}\,A^\nu\,.
\end{aligned}\tag{5.1.6}$$

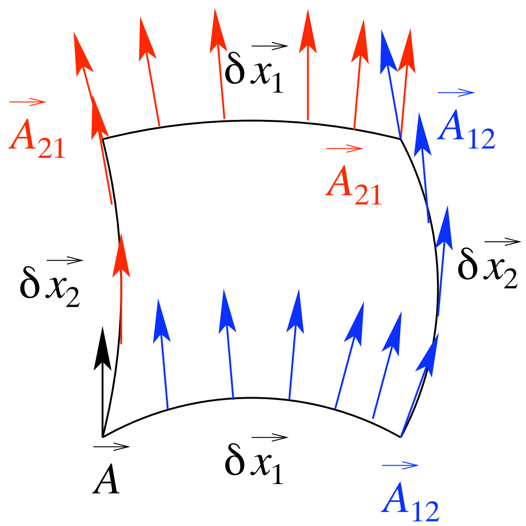

Figure 5.3: Parallel transport around a small loop. The blue shows parallel transport along the $\delta\,\vec{x}_1$ then $\delta\,\vec{x}_2$ directions while the red is along the $\delta\,\vec{x}_2$ then $\delta\,\vec{x}_1$ directions. The result is the difference between $\vec{A}_{12}$ and $\vec{A}_{21}$ at $x^\mu + \delta\,x_1^\mu + \delta\,x_2^\mu$ in the top right corner.

By construction $[\nabla_\sigma, \nabla_\rho]A^\mu$ is a tensor and since A^ν is a contravariant vector then $R^\mu{}_{\nu\sigma\rho}$ is a $\binom{1}{3}$ tensor.

Notice that it does not matter what the vector is. We could have chosen $\vec{A}_{12}(x)$ or $\vec{A}_{21}(x)$ or any other single valued vector function, we would still find the same relation with the Riemann tensor which only depends on the local geometry of the space-time. Also note that if we had chosen $\vec{A} = A_{12}(x)$, and considered x^μ to be the starting point used in the preceding discussion, we would have

$$R^\mu{}_{\nu\sigma\rho}\, A_{12}^\nu\, \delta x_2^\sigma \delta x_1^\rho = [\nabla_\sigma, \nabla_\rho] A_{12}^\mu\, \delta x_2^\sigma \delta x_1^\rho = -\nabla_\rho \nabla_\sigma A_{12}^\mu\, \delta x_2^\sigma \delta x_1^\rho\,, \tag{5.1.7}$$

On the other hand, choosing $\vec{A} = A_{21}(x)$ at the same x^μ gives

$$R^\mu{}_{\nu\sigma\rho}\, A_{21}^\nu\, \delta x_2^\sigma \delta x_1^\rho = [\nabla_\sigma, \nabla_\rho] A_{21}^\mu\, \delta x_2^\sigma \delta x_1^\rho = +\nabla_\sigma \nabla_\rho A_{21}^\mu\, \delta x_2^\sigma \delta x_1^\rho\,, \tag{5.1.8}$$

A nonzero-Riemann tensor tells us that we can't parallel transport a single valued vector function around the loop because it is not possible to set all covariant derivatives to zero.

5.1.2 Properties of the Riemann tensor

Now that we know that the Riemann tensor is indeed a tensor, we can use this fact to establish several important properties about $R^\mu{}_{\nu\sigma\rho}$. For instance, for a global inertial frame the Christoffel symbols are zero everywhere and so $R^\mu{}_{\nu\sigma\rho} = 0$. Hence, there is no curvature. This is why we referred to such a space as *flat* Minkowski space in an earlier lecture. Furthermore, since $R^\mu{}_{\nu\sigma\rho} = 0$ is a tensor equation, all components are zero for any frame on this space-time, even in frames where some of the Christoffel symbols are nonzero, such as the constant accelerating frame. We can also argue backwards to say that if some components of $R^\mu{}_{\nu\sigma\rho}$ are nonzero then there does not exist a frame where all components are zero. Hence, unlike the Christoffel symbols where a frame always exists at any given point where all $\Gamma^\mu_{\nu\lambda}$ are zero, the same is not true for the Riemann tensor. In other words, the curvature is not removed, not even locally, by a good choice of coordinates.

Another property that we can read off immediately from (5.1.6) is that $R^\mu{}_{\nu\sigma\rho}$ is antisymmetric in its last two indices, $R^\mu{}_{\nu\sigma\rho} = -R^\mu{}_{\nu\rho\sigma}$. To find more properties let us consider $R^\mu{}_{\nu\sigma\rho}$ at one point x^μ where the coordinates are a local inertial frame. In this case $g_{\mu\nu} = \eta_{\mu\nu}$ and $\Gamma^\mu_{\nu\sigma} = 0$ for all components at x^μ. Since the metric is covariantly constant this also means that $\partial_\sigma g_{\mu\nu} = 0$ at x^μ. Therefore in the local inertial frame

$$\begin{aligned}
R_{\mu\nu\sigma\rho} &= g_{\mu\lambda}\left(\partial_\sigma\Gamma^\lambda_{\nu\rho} - \partial_\rho\Gamma^\lambda_{\nu\sigma} + \Gamma^\lambda_{\kappa\sigma}\Gamma^\kappa_{\nu\rho} - \Gamma^\lambda_{\kappa\sigma}\Gamma^\kappa_{\nu\rho}\right) \\
&= \eta_{\mu\lambda}\left(\frac{1}{2}\eta^{\lambda\kappa}\partial_\sigma\left(\partial_\nu g_{\kappa\rho} + \cancel{\partial_\rho g_{\kappa\nu}} - \partial_\kappa g_{\nu\rho}\right) - \frac{1}{2}\eta^{\lambda\kappa}\partial_\rho\left(\partial_\nu g_{\kappa\sigma} + \cancel{\partial_\sigma g_{\kappa\nu}} - \partial_\kappa g_{\nu\sigma}\right)\right) \\
&= \frac{1}{2}\left(\partial_\sigma\partial_\nu g_{\mu\rho} - \partial_\sigma\partial_\mu g_{\nu\rho} - \partial_\rho\partial_\nu g_{\mu\sigma} + \partial_\rho\partial_\mu g_{\nu\sigma}\right) \qquad (5.1.9)
\end{aligned}$$

We next notice that for this frame $R_{\mu\nu\sigma\rho} = -R_{\nu\mu\sigma\rho}$ and $R_{\mu\nu\sigma\rho} = R_{\sigma\rho\mu\nu}$. However, if a tensor is symmetric (antisymmetric) under interchange of indices in one frame then it is symmetric (antisymmetric) in all frames. It is also straightforward to check that in (5.1.9)

$$R_{\mu\nu\sigma\rho} + R_{\mu\rho\nu\sigma} + R_{\mu\sigma\rho\nu} = 0\,, \qquad (5.1.10)$$

and so this too is true in all frames. Finally, using $R_{\mu\nu\sigma\rho;\lambda} \equiv \nabla_\lambda R_{\mu\nu\sigma\rho} = \partial_\lambda R_{\mu\nu\sigma\rho}$ for the local inertial frame, we find

$$\begin{aligned}
R_{\mu\nu\sigma\rho;\lambda} + R_{\mu\nu\lambda\sigma;\rho} + R_{\mu\nu\rho\lambda;\sigma} &= \frac{1}{2}\partial_\lambda\left(\partial_\sigma\partial_\nu g_{\mu\rho} - \partial_\sigma\partial_\mu g_{\nu\rho} - \partial_\rho\partial_\nu g_{\mu\sigma} + \partial_\rho\partial_\mu g_{\nu\sigma}\right) \\
&\quad + \frac{1}{2}\partial_\rho\left(\partial_\lambda\partial_\nu g_{\mu\sigma} - \partial_\lambda\partial_\mu g_{\nu\sigma} - \partial_\sigma\partial_\nu g_{\mu\lambda} + \partial_\sigma\partial_\mu g_{\nu\lambda}\right) \\
&\quad + \frac{1}{2}\partial_\sigma\left(\partial_\rho\partial_\nu g_{\mu\lambda} - \partial_\rho\partial_\mu g_{\nu\lambda} - \partial_\lambda\partial_\nu g_{\mu\rho} + \partial_\lambda\partial_\mu g_{\nu\rho}\right) = 0\,,
\end{aligned} \qquad (5.1.11)$$

where one sees that the terms pair up (by color) and cancel each other. This last relation is called the *Bianchi identity*. Actually, there is another relatively simple way of establishing the Bianchi identity that does not use the local inertial frame. Using $(\nabla_\lambda R^\mu{}_{\nu\sigma\rho})A^\nu = \nabla_\lambda(R^\mu{}_{\nu\sigma\rho}A^\nu) - R^\mu{}_{\nu\sigma\rho}\nabla_\lambda A^\nu$ and (5.1.6) we find

$$\begin{aligned}
(R^\mu{}_{\nu\sigma\rho;\lambda} + R^\mu{}_{\nu\lambda\sigma;\rho} + R^\mu{}_{\nu\rho\lambda;\sigma})A^\nu &= [\nabla_\lambda,[\nabla_\sigma,\nabla_\rho]]A^\mu + [\nabla_\rho,[\nabla_\lambda,\nabla_\sigma]]A^\mu + [\nabla_\sigma,[\nabla_\rho,\nabla_\lambda]]A^\mu \\
&= 0 \qquad (5.1.12)
\end{aligned}$$

by the Jacobi identity:

$$[A,[B,C]] + [C,[A,B]] + [B,[C,A]] = 0\,. \qquad (5.1.13)$$

5.1.3 The Ricci tensor, Ricci scalar and Einstein tensor

From the Riemann tensor $R^{\mu}{}_{\nu\sigma\rho}$ we can construct other tensors. The Ricci tensor $R_{\mu\nu}$ is defined as

$$R_{\mu\nu} \equiv R^{\lambda}{}_{\mu\lambda\nu}\,, \tag{5.1.14}$$

and the Ricci scalar R is

$$R \equiv R^{\mu}{}_{\mu}\,. \tag{5.1.15}$$

From these two tensors we can build another tensor $G_{\mu\nu}$ called the *Einstein tensor*,

$$G_{\mu\nu} \equiv R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R\,. \tag{5.1.16}$$

The Einstein tensor satisfies an extremely important property, namely

$$G^{\mu}{}_{\nu;\mu} \equiv \nabla_\mu G^{\mu}{}_{\nu} = 0\,. \tag{5.1.17}$$

Let's now show this:

$$\begin{aligned}
\nabla_\mu G^{\mu}{}_{\nu} &= \nabla_\mu R^{\mu}{}_{\nu} - \tfrac{1}{2}\nabla_\nu R \\
&= \nabla_\mu R^{\lambda\mu}{}_{\lambda\nu} - \tfrac{1}{2}\nabla_\nu R^{\mu\lambda}{}_{\mu\lambda} \\
&= \nabla_\mu R^{\lambda\mu}{}_{\lambda\nu} - \tfrac{1}{2}\left(-\nabla_\lambda R^{\mu\lambda}{}_{\nu\mu} - \nabla_\mu R^{\mu\lambda}{}_{\lambda\nu}\right) && \text{Bianchi Identity} \\
&= \nabla_\mu R^{\lambda\mu}{}_{\lambda\nu} - \tfrac{1}{2}\left(\nabla_\mu R^{\lambda\mu}{}_{\lambda\nu} + \nabla_\mu R^{\lambda\mu}{}_{\lambda\nu}\right) && \text{Relabeling, anti-symmetry} \\
&= 0\,.
\end{aligned} \tag{5.1.18}$$

5.1.4 Independent components

The Riemann tensor comes with 4 indices, so if there are D spatial (or space-time) dimensions, then there are D^4 different ways of choosing the indices. However, because of the symmetries or anti-symmetries, the number of independent components is actually less. We know that $R_{\mu\nu\sigma\rho} = -R_{\nu\mu\sigma\rho} = -R_{\mu\nu\rho\sigma}$, so that there are only $D(D-1)/2$ independent choices for the first two indices and $D(D-1)/2$ independent choices for the last two indices. Moreover, $R_{\mu\nu\sigma\rho} = R_{\sigma\rho\mu\nu}$ so we have to symmetrize over the two sets of $D(D-1)/2$ choices, which gives us $\frac{1}{2}\frac{D(D-1)}{2}(\frac{D(D-1)}{2}+1)$ independent choices. Lastly, we have the relation in (5.1.10) which reduces the number of independent components by 1 for every choice of 4 different indices (if any two indices are the same then one can see that (5.1.10) follows from the other symmetries). For example, consider R_{abcd} where all indices are different. Then

$$R_{abcd} = -R_{acdb} - R_{adbc}\,. \tag{5.1.19}$$

There is no extra relation for, say, R_{bcda} since by the other symmetries we can write this as

$$R_{bcda} = R_{dabc} = -R_{adbc} \tag{5.1.20}$$

which appeared in our relation for R_{abcd}. Hence the number of independent components of the Riemann tensor is reduced by the number of different ways of choosing 4 different indices, that is $\binom{D}{4}$, and so the number of independent components is

$$\begin{aligned} \frac{D(D-1)(D(D-1)+2)}{8} &- \frac{D(D-1)(D-2)(D-3)}{24} \\ &= \frac{1}{12}D^2(D-1)(D+1)\,. \end{aligned} \tag{5.1.21}$$

For $D = 2$ we see there is 1, for $D = 3$ there are 6, for $D = 4$ there are 20, etc.

For the Ricci tensor $R_{\mu\nu}$ and the Einstein tensor $G_{\mu\nu}$ the number of independent components for each is $D(D+1)/2$, since they are symmetric in the indices[1].

5.2 Examples

5.2.1 A round two-dimensional sphere

In this case the metric is

$$ds^2 = R^2\, d\theta^2 + R^2 \sin^2\theta\, d\phi^2\,, \tag{5.2.1}$$

and so the metric components are $g_{\theta\theta} = R^2$, $g_{\phi\phi} = R^2\sin^2\theta$. Since the metric is diagonal, the inverse metric has

$$g^{\theta\theta} = \frac{1}{g_{\theta\theta}} = \frac{1}{R^2} \qquad g^{\phi\phi} = \frac{1}{g_{\phi\phi}} = \frac{1}{R^2\sin^2\theta}\,. \tag{5.2.2}$$

The nonzero Christoffel symbols are

$$\begin{aligned} \Gamma^\theta_{\phi\phi} &= \tfrac{1}{2}g^{\theta\theta}(-\partial_\theta g_{\phi\phi}) = \frac{1}{2R^2}(-2R^2\sin\theta\cos\theta) = -\sin\theta\cos\theta \\ \Gamma^\phi_{\theta\phi} &= \tfrac{1}{2}g^{\phi\phi}(\partial_\theta g_{\phi\phi}) = \frac{1}{2R^2\sin^2\theta}(2R^2\sin\theta\cos\theta) = \frac{\cos\theta}{\sin\theta} = \cot\theta\,. \end{aligned} \tag{5.2.3}$$

Since the space is two-dimensional, the Riemann tensor has only one independent component $R_{\theta\phi\theta\phi}$. All other possibilities are either zero or related to this one by symmetries. We then have

$$\begin{aligned} R_{\theta\phi\theta\phi} &= g_{\theta\theta}R^\theta{}_{\phi\theta\phi} = g_{\theta\theta}\left(\partial_\theta\Gamma^\theta_{\phi\phi} - \partial_\phi\Gamma^\theta_{\phi\theta} + \Gamma^\theta_{\mu\theta}\Gamma^\mu_{\phi\phi} - \Gamma^\theta_{\mu\phi}\Gamma^\mu_{\phi\theta}\right) \\ &= R^2\left((\sin^2\theta - \cos^2\theta) - 0 + 0 - (-\sin\theta\cos\theta)(\cot\theta)\right) = R^2\sin^2\theta\,, \end{aligned} \tag{5.2.4}$$

where the last term has a contribution if $\mu = \phi$. The nonzero components of the Ricci tensor are

$$R_{\theta\theta} = g^{\phi\phi}R_{\theta\phi\theta\phi} = 1\,, \qquad R_{\phi\phi} = g^{\theta\theta}R_{\theta\phi\theta\phi} = \sin^2\theta\,, \tag{5.2.5}$$

[1] If $D = 2$ then the components of the Ricci tensor are related to each other by factors of the metric since there is only one independent component for the Riemann tensor.

while the Ricci scalar is

$$R = g^{\theta\theta} R_{\theta\theta} + g^{\phi\phi} R_{\phi\phi} = \frac{1}{R^2} + \frac{1}{R^2 \sin^2\theta} \sin^2\theta = \frac{2}{R^2}\,. \tag{5.2.6}$$

The Ricci scalar is an invariant, thus it gives us information about the curvature that is independent of the frame. Firstly, we see that R is constant so the curvature is the same everywhere. This seems sensible since a sphere appears equally curved everywhere. Second, we see that the curvature falls off with R. This again seems sensible; the bigger the radius of the sphere the less curved it is (the earth looks pretty flat over distances on orders of meters, but a basketball does not).

5.2.2 Weak gravitational field

In the last chapter we considered the metric

$$ds^2 = -(1 + 2\Phi(\underset{\rightarrow}{x}))dt^2 + d^2x^i\,, \tag{5.2.7}$$

where $|\Phi(\underset{\rightarrow}{x})| << 1$, which was relevant for nonrelativistic free-fall in a weak gravitational field. Let us modify the metric by including small corrections to the spatial part of the metric as well. Assuming small $|\Phi|$ we will consider a metric of the form

$$ds^2 = -(1 + 2\Phi(\underset{\rightarrow}{x}))dt^2 + (1 - 2\,K\,\Phi(\underset{\rightarrow}{x}))d^2x^i\,, \tag{5.2.8}$$

where K is a dimensionless constant that we will leave undetermined for now. The metric is diagonal and so the relevant Christoffel symbols to linear order in Φ are

$$\begin{aligned}
\Gamma^i_{00} &= -\frac{1}{2} g^{ij}\partial_j g_{00} = \partial_i \Phi(\underset{\rightarrow}{x})\,,\\
\Gamma^0_{i0} &= \frac{1}{2} g^{00}\partial_i g_{00} \approx \partial_i \Phi(\underset{\rightarrow}{x})\\
\Gamma^i_{jk} &= \frac{1}{2} g^{il}\left(\partial_j g_{lk} + \partial_k g_{lj} - \partial_l g_{jk}\right) \approx -K\left(\delta_{ik}\partial_j\Phi(\underset{\rightarrow}{x}) + \delta_{ij}\partial_k\Phi(\underset{\rightarrow}{x}) - \delta_{jk}\partial_i\Phi(\underset{\rightarrow}{x})\right)\,.
\end{aligned} \tag{5.2.9}$$

Hence the relevant curvature tensor components are

$$\begin{aligned}
R_{i0j0} &= g_{ik}\left(\partial_j\Gamma^k_{00} - \partial_0\Gamma^k_{0j} + \Gamma^k_{\mu 0}\Gamma^\mu_{0j} - \Gamma^k_{\mu j}\Gamma^\mu_{00}\right)\\
&\approx \partial_i\partial_j\Phi(\underset{\rightarrow}{x})\\
R_{ijkl} &= g_{im}\left(\partial_k\Gamma^m_{jl} - \partial_l\Gamma^m_{jk} + \Gamma^m_{\mu k}\Gamma^\mu_{jl} - \Gamma^m_{\mu l}\Gamma^\mu_{jk}\right)\\
&\approx K\left(\delta_{jl}\partial_i\partial_k\Phi(\underset{\rightarrow}{x}) - \delta_{jk}\partial_i\partial_l\Phi(\underset{\rightarrow}{x}) - \delta_{il}\partial_j\partial_k\Phi(\underset{\rightarrow}{x}) + \delta_{ik}\partial_j\partial_l\Phi(\underset{\rightarrow}{x})\right)\,,
\end{aligned} \tag{5.2.10}$$

where we only kept terms linear in $\Phi(\underset{\rightarrow}{x})$. The components of the Ricci tensor that are linear in Φ are

$$\begin{aligned}
R_{00} &= g^{ij}R_{i0j0} = \nabla^2\Phi(\underset{\rightarrow}{x})\,,\\
R_{ij} &= g^{00}R_{0i0j} + g^{kl}R_{kilj} = -\partial_i\partial_j\Phi(\underset{\rightarrow}{x}) + K\left(\delta_{ij}\nabla^2\Phi(\underset{\rightarrow}{x}) + \partial_i\partial_j\Phi(\underset{\rightarrow}{x})\right)\,,
\end{aligned} \tag{5.2.11}$$

while the Ricci scalar is

$$\begin{aligned} R &= g^{00}R_{00} + g^{ij}R_{ij} = -\nabla^2\Phi(\underset{\rightarrow}{x}) - \nabla^2\Phi(\underset{\rightarrow}{x}) + K\left(3\nabla^2\Phi(\underset{\rightarrow}{x}) + \nabla^2\Phi(\underset{\rightarrow}{x})\right) \\ &= (4K-2)\nabla^2\Phi(\underset{\rightarrow}{x})\,. \end{aligned} \tag{5.2.12}$$

By examining free-fall equations we had previously identified $\Phi(\underset{\rightarrow}{x})$ with the gravitational potential and so $-\vec{\nabla}\Phi(\underset{\rightarrow}{x})$ is proportional to the gravitational force. Here we see that it is not so much the gravitational force, but its divergence, $\vec{\nabla}\cdot(-\vec{\nabla}\Phi(\underset{\rightarrow}{x}))$, that leads to the curvature of space-time.

Finally, for the Einstein tensor we find

$$\begin{aligned} G_{00} &= R_{00} - \tfrac{1}{2}g_{00}R = 2K\,\nabla^2\Phi(\underset{\rightarrow}{x}) \\ G_{ij} &= R_{ij} - \tfrac{1}{2}g_{ij}R = (K-1)\partial_i\partial_j\Phi(\underset{\rightarrow}{x}) + (1-K)\delta_{ij}\nabla^2\Phi(\underset{\rightarrow}{x})\,. \end{aligned} \tag{5.2.13}$$

We will come back to these results to complete the derivation of Einstein's equations.

5.2.3 The Schwarzschild metric

The Schwarzschild metric is given by

$$ds^2 = -\left(1 - \frac{2MG}{r}\right)dt^2 + \left(1 - \frac{2MG}{r}\right)^{-1}dr^2 + r^2\,d\theta^2 + r^2\sin^2\theta\,d\phi^2\,, \tag{5.2.14}$$

where G is Newton's gravitational constant, $G \approx 6.674\times 10^{-11}\,\mathrm{m}^3\,\mathrm{kg}^{-1}\mathrm{s}^{-2} \approx 7.42\times 10^{-28}\,\mathrm{m}\,\mathrm{kg}^{-1}$ and M is a mass. Later we will see that this is a solution to Einstein's equations for $r > r_0$, where r_0 is the radius of a spherically symmetric static object with mass M. If the mass is entirely inside $r = 2MG$ then we will later show that this describes a black hole.

The nonzero metric components are

$$g_{tt} = -\left(1 - \frac{2MG}{r}\right)\,,\quad g_{rr} = \left(1 - \frac{2MG}{r}\right)^{-1}\,,\quad g_{\theta\theta} = r^2 \quad g_{\phi\phi} = r^2\sin^2\theta\,. \tag{5.2.15}$$

Since the metric is diagonal, the inverse metric components are

$$\begin{aligned} g^{tt} &= \frac{1}{g_{tt}} = -\left(1 - \frac{2MG}{r}\right)^{-1}\,, \qquad g^{rr} = \frac{1}{g_{rr}} = \left(1 - \frac{2MG}{r}\right)\,, \\ g^{\theta\theta} &= \frac{1}{g^{\theta\theta}} = \frac{1}{r^2}\,, \qquad g^{\phi\phi} = \frac{1}{g_{\phi\phi}} = \frac{1}{r^2\sin^2\theta}\,. \end{aligned} \tag{5.2.16}$$

The Christoffel symbols $\Gamma^{\theta}_{\phi\phi}$ and $\Gamma^{\phi}_{\theta\phi}$ are the same as in (5.2.3). The other nonzero symbols are (using the rules for diagonal metrics)

$$\begin{aligned}
\Gamma^r_{tt} &= -\tfrac{1}{2}g^{rr}\partial_r g_{tt} = \frac{MG}{r^2}\left(1-\frac{2MG}{r}\right), \qquad \Gamma^t_{rt} = \tfrac{1}{2}g^{tt}\partial_r g_{tt} = \frac{MG}{r^2}\left(1-\frac{2MG}{r}\right)^{-1}, \\
&\qquad \Gamma^r_{rr} = \tfrac{1}{2}g^{rr}\partial_r g_{rr} = -\frac{MG}{r^2}\left(1-\frac{2MG}{r}\right)^{-1} \\
\Gamma^r_{\theta\theta} &= -\tfrac{1}{2}g^{rr}\partial_r g_{\theta\theta} = -r\left(1-\frac{2MG}{r}\right) \qquad \Gamma^\theta_{r\theta} = \tfrac{1}{2}g^{\theta\theta}\partial_r g_{\theta\theta} = \frac{1}{r} \\
\Gamma^r_{\phi\phi} &= -\tfrac{1}{2}g^{rr}\partial_r g_{\phi\phi} = -r\sin^2\theta\left(1-\frac{2MG}{r}\right) \qquad \Gamma^\phi_{r\phi} = \tfrac{1}{2}g^{\phi\phi}\partial_r g_{\phi\phi} = \frac{1}{r}\,. \qquad (5.2.17)
\end{aligned}$$

The components of the Riemann tensor are unfortunately rather tedious to compute. Here we list and compute the independent nonzero components:

$$\begin{aligned}
R_{rtrt} &= g_{rr}\left(\partial_r\Gamma^r_{tt} - \cancel{\partial_t\Gamma^r_{rt}} + \Gamma^r_{\mu r}\Gamma^\mu_{tt} - \Gamma^r_{\mu t}\Gamma^\mu_{tr}\right) = g_{rr}\left(\partial_r\Gamma^r_{tt} + \Gamma^r_{rr}\Gamma^r_{tt} - \Gamma^r_{tt}\Gamma^t_{tr}\right) \\
&= \left(1-\frac{2MG}{r}\right)^{-1}\left(-2\frac{MG}{r^3}\left(1-\frac{2MG}{r}\right) + 2\left(\frac{MG}{r^2}\right)^2 - \left(\frac{MG}{r^2}\right)^2 - \left(\frac{MG}{r^2}\right)^2\right) \\
&= -2\frac{MG}{r^3} = 2\frac{MG}{r^3}g_{rr}g_{tt}\,, \\
R_{r\theta r\theta} &= g_{rr}\left(\partial_r\Gamma^r_{\theta\theta} - \cancel{\partial_\theta\Gamma^r_{r\theta}} + \Gamma^r_{\mu r}\Gamma^\mu_{\theta\theta} - \Gamma^r_{\mu\theta}\Gamma^\mu_{\theta r}\right) = g_{rr}\left(\partial_r\Gamma^r_{\theta\theta} + \Gamma^r_{rr}\Gamma^r_{\theta\theta} - \Gamma^r_{\theta\theta}\Gamma^\theta_{\theta r}\right) \\
&= g_{rr}\left(-1 + \frac{MG}{r} + \left(1-\frac{2MG}{r}\right)\right) = -\frac{MG}{r}\left(1-\frac{2MG}{r}\right)^{-1} = -\frac{MG}{r^3}g_{rr}g_{\theta\theta} \\
R_{r\phi r\phi} &= g_{rr}\left(\partial_r\Gamma^r_{\phi\phi} - \cancel{\partial_\phi\Gamma^r_{r\phi}} + \Gamma^r_{\mu r}\Gamma^\mu_{\phi\phi} - \Gamma^r_{\mu\phi}\Gamma^\mu_{\phi r}\right) = g_{rr}\left(\partial_r\Gamma^r_{\phi\phi} + \Gamma^r_{rr}\Gamma^r_{\phi\phi} - \Gamma^r_{\phi\phi}\Gamma^\phi_{\phi r}\right) \\
&= g_{rr}\left(-1 + \frac{MG}{r} + \left(1-\frac{2MG}{r}\right)\right)\sin^2\theta = -\frac{MG}{r^3}g_{rr}g_{\phi\phi} \\
R_{t\theta t\theta} &= g_{tt}\left(\cancel{\partial_t\Gamma^t_{\theta\theta}} - \cancel{\partial_\theta\Gamma^t_{t\theta}} + \Gamma^t_{\mu t}\Gamma^\mu_{\theta\theta} - \cancel{\Gamma^t_{\mu\theta}\Gamma^\mu_{\theta t}}\right) = g_{tt}\Gamma^t_{rt}\Gamma^r_{\theta\theta} = -\frac{MG}{r^3}g_{tt}g_{\theta\theta} \\
R_{t\phi t\phi} &= g_{tt}\left(\cancel{\partial_t\Gamma^t_{\phi\phi}} - \cancel{\partial_\phi\Gamma^t_{t\phi}} + \Gamma^t_{\mu t}\Gamma^\mu_{\phi\phi} - \cancel{\Gamma^t_{\mu\phi}\Gamma^\mu_{\phi t}}\right) = g_{tt}\Gamma^t_{rt}\Gamma^r_{\phi\phi} = -\frac{MG}{r^3}g_{tt}g_{\phi\phi} \\
R_{\theta\phi\theta\phi} &= r^2\sin^2\theta + g_{\theta\theta}\Gamma^\theta_{r\theta}\Gamma^r_{\phi\phi} = r^2\sin^2\theta - r^2\sin^2\theta\left(1-\frac{2MG}{r}\right) = 2\frac{MG}{r^3}g_{\theta\theta}g_{\phi\phi}\,, \qquad (5.2.18)
\end{aligned}$$

where in the last line we took the two-dimensional sphere result for $R_{\theta\phi\theta\phi}$, replacing R with r, and added an extra term that must be included because r, unlike R, is not held fixed. Notice that the final result (blue) for each term has a very similar form, with the indices on the metric matching the indices on the Riemann tensor.

Because $\Gamma^r_{\phi\phi}$ has both r and θ dependence, the component $R_{r\phi\theta\phi}$ is not obviously zero; let's check that it is

$$\begin{aligned}
R_{r\phi\theta\phi} &= g_{rr}\left(\partial_\theta\Gamma^r_{\phi\phi} - \cancel{\partial_\phi\Gamma^r_{\theta\phi}} + \Gamma^r_{\mu\theta}\Gamma^\mu_{\phi\phi} - \Gamma^r_{\mu\phi}\Gamma^\mu_{\phi\theta}\right) = g_{rr}\left(\partial_\theta\Gamma^r_{\phi\phi} + \Gamma^r_{\theta\theta}\Gamma^\theta_{\phi\phi} - \Gamma^r_{\phi\phi}\Gamma^\phi_{\phi\theta}\right) \\
&= g_{rr}\left(-2\sin\theta\cos\theta + \sin\theta\cos\theta + \sin^2\theta\cot\theta\right)\left(1-\frac{2MG}{r}\right) = 0\,. \qquad (5.2.19)
\end{aligned}$$

We can now take the nonzero components and compute the Ricci tensor. Because the metric is diagonal and because each nonzero component of the Riemann tensor has two repeated indices, the Ricci tensor will also be diagonal. We thus find

$$\begin{aligned}
R_{tt} &= g^{rr}\,R_{rtrt} + g^{\theta\theta}\,R_{\theta t\theta t} + g^{\phi\phi}\,R_{\phi t\phi t} = 2\,\frac{MG}{r^3}g_{tt} - \frac{MG}{r^3}g_{tt} - \frac{MG}{r^3}g_{tt} = 0\,,\\
R_{rr} &= g^{tt}\,R_{trtr} + g^{\theta\theta}\,R_{\theta r\theta r} + g^{\phi\phi}\,R_{\phi t\phi t} = 2\,\frac{MG}{r^3}g_{rr} - \frac{MG}{r^3}g_{rr} - \frac{MG}{r^3}g_{rr} = 0\,,\\
R_{\theta\theta} &= g^{tt}\,R_{t\theta t\theta} + g^{rr}\,R_{r\theta r\theta} + g^{\phi\phi}\,R_{\phi\theta\phi\theta} = -\frac{MG}{r^3}g_{\theta\theta} - \frac{MG}{r^3}g_{\theta\theta} + 2\,\frac{MG}{r^3}g_{\theta\theta} = 0\,,\\
R_{\phi\phi} &= g^{tt}\,R_{t\phi t\phi} + g^{rr}\,R_{r\phi r\phi} + g^{\theta\theta}\,R_{\theta\phi\theta\phi} = -\frac{MG}{r^3}g_{\phi\phi} - \frac{MG}{r^3}g_{\phi\phi} + 2\,\frac{MG}{r^3}g_{\phi\phi} = 0\,.
\end{aligned} \tag{5.2.20}$$

Hence all components of the Ricci tensor are zero! Obviously, the Ricci scalar $R = R^{\mu}{}_{\mu}$ is also zero, and so too is the Einstein tensor $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}\,R$.

It might seem that because $R = 0$ there is no invariant information about the curvature. However, we can also construct an invariant by squaring two Riemann tensors and contracting indices. This gives

$$\begin{aligned}
R^{\mu\nu\sigma\rho}R_{\mu\nu\sigma\rho} &= 4\sum_{a<b} R^{abab}R_{abab} = 4\sum_{a<b}\left(g^{aa}g^{bb}R_{abab}\right)^2\\
&= 4(4+1+1+1+1+4)\left(\frac{MG}{r^3}\right)^2 = 48\,\frac{M^2G^2}{r^6}\,.
\end{aligned} \tag{5.2.21}$$

We have ordered the indices a and b in the sum as t first, r second, θ third and ϕ fourth. As you can see, the invariant is blowing up as $r \to 0$, signaling a real singularity where the space-time is becoming infinitely curved. However, there is no singularity at $r = 2MG$ where the metric is becoming singular. This point is only a coordinate singularity. As $r \to \infty$ the invariant is going to zero, indicating the space-time is becoming flat. This is clear from the metric where in the $r \to \infty$ limit the metric approaches flat Minkowski space.

Chapter 6

The energy-momentum tensor

In this chapter we consider the energy-momentum tensor. We discuss its form in various instances.

6.1 Construction and properties

In the previous two chapters we introduced the tools necessary to describe the geometry of space-time. This will lead to one side of Einstein's equations. On the other side of the equations we need the part that acts as a source for this geometry. This is like electrodynamics, where half of Maxwell's equations can be written as

$$\partial_\mu F^{\mu\nu} = j^\nu \,, \tag{6.1.1}$$

where $F^{\mu\nu}$ contains the field strengths $\vec{E}$ and $\vec{B}$, with $E_i = F_{0i}$ and $B_i = \frac{1}{2}\varepsilon_{ijk}F_{jk}$. The right hand side has the sources, where j^0 is the charge density ρ and j^i are the components of the current. An important feature in these equations is current conservation, $\partial_\nu j^\nu = 0$, which the field equations respect because of the antisymmetry of $F_{\mu\nu}$.

Hence for gravity, we are looking for the analog of j^ν. To this end, let us suppose we are in flat Minkowski space and that we have a uniform distribution of static "dust", where dust can mean any nonrelativistic matter, including hydrogen molecules. In fact, let us assume that the dust *is* made up of hydrogen molecules that are nonrelativistic and so for all intents and purposes can be assumed to be at rest. Each molecule has a rest mass of m and a velocity 4-vector $\vec{U}$ with components $\vec{U} = (1,0,0,0)$ in the rest frame. The momentum 4-vector for each molecule is $\vec{p} = m\,\vec{U}$.

Now let n be the number density of the dust in the dust's rest frame, that is the number of hydrogen molecules per unit volume. We can then construct a number 4-current, $\vec{N} = n\,\vec{U}$, whose spatial components give the rate that the molecules pass through a unit area per unit time, while the number density is $N^0 = n\,U^0$. In the rest frame all spatial components are zero and the number density is $N^0 = n$. If we boost to a new frame $\mathbf{S}'$ which is moving with velocity $v\,\hat{x}$ wrt the rest frame, then the current in this frame is $\vec{N} = n(\gamma, -v\gamma, 0, 0)$. We can easily understand the different factors here. Because of length contraction in the $\hat{x}$ direction, the density will go up by a factor of γ.

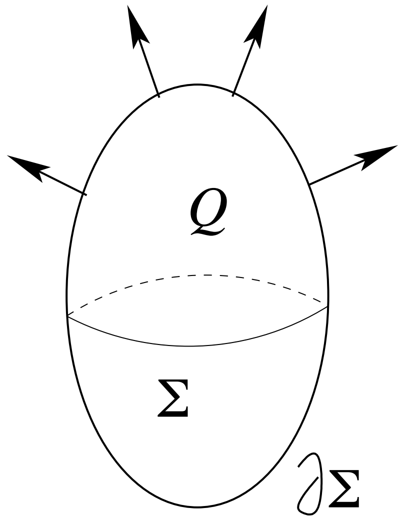

Figure 6.1: The figure shows a charge Q inside a volume Σ. The rate of change of Q with respect to time is the negative of the current flux through the Σ's boundary $\partial\Sigma$.

Also, there is now a current component in the $\hat{x}$ direction because the dust is moving backward with velocity $-v$ in $\mathbf{S}'$. The current is then $-v$ times the density.

The components of the 4-momentum for each molecule in $\mathbf{S}'$ are $\vec{p} = m(\gamma, -v\gamma, 0, 0)$. The energy of each molecule is $m\,U^0$ and so the energy density is $\mathcal{E} = mn\,U^0U^0$. Since this has two Lorentz time-like indices, this suggests that $\mathcal{E} = T^{00}$ where $T^{\mu\nu}$ are the components of a $\binom{2}{0}$ tensor called the *energy-momentum tensor*. In the case of the uniform dust, the full energy momentum tensor is given by $T^{\mu\nu} = mn\,U^\mu U^\nu$. Clearly, $T^{\mu\nu} = T^{\nu\mu}$. We can think of the T^{0i} components as the energy current in the i direction. But we can also think of it as the density of the i^{th} component of the momentum. Either interpretation is consistent, since if there is an energy current then the dust must have some momentum to move the energy from one place to another. The components T^{ij} correspond to the current in the j^{th} direction for the i^{th} component of the momentum, and *vice versa*.

By construction $T^{\mu\nu}$ is symmetric. Since it is a tensor it will be symmetric in any frame. Even if no global inertial frame is present, at every point there exists a local inertial frame where $T^{\mu\nu}$ can be expressed locally in terms of the product of velocity vectors and so $T^{\mu\nu}$ is symmetric in this case as well.

The energy-momentum tensor satisfies another crucial property. It is a conserved current. Let us review current conservation in electrodynamics. The current conservation equation $\partial_\nu j^\nu = 0$ can be rewritten as

$$\frac{\partial}{\partial t}\rho + \vec{\nabla}\cdot\vec{j} = 0\,. \tag{6.1.2}$$

To put this equation into words, it tells us that net electric charge is conserved, it is not created or destroyed out of thin air. If the charge density is changing at one particular point, it must be because charge is flowing into or away from that point. To see this a little differently, consider the situation as shown in figure 6.1. Here we have a charge Q

in a certain volume Σ and with a boundary $\partial\Sigma$. Q is found by integrating the charge density ρ over Σ,

$$Q = \iiint_{\Sigma} d^3x\,\rho\,. \tag{6.1.3}$$

The rate at which the charge in Σ is changing is

$$\frac{d}{dt}Q = \iiint_{\Sigma} d^3x\,\frac{\partial}{\partial t}\rho = -\int_{\Sigma} d^3x\,\vec{\nabla}\cdot\vec{j}\,, \tag{6.1.4}$$

where the last step uses the current conservation in (6.1.2). Then by Gauss' law it follows that the rate at which the charge in Σ is changing is

$$\frac{d}{dt}Q = -\iint_{\partial\Sigma} d\vec{S}\cdot\vec{j}\,, \tag{6.1.5}$$

where the last integral is the negative of the current flux on the boundary of Σ. What we learn from this is that if the charge Q is decreasing in Σ it is because charge is flowing out of the boundary. The larger the rate of change for Q, the larger the rate it flows out.

We can now do the same with energy and momentum. Instead of the charge density ρ we can consider the energy density $\mathcal{E} = T^{00}$ and instead of charge currents j^i we consider energy currents T^{i0}. Then if energy is conserved we must have the same sort of current conservation equation, $\partial_\mu T^{\mu 0} = 0$. We can do the same thing with each component of momentum so we also have $\partial_\mu T^{\mu i} = 0$, hence we have the four equations $\partial_\mu T^{\mu\nu} = 0$. These equations are valid for inertial frames, but the derivatives have to be replaced by covariant derivatives so that we have a tensor equation that applies for any frame. Hence our full set of current conservation equations are

$$T^{\mu\nu}{}_{;\mu} \equiv \nabla_\mu T^{\mu\nu} = 0\,. \tag{6.1.6}$$

As we have emphasized in previous chapters, a tensor where all components are zero in one frame has all components zero in any frame.

6.2 Perfect fluids

When we consider cosmological solutions to Einstein's equations we will assume that the universe is composed of a *perfect fluid*. A perfect fluid could be made up of dust, radiation, or something else. A perfect fluid is assumed to be uniform and isotropic. Uniform means that the components of the energy-momentum tensor are the same everywhere. Isotropic means there is no preferred direction, so the fluid appears the same no matter which direction an observer looks. Because it is isotropic, the energy momentum tensor is invariant under rotations in the spatial directions. Hence, it must have the form

$$T^{\mu\nu} = \begin{pmatrix} \mathcal{E} & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix} \tag{6.2.1}$$

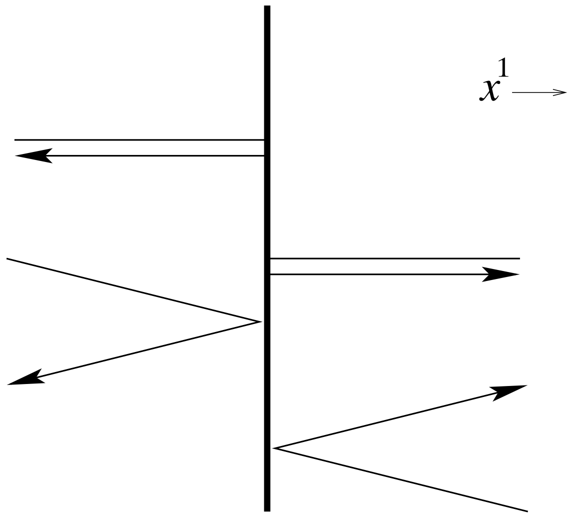

Figure 6.2: A screen inserted in a fluid. The fluid exerts a pressure on the screen.

To understand why this is so, note that we can think of T^{0i} as the three components of a spatial 3-vector. The only vector that is invariant under rotations has all of its components 0. Likewise, the components T^{ij} make up a 3 by 3 matrix. The only matrices that are invariant under general rotations are proportional to the identity matrix. Hence $T^{\mu\nu}$ must have the form in (6.2.1).

We have already discussed the energy density $\mathcal{E} = T^{00}$. What is p? The T^{ii} component is the current for the i^{th} component of momentum in the i^{th} direction. Since we are assuming that the fluid is isotropic, there cannot be any net momentum for the fluid. So any positive momentum flowing in the positive direction is cancelled by negative momentum in the negative direction, but still leaving a net current. Now suppose we insert an opaque screen into this fluid that is orthogonal to the x^1 direction (see figure 6.2.) Then each side of the screen is bombarded by the fluid and thus feels a force on it. Assuming that the fluid bounces off the screen, the impulse from one fluid particle on the negative side of the screen is $2\,p^1$, assuming $p^1 > 0$. Then the force per unit area is $2\,p^1$ times the rate per unit area (aka the *flux*) the particles that make up the fluid hit the screen. However $2\,p^1$ multiplied by the flux is T^{11}, the current in the x^1 direction for the first momentum component. The factor of 2 appears because the backward moving particles contribute equally to the current. p is then the force per unit area, which is the pressure.

6.3 Radiation as a perfect fluid

We can see that static dust in the rest frame is a perfect fluid with $p = 0$. Another example of a perfect fluid is radiation. The radiation is made up of photons and in this case for each photon we have $|\vec{p}| = p^0$. The fluid itself is made up of many photons such that the total spatial momentum is zero and the distribution of the momentum is isotropic. This means that the average value of each component is $\overline{p^i} = 0$ for a given p^0.

However, the average of the squares of the momentum components satisfy

$$(p^0)^2 = \overline{(p^1)^2 + (p^2)^2 + (p^3)^2} = \overline{(p^1)^2} + \overline{(p^2)^2} + \overline{(p^1)^3}\,. \tag{6.3.1}$$

Because the distribution of momentum is isotropic, we have $\overline{(p^1)^2} = \overline{(p^2)^2} = \overline{(p^3)^2}$, and therefore $\overline{(p^i)^2} = \frac{1}{3}\,(p^0)^2$ for each spatial component i. Also, since there is no correlation between the different components, $\overline{p^i p^j} = 0$ if $i \neq j$.

Now let $n(p^0)$ be the photon density per unit energy p^0. Then the energy density is

$$\mathcal{E} = T^{00} = \iint\!\!\Bigg(p^0\, n(p^0)\, dp^0\,. \tag{6.3.2}$$

The contribution to the photon number current in the ith direction is $\frac{p^i}{p^0}\,n(p^0)$ which averages to zero. The components of the momentum currents are then

$$T^{ij} = \int \frac{\overline{p^i p^j}}{p^0} n(p^0)\, dp^0 = \frac{1}{3}\delta_{ij} \iint\!\!\Bigg(p^0\, n(p^0) dp^0 = \frac{1}{3}\delta_{ij}\, \mathcal{E}\,. \tag{6.3.3}$$

Hence we find $p = \frac{1}{3}\,\mathcal{E}$.

For a general perfect fluid, we will write the relation between the pressure and the energy density as $p = w\,\mathcal{E}$. This is referred to as an *equation of state*. Nonrelativistic matter has $w = 0$, while radiation has $w = 1/3$.

Chapter 7

Einstein equations

In this chapter we find Einstein's equations by matching a tensor equation to the Newtonian limit.

7.1 Derivation

Einstein's equations have the schematic form

$$\text{Curvature} = \text{Sources}\,. \tag{7.1.1}$$

We will use as the source term the energy-momentum tensor $T^{\mu\nu}$. As we have just seen, $T^{\mu\nu}$ is symmetric and conserved, meaning that it satisfies ${T^{\mu\nu}}_{;\mu} = 0$. Hence for the curvature part we are looking for a tensor with the same properties so that we can have a consistent tensor equation. In fact, we have seen such a tensor, namely the Einstein tensor, $G_{\mu\nu}$. Hence

$$G_{\mu\nu} = C\,T_{\mu\nu} \tag{7.1.2}$$

is a consistent tensor equation, where C is a constant.

In order to determine C let us consider the weak field limit with gravitational potential $\Phi(\underset{\rightarrow}{x})$ and where $T^{\mu\nu}$ is comprised of nonrelativistic matter. We previously showed that the components of $G_{\mu\nu}$ are given by

$$\begin{aligned} G_{00} &= 2K\,\nabla^2\Phi(\underset{\rightarrow}{x})\,, \\ G_{ij} &= (K-1)\partial_i\partial_j\Phi(\underset{\rightarrow}{x}) + (1-K)\delta_{ij}\nabla^2\Phi(\underset{\rightarrow}{x})\,, \end{aligned} \tag{7.1.3}$$

$$\tag{7.1.4}$$

where K is a yet to be determined constant that appears in the spatial part of the metric. For the energy-momentum tensor, since the motion is nonrelativistic, the energy density $T^{00} = \mathcal{E}$ is approximately the mass density ρ. The other components of the energy momentum tensor are negligible compared to T^{00} so we can set them to zero. Hence, the relation in (7.1.2) gives the equations

$$\begin{aligned} 2K\nabla^2\Phi(\underset{\rightarrow}{x}) &\approx C\,\rho \\ (K-1)\partial_i\partial_j\Phi(\underset{\rightarrow}{x}) + (1-K)\delta_{ij}\nabla^2\Phi(\underset{\rightarrow}{x}) &\approx 0\,. \end{aligned} \tag{7.1.5}$$

Thus the second equation forces us to set $K = 1$.

C is then set by recalling the potential equation for Newtonian gravity. Given a mass density ρ, the potential satisfies the Poisson equation

$$\nabla^2\Phi = 4\pi\, G\, \rho \tag{7.1.6}$$

where G is Newton's gravitational constant. Comparing to the first equation in (7.1.5) we find that $C = 8\pi\, G$ and thus we can write the tensor equation

$$\boxed{G_{\mu\nu} = 8\pi\, G\, T_{\mu\nu}\,.} \tag{7.1.7}$$

Finally, we have derived the long anticipated Einstein equations, one of the great intellectual achievements in science, so far-reaching that we have even put a box around it.

7.2 Einstein's "greatest blunder" — Not!

Actually, Einstein realized that there is another term he could add to the left hand side of his equations that is also a symmetric $\binom{0}{2}$ tensor and whose covariant derivative when contracted with one of the indices is zero. It's the metric multiplied by a constant. Clearly the covariant derivative condition holds since $\nabla_\lambda g_{\mu\nu} = 0$. Hence, we could have written the equations as

$$G_{\mu\nu} + \Lambda\, g_{\mu\nu} = 8\pi\, G\, T_{\mu\nu}\,, \tag{7.2.1}$$

where Λ is a constant called the *cosmological constant*.

In fact Einstein thought he needed this term to explain the universe as it was understood in the early 20th century. At that time the universe was believed to be static, at least on a cosmological scale. But since the universe is also filled with matter, there should be a gravitational attraction pulling the matter inward. However, it turns out that the cosmological constant can counteract this attraction and leave the matter fixed in place.

But in the 1920's Hubble discovered that the universe is not static but is expanding. Hence the gravitational attraction is still there, slowing the expansion down, but the cosmological constant is no longer necessary. When Einstein learned of the expansion he referred to the cosmological constant term as his greatest blunder since without the term he might have been forced to predict an expanding universe before Hubble made his observations. Or, he might have thought that it was a blunder because it sullied his beautiful equations with an extra term.

However, there is another way to look at the cosmological constant. If we bring it to the other side of Einstein's equations so that we get

$$G_{\mu\nu} = 8\pi\, G\, T_{\mu\nu} - \Lambda\, g_{\mu\nu} \tag{7.2.2}$$

then we can think of it not as a correction to the curvature part of Einstein's equations, but instead as an extra contribution to $T^{\mu\nu}$. The cosmological constant contribution $T^{\mu\nu}_{cc}$

is read off from (7.2.2) to be

$$T_{cc}^{\mu\nu} = -\frac{\Lambda}{8\pi G}\, g^{\mu\nu}\,. \tag{7.2.3}$$

If we are in a space-time that is very close to flat such that $g^{\mu\nu} \approx \eta^{\mu\nu}$, then we can see that the energy density is

$$T^{00} = \frac{\Lambda}{8\pi G} \tag{7.2.4}$$

while the pressure is

$$T^{ii} = -\frac{\Lambda}{8\pi G}\,, \qquad i = 1, 2, 3 \tag{7.2.5}$$

and all off-diagonal terms of $T^{\mu\nu}$ are zero. Hence this is a perfect fluid with a rather strange equation of state, $p = -\mathcal{E}$, that is $w = -1$ and either the pressure or the energy density is negative!

But it turns out that Einstein's blunder was in fact a stroke of genius. Observations of distant supernovae over the last 15 years have confirmed that 70% of the energy density in the universe is of this cosmological constant type. The popular name for this energy source is "Dark Energy" or sometimes "Vacuum Energy", and we will explore its consequences more fully in the cosmology part of the course.

Chapter 8

The Schwarzschild solution to Einstein's equations

In this chapter we consider Schwarzschild's solutions to the Einstein equations in empty space. We use this solution to do various tests of general relativity. We will also show that the solution can also describe a black holed.

8.1 The Schwarzschild metric

The Einstein equations are

$$G_{\mu\nu} = 8\pi\, G\, T_{\mu\nu}\,. \tag{8.1.1}$$

One place to look for solutions is in empty space where $T_{\mu\nu} = 0$. In this case $G_{\mu\nu} = 0$ implies $R = 0$ and $R_{\mu\nu} = 0$. To see this consider the contraction of the indices of $G_{\mu\nu}$ (this is also called "taking the trace"),

$$0 = G^{\mu}{}_{\mu} = R^{\mu}{}_{\mu} - \tfrac{1}{2}\, g^{\mu}{}_{\mu}\, R = R - \tfrac{1}{2}\, 4\, R = -R \tag{8.1.2}$$

and so $R = 0$. The factor of 4 comes from $g^{\mu}{}_{\mu} = \delta^{\mu}{}_{\mu} = 4$ since there are 4 indices to sum over. From this, we have

$$R_{\mu\nu} = \tfrac{1}{2}\, g_{\mu\nu}\, R = 0\,. \tag{8.1.3}$$

It might seem that $R_{\mu\nu} = 0$ has only trivial solutions, but this is not true. It is possible to have nonzero Riemann tensor components $R_{\mu\nu\sigma\rho}$ but still have all components of the the Ricci tensor $R_{\mu\nu} = R^{\lambda}{}_{\mu\lambda\nu}$ be zero. In fact, we have already found such a solution. It is the Schwarzschild metric,

$$ds^2 = -\left(1 - \frac{2MG}{r}\right) dt^2 + \left(1 - \frac{2MG}{r}\right)^{-1} dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2) \tag{8.1.4}$$

where it was demonstrated in chapter 5 that all components of the Ricci tensor satisfy $R_{\mu\nu} = 0$, even though many of the Riemann tensor components are nonzero. If we examine the structure of (8.1.4) we see that g_{tt} has the form

$$g_{tt} = -(1 + 2\Phi(r))\,, \qquad \Phi(r) = -\frac{MG}{r}\,, \tag{8.1.5}$$

which is the gravitational potential outside a sperically symmetric massive object with mass M. Hence the Schwarzschild metric gives the general relativistic solution for an object like the sun.

When the gravitational field is weak, we previously showed that to leading order the metric has the form

$$ds^2 = -\,(1+2\Phi)\,dt^2 + (1-2\Phi)\,(dx^i)^2\,. \tag{8.1.6}$$

If we go to polar coordinates and assume that the $\Phi = -\frac{MG}{r}$ the metric becomes

$$ds^2 = -\left(1 - \frac{2MG}{r}\right) dt^2 + \left(1 + \frac{2MG}{r}\right)\left(dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)\right), \tag{8.1.7}$$

which is not exactly the Schwarzschild metric. However, to this order in the expansion we can let $\tilde{r} = r(1 + MG/r) = r + MG$, in which case

$$\begin{aligned} d\tilde{r} &= dr \\ (1+2MG/r) &\approx (1-2MG/r)^{-1} \\ r^2(1+2MG/r) &\approx r^2(1+MG/r)^2 = \tilde{r}^2 \\ (1-2MG/r) &\approx (1-2MG/\tilde{r})\,. \end{aligned} \tag{8.1.8}$$

The metric then becomes the Schwarzschild metric with r replaced by $\tilde{r}$. However, this is just a coordinate transformation so the physics is the same as if we had used r.

8.2 Tests of general relativity

8.2.1 Bending of light

A significant test of general relativity was made in 1919 when Eddington measured the bending of starlight around the sun during a solar eclipse. Figure 8.1 shows the bending of starlight around the sun. This bending leads to an apparent shift outward for the position of the star as seen by an observer on the earth. Notice that the incoming starlight is shifted in the perpendicular direction by a distance equal to the sun's radius R in order to reach the observer. This shift is called the *impact parameter* and is often written as b.

We can compute the bending using the Schwarzschild metric (8.1.4) with M as the mass of the sun, since the light is completely outside the solar mass and the sun is very close to being spherically symmetric. Without any loss of generality we can assume that the light is in the equatorial plane of the sun at $\theta = \pi/2$. Our goal is to compute $\Delta\phi$, the angle at which the light bends from a straight line, as it comes in from $r = \infty$, reaches a minimum distance at the sun's radius R, and then goes back out to $r = \infty$. The angular shift in the sky of the star's apparent position is also $\Delta\phi$, since the star is so much farther from the sun than an observer on earth that all light rays emanating from the star that intersect the solar system can be treated as parallel.

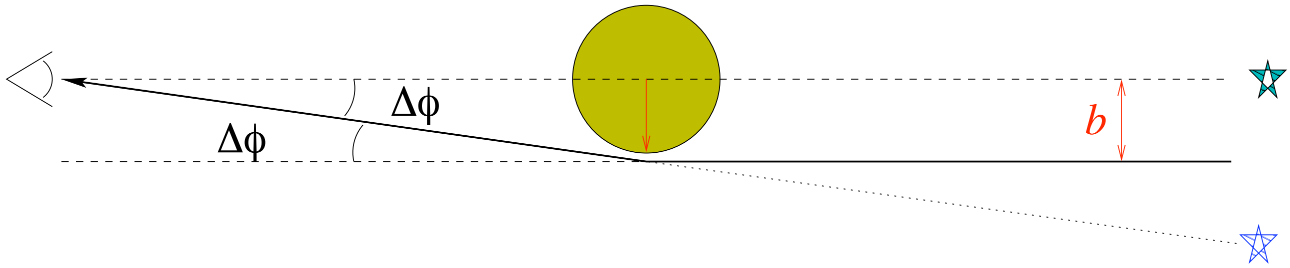

Figure 8.1: The bending of starlight from the sun. Since the star is much farther from the sun than an observer on earth, the two incoming light rays from the star (solid and dashed) can be treated as parallel. The perpendicular separation between the parallel lines is the impact parameter $b = R$, where R is the sun's radius. The apparent angular shift of the star is due to the angular deflection of the light and for an observer on earth it is the same as if there were a star emitting a light ray following the dotted line with no deflection.

Since θ is fixed at $\theta = \pi/2$, we have $\sin\theta = 1$ and $d\theta = 0$, and so along a light-like trajectory the metric satisfies

$$0 = -f(r)\,dt^2 + f^{-1}(r)\,dr^2 + r^2\,d\phi^2 \qquad f(r) = 1 - \frac{2MG}{r}\,, \tag{8.2.1}$$

which can be rewritten as

$$d\phi^2 = r^{-2}f(r)\,dt^2 - r^{-2}f^{-1}(r)\,dr^2\,. \tag{8.2.2}$$

Notice that this expression has the same form as free-fall of a massive body in a radial direction, with ϕ replacing τ and g_{tt} and g_{rr} replaced by $\tilde{g}_{tt} \equiv g_{tt}/r^2$ and $\tilde{g}_{rr} \equiv g_{rr}/r^2$ respectively. Hence, instead of $\dot{t}$ and $\dot{r}$ we will consider $t' \equiv \frac{dt}{d\phi}$ and $r' \equiv \frac{dr}{d\phi}$. Thus, since the metric is invariant under constant shifts of t, we have that $\tilde{g}_{tt}t'$ is a constant of the motion,

$$r^{-2}f(r)\,t' = C\,. \tag{8.2.3}$$

Hence, we have that

$$(r')^2 = f^2(r)(t')^2 - r^2\,f(r) = r^4\,C^2 - r^2\,f(r)\,. \tag{8.2.4}$$

The constant C is then determined by setting $r' = 0$ at the minimum $r = R$, where we find

$$C^2 = R^{-2}\,f(R)\,. \tag{8.2.5}$$

Plugging this into (8.2.4) and solving by quadratures we end up with the integral equation

$$2\int_R^\infty \frac{R\,dr}{r\sqrt{r^2\,f(R) - R^2\,f(r)}} = \int d\phi = \pi + \Delta\phi\,, \tag{8.2.6}$$

where we have the factor of 2 to account for the light coming in and then going out again. Substituting $f(r)$ into the r integral gives

$$\pi + \Delta\phi = 2\int_R^\infty \frac{R\,dr}{\sqrt{r(r-R)\,((r^2 + R\,r)f(R) - 2MGR)}}\,, \tag{8.2.7}$$

Defining the dimensionless constant $\kappa = \frac{2MG}{R}$ and the dimensionless variable $z = r/R$, we can write the integral as

$$\pi + \Delta\phi = 2\int_1^\infty \frac{dz}{\sqrt{z(z-1)(z^2+z-\kappa(z^2+z+1))}}\,. \tag{8.2.8}$$

This integral is an example of an elliptic integral and can be written in terms of special functions called elliptic functions. However, most of us are not that familiar with such functions, so instead we will make an approximation. In a previous chapter we saw that for the sun $R >> MG$ and so $\kappa << 1$. In this case, the integral (8.2.8) can be approximated by

$$\pi + \Delta\phi \approx 2\int_1^\infty \frac{dz}{z\sqrt{z^2-1}} + \kappa\int_1^\infty \frac{(z^2+z+1)dz}{z^2(z+1)\sqrt{z^2-1}} \tag{8.2.9}$$

The integrals are solved in (8.5.1) and (8.5.2) in the appendix, giving us

$$\pi + \Delta\phi \approx \pi + 2\kappa\,. \tag{8.2.10}$$

Hence we find that

$$\Delta\phi \approx 2\kappa = \frac{4GM}{R} \tag{8.2.11}$$

Let us plug in the numbers to see how big this effect is. The solar mass is $M = 1.99\times 10^{30}$ kg, while $G = 6.67\times 10^{-11}\,\mathrm{m}^3\,\mathrm{kg}^{-1}\,\mathrm{s}^{-2}$. Thus using units where $c = 1$, we have that $MG = 1.485\times 10^3$ m. The solar radius is $R = 6.96\times 10^8$ m, therefore

$$\Delta\phi = 4\,\frac{1.485\times 10^3\,\mathrm{m}}{6.96\times 10^8\,\mathrm{m}} = 8.5\times 10^{-6}\,. \tag{8.2.12}$$

This is the result in radians. In arcseconds it is

$$\Delta\phi = 8.5\times 10^{-6}\,\mathrm{rad}\,(360\,\mathrm{deg}/2\pi\,\mathrm{rad})(3600\,\mathrm{arcsec}/\mathrm{deg}) = 1.76\,\mathrm{arcsec}\,. \tag{8.2.13}$$

Eddington's expedition claimed to show this result for the bending, although there was some controversy about systematic errors in the experiment. In fact many claimed that the errors were so big that it was not possible to distinguish between the general relativistic result and the result using Newtonian gravity with special relativity (see exercise). It was not until much later that experiments were able to definitely verify the GR result.

8.2.2 The precession of the perihelion of mercury

Another famous test of Einstein's theory of general relativity involves the precession of Mercury's orbit. The orbit of Mercury around the sun is close to an ellipse, and if Mercury were alone in the solar system with the sun and if Mercury were a point-like object, then Newtonian gravity would predict an elliptical orbit, as in figure 8.2. However, Mercury is not point-like nor is it alone with the sun in the solar system, so the effect of this is for

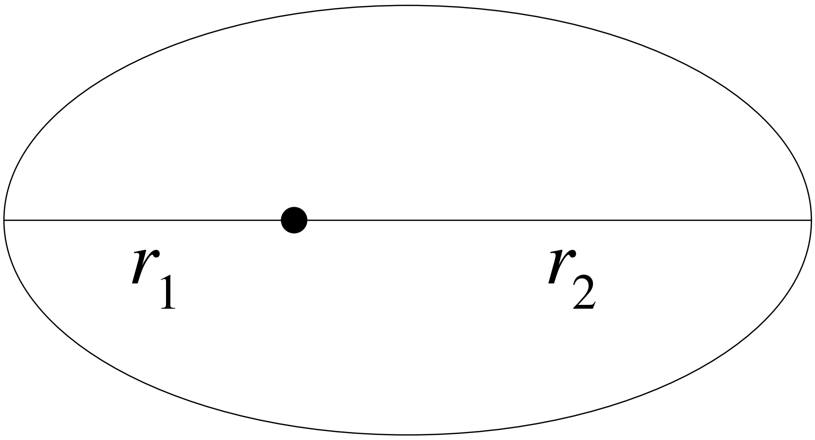

Figure 8.2: Ideal elliptical orbit for a single planet orbiting the sun. r_1 is the distance to the perihelion and r_2 to the aphelion.

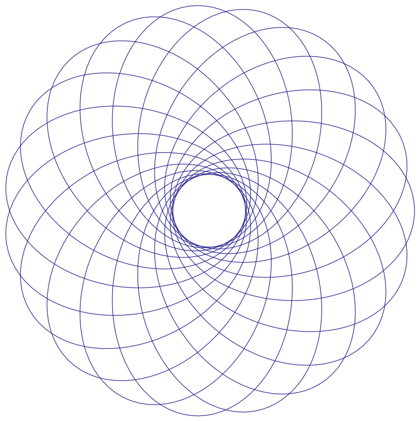

Figure 8.3: Due to precession of equinoxes, the gravitation pull of the other planets and general relativity, the orbit precesses, giving a flower shape. The position of the perihelion (the minimum distance) itself then orbits around the sun. The figure is greatly exaggerated, with an actual precession of 1.5 degrees every 400 revolutions for mercury.

the direction in which the ellipse points to precess, as in figure 8.3. These effects can be accurately computed and it was found that the theoretical precession rate was off by a tiny, but statistically significant amount from the experimental result. There appeared to be an extra 43 arcsecs/century precession that could not be accounted for. Since the precession effects are small, we can treat the different contributions as independent, so we will ignore all precession except for the part coming from general relativity. In other words, we assume that the orbit would have been elliptical if not for general relativity.

We again use the Schwarzschild metric, treating Mercury as a test-body of mass m moving in the sun's equatorial plane ($\theta = \pi/2$.) If the orbit were perfectly elliptical, then the angle ϕ would change by π as r changes from the minimum to the maximum. However, if the orbit is precessing in the same direction as Mercury is orbiting the sun, then the change in ϕ will be slightly greater than π and if it is precessing in the opposite direction then the change in ϕ is slightly less than π. Since the metric is invariant under a constant shift in t or in ϕ, there are two Killing vectors and the constants of the motion are

$$f(r)\,\dot{t} = C \qquad r^2\,\dot{\phi} = L\,. \tag{8.2.14}$$

Hence, we have the equation

$$1 = f^{-1}(r)\,C^2 - f^{-1}(r)\,\dot{r}^2 - \frac{L^2}{r^2}\,, \tag{8.2.15}$$

which we rewrite after multiplying through by a factor of $\frac{1}{2}m$ as

$$\begin{aligned} -\tfrac{1}{2}m(1-C^2) &= \tfrac{1}{2}m\dot{r}^2 - \frac{mMG}{r} + \frac{mL^2}{2r^2} - \frac{mMGL^2}{r^3} \\ &= \tfrac{1}{2}m\dot{r}^2 + V_{eff}(r)\,. \end{aligned} \tag{8.2.16}$$

We immediately notice that this equation is very similar to the usual Newtonian equations of motion if we treat $-\frac{1}{2}m(1-C^2)$ as the total energy, mL as the angular momentum and $\dot{r}$ as the radial derivative with respect to the time t. In fact, this would be the Newtonian equation if it were not for the extra term in the effective potential, $-\frac{mMGL^2}{r^3}$, which we therefore interpret as a general relativistic correction.

We will present two ways of computing the precession, the second of which is given in the next section. The two approaches are complementary to each other in that they make different approximations to do the calculation.

In the first approach an elliptic orbit of arbitrary eccentricity is considered, but the general relativistic correction is assumed to be small. The constants of the motion are determined by the two extrema of the orbit r_1 and r_2 ($r_2 > r_1$) where $\dot{r} = 0$, so that

$$\begin{aligned} 1 &= f^{-1}(r_1)\,C^2 - \frac{L^2}{r_1^2} \\ 1 &= f^{-1}(r_2)\,C^2 - \frac{L^2}{r_2^2}\,. \end{aligned} \tag{8.2.17}$$

This leads to

$$\begin{aligned} L^2 &= \frac{f(r_2) - f(r_1)}{f(r_1)\,r_1^{-2} - f(r_2)\,r_2^{-2}} = r_1^2 r_2^2\,\frac{f(r_2) - f(r_1)}{f(r_1)\,r_2^2 - f(r_2)\,r_1^2} \\ C^2 &= \frac{r_2^2 - r_1^2}{r_2^2\,f^{-1}(r_2) - r_1^2\,f^{-1}(r_1)} = f(r_1)f(r_2)\,\frac{r_2^2 - r_1^2}{f(r_1)\,r_2^2 - f(r_2)\,r_1^2} \end{aligned} \tag{8.2.18}$$

We will also define the constants

$$
\begin{aligned}
\tilde{L}^2 &\equiv r_1^2 r_2^2\,\frac{f(r_2)-f(r_1)}{r_2^2-r_1^2} = 2\,M\,G\,\frac{r_1 r_2}{r_2+r_1} \\
\delta L^2 &\equiv L^2-\tilde{L}^2 \approx (2MG)^2\,\frac{r_2^2+r_1r_2+r_1^2}{(r_1+r_2)^2} \\
\delta C^2 &\equiv C^2-1+\frac{2MG}{r_1+r_2} \\
&\approx \frac{2MG}{r_1+r_2}+2MG\,\frac{r_2^2+r_1r_2+r_1^2}{r_1r_2(r_1+r_2)}-2MG\,\frac{r_1+r_2}{r_1r_2} \\
&\qquad +\frac{(2MG)^2}{r_1r_2}-\frac{(2MG)^2(r_2^2+r_1r_2+r_1^2)}{r_1^2r_2^2}+\frac{(2MG)^2(r_2^2+r_1r_2+r_1^2)^2}{r_1^2r_2^2(r_1+r_2)^2} \\
&= \frac{(2MG)^2}{(r_1+r_2)^2}\,,
\end{aligned}
\tag{8.2.19}
$$

where $m\tilde{L}$ would be Mercury's angular momentum for Newtonian gravity assuming r_2 and r_1 are the maximum and minimum values of the orbit.

Instead of $\dot{r}$, we are really interested in $r'=\frac{dr}{d\phi}$. This is related to $\dot{r}$ and $\dot{\phi}$ by

$$
\begin{aligned}
r'=\frac{\dot{r}}{\dot{\phi}} &= \frac{r^2}{L}\sqrt{C^2-f(r)\left(1+\frac{L^2}{r^2}\right)} \\
&= \frac{1}{L}\sqrt{-\frac{2MG}{r_1+r_2}r^4+2\,M\,G\,r^3-\tilde{L}^2\,r^2+(\delta C^2r^4-\delta L^2r^2+2MGL^2r)} \\
&\approx \frac{1}{L}\sqrt{\frac{2MG}{r_1+r_2}r^2(r_2-r)(r-r_1)-\frac{(2MG)^2}{(r_1+r_2)^2}r(r+r_1+r_2)(r_2-r)(r-r_1)} \\
&\approx \frac{\sqrt{(r_2-r)(r-r_1)}}{\sqrt{r_1r_2}}\left(r-\frac{MG(r+r_1+r_2)}{(r_1+r_2)}-\frac{MG\,r(r_1^2+r_1r_2+r_2^2)}{r_1r_2(r_1+r_2)}\right) \\
&= \frac{\sqrt{(r_2-r)(r-r_1)}}{\sqrt{r_1r_2}}\left(r-MG\left(1+\frac{r(r_1+r_2)}{r_1r_2}\right)\right)
\end{aligned}
\tag{8.2.20}
$$

where in the second to last line the second term inside the parentheses comes from expanding the square root while the third term comes from expanding $\frac{1}{L}$ about $\frac{1}{\tilde{L}}$. In the last line, the first term inside the parentheses gives the Newtonian contribution, while the second term is the leading order general relativistic correction.

For an elliptical orbit, ϕ will change by 2π as r goes from the minimum at r_1 to the maximum at r_2 and back again. Hence to find the precession rate per revolution we now integrate (8.2.20) from r_1 to r_2 and multiply by 2 to get

$$
2\int_{r_1}^{r_2}\frac{\sqrt{r_1r_2}\,dr}{r\sqrt{(r_2-r)(r-r_1)}}\left(1+MG\left(\frac{1}{r}+\frac{r_1+r_2}{r_1r_2}\right)\right)\approx\ 2\pi+\Delta\phi
\tag{8.2.21}
$$

Using the integrals (8.5.4) and (8.5.5) we find that

$$\Delta\phi = 6\pi\, MG\, \frac{1}{\bar{r}}\,, \tag{8.2.22}$$

where $(\bar{r})^{-1}$ is the average of the inverse radius,

$$\frac{1}{\bar{r}} = \frac{1}{2}\left(\frac{1}{r_1} + \frac{1}{r_2}\right) = \frac{r_1 + r_2}{2\, r_1 r_2}\,. \tag{8.2.23}$$

If we now plug in numbers, we have that $MG = 1.485$ km for the sun, while for Mercury, the minimum distance (the distance to the perihelion) is $r_1 = 4.60 \times 10^7$ km, while the maximum distance (the distance to the aphelion) is $r_2 = 6.98 \times 10^7$ km, therefore $\bar{r} = 5.55 \times 10^7$ km. Hence for each revolution the precession is

$$\Delta\phi = 6\pi\, \frac{1.485\,\text{km}}{5.55 \times 10^7\,\text{km}} = 5.04 \times 10^{-7}\,, \tag{8.2.24}$$

The period of mercury's orbit is 88 days= 0.241 years, so over a century the amount of precession is

$$\Delta\phi_{\text{century}} = \frac{100}{0.241}\, 5.05 \times 10^{-7} = 2.09 \times 10^{-4} = 43.1\,\text{arc-secs}\,. \tag{8.2.25}$$

This is exactly what is needed to account for the missing precession factor.

Einstein himself computed this result in 1915 when he formulated general relativity. Note that we could have used the metric in (8.1.7) to this order of the approximation where we would have found the same thing. In fact, this is what Einstein did since at the time he did not know about the Schwarzschild metric!

8.3 The effective potential and the stability of orbits

The effective potential that appears in (8.2.16) can be readily used to look for circular orbits around a massive object and to also decide their stability. The circular orbits have $\dot{r} = 0$, which is consistent when $V_{eff}(r)$ is at an extremum, in other words $\frac{d}{dr}V_{eff}(r) = 0$. Figure 8.4 shows the effective potential for a particular choice of L^2. As one can see, the general relativistic correction is most pronounced at small r, drastically changing the behavior of $V_{eff}(r)$.

The extrema then satisfy the equation

$$\frac{d}{dr}V_{eff}(r) = 0 = \frac{mMG}{r^2} - \frac{mL^2}{r^3} + \frac{3mMGL^2}{r^4}\,, \tag{8.3.1}$$

which has the solutions

$$r_\pm = \frac{1}{2MG}\left(L^2 \pm L\sqrt{L^2 - 12(MG)^2}\right). \tag{8.3.2}$$

Note, that there are no real solutions if $L^2 < 12(MG)^2$, so $L^2 \geq 12(MG)^2$ is a condition to have circular orbits. Assuming that the orbits exist, then from figure 8.4 it is clear

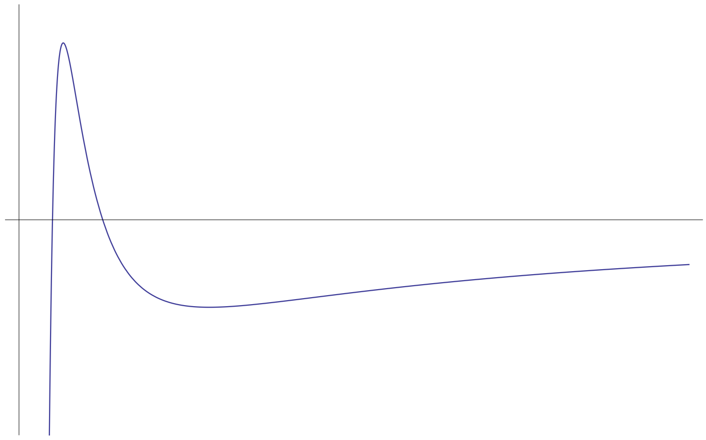

Figure 8.4: The effective potential $V_{eff}(r)$. The Newtonian potential would have $V_{eff} \to \infty$ as $r \to 0$. With the general relativistic correction the behavior is $V_{eff} \to -\infty$ as $r \to 0$..

that the extremum at r_+ is stable and the the extremum at r_- is unstable. This can also be verified by showing that $\frac{d^2}{dr^2}V_{eff}(r) > 0$ at r_+ and $\frac{d^2}{dr^2}V_{eff}(r) < 0$ at r_-. Indeed, we have

$$\frac{d^2}{dr^2}V_{eff}(r) = -\frac{2mMG}{r^3} + \frac{3mL^2}{r^4} - \frac{12mMGL^2}{r^5}\,, \tag{8.3.3}$$

and so substituting $r_\pm$ for r gives

$$\left.\frac{d^2}{dr^2}V_{eff}(r)\right|_{r=r_\pm} = \frac{mL^2}{r_\pm^4}\left(1 - \frac{6MG}{r_\pm}\right). \tag{8.3.4}$$

Since $r_+ > 6MG$ and $r_- < 6MG$ we see that $\frac{d^2}{dr^2}V_{eff}(r)$ has the claimed property. Notice that the presence of the smaller but nonstable orbit is a direct consequence of general relativity.

8.3.1 Precession once again

In the previous section we computed the precession assuming the general relativistic corrections are small. We now present a simpler way of computing the precession of an orbit, although it assumes that the orbit is almost circular. However, it is not necessary to assume that the general relativistic corrections are small, although we will do so at the end to compare our result with that in the previous section.

Suppose that the orbit is very close to being circular, with $r = r_+ + \rho$ where r_+ is the stable extremum and ρ is assumed to be much smaller than r_+. Then we can approximate $V_{eff}(r)$ as

$$\begin{aligned} V_{eff}(r) &\approx V_{eff}(r_+) + \tfrac{1}{2}\rho^2 \left.\frac{d^2}{dr^2}V_{eff}(r)\right|_{r=r_+} \\ &= V_{eff}(r_0) + \tfrac{1}{2}\frac{mL^2}{r_+^4}\left(1 - \frac{6MG}{r_+}\right)\rho^2\,. \end{aligned} \tag{8.3.5}$$

In this approximation, $V_{eff}(r)$ looks like the potential for an harmonic oscillator with angular frequency

$$\omega = \frac{L}{r_+^2}\sqrt{1 - \frac{6MG}{r_+}}\,, \tag{8.3.6}$$

and so $\rho = \rho_0 \cos(\omega\tau + C)$, where C is a constant phase. Since we are very close to a circular orbit, we can approximate $\dot{\phi} \approx \frac{L}{r_+^2}$ and so $\phi = \tau\,\frac{L}{r_+^2} + C'$. Hence,

$$\rho = \rho_0 \cos\left(\phi\sqrt{1 - \frac{6MG}{r_+}} + \phi_0\right), \tag{8.3.7}$$

where we combined the constants into one constant phase.

Note that result in (8.3.7) does not require $r_+ >> MG$. However, in order to compare with the previous result let us now assume that $r_+ >> 6MG$. Then we can approximate ρ as

$$\rho \approx \rho_0 \cos\left(\phi\left(1 + \frac{3MG}{r_+}\right)^{-1} + \phi_0\right)\,, \tag{8.3.8}$$

and so the difference from 2π in one revolution from minimum to minimum for ρ is

$$\Delta\phi = 6\pi\frac{MG}{r_+} \tag{8.3.9}$$

which agrees with our previous result since $\bar{r} = r_+$ in the limit of a circular orbit.

8.4 The horizon

If we examine the Schwarzschild metric in (8.1.4) we notice some similarities to the accelerating frame metric

$$ds^2 = -(1 + 2\alpha_0\tilde{x})d\tilde{t}^{\,2} + \frac{d\tilde{x}^{\,2}}{1 + 2\alpha_0\tilde{x}} + d\tilde{y}^2 + d\tilde{z}^2\,, \tag{8.4.1}$$

which we introduced in the second chapter. In particular, they both have coordinate singularities, in the case of the Schwarzschild metric this happens at $r = 2MG$. Furthermore, just as in the case of the accelerating frame, as a light source gets closer and closer to the singularity, the light as observed by an observer farther out is more and more red-shifted and so the source's clock appears to run slower and slower. Another similarity between the two metrics is that for space-time points beyond the singularity, that is $r < 2MG$ for Schwarzschild and $\tilde{x} < -1/(2\alpha_0)$ for the accelerating frame, the time coordinate becomes space-like, while the spatial coordinate becomes time-like since the sign of the metric changes.

Other things we can investigate with Schwarzschild that we previously studied with the accelerating frame are free-fall of massive objects or light. For example, for a light-ray heading inward we have that

$$\frac{dr}{dt} = -f(r)\,. \tag{8.4.2}$$

Assuming that $r = r_0$ at $t = 0$, we can integrate this equation so that

$$t = -\int_{r_0}^{r} \frac{dr}{f(r)} = -\int_{r_0}^{r} \frac{r\,dr}{r-2MG} = r_0 - r + 2MG\log\frac{r_0-2MG}{r-2MG}\,. \tag{8.4.3}$$

Since the right side diverges logarithmically as $r \to 2MG$, it will take an infinite amount of coordinate time t to reach $r = 2MG$. If the light is heading outward then

$$t = r - r_0 + 2MG\log\frac{r-2MG}{r_0-2MG}\,. \tag{8.4.4}$$

If we consider an infalling massive body then we have that

$$1 = f(r)\dot{t}^2 - f(r)^{-1}\dot{r}^2 = f(r)^{-1}\left(C^2 - \dot{r}^2\right). \tag{8.4.5}$$

If we make the assumption that $\dot{r} = 0$ at $r = r_0$, then $C^2 = f(r_0)$ and so we find

$$\dot{r} = -\sqrt{f(r_0) - f(r)} = -\sqrt{\frac{2MG}{r_0}}\sqrt{\frac{r_0-r}{r}}\,. \tag{8.4.6}$$

We can integrate this equation to give

$$\tau = \sqrt{\frac{2MG}{r_0}}\left(\sqrt{r(r_0-r)} + r_0\left(\frac{\pi}{2} - \arctan\sqrt{\frac{r}{r_0-r}}\right)\right)\,. \tag{8.4.7}$$

The main thing to observe is that if takes a finite amount of proper time τ to reach $r = 2MG$ and that the infalling body continues falling into the region $r < 2MG$. However, it must take an infinite amount of coordinate time to reach this point since the massive body has to move slower than a light ray, which itself takes an infinite amount of coordinate time to reach $r = 2MG$.

As in the case of the accelerating frame, we therefore suspect that $r = 2MG$ is not a real singularity but is only a coordinate singularity. If this is the case then there should be another set of coordinates where there is no singularity at $r = 2MG$. Such a set of coordinates were found by Kruskal in the 1960's and are now called Kruskal coordinates (what else). These coordinates are defined by

$$\begin{aligned} u &\equiv MG\sqrt{\frac{r}{2MG} - 1}\, e^{r/4MG}\cosh\frac{t}{4MG} \\ v &\equiv MG\sqrt{\frac{r}{2MG} - 1}\, e^{r/4MG}\sinh\frac{t}{4MG} \end{aligned} \tag{8.4.8}$$

if $r \geq 2MG$ and

$$\begin{aligned} u &\equiv MG\sqrt{1-\frac{r}{2MG}}\,e^{r/4MG}\sinh\frac{t}{4MG} \\ v &\equiv MG\sqrt{1-\frac{r}{2MG}}\,e^{r/4MG}\cosh\frac{t}{4MG} \end{aligned} \tag{8.4.9}$$

if $r \leq 2MG$. With these coordinates we observe that

$$u^2 - v^2 = (MG)^2\left(\frac{r}{2MG}-1\right)e^{r/2MG}\,, \tag{8.4.10}$$

We also see that $r = 2MG$ corresponds to $u^2 - v^2 = 0$ which has the two solutions, $u = \pm v$. We also have that

$$\begin{aligned} t &= 4MG\,\mathrm{arctanh}\frac{v}{u} \qquad & r \geq 2MG \\ t &= 4MG\,\mathrm{arctanh}\frac{u}{v} \qquad & r \leq 2MG\,. \end{aligned} \tag{8.4.11}$$

The differentials du and dv are related to dt and dr by

$$\begin{aligned} du &= \frac{v}{4MG}\,dt + \left(1+\frac{2MG}{r-2MG}\right)\frac{u}{4MG}\,dr = \frac{v}{4MG}\,dt + f(r)^{-1}\frac{u}{4MG}\,dr \\ dv &= \frac{u}{4MG}\,dt + \left(1+\frac{2MG}{r-2MG}\right)\frac{v}{4MG}\,dr = \frac{u}{4MG}\,dt + f(r)^{-1}\frac{v}{4MG}\,dr \end{aligned} \tag{8.4.12}$$

which is true for any value of r. We then see that

$$\begin{aligned} du^2 - dv^2 &= (u^2-v^2)\frac{1}{(4MG)^2}(f(r)^{-2}\,dr^2 - dt^2) \\ &= \frac{r}{32MG}\,e^{r/2MG}\left(f(r)^{-1}\,dr^2 - f(r)\,dt^2\right). \end{aligned} \tag{8.4.13}$$

Hence, the Schwarzschild metric can be written as

$$ds^2 = \frac{32MG}{r}\,e^{r/2MG}(-dv^2+du^2) + r^2(d\theta^2+\sin^2\theta\,d\phi^2)\,, \tag{8.4.14}$$

where r can be in principle written in terms of $u^2 - v^2$ by inverting (8.4.10). From this form of the metric, we can see that nothing bad happens as $r = 2MG$. We also see that u is always a space-like coordinate while v is a time-like coordinate.

The transformations in (8.4.8) and (8.4.8) are singular transformations at $r = 2MG$. The coordinates u and v are continuous functions of r and t, but their derivatives are not continuous. However, it is this singularity of the transformation that cancels out the coordinate singularity of the metric, leaving a nonsingular metric at $r = 2MG$ for the u and v coordinates

The metric however, is still singular since the point at $r = 0$ is a real singularity. We have previously seen in chapter 5 that the Schwarzschild metric has $R^{\mu\nu\lambda\sigma}R_{\mu\nu\lambda\sigma} =$

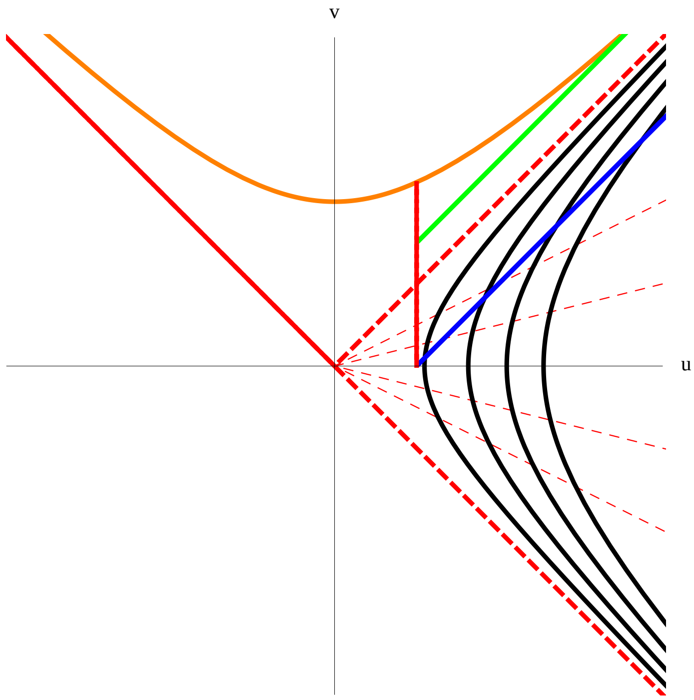

Figure 8.5: The Schwarzschild geometry in terms of Kruskal coordinates. Dashed red lines are constant t, solid black are constant r, solid orange is the real singularity at $r = 0$. The thick dashed red lines are the past and future horizons. The solid red line is a time like line, while the blue and the green solid lines are light-like trajectories outside and inside the future horizon.

$48\frac{(MG)^2}{r^6}$, so the curvature is blowing up at $r = 0$. In terms of u and v we can see from (8.4.10) that this corresponds to the equation $u^2 - v^2 = -(MG)^2$.

Figure 8.5 shows the Schwarzschild geometry in terms of the Kruskal variables u and v. The dashed red lines are curves of constant t. The thicker dashed red lines are the constant t lines at $t = -\infty, \infty$, $r = 2MG$. These lines are light-like and are called the past and future *horizons*. The solid black curves are at constant r, while the solid orange line is at $r = 0$ and is the real singularity of the Schwarzschild metric. The figure also shows a time-like world-line (solid red) which we can think of as some astronaut with a light-source. Before the astronaut crosses the future horizon it emits a light ray (solid blue). As you can see the blue line is crossing the lines of constant r so it is able to go out to $r = \infty$. The astronaut then crosses the future horizon in finite proper time and emits another light ray (solid green). This light ray cannot cross the horizon since it is above the horizon line and the horizon itself is light-like. In fact, as the figure shows, the light ray itself will hit the singularity. If the light ray must hit the singularity, then everything else is doomed to hit as well, including the unfortunate astronaut.

Since no light can escape once the horizon is crossed, the Schwarzschild geometry is also called a *black hole*. The size of the black hole is given by $2MG$, so a black hole with mass equal to the sun's would be 1.5 km in size, smaller than the distance between Polacksbacken and Uppsala C. Even though black holes are black, they are detectable indirectly because they are busy sucking in a large amount of matter that emits x-rays as the matter is pulled in. There is also clear evidence for a black hole at the center of our own Milky Way galaxy which is millions of solar masses in size. Around the black hole one can see stars orbiting an extremely heavy object. The mass is easily determined from the orbits of the stars (you can see the evidence on youtube at http://www.youtube.com/user/deholz).

One irony about black holes is that the curvature at the horizon gets smaller as the black-hole gets bigger. Indeed, we have

$$R^{\mu\nu\lambda\sigma}R_{\mu\nu\lambda\sigma} = 48\frac{(MG)^2}{r^6} = \frac{3}{4(MG)^4} \tag{8.4.15}$$

at $r = 2MG$, so the bigger M, the smaller $R^{\mu\nu\lambda\sigma}R_{\mu\nu\lambda\sigma}$. Recall from chapter 5 that we argued that the curvature was related to the derivative of the gravitational force, so for example for the Schwarzschild metric

$$R^{\mu\nu\lambda\sigma}R_{\mu\nu\lambda\sigma} = 12\left(\frac{d^2}{dr^2}\Phi(r)\right)^2 . \tag{8.4.16}$$

This change in the force is called the tidal force and is responsible for ripping apart bodies of nonzero extent. Without getting into two many details we can determine the size of this tidal force by making a few simplifying assumptions. For example, suppose our astronaut going through the horizon goes in feet first and he is ℓ =2 meters tall. Let us further assume that he weighs m =100 kg and that half the mass is at his feet ($m/2$) and half is at his head (this is obviously not true, but it will not affect the answer beyond an order of magnitude). Then the tension T on the astronaut is the force at his feet minus that at his head which is

$$T \sim \frac{2MG\,\ell\,(m/2)}{r^3} = \frac{\ell\, m}{2(2MG)^2} \tag{8.4.17}$$

For a black hole with mass equal to that of the sun we have

$$T \sim 10^{-4}\,\mathrm{kg\,m^{-1}} \sim 10^{13}\,\mathrm{Newtons} \tag{8.4.18}$$

which is a huge tension, obviously beyond the limits of a human body. However, if the astronaut is crossing the horizon of a black hole with mass equal to 10^6 solar masses, the the tension will come down by a factor of 10^{12}, leaving the long suffering astronaut with a tension he won't even notice.

8.5 Integrals

The following integrals are referred to in the text:

These integrals can be solved by trigonometric substitution, with $z = \sec\xi$, $dz = \sec\xi\,\tan\xi\,d\xi$,

$$2\int_1^\infty \frac{dz}{z\sqrt{z^2-1}} \;=\; 2\int_0^{\pi/2} \frac{\sec\xi\tan\xi\,d\xi}{\sec\xi\,\tan\xi} = 2\int_0^{\pi/2} d\xi = \pi\,, \tag{8.5.1}$$

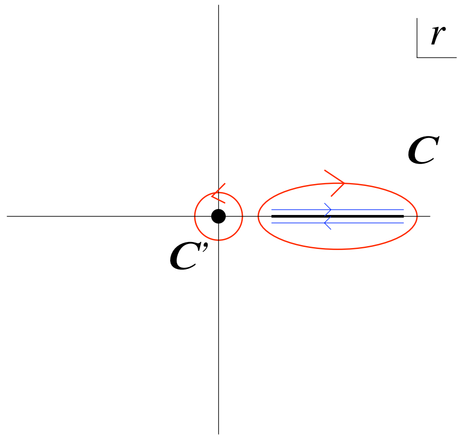

Figure 8.6: Integration contours in the complex r plane. The original integral is the sum of the integral from r_1 to r_2 (blue) minus the integral from r_2 to r_1 just below the cut (blue). This is equivalent to the contour integral on $\mathcal{C}$ which goes around the branch cut at $r_1 \le r \le r_2$. This can be deformed into the contour $\mathcal{C}'$ at $r = 0$ since there is no other singularities in the plane.

$$
\begin{aligned}
\int_1^\infty \frac{(z^2+z+1)dz}{z^2(z+1)\sqrt{z^2-1}} &= \int_0^{\pi/2} \frac{(\sec^2\xi + \sec\xi + 1)\sec\xi\tan\xi\, d\xi}{(\sec\xi+1)\sec^2\xi\ \tan\xi} \\
&= \int_0^{\pi/2} \frac{(1+\cos\xi+\cos^2\xi)\,d\xi}{(1+\cos\xi)} \\
&= \int_0^{\pi/2} \frac{d}{d\xi}\left(\sqrt{\left(\frac{1-\cos\xi}{1+\cos\xi}\right)} + \cos\xi\right) d\xi \\
&= \left(\sqrt{\frac{1-\cos\xi}{1+\cos\xi}} + \sin\xi\right)\Bigg|_0^{\pi/2} = 2\,.
\end{aligned} \tag{8.5.2}
$$

The next integrals we solve by contour integration in the complex r plane. The first is

$$
\begin{aligned}
2\int_{r_1}^{r_2} \frac{\sqrt{r_1r_2}\,dr}{r\sqrt{(r_2-r)(r-r_1)}} &= \int_{r_1+i\epsilon}^{r_2+i\epsilon} \frac{\sqrt{r_1r_2}\,dr}{r\sqrt{(r_2-r)(r-r_1)}} + \int_{r_2-i\epsilon}^{r_1-i\epsilon} \frac{\sqrt{r_1r_2}\,dr}{r\sqrt{(r_2-r)(r-r_1)}} \\
&= \oint_{\mathcal{C}} \frac{\sqrt{r_1r_2}\,dr}{r\sqrt{(r_2-r)(r-r_1)}}\,,
\end{aligned} \tag{8.5.3}
$$

where $\mathcal{C}$ is the contour around the branch cut in figure 8.6. We can deform this contour to the contour $\mathcal{C}'$ that circles the singularity at $r = 0$. Hence we have

$$
2\int_{r_1}^{r_2} \frac{\sqrt{r_1r_2}\,dr}{r\sqrt{(r_2-r)(r-r_1)}} = \oint_{\mathcal{C}'} \frac{\sqrt{r_1r_2}\,dr}{r\sqrt{(r_2-r)(r-r_1)}} = 2\pi\, i\,\frac{\sqrt{r_1r_2}}{\sqrt{-r_1r_2}} = 2\pi\,. \tag{8.5.4}
$$

The other integral is

$$
\begin{aligned}
2\int_{r_1}^{r_2} \frac{\sqrt{r_1 r_2}\,dr}{r^2\sqrt{(r_2-r)(r-r_1)}} &= \oint_{\mathcal{C}'} \frac{\sqrt{r_1 r_2}\,dr}{r^2\sqrt{(r_2-r)(r-r_1)}} \\
&= \oint_{\mathcal{C}'} \frac{\sqrt{r_1 r_2}\,dr}{r^2\sqrt{-r_1 r_2}}\left(1+\frac{1}{2}\,\frac{r_1+r_2}{r_1 r_2}\,r+\dots\right) = 0+\pi\,\frac{r_1+r_2}{r_1 r_2}\,.
\end{aligned}
\tag{8.5.5}
$$

Chapter 9

Introduction to cosmology

In this chapter we start the cosmology part of the course. There has been a huge amount of progress in cosmology over the last 15 years. One reason for this is better and better telescopes and other detection devices have become available, and with this new equipment has come new and often surprising results. Some would say that cosmology is the most exciting field in physics right now.

9.1 The Friedmann-Robertson-Walker Universe

It turns out that we can make a surprising amount of progress if we assume that the universe is homogenous and isotropic. Recall that homogenous means that any one point in the universe is the same as any other point. Now of course, this is not true. For example, the universe at the surface of the earth is drastically different from the universe in the region just outside the solar system. But over large distance scales, and for us large means millions of light years, the average density of stuff is homogenous. In other words, if we were to look at a spherical volume with radius of ten million light years at two different regions in the universe, we would find about the same number of galaxies inside each volume. To say that the universe is isotropic means that no matter what direction we look, it looks the same.

9.1.1 Flat, closed and open

There are three types of three-dimensional spaces that are homogenous and isotropic. All three can be summarized by the metric

$$ds^2 = a^2 \left(\frac{dr^2}{1 - kr^2} + r^2(d\theta^2 + \sin^2\theta\, d\phi^2) \right)\Bigg(, \tag{9.1.1}$$

where a is assumed to be independent of the spatial components. Assuming that r has the dimensions of length then k has dimensions of inverse length squared and a is dimensionless (some books use the convention that r and k are dimensionless). The three possible classes of metrics come from choosing $k = 0$, $k > 0$ or $k < 0$.

$\underline{\ k=0}$: If we absorb the factor of a into r then this is the usual flat three-dimensional metric written in polar coordinates.
$\underline{k>0}$: To see what this is, we let $r=k^{-1/2}\sin\chi$, where χ is an angle. If r is between 0 and $k^{-1/2}$ then χ goes from 0 to $\pi/2$. With this substitution the metric becomes

$$ds^2 = a^2k^{-1}(d\chi^2+\sin^2\chi(d\theta^2+\sin^2\theta\,d\phi^2))\,. \tag{9.1.2}$$

This is the metric for a three-dimensional sphere (called a 3-sphere, which is also written as S^3) with radius $a/\sqrt{k}$. This is a closed space so we would call a universe with this sort of spatial geometry closed. Actually, the range $0\le\chi\le\pi/2$ only covers half the 3-sphere. To cover the entire sphere we have to extend χ to the range $0\le\chi\le\pi$.
$\underline{k<0}$: In this case we let $r=(-k)^{-1/2}\sinh\xi$, where ξ ranges from 0 to ∞. The metric then becomes

$$ds^2 = a^2k^{-1}(d\xi^2+\sinh^2\xi(d\theta^2+\sin^2\theta\,d\phi^2))\,. \tag{9.1.3}$$

This looks a lot like the metric for the 3-sphere, except now we have a sinh function instead of a sin. This is the metric for a *hyperbolic space*, and since r is unbounded this would be the space-time for an open universe.

To become better acquainted with these spaces let us compute their 3-dimensional curvatures. Because the metric in (9.1.1) is diagonal we have that

$$\begin{aligned}
\Gamma^r_{\theta\theta} &= -\tfrac{1}{2}g^{rr}\partial_r g_{\theta\theta} = -(1-kr^2)r\\
\Gamma^r_{\phi\phi} &= -\tfrac{1}{2}g^{rr}\partial_r g_{\phi\phi} = -(1-kr^2)r\sin^2\theta\\
\Gamma^\theta_{r\theta} &= \Gamma^\phi_{r\phi} = \frac{1}{r}\\
\Gamma^r_{rr} &= \tfrac{1}{2}g^{rr}\partial_r g_{rr} = kr(1-kr^2)^{-1}
\end{aligned} \tag{9.1.4}$$

while $\Gamma^\theta_{\phi\phi}$ and $\Gamma^\phi_{\theta\phi}$ are the same as for the 2-sphere (see section 2.1 of chapter 5). We then find that

$$\begin{aligned}
R^r{}_{\theta r\theta} &= \partial_r\Gamma^r_{\theta\theta}-\partial_\theta\Gamma^r_{\theta r}+\Gamma^r_{\mu r}\Gamma^\mu_{\theta\theta}-\Gamma^r_{\mu\theta}\Gamma^\mu_{r\theta} = \partial_r\Gamma^r_{\theta\theta}+\Gamma^r_{rr}\Gamma^r_{\theta\theta}-\Gamma^r_{\theta\theta}\Gamma^\theta_{r\theta}\\
&= -(1-kr^2)+2kr^2-kr^2+(1-kr^2) = kr^2\\
R^r{}_{\phi r\phi} &= \partial_r\Gamma^r_{\phi\phi}-\partial_\phi\Gamma^r_{\phi r}+\Gamma^r_{\mu r}\Gamma^\mu_{\phi\phi}-\Gamma^r_{\mu\phi}\Gamma^\mu_{r\phi} = \partial_r\Gamma^r_{\phi\phi}+\Gamma^r_{rr}\Gamma^r_{\phi\phi}-\Gamma^r_{\phi\phi}\Gamma^\phi r\phi\\
&= -(1-kr^2)\sin^2\theta+2kr^2\sin^2\theta-kr^2\sin^2\theta+(1-kr^2)\sin^2\theta = kr^2\sin^2\theta\\
R^\theta{}_{\phi\theta\phi} &= \tilde{R}^\theta{}_{\phi\theta\phi}+\Gamma^\theta_{r\theta}\Gamma^r_{\phi\phi} = \sin^2\theta-\frac{1}{r}(1-kr^2)r\sin^2\theta = kr^2\sin^2\theta\,,
\end{aligned} \tag{9.1.5}$$

where $\tilde{R}^\theta{}_{\phi\theta\phi}$ is the Riemann tensor for the 2-sphere. It is then clear that for all combinations we have

$$R_{iljm} = \frac{k}{a^2}\left(g_{ij}g_{lm}-g_{im}g_{jl}\right)\,. \tag{9.1.6}$$

from which it follows that

$$R_{ij} = \frac{k}{a^2}(3g_{ij}-g_{ij}) = \frac{2k}{a^2}g_{ij}\,,\qquad R=\frac{6k}{a^2}\,. \tag{9.1.7}$$

Hence we see that the 3-sphere has constant positive scalar curvature, flat space has zero scalar curvature and the hyperbolic space has constant negative spatial curvature.

9.1.2 The Robertson-Walker metric

We now bring the time coordinate into the picture. We still assume that the spatial part of the metric has the form in (9.1.1), but now we assume that a is a time dependent function $a(t)$. Hence the full 4-dimensional space-time metric is

$$ds^2 = -dt^2 + a^2(t)\left(\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)\right). \tag{9.1.8}$$

This is known as the *Robertson-Walker metric*. The spatial coordinates (r, θ, ϕ) are known as the *co-moving coordinates*. We then look for solutions to Einstein's equations assuming that the space is filled with a homogenous isotropic substance, in other words a perfect fluid. The fluid can be one perfect fluid obeying the equation of state $p = w\rho$, or it can be a combination of perfect fluids with different equations of state.

The other Christoffel symbols we need besides the ones in (9.1.4) have the form

$$\Gamma^t_{ii} = -\tfrac{1}{2}g^{tt}\partial_t g_{ii} = \frac{\dot a}{a}\, g_{ii} \qquad \Gamma^i_{ti} = \tfrac{1}{2}g^{ii}\partial_t g_{ii} = \frac{\dot a}{a}\,, \tag{9.1.9}$$

where $\dot a \equiv \frac{d}{dt}a$. Hence, the extra Riemann tensors are

$$R^t{}_{iti} = \partial_t\Gamma^t_{ii} + \Gamma^t_{\mu t}\Gamma^\mu_{ii} - \Gamma^t_{\mu i}\Gamma^\mu_{it} = \frac{\ddot a}{a}\, g_{ii} + \left(\frac{\dot a}{a}\right)^2 g_{ii} + 0 - \left(\frac{\dot a}{a}\right)^2 g_{ii} = \frac{\ddot a}{a}\, g_{ii}\,, \tag{9.1.10}$$

while the Riemann tensor $R^i{}_{jij}$ gets modified to

$$R^i{}_{jij} = \tilde R^i{}_{jij} + \Gamma^i_{ti}\Gamma^t_{jj} = \frac{k}{a^2}\, g_{jj} + \left(\frac{\dot a}{a}\right)^2 g_{jj}\,, \tag{9.1.11}$$

where $\tilde R^i{}_{jij}$ are the Riemann tensor components in (9.1.5).

9.1.3 Scale dependence of the energy density

Recall that the energy-momentum tensor satisfies the conservation equation $\nabla_\mu T^{\mu\nu} = 0$. If we choose $\nu = 0$ then we have

$$\nabla_\mu T^{\mu 0} \;=\; \partial_t T^{00} + \Gamma^\mu_{\mu\lambda}T^{\lambda 0} + \Gamma^0_{\mu\lambda}T^{\mu\lambda} = \dot\rho + \sum_i \left(\Gamma^i_{it}\, T^{00} + \Gamma^t_{ii}\, T^{ii}\right) \tag{9.1.12}$$

where the sum over i is on the spatial coordinates and $\dot\rho = \frac{d}{dt}\rho$. Each component of $g_{ii}T^{ii}$ equals the pressure p, hence, using (9.1.9) we have

$$0 \;=\; \dot\rho + 3\frac{\dot a}{a}\,\rho + 3\frac{\dot a}{a}\,p = \dot\rho + 3(\rho + p)\,\frac{\dot a}{a}\,. \tag{9.1.13}$$

If the perfect fluid has an equation of state $p = w\,\rho$ then (9.1.13) can be rewritten as

$$\frac{d}{dt}\log\rho = -3(1+w)\frac{d}{dt}\log a\,, \tag{9.1.14}$$

which has the solution

$$\rho = \left(\frac{a_0}{a}\right)^{3(1+w)} \rho_0 \,, \tag{9.1.15}$$

where ρ_0 is an integration constant which fixes the density when $a = a_0$.

Let us now give some physical motivation for this result by considering some cases of perfect fluids. For dust, which has $w = 0$, we have nonrelativistic particles. Suppose we have a fixed number of particles N in a volume V_0 when $a = a_0$. If the rest mass is m then the energy density is $\rho = m\frac{N}{V_0} \equiv \rho_0$. If at a later time we have a different value of a, then the number N and mass m stay the same, but the physical volume has changed to $(a/a_0)^3\, V_0$ and so the energy density becomes

$$\rho = m\,\frac{N}{V_0}\left(\frac{a_0}{a}\right)^3 = \left(\frac{a_0}{a}\right)^3 \rho_0 \,, \tag{9.1.16}$$

which agrees with (9.1.15) when $w = 0$.

For radiation we have $w = 1/3$. Let us suppose that we have N photons with wavelength λ_0 inside a volume V_0 when $a = a_0$. In this case the energy density is $\rho = \frac{N\,h}{\lambda_0 V_0} \equiv \rho_0$, where h is Planck's constant. At a later time where we have a different value of a the volume again changes to $(a/a_0)^3\, V_0$ and the wavelength also changes to $(a/a_0)\,\lambda_0$. Hence the energy density becomes

$$\rho = \frac{N\,h}{\lambda_0 V_0}\left(\frac{a_0}{a}\right)^4 = \left(\frac{a_0}{a}\right)^4 \rho_0 \,, \tag{9.1.17}$$

which agrees with (9.1.15) when $w = 1/3$.

For vacuum energy which has $w = -1$, we have that $\rho = -\frac{\Lambda}{8\pi G}\, g^{00} = \frac{\Lambda}{8\pi G} = \rho_0$ for all values of a, where Λ is the cosmological constant. This too agrees with (9.1.15) when $w = -1$.

9.1.4 The Friedmann equations

We are now ready to solve Einstein equations for our homogenous, isotropic universe. We first find the Ricci tensors and Ricci scalar for the Robertson-Walker metric. The various nonzero components of the Ricci tensor are

$$\begin{aligned} R_{tt} &= \sum_i R^i{}_{tit} = \sum_i g^{ii}\,\frac{\ddot a}{a}\, g_{ii} g_{tt} = -3\,\frac{\ddot a}{a} \\ R_{jj} &= R^t{}_{jtj} + \sum_{i\neq j} R^i{}_{jij} = \frac{\ddot a}{a}\, g_{jj} + 2\,\frac{k}{a^2}\, g_{jj} + 2\left(\frac{\dot a}{a}\right)^2 g_{jj} \,, \end{aligned} \tag{9.1.18}$$

where we used that $g_{tt} = -1$ and that $g^{ii} = (g_{ii})^{-1}$, while the Ricci scalar is

$$R = g^{tt} R_{tt} + \sum_j g^{jj} R_{jj} = 6\left(\frac{\ddot a}{a} + \frac{k}{a^2} + \left(\frac{\dot a}{a}\right)^2\right) . \tag{9.1.19}$$

The nonzero components of the Einstein tensor are then

$$\begin{aligned} G_{tt} &= R_{tt} - \tfrac{1}{2}\, g_{tt} R = 3\left(\frac{k}{a^2} + \left(\frac{\dot{a}}{a}\right)^2\right) \\ G_{jj} &= R_{jj} - \tfrac{1}{2}\, g_{jj} R = -\left(2\,\frac{\ddot{a}}{a} + \frac{k}{a^2} + \left(\frac{\dot{a}}{a}\right)^2\right) g_{jj}\,. \end{aligned} \tag{9.1.20}$$

Using $T_{00} = (g_{tt})^2 T^{00} = \rho$ and $T_{jj} = g_{jj} g_{jj} T^{jj} = g_{jj}\, p$, Einstein's equations give us the following equations:

$$\begin{aligned} 3\left(\frac{k}{a^2} + \left(\frac{\dot{a}}{a}\right)^2\right) &= 8\pi\, G\, \rho \\ -\left(2\,\frac{\ddot{a}}{a} + \frac{k}{a^2} + \left(\frac{\dot{a}}{a}\right)^2\right) &= 8\pi\, G\, p\,, \end{aligned} \tag{9.1.21}$$

One can show that the second equation follows from the first equation and eq. (9.1.13) (see exercise). We can rewrite the first equation as

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\,\rho - \frac{k}{a^2}\,, \tag{9.1.22}$$

H is the *Hubble constant* which measures the expansion rate of the universe. We will have much more to say about this later. Equation (9.1.22) is known as the *Friedmann equation*. Notice this equation assumes a perfect fluid but makes no assumption about the composition of the fluid. It could be a mixture of perfect fluids or one perfect fluid with a particular equation of state $p = w\,\rho$.

All indications are that the universe is flat and so $k = 0$. If H_0 is the Hubble constant today, then in this case the energy density has to equal a *critical density* $\rho = \rho_c$ where

$$\rho_c = \frac{3(H_0)^2}{8\pi G}\,. \tag{9.1.23}$$

If $\rho > \rho_c$ then $k > 0$ and the universe is closed. If $\rho < \rho_c$ then $k < 0$ and the universe is open. The present Hubble constant is about 70 km/sec/Mpsc, where Mpsc is "megaparsec". A megaparsec is 3.26×10^6 ly $= 3.09 \times 10^{22}$ m, so plugging in numbers gives today's critical density to be

$$\rho_c = \frac{3(7\times 10^4\,\mathrm{m\,s^{-1}}/3.09\times 10^{22}\ \mathrm{m})^2}{8\pi(6.67\times 10^{-11}\,\mathrm{m^3\,kg^{-1}s^{-2}})} \approx 9.2\times 10^{-27}\,\mathrm{kg\,m^{-3}}\,, \tag{9.1.24}$$

which is a little less than the energy density of 6 hydrogen atoms for every cubic meter. While this might not seem like much, consider how much visible matter is in the universe. There are about 8×10^{10} galaxies in the universe and about 4×10^{11} stars per galaxy. If we assume that every star has the mass of the sun, this gives a total mass of 6×10^{52} kg. The size of the universe is about 14000 Mpsc and so its volume is roughly $3\times 10^{80}\,\mathrm{m}^3$.

This gives an overall density from stars as $2 \times 10^{-28}\,\mathrm{kg\,m^{-3}}$. Hence the stars contain at most 5% of the critical density.

Going back to the Friedmann equation we can reexpress (9.1.22) as

$$\tfrac{1}{2}\,(\dot{a})^2 - \frac{4\pi G}{3} a^2 \rho = -\tfrac{1}{2}\,k\,, \tag{9.1.25}$$

As we have seen, ρ is a function of a and so the equation in (9.1.25) looks like the equation for the total energy for one dimensional motion of a coordinate a, where $\frac{1}{2}\,(\dot{a})^2$ is the kinetic term, $-\frac{4\pi G}{3}a^2\rho \equiv V(a)$ is the potential term and $\frac{1}{2}\,k$ is the total energy. Let us assume that the density is comprised of the three types of fluids, matter (*aka* dust), radiation and vacuum. The density then has the form

$$\rho = \rho_{\mathrm{m}} \left(\frac{a_0}{a}\right)^3 + \rho_{\mathrm{r}} \left(\frac{a_0}{a}\right)^4 + \rho_{\mathrm{v}}\,, \tag{9.1.26}$$

where ρ_{m}, ρ_{r} and ρ_{v} are the energy densities for matter, radiation and vacuum energy respectively, when $a = a_0$. The effective potential is then

$$V(a) = -\frac{4\pi G}{3}\,(a_0)^2 \quad \rho_{\mathrm{m}}\frac{a_0}{a} + \rho_{\mathrm{r}}\left(\frac{a_0}{a}\right)^2 + \rho_{\mathrm{v}}\left(\frac{a}{a_0}\right)^2\Bigg)\,. \tag{9.1.27}$$

Figure 9.1 shows the effective potential as a function of a in the case when $\rho_{\mathrm{v}} = 0$ and for the cases of $k > 0$, $k = 0$ and $k < 0$. We assume that both ρ_{m} and ρ_{r} are nonzero. As you can see from the effective potential, the radiation term dominates for small a and the matter term dominates for large a. We call these regions the *radiation dominated universe* and the *matter dominated universe* respectively. We assume that the universe starts at $t = 0$ with $a = 0$. This is the *big bang* where the universe starts at zero size (at least classically). In the case where $k > 0$, the universe "rolls up the hill", slowing down until it reaches the $k > 0$ line where it stops and turns around. It keeps accelerating downward until it reaches $a = 0$ again. This is the *big crunch* where the universe shrinks to zero size. If $k = 0$ then the universe rolls up the hill, slowing down until reaches zero velocity at $a = \infty$. Hence, in this case the universe expands forever with the Hubble constant decreasing in time until it reaches zero. In the $k < 0$ the universe also expands forever but approaches a nonzero $\dot{a}$ in the limit of large a.

If there is a vacuum energy density present then there can be drastic changes. Figure 9.2 shows the effective potential with $\rho_{\mathrm{v}} > 0$. The figure shows two different values with $k > 0$. In one case with the larger value of k, the universe is not expanding fast enough to overcome the hump, in which case the expansion stops and then the universe contracts back to the big crunch. However, for a smaller value of k, the universe can get over the hump, at which point the expansion starts accelerating outward. This will also be true for the $k = 0$ and the $k < 0$ cases as well.

We can now also see why Einstein needed the cosmological constant to get a static universe. If we look at figure 9.2 again we see that there can be a solution where $\dot{a} = 0$. This is at the top of the hump where $\frac{d}{da}V(a) = 0$.

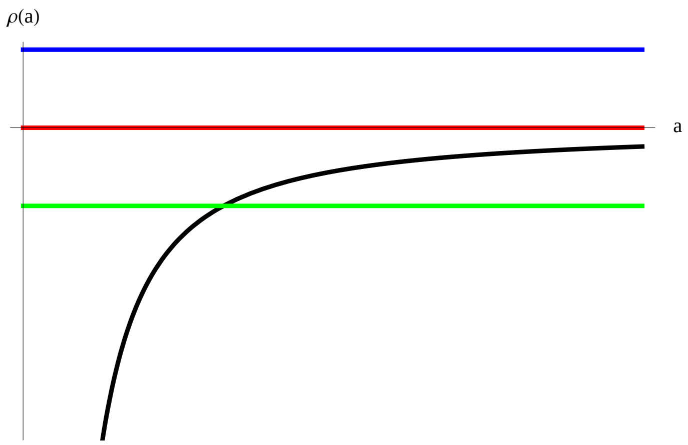

Figure 9.1: The effective potential for a with $\rho_{\rm V}=0$. The green line is $k>0$, the red line is $k=0$ and the blue line is $k<0$.

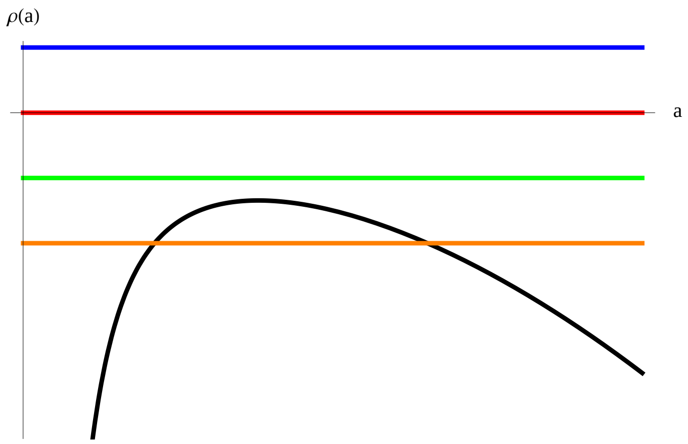

Figure 9.2: The effective potential for a with $\rho_{\rm V}>0$. The green line and orange lines are two different values with $k>0$, the red line is $k=0$ and the blue line is $k<0$.

9.1.5 Solutions for a definite equation of state

Let us now assume that the universe is made up of a material with a definite equation of state $p=w\rho$. Then from (9.1.15) and the Friedmann equation (9.1.22) we have

$$(\dot{a})^2=\frac{8\pi G\rho_0}{3}\,\frac{(a_0)^{3+3w}}{a^{1+3w}}-k\,, \tag{9.1.28}$$

where ρ_0 is the density when $a=a_0$. If we further assume that $k=0$, then ρ_0 is the critical density ρ_c and we can reduce the equation to

$$\dot{a}\,a^{1/2+3w/2}=H_0\,(a_0)^{3/2+3w/2}\,,\qquad H_0=\sqrt{\frac{8\pi G\rho_c}{3}}\,, \tag{9.1.29}$$

which has the solution

$$a=a_0\left(\frac{3(1+w)}{2}\,H_0\,t\right)^{\frac{2}{3(1+w)}}\,, \tag{9.1.30}$$

where we have assumed that $a = 0$ at $t = 0$. The Hubble constant as a function of time is then

$$H(t) = \frac{\dot{a}}{a} = \frac{2}{3(1+w)}\, t^{-1}\,, \tag{9.1.31}$$

which is inversely proportional to the age of the universe t. Another interesting parameter is the *deceleration parameter*

$$q = -\frac{\ddot{a}}{a}\left(\frac{\dot{a}}{a}\right)^{-2} = \frac{1+3w}{2}\,. \tag{9.1.32}$$

Let us now consider a few cases. One is a matter dominated universe with $w = 0$. In this case the Hubble constant and deceleration parameters are

$$H(t) = \frac{2}{3}t^{-1} \qquad q = \frac{1}{2}\,. \tag{9.1.33}$$

As we see, the expansion is positive ($H > 0$), but slowing down ($q > 0$). Using the present measured value of the Hubble constant, $H_0 = 70$ km/sec/Mpsc translates into an age $t_0 = \frac{2}{3}H_0^{-1} = 2.9 \times 10^{17}$ sec $= 9.3 \times 10^9$ years. However, there are galaxies that are older than this, so a matter dominated universe does not seem consistent with this Hubble constant.

For a radiation dominated universe we have $w = 1/3$ and so

$$H(t) = \frac{1}{2}t^{-1} \qquad q = 1\,. \tag{9.1.34}$$

Again this has a positive expansion which is decelerating. Given the present measured value of the Hubble constant, this would correspond to an age of the universe which is only 3/4 the age found for a matter dominated universe, which itself is too small. In any case, we know that there is now far more energy density in matter than in radiation.

Another interesting value for w is $w = -1/3$, in which case $q = 0$. For $w < -1/3$ then $q < 0$, which means the expansion rate is accelerating.

The vacuum dominated case has to be treated separately. In this case $w = -1$ and we see that the equation in (9.1.31) is singular. However, if we go back to (9.1.29) and set $w = -1$, we find that the solution is

$$a(t) = a_{\mathrm{V}} \exp(H_0 t),. \tag{9.1.35}$$

In this case the Hubble constant is really a constant with $H(t) = H_0$. Notice that the vacuum universe never has $a = 0$ in the finite past. However, if we combine a matter dominated or radiation dominated universe at early times with a vacuum dominated universe at later times then we can have a universe that started at $a = 0$ in the finite past.

9.2 Observing the Universe

The expansion of the universe affects the observations we make of stars, supernovae and galaxies. In this section we will study how this happens. We shall see that the expansion leads to a red shift where the red-shift depends on the distance the object is away from us. We will also see how measuring the brightness of supernovae helps us determine the composition of the universe.

9.2.1 Red-shifts, proper distances and the age of the universe

Astronomers describe red-shifts in terms of a quantity Z,

$$Z \equiv \frac{\lambda_{\rm obs}}{\lambda_{\rm em}} - 1\,, \tag{9.2.1}$$

where $\lambda_{\rm em}$ is the emitted wavelength and $\lambda_{\rm obs}$ is the observed wavelength. Let us assume that we are the observers at $r = 0$ and the light source is at comoving radial coordinate r. Before doing anything else, let us quickly show that a body with fixed co-moving coordinates satisfies the geodesic equation. If we assume that $\dot{r} = \dot{\theta} = \dot{\phi} = 0$, then

$$\ddot{r} + \Gamma^r_{tt}\,\dot{t}\dot{t} = \ddot{r} = 0\,, \tag{9.2.2}$$

with a similar equation for $\ddot{\theta}$ and $\ddot{\phi}$. Hence the body stays fixed on the co-moving coordinates.

Assume that the light source emits light with wavelength $\lambda_{\rm em}$. Let us further assume that the light is first emitted at t_1 and one cycle later the emitted time is $t_1 + \lambda_{\rm em}$. The geodesic for the light satisfies $\frac{dr}{dt} = -a^{-1}(t)\sqrt{1-kr^2}$ and so r is related to the time the light is first received, t_0 by

$$\int_0^r \frac{dr}{\sqrt{1-kr^2}} = \int_{t_1}^{t_0} \frac{dt}{a(t)}\,, \tag{9.2.3}$$

while after one cycle the relation is

$$\int_0^r \frac{dr}{\sqrt{1-kr^2}} = \int_{t_1+\lambda_{\rm em}}^{t_0+\lambda_{\rm obs}} \frac{dt}{a(t)}\,. \tag{9.2.4}$$

Assuming that $t_0 - t_1 >> \lambda_{\rm em}, \lambda_{\rm obs}$, and so $a(t_0 + \lambda_{\rm em}) \approx a(t_0)$, $a(t_1 + \lambda_{\rm obs}) \approx a(t_1)$, we then have from (9.2.3) and (9.2.4)

$$\frac{\lambda_{\rm obs}}{a(t_0)} - \frac{\lambda_{\rm em}}{a(t_1)} = 0\,, \tag{9.2.5}$$

and hence it follows that

$$Z = \frac{a(t_0)}{a(t_1)} - 1\,. \tag{9.2.6}$$

Intuitively we can understand this result as follows: While the co-moving distances have stayed constant, the physical distances, also known as the *proper distances* have increased linearly with a as a increased. In particular if the proper wavelength were $a(t_1)\Delta r$ at t_1, then it would be $a(t_0)\Delta r$ at t_0, giving the result in (9.2.6).

Let us now suppose that t_1 is not too far in the distance past, at least compared to the age of the universe. Then we can approximate Z as

$$Z \approx \frac{a(t_0)}{a(t_0)} + (t_0 - t_1)\frac{\dot{a}(t_0)}{a(t_0)} - 1 \approx H_0\, d\,, \tag{9.2.7}$$

where

$$d = a(t_0) \int_0^r \frac{dr}{\sqrt{1 - kr^2}} \tag{9.2.8}$$

is the proper distance between the light source and the observer and H_0 is today's Hubble constant. Over relatively short times the Hubble constant is actually close to being constant and so one sees that the red-shift is proportional to the proper distance between the light source and the observer. This approximation is valid when $Z << 1$. For example, for a source that is 10 Mpscs away we would have $Z = 700\,\mathrm{km/s} = 0.0023$. Hubble first measured H_0 by comparing the red-shifts of light sources to their distances. The original measurements were not very accurate because it is notoriously difficult to make accurate distance measurements. In fact, not that long ago the Hubble constant was usually quoted to be somewhere between 50 and 100 km/s/Mpsc. It is only over the last 15 years that the accuracy has dramatically improved.

For larger red-shifts the relation between the proper distance and the red-shift requires further information about $a(t)$. From (9.2.3) and (9.2.8) we see that the proper distance is given by

$$d = a_0 \int_{t_1}^{t_0} \frac{dt}{a(t)} = a_0 \int_{a_1}^{a_0} \frac{da}{\dot{a}a} = a_0 \int_{a_1}^{a_0} \frac{da}{H(a)a^2}\,, \tag{9.2.9}$$

where $a_0 = a(t_0)$ and $a_1 = a(t_1)$. $H(a)$ is determined from the Friedmann equation in (9.1.22) and hence we are left with an integral over a without any explicit t dependence. The integral will then only depend on ρ and Z. In fact, we can express H as $H(Z)$ by substituting (9.2.6) and (9.1.26) into the Friedmann equation. If we assume that there is only matter and vacuum energy, then we have

$$H^2(Z) = \frac{8\pi G}{3}\left(\rho_{\mathrm{m}}(Z+1)^3 + \rho_{\mathrm{v}}\right) - \tilde{k}(Z+1)^2\,, \tag{9.2.10}$$

where ρ_{m} and ρ_{v} are the present values of the matter and vacuum energy densities and where $\tilde{k} = k/(a_0)^2$. We can then turn the integral over a to an integral over Z by also using the relation between the differentials,

$$da = -a_0\frac{dZ}{(Z+1)^2}\,. \tag{9.2.11}$$

Actually, from (9.2.10) we see that the presence of $\tilde{k}$ has the same effect on $H(Z)$ and $d(Z)$ as having a perfect fluid with $w = -1/3$. To put this term on a more even footing with the matter and vacuum energy contributions we will introduce a quasi energy density ρ_{q} which is related to $\tilde{k}$ by

$$\rho_{\mathrm{q}} = -\frac{3}{8\pi G}\,\tilde{k}\,. \tag{9.2.12}$$

Putting everything together we can express the proper distance d as the integral

$$d = \left(\frac{3}{8\pi G}\right)^{1/2} \int_0^Z \frac{dZ'}{\sqrt{\rho_{\mathrm{m}}(Z'+1)^3 + \rho_{\mathrm{q}}(Z'+1)^2 + \rho_{\mathrm{v}}}}\,. \tag{9.2.13}$$

It is convenient to express these integrals in terms of the ratios of energy densities to the critical density, by defining $\Omega_{\mathrm{m}} \equiv \rho_{\mathrm{m}}/\rho_c$, $\Omega_{\mathrm{q}} \equiv \rho_{\mathrm{q}}/\rho_c$ and $\Omega_{\mathrm{v}} \equiv \rho_{\mathrm{v}}/\rho_c$, and so $\Omega_{\mathrm{m}} + \Omega_{\mathrm{q}} + \Omega_{\mathrm{v}} = 1$. The integral then has the form

$$d = \frac{1}{H_0} \int_0^Z \frac{dZ'}{\sqrt{\Omega_{\mathrm{m}}(Z'+1)^3 + \Omega_{\mathrm{q}}(Z'+1)^2 + \Omega_{\mathrm{v}}}}\,. \tag{9.2.14}$$

For small Z, we see that

$$d \approx \frac{1}{H_0} \int_0^Z \frac{dZ'}{\sqrt{\Omega_{\mathrm{m}} + \Omega_{\mathrm{q}} + \Omega_{\mathrm{v}}}} = \frac{Z}{H_0}\,, \tag{9.2.15}$$

which agrees with (9.2.7). For larger values of Z, the integral in (9.2.14) is in general an elliptic integral, so we will instead consider special cases. For instance, for a matter dominated flat universe with $\Omega_{\mathrm{q}} = \Omega_{\mathrm{v}} = 0$ the integral is

$$d = \frac{1}{H_0}\left(2 - \frac{2}{\sqrt{Z+1}}\right). \tag{9.2.16}$$

Observe that in the limit $Z \to \infty$, $d = \frac{2}{H_0}$. Notice that this is different than the universe's age in a matter dominated universe, $t_0 = \frac{2}{3H_0}$.

Let us next consider the case of a vacuum dominated flat universe with $\Omega_{\mathrm{m}} = \Omega_{\mathrm{q}} = 0$. In this case (9.2.14) is

$$d = \frac{1}{H_0}\,Z\,. \tag{9.2.17}$$

Hence, in this case the large Z behavior is the same as the small Z behavior. We can also see that for a fixed value of Z, d is larger for a vacuum dominated universe than a matter dominated universe.

If we have a combination of these last two examples, we can approximate the integrals by assuming that the universe is vacuum dominated up to a crossover red-shift Z_c and is matter dominated beyond that. Hence for a $Z > Z_c$ we can approximate d as

$$\begin{aligned} d \approx\ &= \frac{1}{H_0}\,Z_c + \frac{1}{H_0}\int_{Z_c}^{Z} \frac{dZ'}{\sqrt{\Omega_{\mathrm{m}}(Z'+1)^3}} \\ &\approx \frac{1}{H_0}\left(Z_c + 2(Z_c+1)\left(1 - \sqrt{\frac{Z_c+1}{Z+1}}\right)\right), \end{aligned} \tag{9.2.18}$$

where in this approximation we used that $\Omega_{\rm m}(1+Z_c)^3 \approx 1$.

Before closing this section let us make a final note about the universe's age in a universe that has a crossover from matter dominated to vacuum dominated at a later time. If the crossover time is at t_c and today's time is t_0, then using that H stays constant during the vacuum dominated period, we have that

$$t_c \approx \frac{2}{3H_0} \qquad\qquad t_0 = t_c + \Delta t\,, \tag{9.2.19}$$

where Δt is however long it has been in the vacuum dominated universe. We already know that $t_c \approx 9.3\times 10^9$ years, so we can add time onto the age of the universe by having an extended vacuum dominated period. The present estimate is that the universe is flat which means $\Omega_{\rm q} = 0$, while the vacuum and matter parts are $\Omega_{\rm v} = 0.74$ and $\Omega_{\rm m} = 0.26$. Assuming at the crossover time the densities were the same, we have that

$$\frac{\Omega_{\rm m}}{\Omega_{\rm v}} = \left(\frac{a(t_c)}{a(t_0)}\right)^3 = \exp(-3H_0\Delta t)\,. \tag{9.2.20}$$

Hence we find that

$$\Delta t \approx \frac{1}{3H_0}\log(7.4/2.6) \approx 4.8\times 10^9\,\text{years}\,. \tag{9.2.21}$$

Which puts the age of the universe to around 14 billion years old. From various experiments the agreed upon age is about 13.7 ± 0.13 billion years. To get this more accurate value, from the Friedmann equation we can express the age as

$$t_0 = \left(\frac{3}{8\pi G}\right)^{1/2}\int_0^{a_0}\frac{da}{a\sqrt{\rho(a)}}\,, \tag{9.2.22}$$

where $\rho(a) = \rho_c(\Omega_{\rm m}\left(\frac{a_0}{a}\right)^3 + \Omega_{\rm v})$. Defining a variable $y = a/a_0$ we can write the integral as

$$t_0 = \frac{1}{H_0}\int_0^1\frac{dy}{\sqrt{\Omega_{\rm m}\,y^{-1} + \Omega_{\rm v}\,y^2}}\,. \tag{9.2.23}$$

Taking the above values for $\Omega_{\rm m}$ and $\Omega_{\rm v}$, as well as the best measurement of the Hubble constant, which is $H_0 = 70.4 \pm 1.5$ km s^{-1}/Mpsc, and inserting these into (9.2.23) we find

$$t_0 = \frac{1}{H_0}\int_0^1\frac{dy}{\sqrt{(0.26)\,y^{-1} + (0.74)\,y^2}} \approx \frac{1}{H_0}(1.003) \approx 13.89 \pm 0.29\times 10^9\,\text{years} \tag{9.2.24}$$

where the integral has been done numerically.

9.2.2 Brightness

Suppose we have a light source with luminosity L, the total energy released by the object per unit time. The brightness as measured by an observer is equal to the flux, the energy per unit time per unit area. Assuming that the light source radiates equally in all directions, then the flux F measured by an observer today is

$$F = \frac{L}{4\pi(a_0 r)^2}\,\frac{1}{1+Z}\,\frac{1}{1+Z}\,. \tag{9.2.25}$$

Let us explain each of these factors. First, the flux should be inversely proportional to the area of a sphere whose physical radius, according to the Robertson-Walker metric in (9.1.8), is $a_0 r$. In the case where $k = 0$ then this is the proper distance d separating the source and the observer. If $k > 0$ then $a_0 r = (\tilde{k})^{-1/2}\sin((\tilde{k})^{1/2}d)$, while for $k < 0$ the relation is $a_0 r = (-\tilde{k})^{-1/2}\sinh((-\tilde{k})^{1/2}d)$. One factor of $(1+Z)^{-1}$ in (9.2.25) accounts for the red-shift of the radiation. Since the energy of a photon is inversely proportional to the wavelength, the energy flux has to be rescaled accordingly. The other factor of $(1+Z)^{-1}$ takes into account time dilation. The luminosity is proportional to the rate that photons are being emitted. But there is a time dilation factor that is proportional to the red-shift, so the photon emission rate goes down by the same factor.

It is convenient to express the flux in terms of the *apparent distance*, $D(Z)$

$$F = \frac{L}{4\pi(D(Z))^2}\,, \tag{9.2.26}$$

where

$$D(Z) = \frac{1}{H_0}(1+Z)|\Omega_{\mathrm{q}}|^{-1/2}S\big(|\Omega_{\mathrm{q}}|^{1/2}H_0\,d(Z)\big) \tag{9.2.27}$$

Here we used that $\tilde{k} = -\Omega_{\mathrm{q}}(H_0)^2$ which follows from (9.2.12). The function $S(x)$ is $S(x) = x$ for $k = 0$, $S(x) = \sin(x)$ for $k > 0$ and $S(x) = \sinh(x)$ for $k < 0$.

As an example, for a matter dominated flat universe we have from (9.2.16) that

$$\begin{aligned} F = \frac{L(H_0)^2/4}{4\pi(1+Z)^2\left(1-(1+Z)^{-1/2}\right)^2} &\approx \frac{L(H_0)^2}{4\pi Z^2} \qquad Z << 1 \\ &\approx \frac{L(H_0)^2}{16\pi Z^2} \qquad Z >> 1\,. \end{aligned} \tag{9.2.28}$$

For a vacuum dominated flat universe we have from (9.2.17) that the flux is

$$\begin{aligned} F = \frac{L(H_0)^2}{4\pi(1+Z)^2Z^2} &\approx \frac{L(H_0)^2}{4\pi Z^2} \qquad Z << 1 \\ &\approx \frac{L(H_0)^2}{4\pi Z^4} \qquad Z >> 1\,. \end{aligned} \tag{9.2.29}$$

Hence, for large values of Z the light source will appear dimmer in a vacuum dominated universe than in a matter dominated universe.

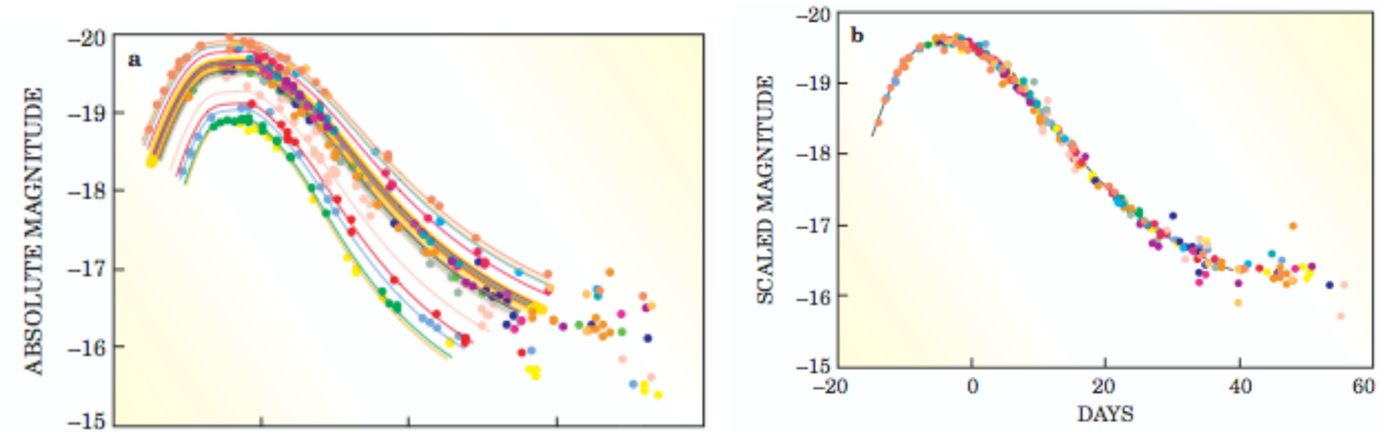

Figure 9.3: Light curves for various supernovae. The left graph shows that the shorter the lifetime the dimmer the supernova. The right graph shows the light curves after rescaling for their lifetime. The results are remarkably uniform. The absolute magnitude is how bright the supernova would appear if it were 10 parsecs away from us. An absolute brightness of -20 means that if the supernova were 10 parsecs away from us it would be more than 600 times brighter than the full moon (which has apparent magnitude -13).

Astronomers give brightness measurements in terms of *apparent magnitude*, m. For historical reasons its definition has now become

$$m \equiv -2.5\,\log_{10}(F/F_0) + m_0 \tag{9.2.30}$$

where m_0 and F_0 are the apparent magnitude and flux for some standard source. So for example, the star Vega has apparent magnitude $m = 0.03$. If we measured a light source with $m = 5.03$ then this object is 100 times dimmer than Vega. Given the form of the flux in (9.2.26), we see that the apparent magnitude for our light source is

$$m = +5\log_{10}\big(D(Z)\big) - 2.5\log_{10}(L/(4\pi F_0) + m_0\,. \tag{9.2.31}$$

Hence, by measuring m and by knowing L we can determine D. If we make the measurements for many different values of Z then $D(Z)$ can be determined and so then can the various energy densities.

The big problem is finding a light source whose luminosity is known. In order to see over significant red-shifts, this light source also has to be very bright. Over the last 15 or 20 years such a light source has been identified. These are what are known as type Ia supernovae. They are a particular type of exploding star whose luminosities as a function of the time following the explosion have become well understood. Some supernovae are brighter than others, but it was observed that the brighter ones stayed brighter longer and that a simple rescaling of luminosity by the supernova lifetime gave a remarkably uniform result. The plots of luminosity versus time for these supernovae are known as their *light curves* (see figure 9.3). So by observing a supernova over its lifetime (typically weeks) one can determine a "standard candle", that is its luminosity. Moreover, supernovae are very bright. They are often brighter than the galaxy they sit in and the Hubble space telescope has been able to observe supernovae with Z up to 4.

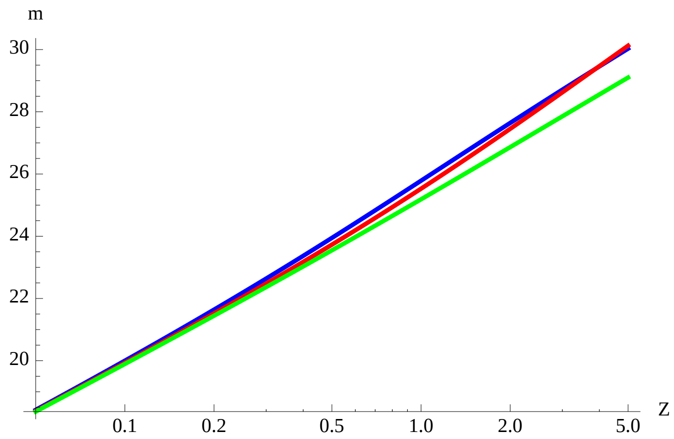

Figure 9.4: Hubble plots for $(\Omega_{\rm m},\Omega_{\rm v})$=(0.3, 0.7) [blue)], (0.3, 0.0] [red], and (1.0, 0.0) [green]. On the vertical axis we show the change in brightness from a source at $Z = 0.05$. The Z axis is shown on a log scale with $0.05 < Z < 5$. The middle case has $k < 0$ while the other two have $k = 0$.

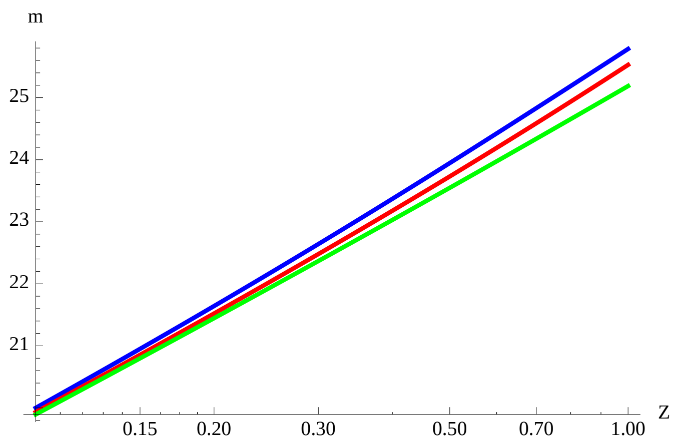

Figure 9.5: The same Hubble plots as in figure 9.4 but over the narrower range $0.1 < Z < 1.0$. Here there is a clearer distinction between the different scenarios.

In figure 9.4 we show plots of magnitude versus Z for 3 possible scenarios for the energy density of the universe. This type of plot is known as a *Hubble plot*. The particular cases here have $(\Omega_{\rm m},\Omega_{\rm v})$ given by $(0.3, 0.7)$, $(0.3, 0)$ and $(1.0, 0.0)$. The first and the third are flat since they have $\Omega_{\rm m} + \Omega_{\rm v} = 1$, while the middle case is open since it has $\Omega_{\rm m} + \Omega_{\rm v} < 1$. The magnitudes are those for rescaled supernovae brightness, where $Z = 0.1$ is about $m = 20$. The plot shows Z on a log scale ranging from 0.05 to 5. Notice that once Z becomes greater than 1 the first and second plots start to approach other. Hence, the best place to distinguish between these two behaviors is for $0.1 < Z < 1$. Figure 9.5 shows the same Hubble plots in this narrower range. Another useful plot is shown in figure 9.6 which shows the differences in magnitude between the three Hubble plots and the middle one.

The actual data is shown in figure 9.7, which combines the results of two major groups, the High-Z Supernova Search Team which uses the Hubble Space Telescope (HST) and

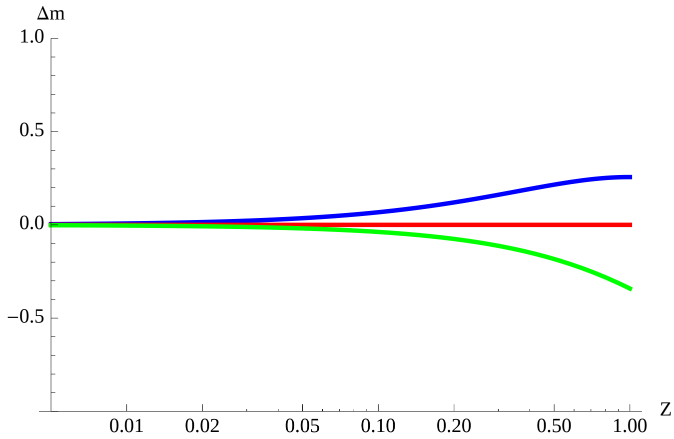

Figure 9.6: The same Hubble plots as in figure 9.4 but shown as a difference between the $(\Omega_{\rm m}, \Omega_{\rm v}) = (0.3,\, 0.0)$ case and the other cases.

the Supernova Cosmology Project using various telescopes including the HST. While it might not seem completely obvious from the plots, there is a huge statistical significance that the points on the plot follow the curve for a vacuum dominated universe with $\Omega_{\rm v} \approx 0.70$. This is made more clear in the difference plot at the bottom, where clearly more points are above the 0.0 line than below it.

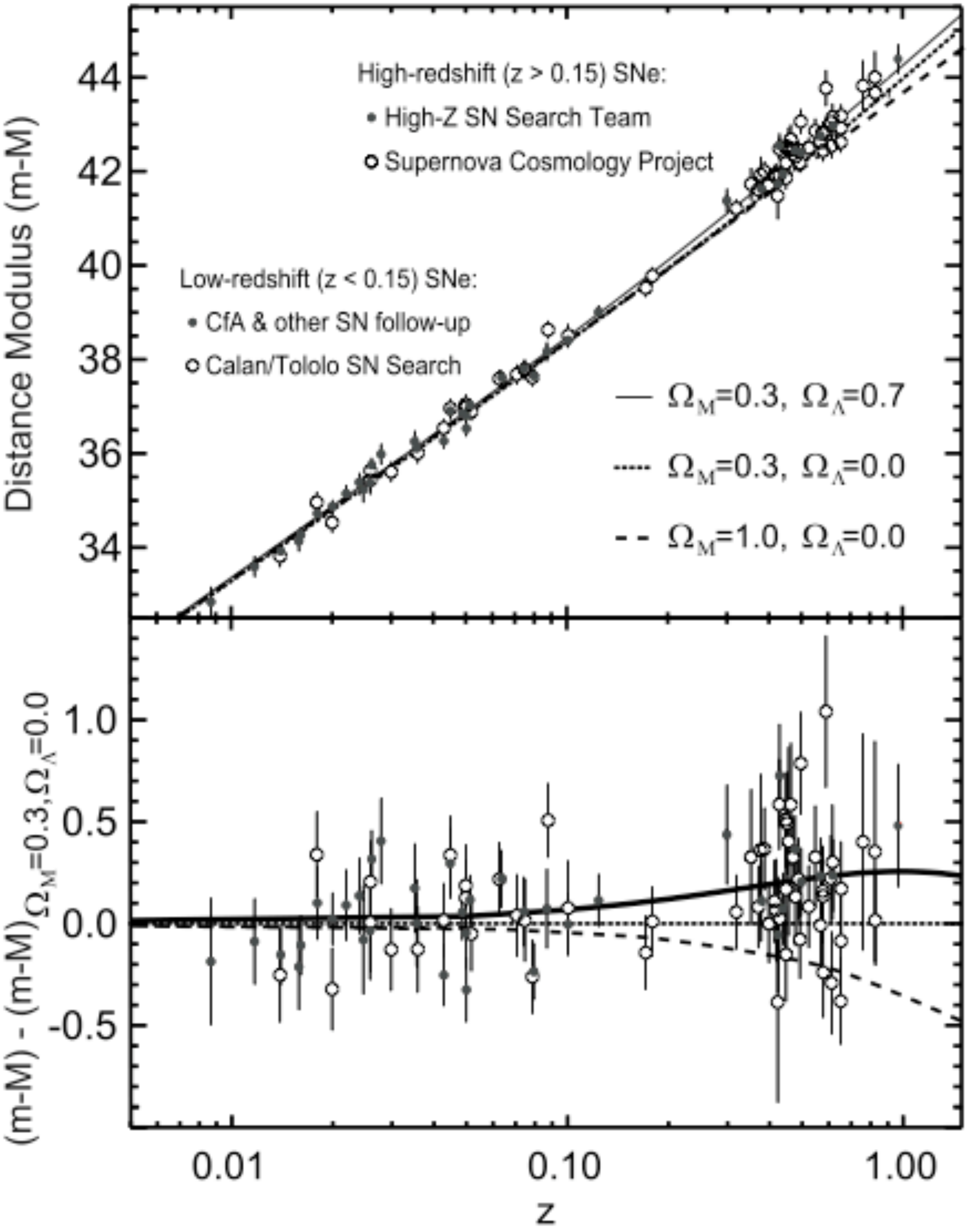

Figure 9.7: Hubble plots from the actual data. The data points track closest to the (0.3, 0.7) plot.

Chapter 10

The thermal history of the universe

In this chapter we discuss the thermal history of the universe. This includes the blackbody spectrum of the cosmic microwave background.

10.1 The thermal universe

10.1.1 Photons in the early universe

Since the universe has been expanding for a long while, at some time in the past it must have been highly compressed. This means that the particles in the universe were very close together and thus interacting with each other. Since the particles could interact they would be continuously exchanging their energies and hence would be expected to be in thermal equilibrium.

Among the particles that were interacting are the photons. Photons are massless bosons that have two degrees of freedom, its two independent polarizations. Assuming that the photons were in a thermal bath with temperature T, then their energy density per wavelength λ would have been given by

$$\rho(\lambda,T)=\frac{4\pi hc\,g_*}{\lambda^5}\frac{1}{\exp(hc/\lambda kT)-1} \tag{10.1.1}$$

where g_* is the number of degrees of freedom which is two if only photons contribute to $\rho(\lambda,T)$, h is Planck's constant and k is the Boltzmann constant. $\rho(\lambda,T)$ is Planck's energy density distribution for a blackbody. $\rho(\lambda,T)$ is peaked where $\frac{\partial}{\partial\lambda}\rho(\lambda,T)=0$, which occurs when

$$\frac{\partial}{\partial\lambda}\rho(\lambda,T)=\frac{4\pi hc\,g_*}{\lambda^5}\frac{1}{\exp(hc/\lambda kT)-1}\left(-\frac{5}{\lambda}+\frac{hc}{\lambda^2 kT}\frac{\exp(hc/\lambda kT)}{\exp(hc/\lambda kT)-1}\right)=0 \tag{10.1.2}$$

This can be solved numerically where one finds that the peak is at

$$\lambda_c=\frac{hc}{bkT}\qquad b\approx 4.9651\,. \tag{10.1.3}$$

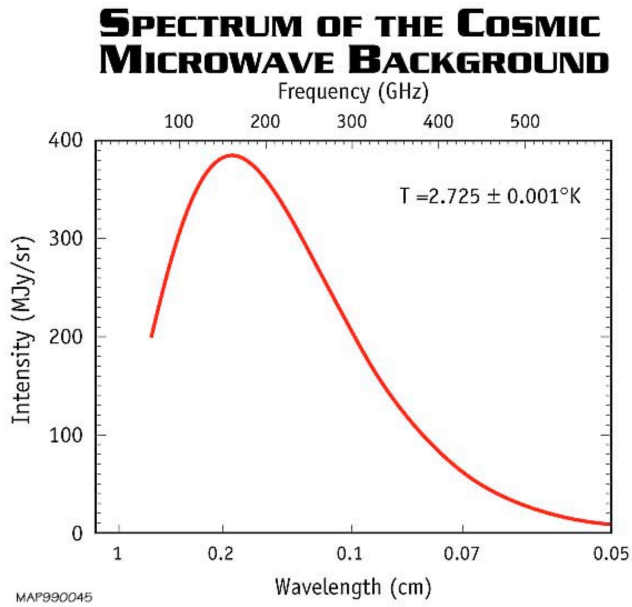

Figure 10.1: The blackbody spectrum of the CMB as observed by Cobe.

The integrated energy density is

$$\rho(T) = \int_0^\infty \rho(\lambda, T)\, d\lambda = \frac{\pi^2 g_*}{30} \frac{k^4}{(\hbar\, c)^3} T^4 = \frac{2 g_*}{c}\, \sigma\, T^4\,, \tag{10.1.4}$$

where $\sigma = 5.67 \times 10^{-8}$ watts m^{-2} K^{-4} is the Stefan-Boltzmann constant.

It is often common to present the data in terms of $\tilde{\rho}(\nu, T)$ where ν is the frequency and $\rho(T) = \int_0^\infty d\nu\, \tilde{\rho}(\nu, T)$. In this case the peak in the frequency is at

$$\nu_{\tilde{c}} = \frac{kT}{\tilde{b}h} \qquad \tilde{b} \approx 2.821\,. \tag{10.1.5}$$

As the universe expands, the wavelengths of the photons are redshifted and hence the photons and the particles they interact with cool down. Eventually, the temperature is low enough that the electrons recombine with the protons to form hydrogen atoms and there are no longer charged particles available to scatter with the photons. These leftover photons can then essentially travel forever without scattering until we see them in our detectors.

Let us say that the temperature that this recombination occurs is at $T = 3740$ K, which corresponds to $kT_{rec} = 0.323$ eV. Let us call the universe's age at this point t_{rec}. Then at time t_{rec}, the photons have the energy density spectrum $\rho(\lambda, T_{rec})$. After this time the photons stop scattering, but their energy density will stay intact although it will be redshifted. In particular, a wavelength λ at t_{rec} becomes $\frac{a(t)}{a(t_{rec})}\,\lambda$ at a later time

t. The combination $\rho(\lambda, T_{rec})\, d\lambda$ then rescales to

$$\begin{aligned}\rho(\lambda, T_{rec})\, d\lambda \ &\rightarrow\ \left(\frac{a(t_{rec})}{a(t)}\right)^4 \rho\left(\frac{a(t_{rec})}{a(t)}\,\lambda, T_{rec}\right)\left(\frac{a(t_{rec})}{a(t)}\, d\lambda\right. \\ &= \frac{4\pi hc\, g_*}{\lambda^5}\,\frac{1}{\exp(hc\, a(t)/a(t_{rec})\lambda k T_{rec}) - 1}\, d\lambda \\ &= \rho\left(\lambda, \frac{a(t_{rec})}{a(t)}\, T_{rec}\right)\left(d\lambda\right., \end{aligned} \tag{10.1.6}$$

where we multiplied by the scale factor for the radiation energy density and we replaced λ in $\rho(\lambda, T)$ with its value at t_{rec}. Hence, the new energy density still has the form of a blackbody, but now the temperature is red-shifted to $\frac{a(t_{rec})}{a(t)}\, T_{rec}$. Hence, today we should expect to see the remnant radiation from this period when the photons were scattering off of the electrons and protons just before recombination. This radiation is called the *cosmic microwave background* (CMB). Figure 10.1 shows the actual spectrum for the CMB as observed by the COBE satellite. What is presented is the intensity per frequency which is given by $\frac{1}{4}\, c\, \tilde\rho(\nu, T)$. Eyeballing the figure you can see that the peak is at $\nu_{\tilde c} = 160$ GHz which corresponds to a wavelength of $\lambda_{\tilde c} = 1.875$ mm. Plugging the frequency into (10.1.3) we find that the temperature today is

$$T_0 = \frac{(6.626\times 10^{-34}\,\text{j-s})(160\times 10^9\,\text{s}^{-1})}{(2.821)(1.381\times 10^{-23}\,\text{j}\,\text{K}^{-1})} = 2.73\,\text{K}\,. \tag{10.1.7}$$

As you can see from the figure, T_0 is known even more accurately than our estimate since the position of the peak can be more accurately determined.

Comparing T_0 to T_{rec}, we then find that Z at recombination is

$$Z_{rec} = \frac{T_{rec}}{T_0} - 1 \approx \frac{3740}{2.73} - 1 \approx 1369\,. \tag{10.1.8}$$

Hence, these photons must have been last scattered when the universe was very young. To find out how young this was, let us suppose that the universe was matter dominated at recombination (we will justify this soon). From the Friedmann equation and analogously to eq. (2.22) in chapter 9, we have that t_{rec} is

$$\begin{aligned} t_{rec} \ &=\ \left(\frac{3}{8\pi G}\right)^{1/2}\int_0^{a_{rec}} \frac{da}{a\sqrt{\rho(a)}} \\ &= \frac{1}{H_0}\int_0^{a_{rec}/a_0} \frac{dy}{\sqrt{\Omega_{\rm m}\, y^{-1} + \Omega_{\rm v}\, y^2}} \\ &\approx \frac{1}{H_0\sqrt{\Omega_{\rm m}}}\int_0^{Z_{rec}^{-1}} y^{1/2}\, dy = \frac{2}{3H_0\sqrt{\Omega_{\rm m}}\, Z_{rec}^3}\,. \end{aligned} \tag{10.1.9}$$

Using $H_0 = 70$ km/s/Mpsc, $\Omega_{\rm m} = .26$ and $Z_{rec} = 1369$ gives

$$t_{rec} = \frac{2\,(3\times 10^8\,\text{m/s})}{3(7\times 10^4\,\text{m/s}\,\text{Mpsc}^{-1}\sqrt{.26(1369)^3}} \approx 3.6\times 10^5\,\text{years}\,. \tag{10.1.10}$$

Hence the age of the universe at recombination is a small fraction of its age today.

10.1.2 Other contributions to g_*

Up to now we have been assuming that the radiation inside the universe is only comprised of photons. But there are other light particles in nature that are relativistic throughout the history of the universe. Their contribution to the energy density is partly determined by when these particles fell out of thermal equilibrium. For example, in addition to the photons there should be neutrinos present as well. We can take account of these extra degrees of freedom by modifying the g_* that appears in (10.1.4) to

$$g_* = \sum_{\text{bosons}} g_i \left(\frac{T_i}{T_0}\right)^4 + \frac{7}{8} \sum_{\text{fermions}} g_i \left(\frac{T_i}{T_0}\right)^4 \tag{10.1.11}$$

where g_i is the number of degrees of freedom of each species and T_i is its temperature today. Fermions have the extra factor of 7/8 because they have Fermi-Dirac statistics and so their spectra have the form

$$\rho_i(\lambda, T) = \frac{4\pi hc\, g_i}{\lambda^5} \frac{1}{\exp(hc/\lambda kT) + 1} \,. \tag{10.1.12}$$

The + sign in the denominator then leads to the 7/8 factor after doing the integral in (10.1.4). Note that the particles can have different temperatures today because they are no longer interacting with each other, so they cannot exchange energies to equalize their temperatures if they are different.

Aside from the photons, the only other particles contributing to g_* today are the neutrinos. There are three types of neutrinos (ν_e, ν_μ, ν_τ) and antineutrinos ($\bar{\nu}_e$, $\bar{\nu}_\mu$, $\bar{\nu}_\tau$), where the neutrinos are lefthanded meaning that their spin is antialigned with their momentum while the antineutrinos are righthanded meaning that their spin is aligned with its momentum. Hence the number of degrees of freedom from all neutrino and antineutrino species is $g_i = 6$.

We next derive the neutrino temperature. The neutrinos are weakly interacting and so they stopped interacting with other particles at a temperature around $kT = 1$ MeV. At temperatures above 1 MeV electrons and positrons are highly relativistic since their mass is 0.511 MeV and so they would have been contributing to g_*. Now the entropy density s is given by

$$s(T) = \frac{1}{T}\big(\rho(T) - f(T)\big) \tag{10.1.13}$$

where $f(T)$ is the free energy density. For highly relativistic particles $f_i(T) = -\frac{1}{3}\rho_i(T)$, where i labels the particle species. Hence the contribution to the entropy from species i is $s_i(T) = \frac{4}{3}\rho_i(T)/T$. At temperatures above $kT = 1$ MeV, the contribution to the entropy density of the electrons and positrons will therefore be

$$s_{ep}(T) = \tfrac{7}{8}\tfrac{4}{2}\, s_\gamma(T) \,, \tag{10.1.14}$$

where $s_\gamma(T)$ is the entropy density of the photons and the factor of 4 counts the number of degrees of freedom for the electrons and positrons (2 spins for each). As kT drops below

1 MeV, it is no longer energetically favorable to have electron positron pairs, so they will annihilate into photons. Let us say that this conversion takes place at temperature T_a. However, during this annihilation process the entropy density cannot drop so all of the entropy goes into the photons, forcing them to heat up. Hence we have

$$s_{ep}(T_a) + s_\gamma(T_a) = \frac{11}{4} s_\gamma(T_a) = s_\gamma(T_{rh})\,, \tag{10.1.15}$$

where T_{rh} is the temperature of the reheated photons after the annihilation. Since the entropy densities are proportional to T^3, we have that $T_{rh} = \left(\frac{11}{4}\right)^{1/3} T_a$. However, the neutrinos, which are no longer in thermal equilibrium will remain at the temperature T_a since there cannot be any energy transfer over to them. Hence we learn that for the neutrinos, $T_i = \left(\frac{11}{4}\right)^{-1/3} T_0$. Applying this result, we find that the effective number of degrees of freedom g_* is

$$g_* = 2 + \frac{7}{8} 6 \left(\frac{4}{11}\right)^{4/3} = 3.36\,. \tag{10.1.16}$$

Let us now justify that the universe was matter dominated at recombination. The energy density of the photons and neutrinos at a temperature T_{rec} is given by

$$\begin{aligned} \rho_{\mathrm{r}}(T_{rec}) &= \frac{2g_*}{c}\,\sigma\, T^4_{rec} = \frac{2(3.36)(5.67\times 10^{-8}\,\mathrm{watts\,m^{-2}\,K^{-4}})(3740\,\mathrm{K})^4}{3\times 10^8\,\mathrm{m\,s^{-1}}} \\ &= 0.249\,\mathrm{joules\,m^{-3}} = 2.76\times 10^{-18}\,\mathrm{kg\,m^{-3}}\,. \end{aligned} \tag{10.1.17}$$

The matter density at recombination was

$$\rho_{\mathrm{m}}(T_{rec}) = \rho_c\,\Omega_{\mathrm{m}}\,Z^3_{rec} = 9.2\times 10^{-27}\,\mathrm{kg\,m^{-3}}(0.26)(1369)^3 \approx 6.2\times 10^{-18}\,\mathrm{kg\,m^{-3}} \tag{10.1.18}$$

and so $\rho_{\mathrm{m}}(T_{rec}) > \rho_{\mathrm{r}}(T_{rec})$, although not by much.

To see when there was crossover to the radiation dominated period at time t_{rc}, we note that this occurs when

$$\rho_{\mathrm{r}}(T) = \rho_{\mathrm{m}}(T) \qquad \Rightarrow \qquad \frac{2g_*}{c}\sigma\, T^4 = \rho_c\Omega_{\mathrm{m}}\left(\frac{T}{T_0}\right)^3 \tag{10.1.19}$$

and so

$$Z \approx \frac{T}{T_0} = \frac{c\rho_c\Omega_{\mathrm{m}}}{2g_*\sigma(T_0)^4} = \frac{(3\times 10^8\,\mathrm{m\,s^{-1}})(9.2\times 10^{-27}\,\mathrm{kg\,m^{-3}})(0.26)}{2(3.36)(5.67\times 10^{-8}\,\mathrm{watts\,m^{-2}\,K^{-4}})(2.73\,\mathrm{K})^4} \approx 3075 \tag{10.1.20}$$

Hence, if we do the same calculation as we did for t_{rec} we find

$$t_{rc} \approx \frac{2(3\times 10^8)\,\mathrm{m/s}}{3(7\times 10^4\,\mathrm{m/s\,Mpsc^{-1}}\sqrt{.26(3075)^3}} \approx 1.1\times 10^5\,\mathrm{years}\,. \tag{10.1.21}$$

10.1.3 The recombination temperature

In a previous section we stated that $kT_{rec} = 0.323$ eV. Here we compute this result. The idea is that the matter in the universe, at least the matter that we can see is mainly made up of electrons, protons and neutrons. We will ignore the neutrons for now and consider the electrons and protons only.

An electron can combine with the proton to form a hydrogen atom, and in order to conserve energy and momentum a photon must also be created. Likewise, an energetic photon can hit an hydrogen atom, ionizing it into a separate electron and proton. Hence we have the chemical reactions

$$\mathbf{H} + \gamma \leftrightarrow \mathbf{e}^- + \mathbf{p}\,. \tag{10.1.22}$$

The number densities of the $\mathbf{H}$, $\mathbf{e}^-$ and $\mathbf{p}$ are all given by the Boltzmann density

$$n_i(T) = \iiint \frac{d^3p}{(2\pi\hbar)^3} e^{-p^2/2m_i kT} e^{(\mu_i - m_i)/kT} = \left(\frac{m_i kT}{2\pi\hbar^2}\right)^{3/2} e^{(\mu_i - m_i)/kT}\,, \tag{10.1.23}$$

where m_i is the mass and μ_i is the chemical potential for whichever of the three we are discussing. We could also include a factor for the spin degrees of freedom, but its effect will divide out below so we ignore it. The chemical potential of the photons is zero since they are not conserved. Thus, if the particles are in chemical equilibrium then their chemical potentials satisfy the relation $\mu_{\mathrm{H}} = \mu_{\mathrm{e}} + \mu_{\mathrm{p}}$. Therefore, we find that

$$\frac{n_{\mathrm{H}}(T)}{n_{\mathrm{e}}(T)n_{\mathrm{p}}(T)} = \left(\frac{2\pi\hbar^2 m_{\mathrm{H}}}{m_{\mathrm{e}} m_{\mathrm{p}} kT}\right)^{3/2} e^{Q/kT}\,, \tag{10.1.24}$$

where $Q = m_{\mathrm{e}} + m_{\mathrm{p}} - m_{\mathrm{H}} = 13.6$ eV is the binding energy of the hydrogen atom.

We next define the quantity $X = \frac{n_{\mathrm{p}}(T)}{n_{\mathrm{B}}(T)}$, where $n_{\mathrm{B}}(T) = n_{\mathrm{p}}(T) + n_{\mathrm{H}}(T)$ is the baryon number density (protons and neutrons both have baryon number 1, as does the hydrogen atom since it has one proton in the nucleus). We further assume that $n_{\mathrm{e}} = n_{\mathrm{p}}$ since the total charge is zero and also that $m_{\mathrm{H}} \approx m_{\mathrm{p}}$. Hence, we find the equation for X,

$$\frac{1-X}{X^2} = n_{\mathrm{B}}(T)\left(\frac{2\pi\hbar^2}{m_{\mathrm{e}} kT}\right)^{3/2} e^{Q/kT}\,. \tag{10.1.25}$$

This equation is known as the *Saha equation.* We then assume that at recombination, $n_{\mathrm{H}} = n_{\mathrm{p}}$, and so $X = 1/2$. The energy density today of the baryons is about 4-5% of ρ_c, which gives a baryon number density of about

$$n_{\mathrm{B}}(T) \approx (0.04)\frac{\rho_c}{m_{\mathrm{p}}}\left(\frac{T}{T_0}\right)^3 = (0.04)\frac{9.2\times10^{-27}\,\mathrm{kg\,m^{-3}}}{1.67\times10^{-27}\,\mathrm{kg}}\left(\frac{T}{T_0}\right)^3 \approx 0.22\,\mathrm{m}^{-3}\left(\frac{T}{T_0}\right)^3, \tag{10.1.26}$$

although the exact number is not terribly important for determing T_{rec}. Plugging in the numbers for $m_{\mathrm{e}} = 9.11\times10^{-31}$ kg, $\hbar = 1.05\times10^{-34}$ joules-s, and $kT_0 = 2.35\times10^{-4}$ eV we find that $\frac{2\pi\hbar^2}{m_{\mathrm{e}}(kT_0)^2} = 8.6\times10^{-12}\,\mathrm{eV}^{-1}\,\mathrm{m}^2$. This leaves us with the equation

$$2 = (0.22)(8.6\times10^{-12}\,\mathrm{eV}^{-1})^{3/2}(kT_{rec})^{3/2}\exp((13.6\,\mathrm{eV})/kT_{rec})\,, \tag{10.1.27}$$

which can be solved numerically to give $kT_{rec} \approx 0.323$ eV. Note that the actual baryon density does not have to be too accurate because of the sharp exponential behavior coming from the $e^{Q/kT}$ term. If we had used 5% instead of 4% the answer would have been $kT_{rec} = 0.325$ eV.

10.1.4 Nucleosynthesis

Not all of the baryons in the universe are protons. There are also neutrons, all of which are contained in the nuclei of heavier atoms. Most of the neutrons are in fact in helium atoms. Helium is a byproduct of stellar burning, but most of the helium in the universe was produced soon after the big bang during a stage called nucleosynthesis. In this section we will estimate the density of this *primordial* helium.

When the universe was very hot, there was thermal equilibrium between neutrons, protons, neutrinos, electrons and positrons. Among the interactions that thermalized this soup were

$$\begin{aligned} \mathbf{n} + \nu_e &\leftrightarrow \mathbf{e}^- + \mathbf{p} \\ \mathbf{n} + \mathbf{e}^+ &\leftrightarrow \bar{\nu}_e + \mathbf{p} \end{aligned} \tag{10.1.28}$$

where ν_e and $\bar{\nu}_e$ represent a neutrino and antineutrino respectively of electron type. While in thermal equilibrium, the number density of neutrons to protons is

$$\frac{n_{\mathrm{n}}}{n_{\mathrm{p}}} = e^{-Q/kT}\,, \tag{10.1.29}$$

where Q is the mass difference, $Q = m_{\mathrm{n}} - m_{\mathrm{p}} = 1.29$ MeV. Hence, it would seem that once kT fell well below $1.29MeV$ the number of neutrons to protons would be negligibly small.

However, the interactions in (10.1.28) are weak interactions, meaning that do not occur too often. In fact the interaction rate Γ for a neutron to turn into an electron and a proton is given by

$$\Gamma = n_\nu \sigma_{weak} c\,, \tag{10.1.30}$$

where n_ν is the number density of the neutrinos and σ_{weak} is the weak cross-section for this interaction. As the universe cools the neutrino number density falls off as $n_\nu \sim T^3$ while the cross-section falls off as $\sigma_{weak} \sim T^2$, hence the rate falls off as $\Gamma \sim T^5$. Since the universe is radiation dominated during this epoch, we have that $a(t) \sim t^{1/2}$ (see section (1.5) in chapter 9). Therefore, $\Gamma \sim t^{-5/2}$ since $a(t)$ is inversely proportional to T. Because the rate falls off faster than $1/t$, only a finite number of the neutrons will interact with the neutrinos beyond a certain time. This time can be reasonably estimated as the point where the interaction rate Γ equals the expansion rate $H = \frac{1}{2}t^{-1}$ for a radiation dominated universe. After this point the neutrons stop interacting and they essentially *freeze out*.

To estimate this temperature, we have that the cross-section is approximately,

$$\sigma_{weak} \sim 10^{-47}\,\mathrm{m}^2 \left(\frac{kT}{1\,\mathrm{MeV}}\right)^2 \tag{10.1.31}$$

while the electron neutrino density is

$$n_\nu = \int_0^\infty d\lambda \frac{4\pi}{\lambda^4}\frac{1}{\exp(hc/\lambda kT)+1} = \frac{3\,\zeta(3)}{4\pi^2}\left(\frac{kT}{\hbar c}\right)^3 . \tag{10.1.32}$$

From the Friedmann equation we also have that

$$H = \sqrt{\frac{8\pi G}{3}\rho_{\mathrm{r}}(T)}\,, \tag{10.1.33}$$

where $\rho_{\mathrm{r}}(T)$ has the form in (10.1.4). At these temperatures g_* is approximately

$$g_* \approx 2 + \tfrac{7}{8}(6+2+2) \approx 11\,, \tag{10.1.34}$$

where we have counted 6 neutrino species and the two spins for the electrons and the positrons. Hence, we find that

$$(kT)^3 \approx \sqrt{\frac{4\pi^3 G\, g_*\, \hbar^3}{45\, c}}\,\frac{4\pi^2}{3\,\zeta(3)} \times 10^{47}\,\mathrm{MeV}^2\,\mathrm{m}^{-2}\,, \tag{10.1.35}$$

where we have inserted some factors of c to convert kg to joules. Plugging in numbers we find that

$$(kT)^3 \approx 19\,\mathrm{MeV}^3 \qquad kT \approx 2.7\,\mathrm{MeV}\,. \tag{10.1.36}$$

Actually, our approximations have been a little too crude, the actual temperature for the freezeout is closer to $kT \approx 0.8$ MeV. Hence, we expect that the number density of neutrons to protons is $n_{\mathrm{n}} \approx n_{\mathrm{p}}\exp(-1.29/0.8) \approx 0.2\, n_{\mathrm{p}}$.

We can estimate the age of the universe when the freeze-out occurred. Using that $t = \frac{1}{2}H^{-1}$ we have that

$$\begin{aligned} t &= \frac{1}{2}\sqrt{\frac{45\,\hbar^3\, c^5}{4\pi^3 G\, g_*\,(kT)^4}} = \frac{1}{2}\sqrt{\frac{45\,(1.05\times 10^{-34}\,\text{j-s})^3\,(3\times 10^8\,\mathrm{m\,s^{-1}})^5}{4\pi^3(6.67\times 10^{-11}\,\mathrm{m^3\,kg^{-1}\,s^{-2}}\,(11)\,(1.29\times 1.6\times 10^{-13}\,\mathrm{j})^4}} \\ &\sim 1\ \mathrm{sec}\,. \end{aligned} \tag{10.1.37}$$

After the freeze-out the neutrons can interact with the protons to form heavier nuclei. The process goes forward, first making deuterium nuclei and then tritium, before producing helium. After helium the process more or less stops since there are no stable nuclei with 5 nucleons. Hence, we expect that most of the neutrons end up in helium. Since helium has 2 protons and 2 neutrons, we expect that the number of helium atoms to hydrogen atoms is approximately

$$n_{\mathrm{He}} \approx \frac{n_{\mathrm{n}}/2}{n_{\mathrm{p}} - n_{\mathrm{n}}} n_{\mathrm{H}} \approx \frac{1}{8}\, n_{\mathrm{H}}\,. \tag{10.1.38}$$

This puts the mass fraction of helium for all baryons to be

$$Y \equiv \frac{4n_{\mathrm{He}}}{4n_{\mathrm{He}} + n_{\mathrm{H}}} = \frac{1}{3}\,. \tag{10.1.39}$$

The observed value is about $Y \approx 0.24$. With a more careful analysis along the lines of the recombination computations it is possible to get this value, but we won't do this here.

Chapter 11

Cosmology continued

In this chapter we discuss the uniformity of the cosmic microwave background and introduce the concept of inflation.

11.1 Inflation

11.1.1 The particle horizon

Recall from chapter 9 that the proper distance in a spatially flat universe is given by

$$d = \frac{1}{H_0}\int_0^Z \frac{dZ'}{\sqrt{\Omega_{\rm m}(Z'+1)^3+\Omega_{\rm v}}}\,. \tag{11.1.1}$$

The maximum proper distance, d_0 that a light ray can travel over the age of the universe is then found by setting $Z=\infty$. If we use the values $\Omega_{\rm m}=0.26$ and $\Omega_{\rm v}=0.74$ then we find that $d_0 = 3.50\,H_0^{-1} = 3.50\,(3\times10^8\,\mathrm{m\ s^{-1}})/(7\times10^4\,\mathrm{m\ s^{-1}\,Mpsc^{-1}}) \approx 15,000$ Mpsc. We call this the *particle horizon* of the universe today, or just the horizon.

Now let us go back to the time of recombination and compute the particle horizon then. In this case the proper distance traveled is

$$\begin{aligned} d_{rec} &= \frac{1}{H_0}\frac{a_{rec}}{a_0}\int_{Z_{rec}}^{\infty}\frac{dZ'}{\sqrt{\Omega_{\rm m}(Z'+1)^3+\Omega_{\rm v}}} \\ &\approx \frac{1}{Z_{rec}H_0\sqrt{\Omega_{\rm m}}}\int_{Z_{rec}}^{\infty} Z^{-3/2}dZ = \frac{2}{H_0\sqrt{\Omega_{\rm m}Z_{rec}^3}}\,. \end{aligned} \tag{11.1.2}$$

If we compare the coordinate distances of these particle horizons, we find that today it is given by $\Delta r_0 = d_0/a_0$, while at recombination it was $\Delta r_{rec} = d_{rec}/a_{rec}$. Hence their ratios is given by

$$\delta = \frac{\Delta r_{rec}}{\Delta r_0} = \frac{2}{(3.5)\sqrt{\Omega_{\rm m}Z_{rec}}} \approx \frac{2}{(3.5)\sqrt{(0.26)(1369)}} \approx 3.0\times10^{-2}\,. \tag{11.1.3}$$

The result in (11.1.3) presents a problem. Supposedly at recombination the photons were in thermal equilibrium. But in order for the photons in one region of the universe

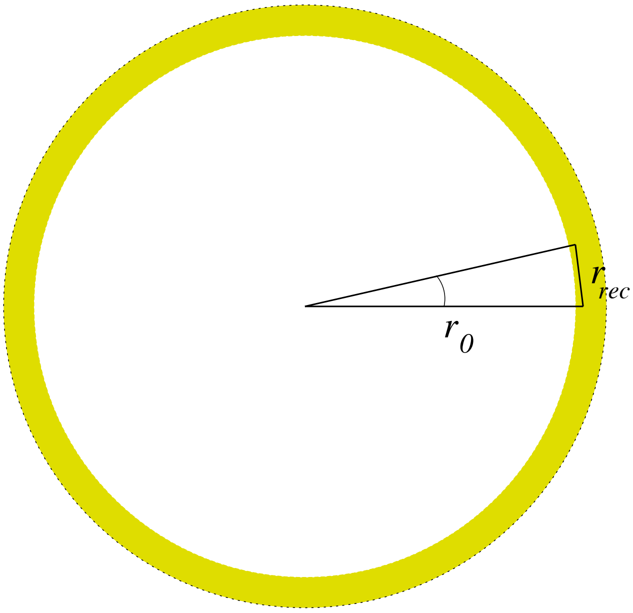

Figure 11.1: Particle horizons at recombination and today in terms of the comoving coordinates. The angle in the sky inside the recombination particle horizon is δ.

to have the same temperature as those in another region, it must have been possible for some particles to travel from one region to another, or to at least scatter off particles that scattered off of other particles, etc. until energy was transfered from the first region to the second or *vice versa*. However, the absolute fastest that such an energy transfer can happen is the speed of light. Hence, we seem to have no reason to expect that over coordinate distances greater than r_{rec} that the photon spectrum would have the same temperature. The radiation after the photons last scatter emerge from regions of the universe that are more or less fixed in the comoving coordinates. This means that for an observer looking at the CMB today, we would expect the temperature to be uniform only over an angular separation δ or less (see figure 11.1). From (11.1.3) we find that $\delta \approx 1.7$ degrees. But what do we see? The temperature is remarkably uniform over the entire sky, with a relative deviation in temperature $\Delta T \sim 10^{-5} T_0$. Therefore, it seems that the entire region of the sky was at one time in thermal equilibrium, even though this does not seem possible for the scenario we have described.

11.1.2 Inflation and the particle horizon

To find a way around this dilemma, let us consider what happens if there were a short period at a very early stage when the universe was vacuum dominated. Assume that at $t = 0$ the universe starts out radiation dominated until a time t_1. Then at $t = t_1$ the particle horizon is at

$$d_1 = a_1 \int_0^{t_1} \frac{dt}{a(t)} = \int_0^{t_1} \sqrt{\frac{t_1}{t}}\, dt = 2\, t_1\,. \tag{11.1.4}$$

The coordinate size of the universe at this point is $\Delta r_1 = 2t_1/a_1$. Then at t_1 the universe switches over to a vacuum dominated universe until a time $t = t_2$. The particle horizon then "inflates" to a new value d_2,

$$d_2 = d_1 e^{H_1(t_2 - t_1)}\,, \tag{11.1.5}$$

where H_1 is the hubble constant during this expansion. After time t_2 the universe goes back to a radiation dominated universe with expansion

$$a(t) = e^{H_1(t_2 - t_1)} a_1 \left(\frac{t}{t_2}\right)^{1/2}\,. \tag{11.1.6}$$

Now let us say we knew nothing about the inflationary period and we just observed the expansion as in (11.1.6). Then we would say that the particle horizon at time t is at $d = 2\,t$ and that the coordinate horizon is at

$$\Delta r = e^{-H_1(t_2 - t_1)} a_1^{-1} \left(t_2\, t\right)^{1/2}\,. \tag{11.1.7}$$

Thus it is possible for $\Delta r << \Delta r_1$.

For the sake of argument, let us suppose that we live in a radiation dominated universe. This will get the age of the universe wrong by a factor of two, but that won't matter much for our discussion but it will simplify the calculations. Then the coordinate distance of the particle horizon today is

$$\Delta r_0 = e^{-H_1(t_2 - t_1)} a_1^{-1} \left(t_2\, t_0\right)^{1/2}\,. \tag{11.1.8}$$

Now if $\Delta r_1 > \Delta r_0$, then the entire visible universe could have been in thermal equilibrium *before* the inflationary period. In order for this to be true, we must have that

$$e^{H_1(t_2 - t_1)} > \left(\frac{t_2\, t_0}{t_1^2}\right)^{1/2}\,. \tag{11.1.9}$$

If we assume that $t_2 \sim t_1$, and using that the universe is radiation dominated afterward, we find that

$$e^{H_1(t_2 - t_1)} > \left(\frac{t_0}{t_1}\right)^{1/2} = Z_2\,, \tag{11.1.10}$$

where Z_2 is the Z factor at t_2. For various particle physics reasons, it is often assumed that the temperature during inflation is $kT_1 \sim 10^{14}$ GeV. The temperature today is $kT_0 = 2.35 \times 10^{-4}$ eV, thus we have

$$H(t_2 - t_1) > \log\left(\frac{10^{14}\,\mathrm{GeV}}{2.35 \times 10^{-4}\,\mathrm{eV}}\right)\Bigg(\sim 60\,. \tag{11.1.11}$$

Hence, the inflationary period has to go through 60 *e-foldings* to perhaps explain the uniform temperature of the universe.

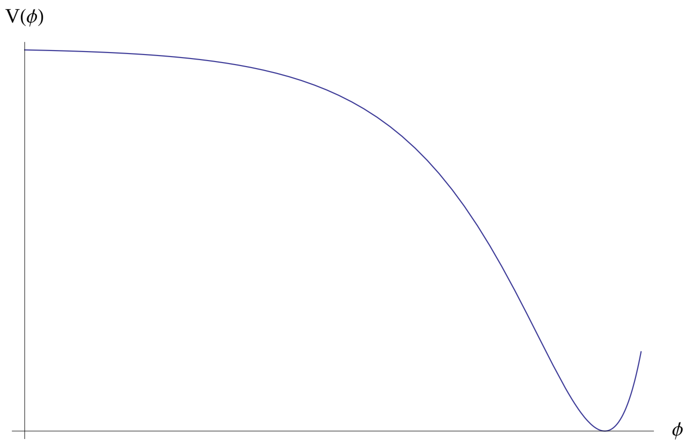

Figure 11.2: A slowly rolling potential that can lead to inflation in the early universe.

11.1.3 The free lunch

A little thought shows that there could be a problem with having an inflationary period. The problem is that during inflation the radiation in the universe rapidly cools due to the exponential expansion. The universe would thus be much too cold since after 60 e-foldings the universe would have dropped to the present temperature, even though we are assuming that the inflation ends a tiny fraction of a second after the beginning of the universe. Actually, this problem is related to a question: How does one get out of inflation, since up to now we have been thinking that once we are in a vacuum dominated universe the vacuum energy density will dominate at all later times? It turns out that the answer to this question will also resolve our problem.

Let us suppose that we have a scalar field ϕ which has a potential $V(\phi)$. The contribution to the energy momentum tensor from this potential is $\Delta T_{\mu\nu} = -g_{\mu\nu}V(\phi)$. Hence this potential is a source of vacuum energy. Let us suppose that the potential has the form as shown in figure 11.2 and let us suppose that the field starts at $\phi = 0$. At some point in time, say t_1 this term will dominate the energy density and the universe will start to inflate. However, let us also assume that during the inflation, the field ϕ is rolling down the hill. The rolling starts slowly since the slope is very gentle at the top, but eventually it approaches the edge and rolls down to the bottom. This is where the inflation is ending since the vacuum energy is disappearing. But this energy must go somewhere, and what happens is that the energy is transferred to all of the radiation. Hence the energy density of the radiation after this *graceful exit* is roughly the energy density that the vacuum started with at t_1 which was also the energy density of the radiation, since the inflation started when their energy densities were equal. Hence, the temperature of the radiation after the graceful exit is the same as at the beginning of inflation and so we have solved our problem. This resolution was originally called the "free lunch" by Alan Guth, the inventor of the inflationary paradigm.

Let us make this story a little more concrete. A scalar field has an action given by

$$S = -\int d^4x\sqrt{-g}\left[\tfrac{1}{2}g^{\mu\nu}\partial_\mu\phi\,\partial_\nu\phi + V(\phi)\right] \qquad (11.1.12)$$

The equation of motion is then given by

$$-\frac{1}{\sqrt{-g}}\,\partial_\mu g^{\mu\nu}\,\partial_\nu\phi + V'(\phi) = 0\,. \tag{11.1.13}$$

If we assume that we have a spatially flat space with the Robertson-Walker metric, then $g_{00} = -1$ and $\sqrt{-g} = a^3(t)$. If we further assume that ϕ has only time dependence, then the equation of motion reduces to

$$\begin{aligned}\frac{1}{a^3}\partial_0(a^3\partial_0\phi) + V'(\phi) &= 0\\ \ddot{\phi} + 3H\dot{\phi} + V'(\phi) &= 0\,,\end{aligned} \tag{11.1.14}$$

where we used that $H = \frac{\dot{a}}{a}$. Examining the second equation in (11.1.14), we see that it looks like the motion of a particle in one dimension, with H acting like a friction term.

If the scalar field is slowly rolling, then $|\ddot{\phi}| \ll 3H|\dot{\phi}|$ and (11.1.14) reduces to

$$\dot{\phi} \simeq -\frac{V'(\phi)}{3H} \tag{11.1.15}$$

We can find a condition between $V(\phi)$ and the Hubble parameter so that this slow rolling holds. If we take another time derivative on (11.1.14) we find

$$\dddot{\phi} + 3H\ddot{\phi} + \dot{\phi}\,V''(\phi) \;=\; 0\,, \tag{11.1.16}$$

where we assumed that $\dot{H} = 0$. Ignoring the $\dddot{\phi}$ term and using that $|\ddot{\phi}| << |3H\dot{\phi}|$ leads to the condition

$$|\dot{\phi}\,V''(\phi)| = |3H\ddot{\phi}| \ll 9H^2|\dot{\phi}| \qquad \Rightarrow \qquad |V''(\phi)| << 9H^2\,. \tag{11.1.17}$$

This is known as the *slow roll condition*. From the Friedman equation we can write this another way. Namely, at the top of the hill, $\rho = V(\phi)$, therefore $H^2 = \frac{8\pi G}{3}V(\phi)$ and so the slow roll condition becomes

$$|V''(\phi)| \ll 24\pi\, G\, V(\phi)\,. \tag{11.1.18}$$

It turns out that there is another slow-roll condition. The changing of the scalar field also contributes to the energy momentum tensor, such that the energy density and pressure are

$$\rho = \tfrac{1}{2}\dot{\phi}^2 + V(\phi)\,, \qquad\qquad p = \tfrac{1}{2}\dot{\phi}^2 - V(\phi)\,. \tag{11.1.19}$$

Hence we want $\frac{1}{2}\dot{\phi}^2 \ll V(\phi)$ in order to have a vacuum dominated energy density. From the equation in (11.1.15) we then have that

$$\tfrac{1}{2}\dot{\phi}^2 \approx \frac{1}{18H^2}\,(V'(\phi))^2 = \frac{1}{48\pi\, G}\frac{(V')^2}{V}\,, \tag{11.1.20}$$

and so

$$|V'(\phi)| \ll (48\pi\, G)^{1/2}|V(\phi)|,. \tag{11.1.21}$$

These two slow roll conditions are normally expressed in terms of two dimensionless parameters,

$$
\begin{aligned}
\varepsilon &= \frac{1}{16\pi\, G}\left(\frac{V'(\phi)}{V(\phi)}\right)^2 \\
\eta &= = \frac{1}{8\pi\, G}\left(\frac{V''(\phi)}{V(\phi)}\right)\Bigg(
\end{aligned}
\qquad (11.1.22)
$$

If the slow roll conditions are met then $\varepsilon, |\eta| \ll 1$.

11.2 Other things to do

Given more time, I would eventually like to include discussions on the following topics:

- More thorough discussion of nucleosynthesis
- Jeans instability and galaxy formation
- Fluctuations in CMB
- Spectrum of multipole moments in CMB
- Combining supernova, CMB data to get the best fit for the present cosmology.
- Gaussian fluctuations in inflation; ”ringing of the bell”
- Other black holes including Reisner-Nö rdst”om and Kerr.
- Gravitational waves