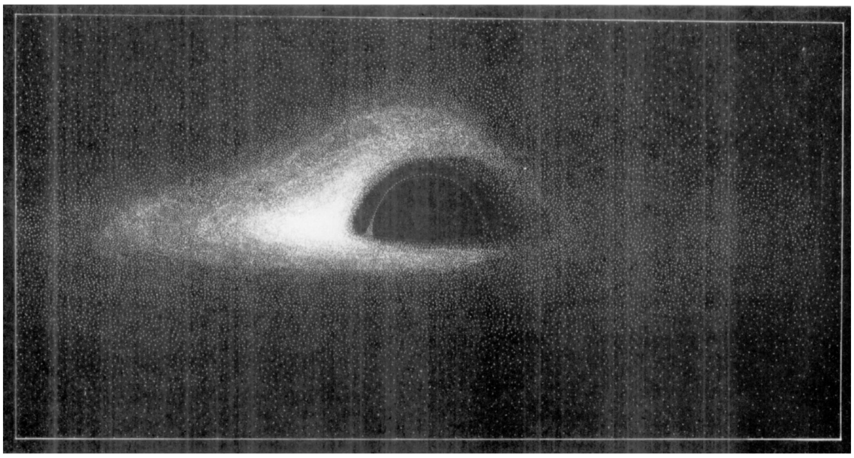
PHY405 Relativity



Lecture notes

Yen-Kheng Lim

Xiamen University Malaysia

Course synopsis

This course is an introduction to General Relativity. We start with a review of Special Relativity along with an introduction of tensor calculus and tensor notation. Subsequently the relevant notions of differential geometry such as geodesics, Riemann and Ricci curvatures are covered. This is followed by a relativistic formulation of the stress-energy tensor. The notion of stress-energy causing curvature is given by the Einstein equations. Consequences and solutions of the Einstein equations are then discussed, particularly the Schwarzschild solution and FLRW cosmology. [Cover image: J. P. Luminet, 'Image of a spherical black hole with thin accretion disk', Astron. Astrophys. **75** 228–235 (1979).]

Contents

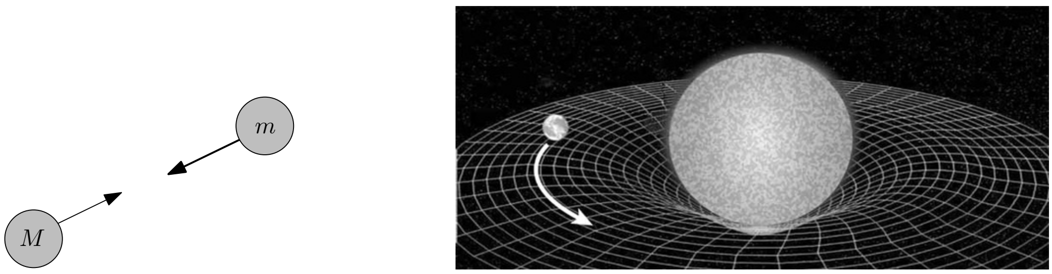
Part I

Relativity and geometry

Chapter 1 Introuction, Special Relativity, and spacetime

1.1 Introduction: A critical look at Newtonian physics

Even though a physicist might refer to General Relativity (GR) as a shorthand for 'Einstein's theory of gravity', it is technically not an accurate description. GR is a theory of *spacetime*. Indeed, the formulation of the theory started when Einstein attempted to account for gravitational effects in his theory of Special Relativity (SR). What he ended up with was a paradigm shift that gravity is not 'really a force', but a manifestation of curved spacetime. So in GR, the concept of 'gravitational force' completely disappeared! That is, the Newtonian framework of gravity being a force is replaced by the notion of spacetime curvature.



(a) Gravity as a force.

(b) Gravity as spacetime curvature.

Figure 1.1

We will spend this introduction briefly reviewing the Newtonian concept of gravity, depicted in Fig. 1.1a. Then we will spend almost the entire semester understanding gravity as spacetime curvature as in Fig. 1.1b.

First, recall Newton's laws of motion [1]

- **First law.** *A body remains at rest or in uniform motion unless acted upon by a force.*
- **Second law.** *A body acted upon by a force moves in such a manner that the time rate of change of momentum equals the force.*
- **Third law.** *If two bodies exert forces on each other, these forces are equal in magnitude and opposite in direction.*

In other words, determining the motion of an object requires the knowledge of all *forces* acting on it. But what is 'force'?

Seriously, what is the definition of 'force'? Viewed from another angle, we may take Newton's three laws as statements which defines the concept of force. In other words, *a force is any physical phenomena which causes acceleration.*

The first law, states that if no (net) force is acting on an object, it will remain at uniform motion. In other words, this is the *principle of inertia*, saying that objects tend to resist changes to its state of motion. More mathematically, objects tend to remain at constant velocity, unless a net force changes it. Yet another way to say this is that objects tend to *resist acceleration.*

To specify a force, one may need a separate *theory* that provides a formula for the force. For instance, Coulomb's law gives us the electrostatic force:

$$\vec{F} = \frac{1}{4\pi\epsilon_0}\frac{qQ}{r^2}\,\hat{r}. \tag{1.1}$$

This is the force on a particle of charge q at position $\vec{r}$, experiencing a force due to another charge Q located at the origin. (That is, Q is at position $\vec{r} = \vec{0}$.)

If we wish to determine the particle's acceleration, we need to use $\vec{F} = m\vec{a}$. So knowledge of the mass m is required to determine the motion. This raises another question: *What is mass?* From Newton's laws we have the concept of *inertia*, which is the tendancy of an object to resist changes in velocity. Then $\vec{F} = m\vec{a}$ tells us that objects with larger m will have higher inertia. Therefore, an object's mass m is the measure of its inertia. We now need to be precise with the concept of mass, and therefore we define:

> **Inertial mass.** *The measure of inertia, its numerical value is represented by 'm' in the equations of Newton's three laws of motion.*

Now, $\vec{F} = m\vec{a}$ applies for arbitrary forces. In order to determine the acceleration of a mass m, we need to know the expression for $\vec{F}$ on the left hand side. For example, if m carries an electric charge q, and it's in the presence of another charge Q at distance r away, the force is given by Coulomb's law

$$\vec{F} = \frac{Qq}{4\pi\epsilon_0 r^2}\,\hat{r}. \tag{1.2}$$

Substituting this into the left-hand side of $\vec{F} = m\vec{a}$, we can determine the particle's acceleration by

$$\vec{a} = \frac{Qq}{4\pi\epsilon_0 m r^2}\hat{r}. \tag{1.3}$$

So the acceleration depends on the distance r from Q, its own charge q, as well as its mass m. Changing Q, q, or m will change the resulting acceleration.

Newtonian gravity is already a bit weirder than Coulomb. Let us now put ourselves in the shoes of Newton while he's sitting under that apple tree that was when noticed that planets separated at distance r will exert a force proportional to $1/r^2$ on each other. It's similar to Coulomb's law,[1] but clearly planets are electrically neutral. So the force law acting on m can be written in the form

$$\vec{F} = \frac{GNn}{r^2}\hat{r}, \tag{1.4}$$

where N and n are the 'gravitational charge' carried by the two planets (in analogy to electrical charges Q and q), and G is some constant (in analogy to $1/4\pi\epsilon_0$). At this stage, there is no reason to assume that the 'gravitational charge' is equal to inertial mass. This situation is sketched in Fig. 1.2.

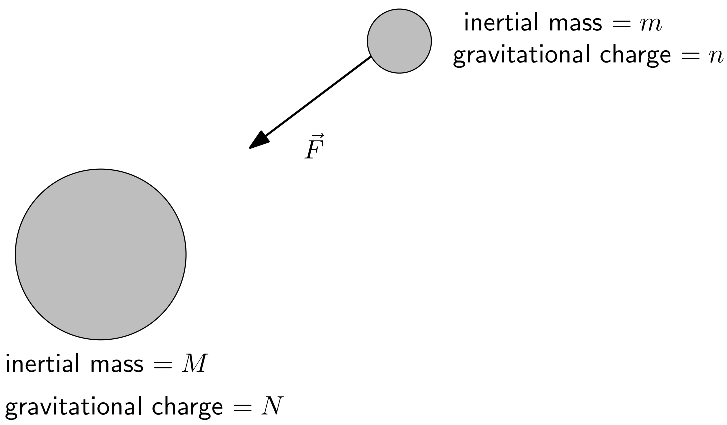


Figure 1.2

Substituting this into $\vec{F} = m\vec{a}$ to obtain

$$\vec{a} = \frac{GNn}{mr^2}\hat{r}.$$

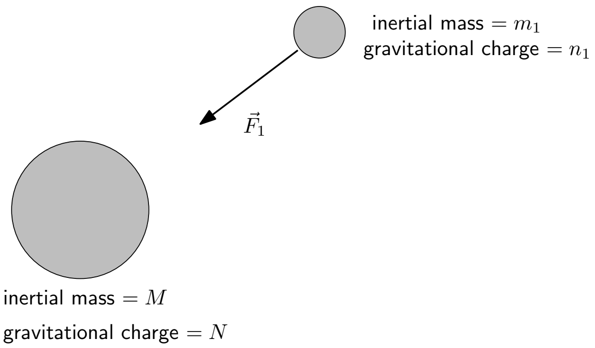
This is a reasonable first guess for a theory of gravitation.

How do we measure these quantities? Hypothetically, we can measure m by taking the object and applying it to other known forces (such as drag forces, Coulomb forces, pulleys,

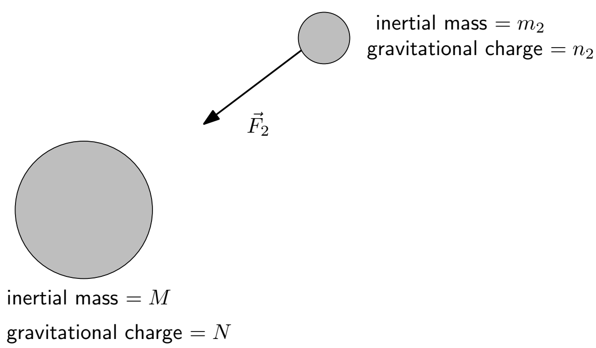
[1]Strictly speaking, Coulomb's law was officially published later than Newton, in 1785, but it's known earlier.

etc.). Then we release m under the influence of the gravity of the larger object and measure its acceleration. At best, this experiment can only give the product GNn.

We can get GN by repeating the experiment with different small masses. Suppose we take two different objects masses m_1 and m_2. We expect them to carry some (unknown) gravitational charges n_1 and n_2, respectively, as shown in Fig. 1.3.



(a) Experiment with m_1.



(b) Experiment with m_2.

Figure 1.3

We make sure to measure the acceleration where both cases have the same distance r from M. Therefore the acceleration for both cases is expected to be

$$\vec{a}_1 = \frac{GNn_1}{m_1 r^2}\,\hat{r}, \quad \vec{a}_2 = \frac{GNn_2}{m_2 r^2}\,\hat{r}. \tag{1.5}$$

This experiment can be done by taking M as the earth and dropping balls of different masses m_1, m_2 made from different materials (hopefully so that n_1 and n_2 are different).

Galileo, for instance, has done experiments like these. But remarkably, it turns out that for fixed r, the accelerations are always the same regardless of mass! That is, $\vec{a}_1 = \vec{a}_2$ for any two objects so Eq. (1.5) implies

$$\frac{n_1}{m_1} = \frac{n_2}{m_2}. \tag{1.6}$$

Experimentalists like Galileo, Hooke, and others found that the acceleration is always the same regardless of what mass used. So the only way for Eq. (1.6) to be true, we have[2]

$$n_1 = m_1 \quad n_2 = m_2,$$

[2]Actually Eq. (1.6) is still true if $n_i = km_i$ for some constant k, but this constant can be absorbed into the definition of G in the force law. Therefore we do not lose generality by assuming $k = 1$.

or generally,

$$\text{Gravitational charge} = \text{inertial mass.} \tag{1.7}$$

This is what we call the *universality of gravitation.* That is, the gravitational force is universal – it produces the same acceleration for any (test) mass.

Hence the gravitational force law can now be revised to

$$\vec{F} = \frac{GMm}{r^2}\,\hat{r}. \tag{1.8}$$

Subsequent experiments and observations have fixed the proportionality constant to be

$$\boxed{G = 6.6743 \times 10^{-11}\ \mathrm{kg}^{-1}\,\mathrm{m}^3\,\mathrm{s}^{-2}.} \tag{1.9}$$

This is called *Newton's gravitational constant*, or simply *Newton's constant* or *gravitational constant.*

The equivalence between gravitational charge and inertial mass is a non-trivial coincidence. At first, we have no philosophical reason to claim that they are the same. To reiterate, the inertial mass m is a measure of inertial, it's resistance to forces. The gravitational charge is supposed to determine the strength of the gravitational force. Yet we found that the gravitational force is the same for all objects regardless of its mass.

Now, there is another way where particles of different mass experience the same acceleration. Suppose we start with a set of particles (of different mass) that is stationary in some coordinate system $\mathcal{O}$. Then we transform to another coordinate system $\mathcal{O}'$ that is accelerating relative to $\mathcal{O}$. Under this new coordinate system, all the particles will undergo the same acceleration!

If you think this scenario seems too artificial, consider putting particles (say, marbles of different mass) inside a frictionless box. We draw a Cartesian axis on the side of the box, then we push it so the whole box undergoes an acceleration. The marbles inside the box will accelerate relative to the Cartesian axes drawn on the box. This can be clearly observed when the marbles accumulate on the side of the box opposite to the direction of acceleration.

The preceding discussion tells us that, if we see different masses experiencing the same acceleration there are two possibilities: Either

(i) we are in an accelerated frame, or,

(ii) there is gravity present.

In formulating his theory of GR, Einstein makes the bold step in asserting that (i) and (ii) are completely identical! Gravity is nothing but an accelerated frame. In this sense, gravity is not really a 'force', but an artefact of a 'coordinate transformation' into an accelerated frame. We will spend the first half of this course trying to develop this claim into mathematically precise equations. We will see that the notion of gravity as an accelerated frame is realised by the concept of a curved spacetime, which is roughly like the picture in Fig. 1.1b.

Poisson's equation. Since gravity is a conservative force, it can be written as the gradient of the gravitational potential

$$\vec{F} = -m\vec{\nabla}\Phi. \tag{1.10}$$

(This is again, in analogy to electrostatics as the electric field is defined by $\vec{F} = q\vec{E}$, and $\vec{E}$ is the gradient of the electric potential $\vec{E} = -\vec{\nabla}V$.)

The gravitational potential due to a point mass M located at position $\vec{r}'$ is

$$\Phi = -\frac{GM}{|\vec{r} - \vec{r}'|}.$$

If we have a distribution of N point masses located at $\vec{r}_i$, the total potential is just the sum

$$\Phi = -G\sum_{i=1}^{N} \frac{M_i}{|\vec{r} - \vec{r}_i|}.$$

The gravitational potential due to a continumm distribution of masses is just the continumm version of the sum, i.e., an integral

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

Now if we apply the Laplacian to both sides of this equation,

$$\nabla^2\Phi = -G\int \mathrm{d}^3\vec{r}'\rho(\vec{r}')\ \underbrace{\nabla^2\frac{1}{|\vec{r} - \vec{r}'|}}_{=-4\pi\delta^{(3)}(\vec{r}-\vec{r}')},$$

where $\delta^{(3)}(\vec{r} - \vec{r}')$ is the three-dimensional delta function. Integrating out the $\vec{r}'$ variable, we get

$$\boxed{\nabla^2\Phi = 4\pi G\rho.} \tag{1.11}$$

This is the differential equation which determines the gravitational potential Φ, given a mass distribution ρ. This is called the *Poisson equation.* The Poisson equation is the governing equation for gravitational theory. The main goal of General Relativity is to generalise the Poisson equation. The question to be answered by GR is: *Given some distribution of matter, what would be the resulting shape of spacetime?*

Some suggested references on relativity.

- *General Relativity: An introduction for physicists* by Hobson, Efstathiou, and Lasenby [2]. One of the clearest intuitive introduction to the basic concepts while not skipping on the mathematical machinery, which is often important.
- *A First Course in General Relativity* by Bernard Schutz [3]. Perhaps one of the most popular undergraduate introductions to the topic. A good balance of physical intuition and mathematical rigour. However, the notation used in their book will be quite different from the present lectures.
- *Gravity: An Introduction to Einstein's General Relativity* by J. B. Hartle [4]. Another commonly used undergraduate textbook in many universities. This one has great emphasis on physics and less on mathematics. Probably favoured by astrophysics.
- *A short course in General Relativity* by Foster and Nightingale [5]. A thin and brief textbook. A good and easy read for undergraduates.
- *Spacetime and Geometry* by Sean Carroll [6]. This is typically recommended as a graduate-level text, but it should be simple and clear enough for undergraduates to understand. This books also serves as a good introduction to manifolds and differential geometry.
- *General Relativity* by Robert Wald [7]. Another graduate text, perhaps the next step after Carroll. It has a good coverage on causal structures and singularity theorems.
- *The Mathematical Theory of Black holes* by Subrahmanyan Chandrasekhar [8]. A fairly advanced book which includes calculational details in the analysis of various black holes. Quite comprehensive as for each black hole solution, he analyses the geodesics, stability, and quasinormal modes.
- *The Large Scale Structure of Spacetime* by Stephen Hawking and George Ellis (Fig. 1.4). Its main emphasis is on global structures such as singularity theorems and makes frequent use of Penrose diagrams. Has a good introduction to common spacetimes in relativity, such as Schwarzschild, de Sitter, and Anti-de Sitter.

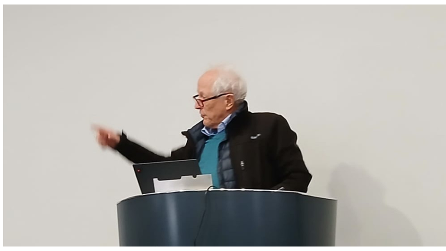


Figure 1.4: George Ellis on 22nd August 2024, recounting his career during a celebration of his 85th birthday in conjunction with the South African Gravity Society (SAGS) 2024 conference.

- *Gravitation* by Misner, Thorne, and Wheeler [9]. The famous 'MTW book'. One of the main references used by relativists worldwide.
- *Exact Space-Times in Einstein's General Relativity* by Griffiths and Podolský [10]. This is one of the most comprehensive handbooks of exact solutions in relativity. Almost all known exact solutions to the Einstein–Maxwell equations in four dimensions are listed and described here.
- *Relativists' Toolkit* by E. Poisson [11]. As the title says, it is indeed a toolkit for researchers in relativity. It contains developments of advanced methods such as action principles, Lagrangian/Hamiltonian formulation of the field equations, and introduction to the laws of black-hole mechanics.
- **Online resources:** A useful collection of GR-related formula can be found in Robert McNees' website [12]. A fairly comprehensive graduate-level lecture notes by Mathias Blau [13] and Daniel Baumann [14]. Lectures on quantum gravity by Tom Hartmann [15]. David Tong famously has excellent lecture notes various topics, including relativity [16].

1.2 Inertial frames, spacetimes, and worldlines

As seen in the previous discussion, the idea of an accelerated frame will be important in relativity. A frame that is NOT accelerated is something called an *inertial frame*. We start this section by introducing this concept more precisely.

Inertial frames and the principle of (Galilean) relativity

[*Parts of this section is similar to PHY201 Theoretical Mechanics. Indeed, in order to understand any theory of mechanics (be it Newtonian or relativistic), one has to dive deep* *into the concepts of inertia.*]

Newton's first two laws actually singles out the concept of *inertial frames*. Here, a *frame* refers to a coordinate system used by an observer or experimentalist. Different observers might use different coordinate systems/frames convenient to them. However, there must be some way for different observers to *agree* on the observed phenomena. (e.g., what is the net force acting on a particle.) Otherwise, different people claim different results, and everything is chaos!

Can two observers agree on the most basic situation: Is a particle free? (Zero force.) Consider Alice and Bob on different spaceships. Both are observing an asteroid. Suppose the asteroid is at fixed constant distance from Bob. So Bob would say the asteroid is at rest; no forces are acting on the asteroid. If Alice is flying away from Bob, she would see the asteroid going away form her as well.

Alice observation may sound different from Bob, but if Alice is moving at *constant velocity* away from Bob, she sees the asteroid moving at constant velocity as well. If Alice uses Newton's first law, she would be correct to say that the asteroid is experiencing zero net force.

Newton's second law states $\vec{F} = m\vec{a}$. So if an object accelerates, there must be a force. Now, if Alice is *accelerating* away from Bob but is unaware of this, and she sees the asteroid accelerating away as well, she concludes that there's a force acting on the Asteroid. In this case she disagrees with Bob!

In what situations would Alice and Bob agree with each other? Let us first consider one-dimensional motion. Alice uses coordinate x, and Bob uses coordinate X.

- In Alice's coordinates, the asteroid is moving as $x(t)$.
- In Bob's coordinates, the asteroid is moving at $X(t) = f(x, t)$. Where f is some function depending on Alice's observations.

We wish to determine the relation between X and x which will determine a formula for f. Alice observes a moving asteroid, and concludes that the force on it is

$$F = m\frac{\mathrm{d}^2x}{\mathrm{d}t^2}. \tag{1.12}$$

Bob observes the same asteroid and concludes that there is a force acting on it is

$$F = m\frac{\mathrm{d}^2X}{\mathrm{d}t^2} = m\frac{\mathrm{d}^2}{\mathrm{d}t^2}f(x, t).$$

We compute the total time derivatives of $X = f(x,t)$ as follows:

$$\begin{aligned}
\frac{\mathrm{d}X}{\mathrm{d}t} &= \frac{\partial f}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial f}{\partial t}, \\
\frac{\mathrm{d}^2X}{\mathrm{d}t^2} &= \frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{\partial f}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial f}{\partial t}\right] \\
&= \frac{\partial^2 f}{\partial x^2}\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 + \frac{\partial f}{\partial x}\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + 2\frac{\partial^2 f}{\partial x \partial t}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial^2 f}{\partial t^2}.
\end{aligned}$$

By Newton's second law, Bob concldues that the force on the asteroid must be

$$F = m\frac{\mathrm{d}^2X}{\mathrm{d}t^2} = m\left[\frac{\partial^2 f}{\partial x^2}\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 + \frac{\partial f}{\partial x}\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + 2\frac{\partial^2 f}{\partial x \partial t}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial^2 f}{\partial t^2}\right].$$

For two observers to agree on the same force, Eq. (1.12) must be equal to (1.13). Therefore we have

$$\frac{\partial^2 f}{\partial x^2} = 0, \quad \frac{\partial f}{\partial x} = 1, \quad \frac{\partial^2 f}{\partial x \partial t} = 0, \quad \frac{\partial^2 f}{\partial t^2} = 0. \tag{1.13}$$

To satisfy all these conditions, the formula f must be

$$X = f(x,t) = x + ut, \tag{1.14}$$

where u is some constant.

We repeat the same arguments for the y and z directions. So generally, in three dimensions, the transformation between Alice's coordinates $\vec{r}$ and Bob's coordinates $\vec{R}$ must be

$$\vec{R} = \vec{r} + \vec{u}t, \tag{1.15}$$

where $\vec{u}$ is a constant vector. By setting $\vec{r} = 0$, we see that $\vec{u}$ represents the velocity of Alice's origin O_A with respect to Bob's coordinate system O_B. We call this the *relative velocity* between Alice and Bob. This situation is shown in Fig. 1.5.

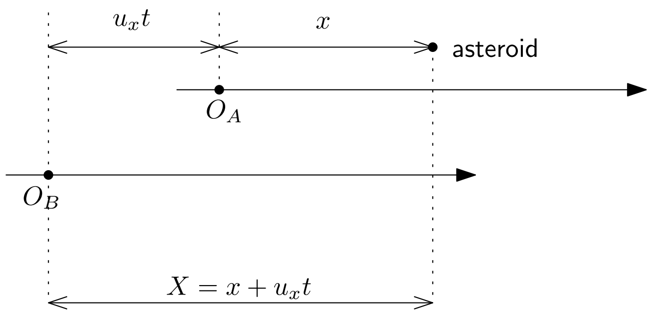


Figure 1.5: An inertial coordinate transformation between Alice and Bob. Only the x direction is shown here.

Eq. (1.15) is called a *Galilean transformation*, describing the set of coordinate frames where two observers are able to agree on a phenomena. Frames that obey these conditions are called *inertial frames*. From these discussions we can now state the Galilean Principle of Relativity: The laws of physics are the same for all inertial frames.

Worldlines

In **Newtonian Mechanics**, the evolution of a particle's position is described by a vector

$$\vec{r} = (x^1, x^2, x^3) = (x, y, z). \tag{1.16}$$

Therefore this is a curve parametrised by time t. Its velocity is the time derivative

$$\vec{v}(t) = \frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = (\dot{x}(t), \dot{y}(t), \dot{z}(t)), \tag{1.17}$$

where overdots denote derivatives with respect to t: $\dot{x}(t) = \frac{\mathrm{d}x}{\mathrm{d}t}$, etc.

In relativity, we include t as a fourth coordinate as well. We define an *event* as an object with four components,

$$x^\mu = (x^0, x^1, x^2, x^3) = (ct, x, y, z), \tag{1.18}$$

where c is the speed of light,

$$\boxed{c = 2.99\,792\,458 \times 10^8 \text{ m s}^{-1}.} \tag{1.19}$$

The first entry is multiplied by c so that $x^0 = ct$ has the dimensions of length:

$$[ct] = [c][t] = LT^{-1}T = L.$$

Therefore all components of (x^0, x^1, x^2, x^3) can be described in units of metres, or other length units. Since there are four components, we shall view these as coordinates of a *four*-dimensional *spacetime*. Since this consists of three dimensions of space and one dimension of time.

In **relativistic mechanics**, the evolution of a particle is described by a curve in spacetime. Let us ignore the y and z coordinates at the moment, and focus on x and t only. Then, we have a two-dimensional spacetime. At this stage, spacetime is nothing special – we're just plotting a particle's position x given a time t. As the particle moves, we then have a curve in spacetime, called a *worldline*.

For example, in the *Lord of the Rings* trilogy, Frodo's worldline may be depicted as in Fig. 1.6. Here, event P is the moment Frodo throws the ring into Mt. Doom.

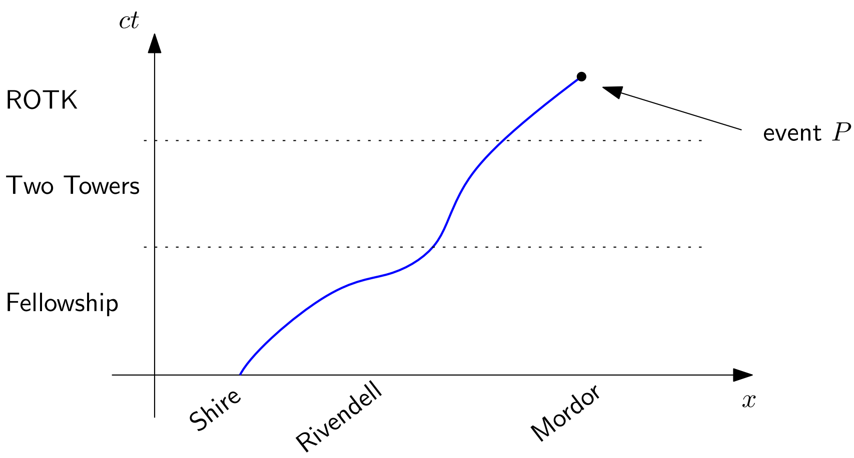


Figure 1.6: Spacetime of Middle Earth during the *Lord of the Rings* trilogy. The worldline of Frodo Baggins is depicted as the blue curve. Event P is when Frodo throws the ring into Mt. Doom.

XCKD's Movie Narrative Chart (Fig. 1.7) is also a spacetime diagram, but where the time axis is horizontal and the space axis is vertical. In the XKCD chart, each curve represents the time and place of a particular character. So each curve in this chart is also a worldline.

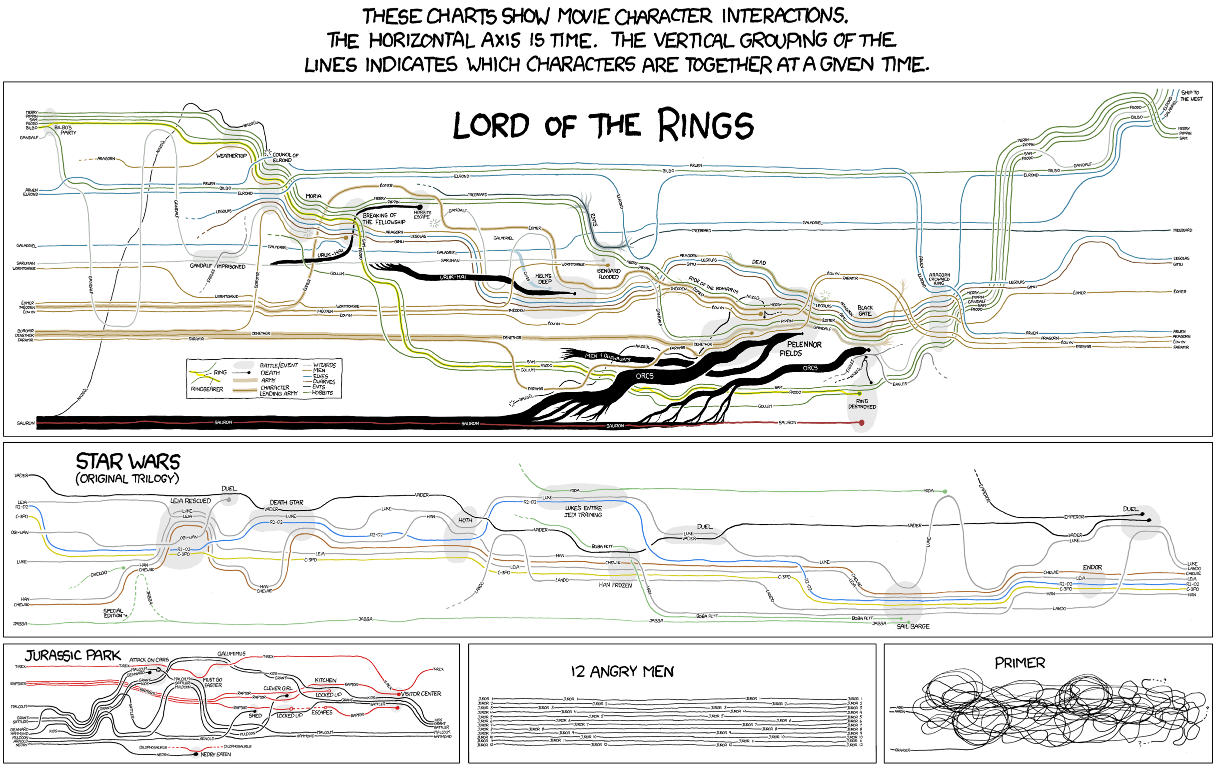


Figure 1.7: https://xkcd.com/657/. The zoomable full-sized version is in https://xkcd.com/657/large/.

When represented on a worldline, the velocity of a particle at an event $P(ct, x)$ is the slope of the tangent at that point, as shown in Fig. 1.8.

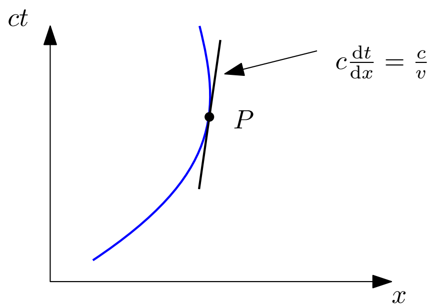


Figure 1.8: Tangent to a worldline at P.

The velocity of the particle is

$$v = \frac{\mathrm{d}x}{\mathrm{d}t} \quad \to \quad c\frac{\mathrm{d}t}{\mathrm{d}x} = \frac{c}{v}. \tag{1.20}$$

We see that $|\frac{c}{v}| = \sigma$ gives three classifications on the velocity of the particle:

$$\left|\frac{c}{v}\right| = \sigma \begin{cases} > 1 & \text{if slower than light. (Time-like)} \\ = 1 & \text{if equal to light. (Light-like/Null)} \\ < 1 & \text{if faster than light. (Space-like).} \end{cases} \tag{1.21}$$

In particular, anything travelling slower than the speed of light is a *time-like* particle. Light itself, photons, or generally mass-less particles can travel at speed c. These are called *light-like* or *null* particles. As we will prove below, no time-like particle can reach the speed of light; and nothing can travel faster than light. That is, space-like particles are impossible.

Since we use $x^0 = ct$ as the time (multiplied with c) axis as the vertical axis, worldlines of light are always 45°-lines. Fig. 1.9 shows various examples of time-like, null, and space-like worldlines.

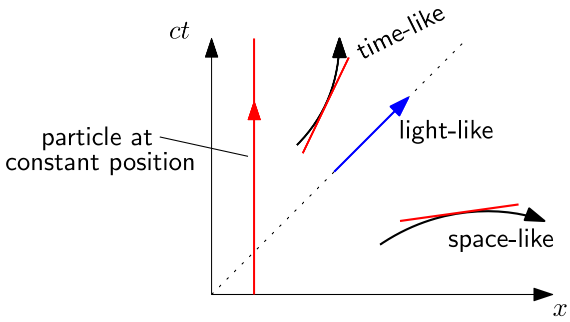


Figure 1.9: Examples of various worldlines

1.3 Postulates of SR and Lorentz transformations

Special relativity concerns observers related by inertial frames. We define an **inertial frame** as a system recording events by recording positions in Cartesian coordinates (x, y, z) and time (ct), which obeys the following properties:

1. The distance between any two points $P_1(x_1, y_1, z_1)$ and $P_2(x_2, y_2, z_2)$ is independent of time.

2. Clocks that sit at every point (x, y, z) are synchronised and run at the same rate.

3. The geometry of space at any fixed time t is Euclidean $(\mathbb{R}^3)$.

For observers in inertial frames, Special Relativity (SR) is based on two basic postulates.

Postulate I. *The laws of physics are the same in all inertial frames.*

Postulate II. *The speed of light is the same in all inertial frames, and is equal to c.*

Suppose we have two observers:

- Alice, uses frame S with coordinates (ct, x, y, z).
- Bob, uses frame S' with coordinates (ct', x', y', z').

Furthermore, we suppose Bob moves with velocity $\vec{u} = \frac{\mathrm{d}\vec{r}}{\mathrm{d}t}$ relative to Alice.

By PII, the first coordinate for Alice and Bob, ct and ct', uses the same c. Therefore, worldlines of light will travel at 45°-degree lines for both Alice and Bob.

Time dilation. Consider two observers, Alice using frame S standing on the ground, and Bob using frame S' on a train. The train is moving at speed u relative to Alice. On the train, there is a laser pointed at a mirror on the ceiling, at height h above the laser.

We first consider Alice's observation using frame S, as shown in Fig. 1.10. At time $t = t_1$, the laser is emitted. Then, at some time $\Delta t = t_2 - t_1$ later, the light was reflected by the mirror and returns to the laser source.

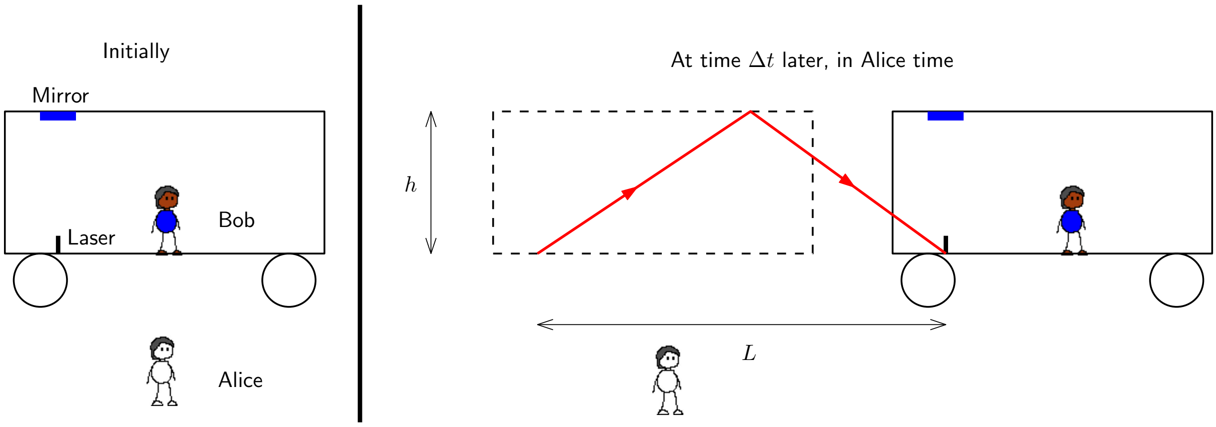


Figure 1.10: The path of the laser light, as observed by Alice.

From Alice's perspective, during this time, the train has travelled a horizontal distance of L. Then we have

$$L = u\Delta t. \tag{1.22}$$

The distance travelled by light, according to Alice is (using Phythagoras' theorem)

$$\Delta\ell_A = 2\sqrt{h^2 + \left(\frac{L}{2}\right)^2} = 2\sqrt{h^2 + \left(\frac{u\Delta t}{2}\right)^2}.$$

The speed of light measured by Alice is

$$c_A = \frac{\Delta\ell_A}{\Delta t} = \frac{2}{\Delta t}\sqrt{h^2 + \left(\frac{u\Delta t}{2}\right)^2}.$$

Next we donsider Bob's perspective, as shown in Fig. 1.11. Initially, at time $t' = t'_1$, the laser source emits the light. The light travels to the mirror and gets reflected back towards the source at time $\Delta t' = t'_2 - t'_1$ later. From Bob's perspective, the light only travels a distance of $2h$. From Bob's perspective, the distance travelled by light is just $\Delta\ell_B = 2h$. Therefore the speed of light measured by Bob is

$$c_B = \frac{\Delta\ell_B}{\Delta t'} = \frac{2d}{\Delta t'} \tag{1.23}$$

$$d = \frac{c_B\Delta t'}{2}. \tag{1.24}$$

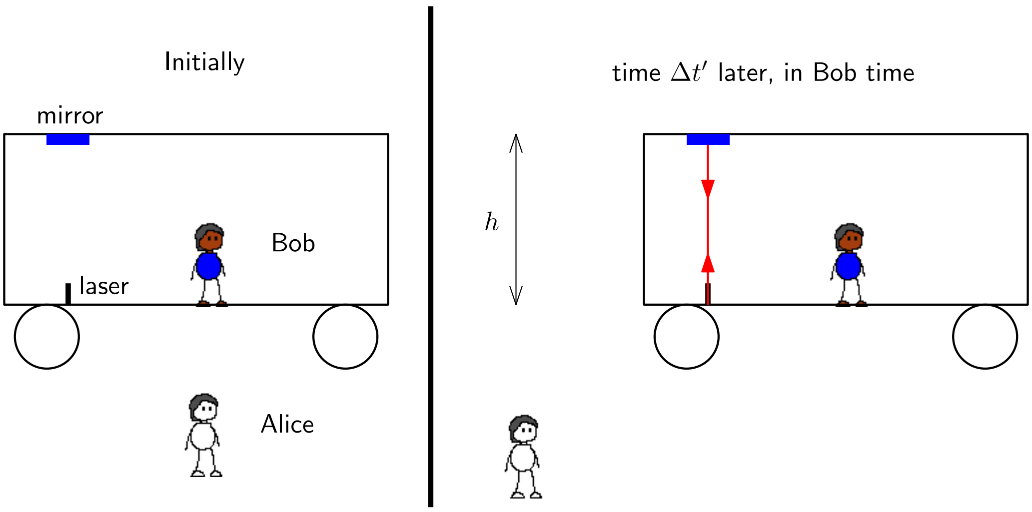


Figure 1.11

Now, we substitute Eq. (1.24) into (1.23),

$$c_A = \frac{2}{\Delta t}\sqrt{\left(\frac{c_B\Delta t'}{2}\right)^2 + \left(\frac{u\Delta t}{2}\right)^2}.$$

By Postulate 2, all observers should measure the same speed of light. Therefore $c_A =$

$c_B = c$. So the equation becomes

$$\begin{aligned} c &= \frac{2}{\Delta t}\sqrt{\frac{c^2\Delta t'^2}{4} + \frac{c^2\Delta t^2}{4}} \\ c^2\Delta t^2 &= c^2\Delta t'^2 + u^2\Delta t^2 \\ \left(1 - \frac{u^2}{c^2}\right)\Delta t^2 &= \Delta t'^2 \\ \Delta t &= \frac{\Delta t'}{\sqrt{1-\frac{u^2}{c^2}}}. \end{aligned}$$

It is convenient to define the **Lorentz factor** by $\gamma = \frac{1}{\sqrt{1-u^2/c^2}}$. The factor satisfies $\gamma \geq 1$ for $0 \leq u < c$. Then we have

$$\Delta t = \gamma\Delta t', \quad \gamma = \frac{1}{\sqrt{1-\frac{u^2}{c^2}}}. \tag{1.25}$$

So the time interval between the laser leaving and returning to the source is different according to Alice and Bob. In particular, Alice's time interval is longer compared to Bob. This is the phenomena of **time dilation**.

Whose time is dilated? – Proper time. If there exist a frame such that two events occur at the **same position**, the time interval between the two events is the **proper time** $\Delta t'$.

For Bob, the laser emited and returning happens on the same position in Bob's coordinates S'. Therefore the time interval of this process is the proper time $\Delta t'$.

For Alice, the laser emitted and laser returning occurs at **different positions** of her coordinate system. Therefore her time measurement is *not* the proper time. We can calculate her time interval based on Bob's proper time by the time dilation formula,

Spacetime intervals. On spacetime, we can measure length intervals $(\Delta x, \Delta y, \Delta z)$ as well as time intervals $(c\Delta t)$. Now, the speed of a particle is the ratio of its length interval over its time interval:

$$\frac{\Delta r}{\Delta t} = v, \quad \Delta r = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}. \tag{1.26}$$

Dividing both sides by c, we obtain $1/\sigma$,

$$\frac{\Delta r}{c\Delta t} = \frac{v}{c} = \frac{1}{\sigma}\begin{cases} < 1 & \text{time-like} \\ = 1 & \text{null} \\ > 1 & \text{space-like} \end{cases} \tag{1.27}$$

Let us square both sides of this equation and rearrange,

$$\Delta r^2 = \frac{1}{\sigma^2} c^2 \Delta t^2.$$

Subtract by $c^2\Delta t^2$ from both sides,

$$-c^2\Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2 = \left(\frac{1}{\sigma^2} - 1\right) c^2 \Delta t^2 \begin{cases} < 0 & \text{time-like} \\ = 0 & \text{null} \\ > 0 & \text{space-like.} \end{cases}$$

This suggests that it's convenient to define a **spacetime interval**

$$\Delta s^2 = -c^2\Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2, \tag{1.28}$$

so $\Delta s^2 < 0$ for time-like intervals, $\Delta s^2 = 0$ for null, and $\Delta s^2 > 0$ for space-like intervals. Figure 1.12 shows examples of various intervals as they appear in spacetime.

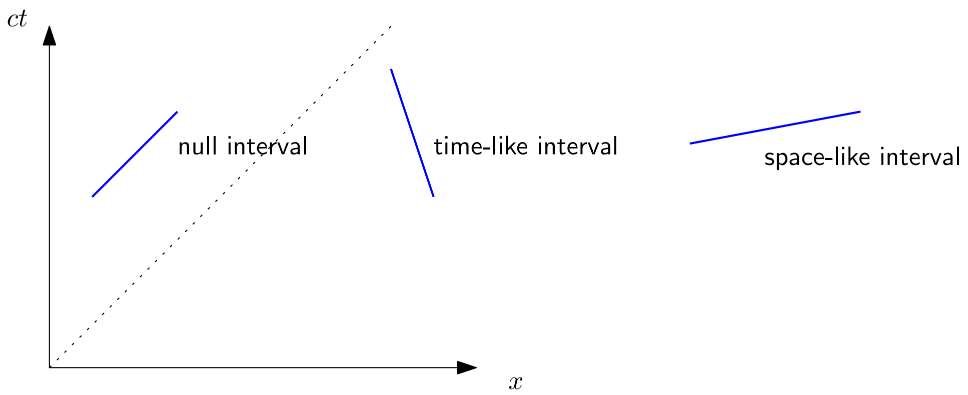


Figure 1.12: Examples of null, time-like, and space-like intervals.

Suppose Alice measures an interval between two events, P and Q to be

$$\Delta s^2 = -c^2\Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2. \tag{1.29}$$

Bob, moving $\vec{u} = \frac{\mathrm{d}\vec{r}}{\mathrm{d}t}$ relative to Alice, measures the interval between the same two events P and Q to be

$$\Delta s'^2 = -c^2\Delta t'^2 + \Delta x'^2 + \Delta y'^2 + \Delta z'^2. \tag{1.30}$$

What is the relation between $\Delta s'^2$ and Δs^2? First, it clearly depends on the relation between (ct, x, y, z) and (ct', x', y', z'). We assume there must be some coordinate transformation between the two, which we assume to be linear:

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \Lambda_{00} & \Lambda_{01} & \Lambda_{02} & \Lambda_{03} \\ \Lambda_{10} & \Lambda_{11} & \Lambda_{12} & \Lambda_{13} \\ \Lambda_{20} & \Lambda_{21} & \Lambda_{22} & \Lambda_{23} \\ \Lambda_{30} & \Lambda_{31} & \Lambda_{32} & \Lambda_{33} \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \tag{1.31}$$

we shall denote the 4×4 matrix by Λ. The interval is

$$\Delta s^2 = -c\Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2 \tag{1.32}$$

$$= \begin{pmatrix} -c\Delta t & \Delta x & \Delta y & \Delta z \end{pmatrix} \begin{pmatrix} c\Delta t \\ \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} \tag{1.33}$$

Correspondingly, the interval measured by Bob would be

$$\Delta s'^2 = \begin{pmatrix} -c\Delta t' & \Delta x' & \Delta y' & \Delta z' \end{pmatrix} \begin{pmatrix} c\Delta t' \\ \Delta x' \\ \Delta y' \\ \Delta z' \end{pmatrix} \tag{1.34}$$

$$= \begin{pmatrix} -c\Delta t & \Delta x & \Delta y & \Delta z \end{pmatrix} \Lambda^{\mathrm{T}} \Lambda \begin{pmatrix} c\Delta t \\ \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} \tag{1.35}$$

$$= \sum_{\mu,\nu=0}^{3} M_{\mu\nu} \Delta x^\mu \Delta x^\nu. \tag{1.36}$$

Working out the matrix multiplications, we obtain a quadratic sum. First, we denote

$$x^0 = ct, \quad x^1 = x, \quad x^2 = y, \quad x^3 = z. \tag{1.37}$$

(Note that the superscripts are labels, not powers.) An arbitrary spacetime coordinate is x^μ, where μ can take values $\{0, 1, 2, 3\}$.

Then Bob's interval is

$$\Delta s'^2 = \sum_{\mu,\nu=0}^{3} M_{\mu\nu} \Delta x^\mu \Delta x^\nu. \tag{1.38}$$

Since M always appear in the combination $M_{\mu\nu} + M_{\nu\mu}$, every term of the sum would be the same if we swapped the indices $(\mu\nu) \leftrightarrow (\nu\mu)$. Therefore we may assume the components of M are symmetric.

Suppose both Alice and Bob both measure the worldline of an arbitrary photon, emitted at event P, and absorbed at event Q. Since this is light, the interval is zero. For Alice,

$$\begin{aligned}\Delta s^2 = 0 &= -c^2\Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2 \\ c^2\Delta t^2 &= \Delta x^2 + \Delta y^2 + \Delta z^2.\end{aligned} \tag{1.39}$$

Note that the photon travelled the same distance, but in the opposite direction, we have $(\Delta x, \Delta y, \Delta z) \to (-\Delta x, -\Delta y, -\Delta z)$ but Eq. (1.39) remains unchanged.

By PII, Bob should also measure a zero interval. Therefore

$$\begin{aligned}\Delta s'^2 = 0 &= \sum_{\mu,\nu=0}^{3} M_{\mu\nu}\Delta x^\mu \Delta x^\nu \\ &= M_{00}(\Delta x^0)^2 + 2\Delta x^0 \sum_{i=1}^{3} M_{0i}\Delta x^i + \sum_{i,j=1}^{3} M_{ij}\Delta x^i \Delta x^j.\end{aligned} \tag{1.40}$$

In the second line, we have explicitly written out the terms with $\mu = 0$ (the time direction). The indices $\mu = i, j$ are when they take values $1, 2, 3$, hence x^i or x^j represent the spatial directions. If this arbitrary photon travelled in the opposite direction, $(\Delta x, \Delta y, \Delta z) \to (-\Delta x, -\Delta y, -\Delta z)$, and Bob's interval becomes

$$\Delta s'^2 = 0 = M_{00}(\Delta x^0)^2 - 2\Delta x^0 \sum_{i=1}^{3} M_{0i}\Delta x^i + \sum_{i,j=1}^{3} M_{ij}\Delta x^i \Delta x^j. \tag{1.41}$$

Adding Eq. (1.40) to (1.41),

$$0 = 4\Delta x^0 \sum_{i=1}^{3} M_{0i}\Delta x^i. \tag{1.42}$$

Since Δx^i is arbitrary, we conclude that $M_{0i} = 0$. Knowing this, we now have, for Bob's measured interval for the photon,

$$
\begin{aligned}
0 &= M_{00}(\Delta x^0)^2 + \sum_{i,j=1}^{3} M_{ij}\Delta x^i \Delta x^j \\
&= M_{00}(\Delta x^0)^2 + \sum_{i=j} M_{ii}(\Delta x^i)^2 + \sum_{i\neq j} M_{ij}\Delta x^i \Delta x^j \\
&= M_{00}(\Delta x^0)^2 + \sum_{i=j} \left(M_{00} + M_{ii} - M_{00}\right)(\Delta x^i)^2 + \sum_{i\neq j} M_{ij}\Delta x^i \Delta x^j \\
&= M_{00}(\Delta x^0)^2 + \sum_{i}(-M_{00})(\Delta x^i)^2 + \sum_{i} \left(M_{ii} + M_{00}\right)(\Delta x^i)^2 + \sum_{i\neq j} M_{ij}\Delta x^i \Delta x^j \\
&= M_{00}\underbrace{\left[(\Delta x^0)^2 - (\Delta x^1)^2 - (\Delta x^2)^2 - (\Delta x^3)^2\right]}_{-\Delta s^2 = 0} + \sum_{i} \left(M_{00} + M_{ii}\right)(\Delta x^i)^2 \\
&\qquad + \sum_{i\neq j} M_{ij}\Delta x^i \Delta x^j
\end{aligned}
\tag{1.43}
$$

The term in the square brackets is just the negative of Alice's interval for the photon, which is zero. Therefore we are left with

$$
0 = \sum_{i} \left(M_{00} + M_{ii}\right)(\Delta x^i)^2 + \sum_{i\neq j} M_{ij}\Delta x^i \Delta x^j. \tag{1.44}
$$

Again, the photon is arbitrary. That means this relation should hold regardless of which direction the photon propagates, Δx^i. So for the above equation to hold true always,

$$
\begin{aligned}
M_{ii} &= -M_{00}, \\
M_{ij} &= 0 \quad \text{if } i \neq j.
\end{aligned}
$$

We conclude that the transformation between Alice's and Bob's interval is

$$
\begin{aligned}
\Delta s'^2 &= M_{00}(\Delta x^0)^2 - M_{00}\left((\Delta x^1)^2 + (\Delta x^2)^2 + (\Delta x^3)^2\right) \\
&= -M_{00}\left(-(\Delta x^0)^2 + (\Delta x^1)^2 + (\Delta x^2)^2 + (\Delta x^3)^2\right) \\
&= \phi(\vec{v})\Delta s^2.
\end{aligned}
\tag{1.45}
$$

In other words, Bob's interval is related to Alice's by an overall factor $\phi(\vec{v})$. Since Bob is moving at velocity $\vec{v}$ relative to Alice, there's a chance that this factor might depend on $\vec{v}$. However we will next show that in fact, $\phi = 1$.

Consider the time dilation thought experiment again. In the experiment, Bob measures the interval of two events. Event P is when the laser is emitted from the source, and event Q is when the laser returns to the source. To Bob, this occurs at the same position.

Therefore his interval is

$$\Delta s'^2 = -c^2 \Delta t'^2. \tag{1.46}$$

Alice measures the interval between P and Q in her frame. To Alice, these two events occur at different places, and different times. The interval she obtains is

$$\Delta s^2 = -c^2 \Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2. \tag{1.47}$$

Then Eq. (1.45) is

$$-c^2 \Delta t'^2 = -c^2 \Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2.$$

In Eq. (1.25), we have derived the time dilation formula, which says $\Delta t = \frac{\Delta t'}{\sqrt{1-\frac{u^2}{c^2}}}$, or $\Delta t'^2 = \left(1 - \frac{u^2}{c^2}\right) \Delta t^2$

$$\begin{aligned} -c^2 \Delta t^2 \left(1 - \frac{u^2}{c^2}\right) &= \phi \left(-c^2 \Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2\right) \\ -c^2 \cancel{\Delta t^2} \left(1 - \frac{u^2}{c^2}\right) &= -c^2 \phi \cancel{\Delta t^2} \underbrace{\left(1 - \frac{\Delta x^2 + \Delta y^2 + \Delta x^2}{\Delta t^2}\right)}_{1-\frac{u^2}{c^2}} \\ \phi &= 1. \end{aligned} \tag{1.48}$$

We then arrive at an important result: When transforming to a different inertial frame, the spacetime interval is the same,

$$\Delta s'^2 = \Delta s^2. \tag{1.49}$$

This property is known as **Lorentz invariance**. Recalling the assumptions we used to arrive at Eq. (1.49), we conclude that Lorentz invariance is a consequence of the postulates of SR.

Minkowski metric. Turning now to infinitesimal displacements in time and position, the spacetime interval is now written as

$$\mathrm{d}s^2 = -c^2 \mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2. \tag{1.50}$$

This is known as the *Minkowski metric*. Note that the right-hand-side of Eq. (1.50) can be viewed as a matrix multiplication:

$$
\begin{aligned}
-c^2\mathrm{d}t^2+\mathrm{d}x^2+\mathrm{d}y^2+\mathrm{d}z^2 &= \begin{pmatrix} c\mathrm{d}t & \mathrm{d}x & \mathrm{d}y & \mathrm{d}z \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} c\mathrm{d}t \\ \mathrm{d}x \\ \mathrm{d}y \\ \mathrm{d}z \end{pmatrix} \\
&= (\mathrm{d}x)^{\mathrm{T}}\eta(\mathrm{d}x) = \sum_{\mu,\nu=0}^{3} \eta_{\mu\nu}\mathrm{d}x^{\mu}\mathrm{d}x^{\nu}, \qquad (1.51)
\end{aligned}
$$

where $(\mathrm{d}x)$ and $(\mathrm{d}x)^{\mathrm{T}}$ represent the corresponding column vector and its transpose, respectively.

Earlier, we considered the fact that Bob's coordinates are related to Alice's coordinates by a linear transformation. For infinitesimal displacements, it would be

$$
\begin{aligned}
\begin{pmatrix} c\mathrm{d}t' \\ \mathrm{d}x' \\ \mathrm{d}y' \\ \mathrm{d}z' \end{pmatrix} &= \begin{pmatrix} \Lambda_{00} & \Lambda_{01} & \Lambda_{02} & \Lambda_{03} \\ \Lambda_{10} & \Lambda_{11} & \Lambda_{12} & \Lambda_{13} \\ \Lambda_{20} & \Lambda_{21} & \Lambda_{22} & \Lambda_{23} \\ \Lambda_{30} & \Lambda_{31} & \Lambda_{32} & \Lambda_{33} \end{pmatrix} \begin{pmatrix} c\mathrm{d}t \\ \mathrm{d}x \\ \mathrm{d}y \\ \mathrm{d}z \end{pmatrix} \\
(\mathrm{d}x') &= \Lambda(\mathrm{d}x).
\end{aligned}
$$

In matrix notation, Bob's spacetime interval is

$$
\mathrm{d}s'^2 = (\mathrm{d}x')^{\mathrm{T}}\eta(\mathrm{d}x') = (\mathrm{d}x)^{\mathrm{T}}\Lambda^{\mathrm{T}}\eta\Lambda(\mathrm{d}x). \qquad (1.52)
$$

But Lorentz invariance state that this is equal to Alice's $\mathrm{d}s^2$. Comparing Eq. (1.52) with (1.51) leads to

$$
\Lambda^{\mathrm{T}}\eta\Lambda = \eta. \qquad (1.53)
$$

We focus on one specific matrix that obey the above relation:

$$
\Lambda_x(\theta) = \begin{pmatrix} \cosh\vartheta & -\sinh\vartheta & 0 & 0 \\ -\sinh\vartheta & \cosh\vartheta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.
$$

This transformation is called a *Lorentz boost* in the x-direction, because it leaves the y and z directions unchanged. It can be checked directly that $\Lambda_x(\vartheta)^{\mathrm{T}}\eta\Lambda_x(\vartheta) = \eta$. The coordinate transformation from Alice to Bob reads

$$\begin{pmatrix} c\mathrm{d}t' \\ \mathrm{d}x' \\ \mathrm{d}y' \\ \mathrm{d}z' \end{pmatrix} = \begin{pmatrix} \cosh\vartheta & -\sinh\vartheta & 0 & 0 \\ -\sinh\vartheta & \cosh\vartheta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} c\mathrm{d}t \\ \mathrm{d}x \\ \mathrm{d}y \\ \mathrm{d}z \end{pmatrix} = \begin{pmatrix} c\mathrm{d}t\cosh\vartheta - \mathrm{d}x\sinh\vartheta \\ -c\mathrm{d}t\sinh\vartheta + \mathrm{d}x\cosh\vartheta \\ \mathrm{d}y \\ \mathrm{d}z \end{pmatrix}$$

What is the interpretation of ϑ? We can relate this to Bob's velocity relative to Alice. Suppose Bob sits on the origin of his own coordinate system. And Alice is measuring Bob's motion. Then she would obtain $c\mathrm{d}t$ and $\mathrm{d}x$ as Bob's infinitesimal displacement, and that Bob is moving at $\frac{\mathrm{d}x}{\mathrm{d}t} = u$. Bob, of course, would measure himself to be having zero displacement relative to his coordinate system. Therefore $\mathrm{d}x' = 0$. But Bob certainly experiences passage of time, so $c\mathrm{d}t' \neq 0$. The above equation is now

$$\begin{aligned} c\mathrm{d}t' &= c\mathrm{d}t\cosh\vartheta - \mathrm{d}x\sinh\vartheta, \\ 0 &= -c\mathrm{d}t\sinh\vartheta + \mathrm{d}x\cosh\vartheta \end{aligned}$$

Solving for $c\mathrm{d}t$ and $\mathrm{d}x$, we find

$$c\mathrm{d}t = \cosh\vartheta \mathrm{d}t', \quad \mathrm{d}x = \sinh\vartheta\ c\mathrm{d}t'.$$

Eliminating $\mathrm{d}t'$,

$$\frac{1}{c}\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{u}{c} = \tanh\vartheta. \tag{1.54}$$

Using identities of hyperbolic functions, we have

$$\cosh\vartheta = \frac{1}{\sqrt{1-u^2/c^2}}, \quad \sinh\vartheta = \frac{u}{c}\frac{1}{\sqrt{1-u^2/c^2}}.$$

Recognising the appearance of the Lorentz factor,

$$\gamma = \frac{1}{\sqrt{1-u^2/c^2}},$$

the general transformation between Alice's and Bob's coordinates are

$$\begin{aligned} c\mathrm{d}t' &= \gamma\left(c\mathrm{d}t - \frac{u\mathrm{d}x}{c}\right), \\ \mathrm{d}x' &= \gamma\left(\mathrm{d}x - u\mathrm{d}t\right), \\ \mathrm{d}y' &= \mathrm{d}y, \\ \mathrm{d}z' &= \mathrm{d}z. \end{aligned}$$

The corresponding point-wise transformation is

$$t' = \gamma\left(t - \frac{ux}{c^2}\right), \tag{1.56a}$$

$$x' = \gamma\left(x - ut\right), \tag{1.56b}$$

$$y' = y, \tag{1.56c}$$

$$z' = z, \tag{1.56d}$$

thereby recovering the familiar Lorentz transformation from first-year undergraduate courses. We see that the transformation $\Lambda_x(\vartheta)$ corresponds to the case where Bob is moving along Alice's x-direction. So their $y = y'$ and $z = z'$ are the same in both systems. We say that $\Lambda_x(\vartheta)$ is a **Lorentz boost** along the x-direction, and the parameter ϑ is called the **rapidity** which is related to the relative speed u by

$$\tanh\vartheta = \frac{u}{c}. \tag{1.57}$$

The matrices for Lorentz boosts along the y- and z- directions are defined similarly, and we collect all of them here:

$$\Lambda_x(\vartheta) = \begin{pmatrix} \cosh\vartheta & \sinh\vartheta & 0 & 0 \\ \sinh\vartheta & \cosh\vartheta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{1.58a}$$

$$\Lambda_y(\vartheta) = \begin{pmatrix} \cosh\vartheta & 0 & \sinh\vartheta & 0 \\ 0 & 1 & 0 & 0 \\ \sinh\vartheta & 0 & \cosh\vartheta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{1.58b}$$

$$\Lambda_z(\vartheta) = \begin{pmatrix} \cosh\vartheta & 0 & 0 & \sinh\vartheta \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \sinh\vartheta & 0 & 0 & \cosh\vartheta \end{pmatrix}. \tag{1.58c}$$

There are three more transformations that ensures $\mathrm{d}s'^2 = \mathrm{d}s^2$. These are

$$R_x(\varphi) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\varphi & -\sin\varphi \\ 0 & 0 & \sin\varphi & \cos\varphi \end{pmatrix} \tag{1.59a}$$

$$R_y(\varphi) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\phi & 0 & -\sin\phi \\ 0 & 0 & 1 & 0 \\ 0 & \sin\phi & 0 & \cos\phi \end{pmatrix}, \tag{1.59b}$$

$$R_z(\varphi) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi & 0 \\ 0 & \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{1.59c}$$

These transformation are only rotating the spatial axes, and does not transform the time parts. It can be shown that the above transformations obey $R_x(\varphi)^{\mathrm{T}} R_x(\varphi) = R_y(\varphi)^{\mathrm{T}} R_y(\varphi) = R_z(\varphi)^{\mathrm{T}} R_z(\varphi) = \mathbb{1}$, with the consequence that they preserve spatial lengths:

$$\mathrm{d}x'^2 + \mathrm{d}y'^2 + \mathrm{d}z'^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2.$$

These are members of the rotation group $O(3)$, called the *orthogonal group* of dimension 3. (Actually these belong to a smaller subset of elements with determinant $+1$. So they are the *special orthogonal group*, $SO(3)$.) All six transformations that preserves the spacetime interval,

$$SO(3,1) = \{\Lambda_x(\vartheta), \Lambda_y(\vartheta), \Lambda_z(\vartheta), R_x(\varphi), R_y(\varphi), R_z(\varphi)\}, \tag{1.60}$$

constitute the elements of the *Lorentz group*.

1.4 Tensors and the summation convention

In the following we will measure time in length units. That sounded weird, but it simply means taking $x^0 = ct$, where the dimension of x^0 is

$$\left[x^0\right] = [ct] = [c][t] = LT^{-1}T = L.$$

So that the Minkowski metric appears in a more symmetric form

$$\begin{aligned} ds^2 = \sum_{\mu,\nu=0}^{3} \eta_{\mu\nu} dx^\mu dx^\nu &= -\left(dx^0\right)^2 + dx^2 + dy^2 + dz^2. \\ &= -\left(dx^0\right)^2 + \left(dx^1\right)^2 + \left(dx^2\right)^2 + \left(dx^3\right)^2 \end{aligned} \tag{1.61}$$

Einstein summation convention. In the Minkowski metric, we have seen earlier that $\sum_{\mu,\nu=0}^{3} \eta_{\mu\nu} dx^\mu dx^\nu$ can be viewed as a matrix multiplication of a column vector, a 4×4 matrix, and a row vector.

Let A^μ be a column vector, and B_μ be a row vector. Then we have

$$\sum_{\mu=0}^{3} B_\mu A^\mu = \begin{pmatrix} B_0 & B_1 & B_2 & B_3 \end{pmatrix} \begin{pmatrix} A^0 \\ A^1 \\ A^2 \\ A^3 \end{pmatrix}$$

$$\sum_{\mu=0}^{3} B_\mu A^\mu = B_0 A^0 + B_1 A^1 + B_2 A^2 + B_3 A^3.$$

These types of expression appear often in relativistic calculations. Therefore, it is convenient to adopt **Einstein's summation convention**. The convention is simple – we notice that summation occurs when there's repeated indices in the expression. Therefore we simply *neglect writing the* $\sum_\mu$ *symbol*:

$$\sum_{\mu=0}^{3} B_\mu A^\mu = B_\mu A^\mu$$

The Minkowski metric involves a summation over two indices. So by the summation convention, we write

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu. \tag{1.62}$$

In its full meaning, it is a sum

$$\begin{aligned}
\mathrm{d}s^2 = \eta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu &= \sum_{\mu=0}^{3}\sum_{\nu=0}^{3}\eta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu \\
&= \eta_{00}\mathrm{d}x^0\mathrm{d}x^0 + \eta_{01}\mathrm{d}x^0\mathrm{d}x^1 + \eta_{02}\mathrm{d}x^0\mathrm{d}x^2 + \eta_{03}\mathrm{d}x^0\mathrm{d}x^3 \\
&\quad + \eta_{10}\mathrm{d}x^1\mathrm{d}x^0 + \eta_{11}\mathrm{d}x^1\mathrm{d}x^1 + \eta_{12}\mathrm{d}x^1\mathrm{d}x^2 + \eta_{13}\mathrm{d}x^1\mathrm{d}x^3 \\
&\quad + \eta_{20}\mathrm{d}x^2\mathrm{d}x^0 + \eta_{21}\mathrm{d}x^2\mathrm{d}x^1 + \eta_{22}\mathrm{d}x^2\mathrm{d}x^2 + \eta_{23}\mathrm{d}x^2\mathrm{d}x^3 \\
&\quad + \eta_{30}\mathrm{d}x^3\mathrm{d}x^0 + \eta_{31}\mathrm{d}x^3\mathrm{d}x^1 + \eta_{32}\mathrm{d}x^3\mathrm{d}x^2 + \eta_{33}\mathrm{d}x^3\mathrm{d}x^3.
\end{aligned} \tag{1.63}$$

For the Minkowski metric, we know that $\eta_{00} = -1$, $\eta_{ij} = \delta_{ij}$ for $i, j = 1, 2, 3$. Therefore the sum reduces to the four terms we see in Eq. (1.61).

Contravariant and covariant vectors. Looking at the Minkowski sum again, we notice that

$$\sum_{\mu=0}^{3}\sum_{\nu=0}^{3}\eta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = \sum_{\nu=0}^{3}\left(\sum_{\mu=0}^{3}\eta_{\mu\nu}\mathrm{d}x^\mu\right)\mathrm{d}x^\nu \tag{1.64}$$

The quantity in the brackets, after μ is summed over, has index ν. So this is itself a vector with four components:

$$\left(\sum_{\mu=0}^{3}\eta_{\mu\nu}\mathrm{d}x^\mu\right) = \eta_{\mu\nu}\mathrm{d}x^\mu = \mathrm{d}x_\nu, \quad \nu = 0, 1, 2, 3.$$

Indices that are not yet summed over will be called *free indices*. This can be applied to any vector u^μ. Operating η on this vector,

$$u_\mu = \eta_{\mu\nu}u^\nu = \sum_{\nu=0}^{3}\eta_{\mu\nu}u^\nu, \quad \mu = 0, 1, 2, 3. \tag{1.65}$$

So every vector with an upper index u^μ can be converted into another 'vector' with lower index $u_\mu = \eta_{\mu\nu}u^\nu$. Here, those with upper indices u^μ are called **contravariant vectors** and those with lower indices u_μ are called **covariant vectors**. A contravariant vector can be **contracted** with a covariant vector, which sums to a scalar number:

$$u_\mu u^\mu = \eta_{\mu\nu}u^\mu u^\nu = u_0u^0 + u_1u^1 + u_2u^2 + u_3u^3 = \text{a number}.$$

Example 1.4.1. In Minkowski spacetime, let a contravariant vector u^μ be given by components

$$u^0 = a, \quad u^1 = b, \quad u^2 = c, \quad u^3 = d.$$

For simplicity we may write $u^\mu = (a, b, c, d)$. Then, the covariant version is

$$u_\mu = (u_0, u_1, u_2, u_3) = \eta_{\mu\nu} u^\nu.$$

Note the summation convention. To obtain, say u_2, we compute the sum

$$\begin{aligned} u_2 = \eta_{2\nu} u^\nu &= \sum_{\nu=0}^{3} \eta_{2\nu} u^\nu = \eta_{20} u^0 + \eta_{21} u^1 + \eta_{22} u^2 + \eta_{23} u^3 \\ &= 0 + 0 + 1(c) + 0 \\ &= c. \end{aligned}$$

In this case, $u_2 = c$ happens to be the same as u^2 unchanged. But for the component u_0,

$$\begin{aligned} u_0 = \eta_{0\nu} u^\nu &= \sum_{\nu=0}^{3} \eta_{0\nu} u^\nu = \eta_{00} u^0 + \eta_{01} u^1 + \eta_{02} u^2 + \eta_{03} u^3 \\ &= (-1)a + 0 + 0 + 0 \\ &= -a. \end{aligned}$$

Computing all the rest, we find that

$$u_\mu = (-a, b, c, d), \quad u^\mu = (a, b, c, d).$$

And the contraction is

$$u_\mu u^\mu = -a^2 + b^2 + c^2 + d^2.$$

Using the summation convention, Lorentz transformations are written as

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

where x'^μ are Bob's coordinates and x^μ are Alice's coordinates. For the Lorentz boost in the x-direction of Alice, the components of $\Lambda^\mu{}_\nu$ are explicitly

$$\Lambda^0{}_0 = \Lambda^1{}_1 \cosh\vartheta, \quad \Lambda^0{}_1 = \Lambda^1{}_0 = -\sinh\vartheta, \quad \Lambda^2{}_2 = \Lambda^3{}_3 = 1, \tag{1.66}$$

and all the other components are zero.

Let $A^\mu{}_\nu$ and $B^\mu{}_\nu$ be two quantities, each with two indices. Therefore we can view them as two 4×4 matrices. The matrix multiplication between these two results in a new 4×4 matrix, say, $C^\mu{}_\nu$. Using our summation convention, this matrix multiplication is written as

$$\begin{aligned} C &= A \cdot B \\ C^\mu{}_\nu &= A^\mu{}_\lambda B^\lambda{}_\nu. \end{aligned} \tag{1.67}$$

Recall that under the summation convention, we have a repeated index λ, which is summed over. The remaining indices are μ and ν, which are two free indices. Objects with two free indices are viewed in the present context as a 4×4 matrix.

Now, the Minkowski coefficients is a 'matrix' with components $\eta_{\mu\nu}$. Its determinant is non-zero. So as a 4×4 matrix, it should have an inverse. We'll denote this inverse by $\eta^{\mu\nu}$ (with two upper indices). The multiplication of a matrix with its inverse gives an identity matrix:

$$\eta^{\mu\lambda}\eta_{\lambda\nu} = \delta^\mu_\nu, \tag{1.68}$$

where δ^μ_ν is the Kroenecker delta. Its components are 1 if $\mu = \nu$, and 0 otherwise.

We can generalise all these concepts to mathematical objects with arbitrary number of indices. An (r, s)-*tensor* is a tensor with r contravariant indices and s covariant indices:

$$T^{\mu_1\cdots\mu_s}{}_{\nu_1\cdots\nu_r}.$$

Given any tensor, its indices can be raised or lowered using $\eta^{\mu\nu}$ or $\eta_{\mu\nu}$. For instance,

$$\eta^{\mu\alpha}M_{\alpha\beta} = M^\mu{}_\beta, \quad \eta_{\mu\nu}M^{\nu\beta} = M_\mu{}^\beta.$$

We have seen that trying to raise one index of $\eta_{\mu\nu}$ using its inverse results in the Kroenecker delta, or the identity matrix,

$$\eta^{\mu\alpha}\eta_{\alpha\nu} = \delta^\mu_\nu.$$

We denote the partial derivative with respect to coordinate x^μ as

$$\frac{\partial}{\partial x^\mu} = \partial_\mu. \tag{1.69}$$

Let $f(x^0, x^1, x^2, x^3)$ be a function of spactime coordinates. Therefore the *gradient* of f would be

$$\begin{aligned}\partial_\mu f &= (\partial_0 f, \partial_1 f, \partial_2 f, \partial_3 f)\\ &= (\partial_0 f, \vec{\nabla} f),\end{aligned} \tag{1.70}$$

where $\vec{\nabla} = (\partial_1, \partial_2, \partial_3)$ is the ordinary gradient in the spatial coordinates.

If A^μ is a contravariant vector, then the **divergence** in spacetime is

$$\partial_\mu A^\mu = \partial_0 A^0 + \partial_1 A^1 + \partial_2 A^2 + \partial_3 A^3 = \partial_0 A^0 + \vec{\nabla} \cdot \vec{A}, \tag{1.71}$$

where the spatial components form a Cartesian vector $\vec{A} = (A^1, A^2, A^3)$, and $\vec{\nabla} \cdot \vec{A}$ is the usual divergence.

So far we have said that x^μ is a contravariant vector. But more precisely, they are *components* of a contravariant vector. Components of a vector must come with basis vectors. Recall in $\mathbb{R}^3$, a vector is

$$\vec{r} = x\,\hat{\imath} + y\,\hat{\jmath} + z\,\hat{k}.$$

So the components are (x, y, z) and the basis vectors are $\hat{\imath}$, $\hat{\jmath}$, and $\hat{k}$.

Example 1.4.2. Consider two-dimensional Minkowski spacetime (ignoring the y and z directions) with metric

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + \mathrm{d}x^2 = -\left(\mathrm{d}x^0\right)^2 + \left(\mathrm{d}x^1\right)^2.$$

A $(1,1)$-tensor $M^\mu{}_\nu$ is given with components

$$M^\mu{}_\nu = \begin{pmatrix} M^0{}_0 & M^0{}_1 \\ M^1{}_0 & M^1{}_1 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Determine the components of $M_{\mu\nu}$.

Solution. *Since* $M_{\mu\nu} = \eta_{\mu\lambda} M^{\lambda}{}_{\nu}$,

$$\begin{aligned} M_{00} &= \eta_{0\lambda} M^{\lambda}{}_{0} = \eta_{00} M^{0}{}_{0} = -a, \\ M_{01} &= \eta_{0\lambda} M^{\lambda}{}_{1} = \eta_{00} M^{0}{}_{1} = -b, \\ M_{10} &= \eta_{1\lambda} M^{\lambda}{}_{0} = \eta_{11} M^{1}{}_{0} = c, \\ M_{11} &= \eta_{1\lambda} M^{\lambda}{}_{1} = \eta_{11} M^{1}{}_{1} = d, \end{aligned}$$

Therefore

$$M_{\mu\nu} = \begin{pmatrix} M_{00} & M_{01} \\ M_{10} & M_{11} \end{pmatrix} = \begin{pmatrix} -a & -b \\ c & d \end{pmatrix}$$

1.5 Particle motion in Special Relativity

Proper time. Suppose Alice is observing Bob, now moving at arbirary velocities. (Not necessarily constant velocity anymore.) Therefore now Bob's coordinate system is not inertial anymore. But, suppose Bob carries a clock with him. And at any instant, we can attach a three-dimensional coordinate system to Bob instantaneously, as shown in Fig. 1.13

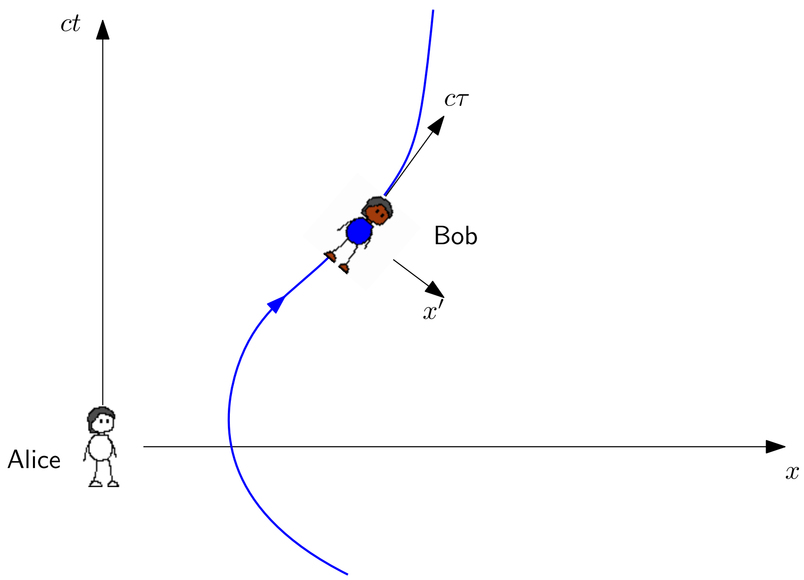


Figure 1.13: Bob is moving at non-constant velocity relative to Alice. However, we construct an instantaneous inertial frame at the moment he is at point P.

During that instant, we suppose that Bob's velocity is some value $\vec{v} = \frac{\mathrm{d}\vec{r}}{\mathrm{d}t}$ according to Alice. Therefore we can now use v as an instantaneous inertial frame.

During an infinitesimal time interval $\mathrm{d}t$ for Alice, she measures Bob to have moved a distance $\sqrt{\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2}$ and she measures a spacetime interval

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2. \tag{1.72}$$

During this interval, Bob measures himself to have an interval

$$\mathrm{d}s'^2 = -c^2\mathrm{d}t'^2. \tag{1.73}$$

By Lorentz invariance, $\mathrm{d}s'^2 = \mathrm{d}s^2$ and we have

$$\mathrm{d}t' = \mathrm{d}t\sqrt{1 - \frac{\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2}{c^2\mathrm{d}t^2}} = \mathrm{d}t\sqrt{1 - \frac{v^2}{c^2}} \tag{1.74}$$

So, even though Bob's is not inertial relative to Alice, we regard his motion as a sequence of infinitesimal time intervals $\mathrm{d}t'$, where t' is the time recorded by Bob's clock. By definition, this is Bob's proper time, and the time taken of his motion measured by Bob is

$$\tau = \int \mathrm{d}t' = \int \mathrm{d}t\,\sqrt{1 - \frac{v^2}{c^2}} = \frac{1}{c}\int \mathrm{d}s, \tag{1.75}$$

where $\mathrm{d}s$ is the infinitesimal spacetime interval as measured by Alice.

In Newtonian mechanics, a particle's motion traces out a curve in $\mathbb{R}^3$. Its velocity at each point on the curve is the corresponding tangent vector. We extend this definition to worldlines in four-dimensional spacetime. Suppose the motion of particle is being observed by Alice, using coordinates (ct, x, y, z). Now, if Bob plays the role of the moving particle, and τ is his proper time, we may represent the worldline as a curve in spacetime parametrised by τ:

$$x^\mu(\tau) = (ct(\tau), x(\tau), y(\tau), z(\tau)) \tag{1.76}$$

The **four-velocity** of a particle's worldline is the tangent vector

$$u^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} = \dot{x}^\mu, \tag{1.77}$$

By the invariance of the spacetime interval, (with $t' = \tau$ as Bob's time),

$$\begin{aligned}
\mathrm{d}s'^2 &= \mathrm{d}s^2 \\
-c^2\mathrm{d}\tau^2 &= -c^2\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 \\
-c^2 &= -c^2\left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^2 + \left(\frac{\mathrm{d}x}{\mathrm{d}\tau}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}\tau}\right)^2 + \left(\frac{\mathrm{d}z}{\mathrm{d}\tau}\right)^2 \\
-c^2 &= -c^2\dot{t}^2 + \dot{x}^2 + \dot{y}^2 + \dot{z}^2 \\
-c^2 &= \eta_{\mu\nu}\dot{x}^\mu\dot{x}^\nu = u_\mu u^\mu.
\end{aligned} \tag{1.78}$$

We see that a (time-like) particle's four-velocity must obey

$$\boxed{u_\mu u^\mu = -c^2.} \tag{1.79}$$

We can also obtain the inverse transformation of the time dilation as follows. If Bob only moves in the x-direction according to Alice's coordinate system, his 4-velocity is

$$u^\mu = \left(c\dot{t}, \dot{x}, 0, 0\right).$$

Then, the condition $u_\mu u^\mu = -c^2$ leads to

$$\begin{aligned}
-c^2\dot{t}^2 + \dot{x}^2 &= -c^2 \\
c^2\left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^2 - \left(\frac{\mathrm{d}x}{\mathrm{d}\tau}\right)^2 &= c^2 \\
c^2\left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^2 - \left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2\left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^2 &= c^2 \\
\left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^2\left[c^2 - u^2\right] &= c^2,
\end{aligned}$$

where $u = \frac{\mathrm{d}x}{\mathrm{d}t}$ is Bob's velocity measured according to Alice. Further rearranging and taking the positive square root,

$$\frac{\mathrm{d}t}{\mathrm{d}\tau} = \frac{1}{\sqrt{1 - u^2/c^2}} = \gamma, \tag{1.80}$$

where we recognised the presence of the Lorentz factor γ. Then, the time interval measured by Alice is

$$t = \int \gamma\, \mathrm{d}\tau = \int \frac{\mathrm{d}\tau}{\sqrt{1 - u^2/c^2}}. \tag{1.81}$$

Compare this equation with Eq. (1.75). This is a generalisation of the time dilation formula, as this is still valid even if Bob is moving at non-constant speed u. If u is constant, we recover $\Delta t = \Delta\tau/\sqrt{1-u^2/c^2}$.

Lagrangian formulation. We can use the Lagrangian formulation that is consistent with SR. Recall that a classical Lagrangian $\mathcal{L}(q,\dot{q})$ is a function of generalised coordinates q and velocities $\dot{q}$. For SR, we take the generalised coordinates to be $(x^0, x^1, x^2, x^3) = (ct, x, y, z)$. This means $x^0 = ct$ is a fourth generalised coordinate. The velocities are derivatives with respect to Bob's proper time,

$$\dot{x}^0 = c\dot{t} = c\frac{\mathrm{d}t}{\mathrm{d}\tau}, \quad \dot{x} = \frac{\mathrm{d}x}{\mathrm{d}\tau}, \quad \dot{y} = \frac{\mathrm{d}y}{\mathrm{d}\tau}, \quad \dot{z} = \frac{\mathrm{d}z}{\mathrm{d}\tau}. \tag{1.82}$$

Let m be the mass of Bob (the particle.) The Lagrangian is

$$\mathcal{L} = \frac{1}{2}m\left(-c^2\dot{t}^2 + \dot{x}^2 + \dot{y}^2 + \dot{z}^2\right). \tag{1.83}$$

The canonical momenta is defined to be a *covariant* vector $p_\mu = \frac{\partial\mathcal{L}}{\partial x^\mu}$. We find that $p_\mu = mu_\mu$. Writing out the components explicitly,

$$p_t = -mc\dot{t}, \quad p_x = m\dot{x}, \quad p_y = m\dot{y}, \quad p_z = m\dot{z}. \tag{1.84}$$

Note that p_x, p_y, and p_z are the same as the usual Newtonian momentum. So this quantity p_t is a new concept in SR. The contravariant version of this momentum is

$$p^0 = -p_0 = mc\dot{t}, \quad p^x = m\dot{x}, \quad p^y = m\dot{y}, \quad p^z = m\dot{z}. \tag{1.85}$$

(Only p^0 is different from p_0.) We note also the contraction

$$\begin{aligned} p_\mu p^\mu &= -m^2c^2\dot{t}^2 + m^2\dot{x}^2 + m^2\dot{y}^2 + m^2\dot{z}^2 = m^2\left(-c^2\dot{t}^2 + \dot{x}^2 + \dot{y}^2 + \dot{z}^2\right) \\ &= m^2u_\mu u^\mu = -m^2c^2. \end{aligned} \tag{1.86}$$

In SR, the results of an observer's measurement depends on the observer's motion. Let V^μ be the four-velocity of the observer. We require this observer to be time-like, since the observer themselves should not travel faster than light. Hence V^μ must also satisfy $V^\mu V_\mu = -c^2$. We *define* the energy of Bob measured by this observer to be

$$E = -V^\mu p_\mu. \tag{1.87}$$

If this observer is static in Alice's coordinate system, $V^\mu = (c, 0, 0, 0)$. This gives

$$E = -cp_t = -c(-mc\dot{t}) = mc^2\dot{t} \quad \rightarrow \quad \dot{t} = \frac{E}{mc^2}.$$

So, if E is the energy measured by an observer static relative to Alice,

$$p_t = mc\dot{t} = \frac{E}{c}, \tag{1.88}$$

and we find that p_t is the energy of the particle m. Now from Eq. (1.80), we have $\dot{t} = \gamma = 1/\sqrt{1\frac{u^2}{c^2}}$, then

$$\dot{t} = \gamma = \frac{E}{mc^2} \quad \rightarrow \quad E = \gamma mc^2 = \frac{mc^2}{\sqrt{1 - u^2/c^2}},$$

we recover the expression for relativistic total energy of a particle measured by Alice. If Bob is moving at speed slow compared to c, we expand

$$\begin{aligned} E &= mc^2\left(1 + \frac{1}{2}\frac{u^2}{c^2} + \frac{3}{8}\frac{u^4}{c^4} + \dots\right) \\ E &= mc^2 + \underbrace{\frac{1}{2}mu^2 + \frac{3}{8}\frac{u^4}{c^2} + \dots}_{\text{kinetic energy}} \end{aligned}$$

We see that the total energy is mc^2+velocity-dependent terms, which we interpret as the *relativistic kinetic energy*, or $K = E - mc^2$. In the regime where $u \ll c$, the higher order terms are negligible and we recover the usual $K = \frac{1}{2}mu^2$.

Next, applying Eq. (1.88) to Eq. (1.86),

$$\begin{aligned} p_\mu p^\mu = -m^2c^2 &= p_t p^t + p_x p^x + p_y p^y + p_z p^z \\ -m^2c^2 &= -\frac{E^2}{c^2} + p_x^2 + p_y^2 + p_z^2 \\ -m^2c^4 &= -E^2 + p^2c^2, \end{aligned}$$

where $p^2 = \vec{p}\cdot\vec{p} = p_x^2 + p_y^2 + p_z^2$ is the magnitude of the standard Newtonian momentum $\vec{p} = m\vec{u}$. Rearranging this equation and taking the positive square root,

$$E = \sqrt{(mc^2)^2 + (pc)^2}. \tag{1.89}$$

If Bob has zero velocity in Alice's coordinate system, $p = 0$, and we recover the famous formula $E = mc^2$.

Euler–Lagrange equations. The Euler–Lagrange equation is

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\frac{\partial \mathcal{L}}{\partial \dot{x}^\mu} = \frac{\partial \mathcal{L}}{\partial x^\mu}$$

for each component $\mu = 0, 1, 2, 3$. Explicitly, for $\mu = 0$,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\tau}\frac{\partial \mathcal{L}}{\partial (c\dot{t})} &= \frac{\partial \mathcal{L}}{\partial (ct)} \\ \ddot{t} &= 0. \end{aligned}$$

For $\mu = 1$,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\tau}\frac{\partial \mathcal{L}}{\partial \dot{x}} &= \frac{\partial \mathcal{L}}{\partial x} \\ \ddot{x} &= 0. \end{aligned}$$

Similarly we can calculate $\ddot{y} = 0$ and $\ddot{z} = 0$. Now, the three equations $\ddot{x} = 0$, $\ddot{y} = 0$, and $\ddot{z} = 0$ is expected and consistent with traditional Newton's laws; when the force is zero, the trajectories are straight lines. The equation that comes new with SR is $\ddot{t} = 0$. This means

$$t = a\tau + b,$$

where a and b are constants. In the absence of forces, proper time τ and coordinate time t are related linearly, though they are not equal to each other.

Particles experiencing forces can be studied by including a potential into the Lagrangian,

$$\mathcal{L} = \frac{1}{2}m\left(-c^2\dot{t}^2 + \dot{x}^2 + \dot{y}^2 + \dot{z}^2\right) - V(x). \tag{1.90}$$

Then the Euler–Lagrange equation would give

$$m\ddot{x}^\mu = -\frac{\partial V}{\partial x^\mu} = f^\mu,$$

where f^μ is the *four-force*.

1.6 Causal structure and light cones

From Eq. (1.89), we see that $E \to \infty$ as $u \to c$, regardless of particle mass m. Therefore it takes infinite energy for any particle to reach the speed of light. Light, which consists of photons, travel exactly at the speed of light in all frames. Nothing can travel faster than light.

This brings us to the concept of *light cones*. Consider an event P. Whatever that happens at P can only affect future events in the future light cone of P, shown in Fig. 1.14.

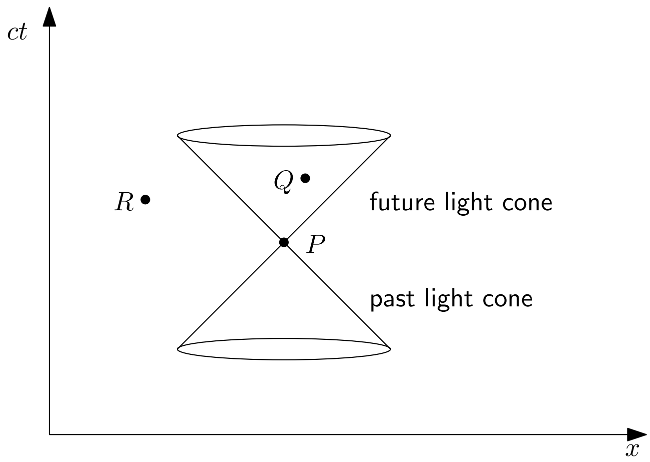


Figure 1.14: The light cone of P.

The light cone of P consists of two parts, the upper part called the *future light cone*, and the lower part *past light cone*. P can only influence and communicate with any events within the future light, such as Q. We see that R lies outside the light cone of P. So P can never influence anything at R, since nothing can travel faster than light. We say that R and P are *causally disconnected*.

Similarly, only events in the past light cone of P can influence P. In other words, P can only receive information and signals only from events in the past light cone.

1.7 The twin paradox*

Alice and Bob are a pair of fratenal twins born on the same day. Alice works as a NASA mission controller, while Bob is an astronaut. On their 30th birthday. Bob boards a spaceship to reach a distant star. Assume that the spaceship always travels at constant speed $u = 0.9c$ relative to the Earth. Therefore, from the perspective of the Earth, Bob ship will take 10 years to reach the star, and another 10 more for the return trip.

Question: When Bob returns to Earth and reunites with Alice, will there be an age difference between them? There are two possibilities.

Possibility 1. From Alice's point of view Bob is moving at $u = 0.9c$. Obviously Bob's age is Bob's proper time, $\tau = t_B$. The time taken for Bob's journey, according to Alice's measurement, is

$$\Delta t_A = \frac{\Delta t_B}{\sqrt{1 - u^2/c^2}}.$$

So, Bob has grown older by $T_B = \Delta t_B$ while Alice has grown older by $T_A = \Delta t_A$. In this case, since $\gamma > 1$, we have $T_A > T_B$. So Alice has become older than Bob.

Possibility 2. From Bob's point of view, Alice is moving away from Bob's frame at speed $u = 0.9c$. Now Bob will clearly see that Alice's age is Alice's proper time. During the journey, the time taken according to Bob is

$$\Delta t_B = \frac{\Delta t_A}{\sqrt{1 - u^2/c^2}}.$$

So Alice grows older by $T_A = \Delta t_A$ and Bob grows older by $T_B = \Delta t_B$. But in this case, $T_B < T_A$. Alice is younger than Bob.

Alice calculates $T_A > T_B$ but Bob calculates $T_B > T_A$. Obviously one of them has to be wrong. Which is it? In a sense, this problem can be easily resolved by observation. When Bob reunites with Alice, we'll just see who is the elderly person!

The true answer to this paradox is that **Possibility 2 is WRONG.** Note that the Postulates of SR requires that all laws of physics are the same for all inertial frames. But Bob gets on a spaceship and makes a return trip. So he changes direction and hence there's an acceleration! *Bob's frame is not inertial.* Therefore the Postulates of SR does not require Bob to apply the same equation as Alice!

In this scenario, only Alice has an inertial frame. We can only use the Minkowski metric in her frame,

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2.$$

The motion of Bob is a worldline in Alice's spacetime diagram. Since Bob's frame is non-inertial, the only valid measurement from Bob the measurement on himself. After an infinitesimal displacement, the spacetime intervals measured by both are

$$\begin{aligned}
\mathrm{d}s_A^2 &= \mathrm{d}s_B^2 \\
-c^2\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 &= -c^2\mathrm{d}\tau^2 \\
-c^2 + \left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2 + \left(\frac{\mathrm{d}z}{\mathrm{d}t}\right)^2 &= -c^2\left(\frac{\mathrm{d}\tau}{\mathrm{d}t}\right)^2.
\end{aligned}$$

Let $u = \sqrt{\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2 + \left(\frac{\mathrm{d}z}{\mathrm{d}t}\right)^2}$ be Bob's speed measured by Alice. Then

$$c^2 - u^2 = c^2 \frac{\mathrm{d}\tau}{\mathrm{d}t} \quad \rightarrow \quad \int \mathrm{d}\tau = \int \sqrt{1 - \frac{u^2}{c^2}} \mathrm{d}t$$

Since the speed u is constant in both directions, and knowing that Bob's ship travels for 20 years in Alice's perspective, Bob's perspective of time is

$$\begin{aligned} \Delta\tau &= \int_0^{\Delta t} \sqrt{1 - \frac{u^2}{c^2}}\, \mathrm{d}t = \Delta t \sqrt{1 - \frac{u^2}{c^2}} \\ &= (20\text{ yrs})\sqrt{1 - (0.9)^2} \\ &= 8.71\text{ yrs.} \end{aligned}$$

For Bob, only 8.7 years has passed. So, when they reunited, Bob is 38.7 years old but Alice is now 50 years old.

Chapter 2 Manifolds and tensors

In Relativity, it is important to remember the distinction between the *spacetime* itself vs the *coordinates* used to represent points of a spacetime. A simple analogy is the following: Suppose two people want to discuss various landmarks in Paris. One obvious way to do it is to actually walk around the streets of Paris to see the landmarks themselves. However, if these two people are currently in Kuala Lumpur, they can't do that easily. What they *can* do is to pull up Google maps (or a conventional paper map) and point to the landmarks on the map.

Clearly, pointing to a map is not the same as actually being in Paris. Yet, the maps *represent* various points on Paris. In relativity, *coordinate systems* play the role of maps. They are representation of some actual physical event in spacetime. The distinction between coordinates and physical points is important once we remember we can use different coordinate systems to represent the same point. The same point can have different values in Cartesian coordinates (x, y, z) or spherical coordinates (r, θ, ϕ).

To handle this in relativity, we need to use the language of *differential geometry*, which (among many other things) systematically treats the concepts of coordinate systems vs spacetime points.

2.1 Manifolds

In relativity, the basic physical quantity is an *event*, which is a point in spacetime: A time and place where something happens. In the previous chapter, we would have written an event as[1] (t, x, y, z), but the four numbers here depend on the coordinate system used.

This situation is depicted in Fig. 2.1. The spacetime is a set of events which we denote as $\mathcal{M}$. A particular event is a point p. When we wish to write down the coordinates corresponding to an event, we use the coordinate (t, x, y, z), which is a point in four-dimensional vector space $\mathbb{R}^4$.

[1] A reminder that from this point forward we will be using units where $t = ct_{\mathrm{SI}}$, or in units where $c = 1$.

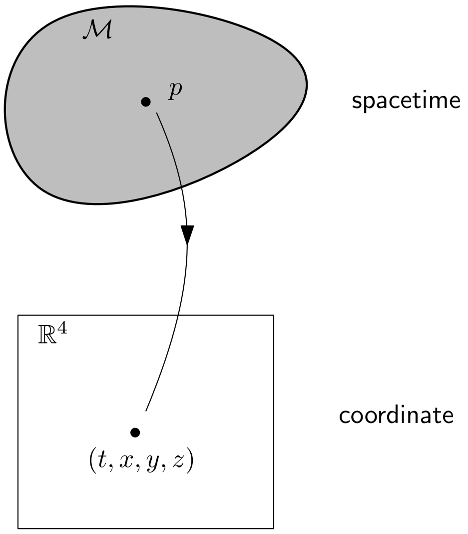


Figure 2.1: The event p in spacetime $\mathcal{M}$ and its coordinate representation (t, x, y, z).

Here, we have stepped into a field of *differential geometry*, whose goal (among other things) is to study smooth/curved sets like $\mathcal{M}$ in Fig. 2.1. Relativity is an application of differential geometry in the case where $\mathcal{M}$ is spacetime. In differential geometry, one learns how to define and calculate the *curvature*. General Relativity is famously a theory about how spacetime is curved, and we often see pictures like Fig. 2.2 to describe 'curved spacetime'. In order to wield General Relativity proficiently, we have to learn how to describe the so-called 'curved spacetime' in a mathematically precise way.

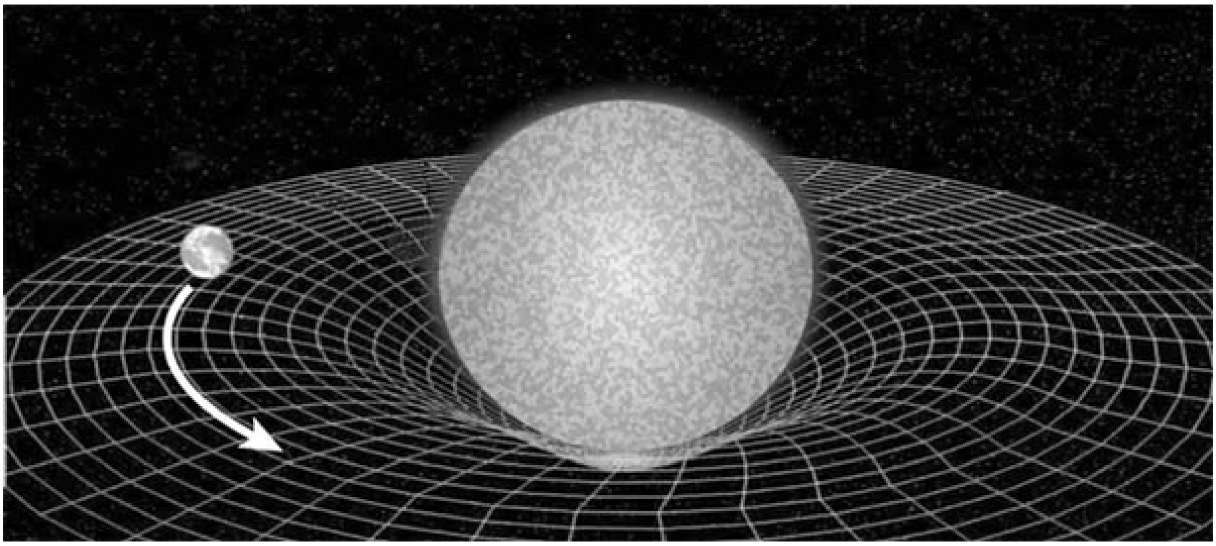


Figure 2.2: Cartoon depiction of spacetime curvature. Image from `https://uh.edu/~jclarage/astr3131/lectures/4/einstein/Einstein_stanford_Page7.html`

We start with a *manifold* $\mathcal{M}$, which is generally some kind of smooth, curved object.[2] As discussed earlier, in order to describe points on a manifold, we need to specify a coordinate

[2]For concreteness, physicists may imagine the manifold as a spacetime.

system. A coordinate system is a map which assign coordinates to each point on $\mathcal{M}$. This system of assigning coordinates is called *a chart*. The coordinates are typically a set of n real numbers, $(x^1, \ldots, x^n)$. Therefore we denote the set of coordinates by $\mathbb{R}^n$, as shown in Fig. 2.3. If a manifold $\mathcal{M}$ requires n numbers to describe its points, we say that $\mathcal{M}$ is an n-dimensional manifold, and we write

$$\dim \mathcal{M} = n.$$

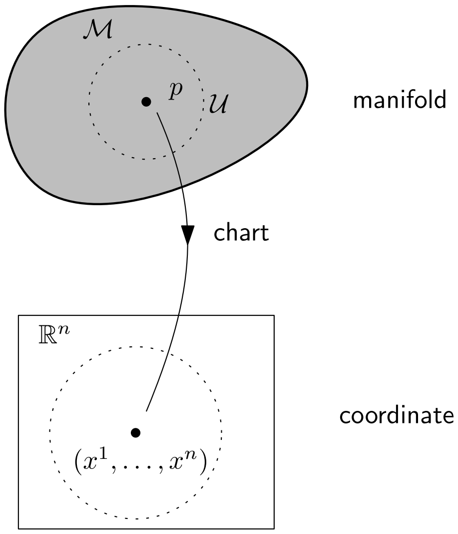


Figure 2.3: A manifold $\mathcal{M}$ and its coordinate chart.

Example 2.1.1. A familiar example of a manifold is the case $\mathcal{M} = \mathbb{R}^3$, the familiar three-dimensional space in Newtonian mechanics. This manifold is already $\mathbb{R}^3$ therefore a coordinate chart to Cartesian coordinate is simply the identity map, marked as 'chart 1' in Fig. 2.4. So any point $p \in \mathbb{R}^3$ can be immediately identified with its Cartesian coordinates (x, y, z). On the other hand, three-dimensional space can also be covered in spherical coordinates, marked 'chart 2' in Fig. 2.4. So a point can be represented as coordinates (r, θ, ϕ). The transformation between the two different coordinate system is

$$x = r \sin\theta \cos\phi, \quad y = r \sin\theta \sin\phi, \quad z = r \cos\theta. \tag{2.1}$$

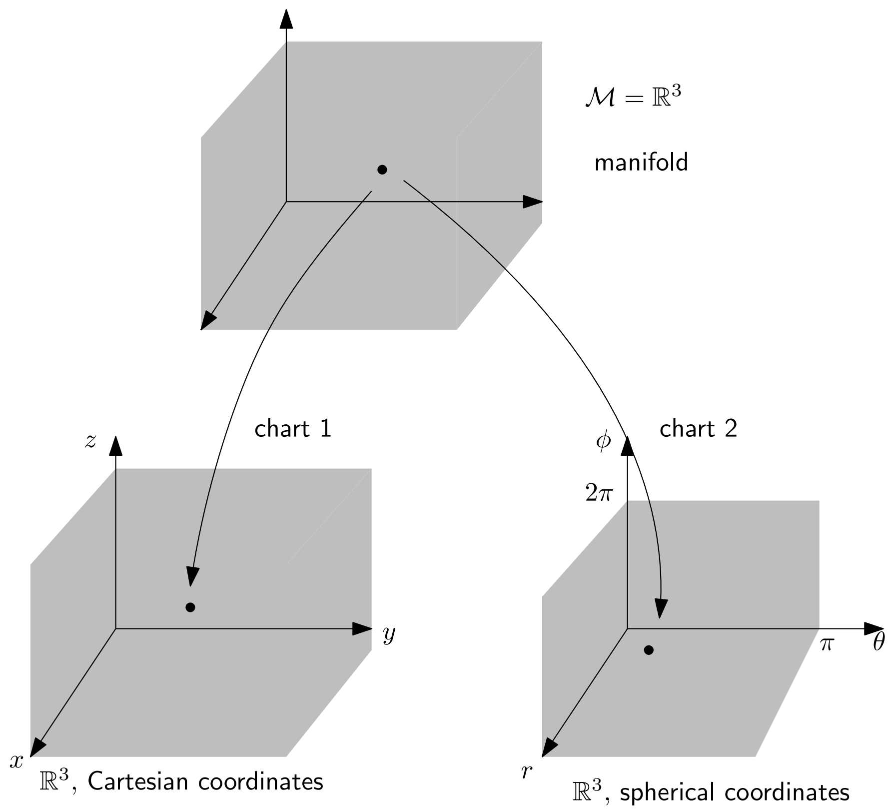


Figure 2.4: The manifold $\mathcal{M} = \mathbb{R}^3$ and two different coordinate charts.

As we seen in the previous example, one can use different coordinate systems to describe the same manifold. Suppose in an n-dimensional manifold $\mathcal{M}$, a point p is given under a chart as

$$\left(x^1, \ldots, x^n\right) . \tag{2.2}$$

In a second chart, the same point can be given as $(y^1, \ldots, y^n)$. Each y^μ can be written expressed as a function of the previous coordinates $(x^1, \ldots, x^n)$. (For instance, Eq. (2.1) in the example $\mathcal{M} = \mathbb{R}^3$.) In other words,

$$y^\mu = y^\mu \left(x^1, \ldots, x^n\right) . \tag{2.3}$$

An important quantity is the *transformation matrix*

$$\frac{\partial y^\mu}{\partial x^\nu} = \begin{pmatrix} \frac{\partial y^1}{\partial x^1} & \frac{\partial y^1}{\partial x^2} & \cdots & \frac{\partial y^1}{\partial x^n} \\ \frac{\partial y^2}{\partial x^1} & \frac{\partial y^2}{\partial x^2} & \cdots & \frac{\partial y^2}{\partial x^n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y^n}{\partial x^1} & \frac{\partial y^n}{\partial x^2} & \cdots & \frac{\partial y^n}{\partial x^n}. \end{pmatrix} \tag{2.4}$$

Note that the rows are labelled by the index in the numerator of the partial derivative and columns by the index of the denominator.

In most situations, we expect to be able to invert the transformation. So viewing in the other direction, we can express x^μ as a function of $(y^1, \ldots, y^n)$. The transformation matrix of the inverse transformation would be

$$\frac{\partial x^\mu}{\partial y^\nu} = \begin{pmatrix} \frac{\partial x^1}{\partial y^1} & \frac{\partial x^1}{\partial y^2} & \cdots & \frac{\partial x^1}{\partial y^n} \\ \frac{\partial x^2}{\partial y^1} & \frac{\partial x^2}{\partial y^2} & \cdots & \frac{\partial x^2}{\partial y^n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x^n}{\partial y^1} & \frac{\partial x^n}{\partial y^2} & \cdots & \frac{\partial x^n}{\partial y^n}. \end{pmatrix} \tag{2.5}$$

By applying the chain rule, we find

$$\frac{\partial x^\mu}{\partial x^\nu} = \frac{\partial x^\mu}{\partial y^\alpha}\frac{\partial y^\alpha}{\partial x^\nu}.$$

Clearly the expression on the left hand side equals 1 when $\mu = \nu$, and zero otherwise. (Partial derivatives of a variable with any other variable in the same coordinate system is zero.) Therefore we have

$$\frac{\partial x^\mu}{\partial y^\alpha}\frac{\partial y^\alpha}{\partial x^\nu} = \delta^\mu_\nu. \tag{2.6}$$

In other words, the transformation matrix $\frac{\partial x^\mu}{\partial y^\alpha}$ is the inverse of the inverse transformation matrix $\frac{\partial y^\alpha}{\partial x^\nu}$.

The determinant of the transformation matrix is the *Jacobian*

$$J = \det\left(\frac{\partial y^\mu}{\partial x^\nu}\right). \tag{2.7}$$

Note that the value of J may depend on the position of the point in $\mathcal{M}$. As long as $J \neq 0$, the transformation from x^μ to y^μ is well-defined. In particular, one can invert the transformation. From Eq. (2.6), we can apply theorems of linear algebra to find

$$\begin{aligned} \det\left[\frac{\partial x^\mu}{\partial y^\alpha}\frac{\partial y^\alpha}{\partial x^\nu}\right] &= \det\delta^\mu_\nu \\ \det\left(\frac{\partial x^\mu}{\partial y^\alpha}\right)\det\left(\frac{\partial y^\alpha}{\partial x^\nu}\right) &= 1 \\ \det\left(\frac{\partial x^\mu}{\partial y^\alpha}\right) &= \frac{1}{\det\left(\frac{\partial y^\alpha}{\partial x^\nu}\right)}. \end{aligned}$$

In other words, the Jacobian of the inverse transformation is the reciprocal of the Jacobian of the forward transformation.

Example 2.1.2. Let us take the case $\mathcal{M} = \mathbb{R}^2$, the two-dimensional space. Two possible coordinate charts are Cartesian $y^\mu = (x, y)$ in the domain $x, y \in (-\infty, \infty)$, and polar $x^\mu = (r, \phi)$ with domain $r \in [0, \infty)$, $\phi \in [0, 2\pi)$. The relation between the two systems is

$$x = r\cos\phi, \quad y = r\sin\phi.$$

The derivatives are

$$\frac{\partial x}{\partial r} = \cos\phi, \quad \frac{\partial x}{\partial \phi} = r\sin\phi, \quad \frac{\partial y}{\partial r} = \sin\phi, \quad \frac{\partial y}{\partial \phi} = r\cos\phi.$$

Therefore the transformation matrix is

$$\begin{pmatrix} \cos\phi & -r\sin\phi \\ \sin\phi & r\cos\phi \end{pmatrix}$$

The Jacobian is

$$J = \det\begin{pmatrix} \cos\phi & -r\sin\phi \\ \sin\phi & r\cos\phi \end{pmatrix} = r\cos^2\phi + r\sin^2\phi = r.$$

We see that as long as $\neq 0$, J is non-zero and the transformation is well defined. Now, the origin $(x, y) = (0, 0)$ correspond to $r = 0$. At the origin, ϕ is undefined (any value of ϕ, with $r = 0$ represents the origin. So ϕ is not unique at the origin.) Therefore $J = 0$ at $r = 0$ and the polar coordinates is undefined there. This is one of the early instances where we encounter the concept of a *coordinate singularity*.

2.2 Distances on manifolds: The metric

By this stage, every student is aware of the *Phythogrean theorem*. In the language of this chapter, the Phythagorean theorem is a statement about the distance between two points on $\mathcal{M} = \mathbb{R}^2$. Suppose we have two points A and B, given in Cartesian coordinates as

$$A(x_1, y_1), \quad B(x_2, y_2).$$

By applying the Phythogrean theorem, the distance from A to B is

$$\Delta s = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2},$$

or,

$$\Delta s^2 = \Delta x^2 + \Delta y^2,$$

where $\Delta x = x_1 - x_2$ and $\Delta y = y_1 - y_2$.

In differential geometry, we study the local behaviour around a given point on the manifold. Therefore we are interested in *infinitesimal distances*. In the case $\mathcal{M} = \mathbb{R}^2$, we consider $\Delta x \simeq \mathrm{d}x$ and $\Delta y \simeq \mathrm{d}y$ to be infinitesimally small. The Phythogrean theorem gives

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2. \tag{2.8}$$

The generalisation of Eq. (2.8) to higher dimensions is straightforward. In $\mathbb{R}^n$,

$$\mathrm{d}s^2 = \left(\mathrm{d}x^1\right)^2 + \ldots + (\mathrm{d}x^n)^2 \tag{2.9}$$

Now, we wish to generalise Eq. (2.9) to general curved manifolds in arbitrary dimensions n. (See Fig. 2.5.)

Figure 2.5

In Chapter 1, we have already encountered the interval for Minkowski spacetime,

$$\mathrm{d}s^2 = \eta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = -\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2. \tag{2.10}$$

This almost looks similar to a distance in $\mathbb{R}^4$, except for a minus sign in the coefficient of $\mathrm{d}t^2$. This is an indication that time is to be distinguished from the spatial directions. Hence we denote Minkowski spacetime by $\mathbb{R}^{3,1}$.

Skipping many details typically covered in a mathematical differential geometry course, we proceed to state that distances on a general manifold $\mathcal{M}$ is written in the form

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu, \tag{2.11}$$

where $\mathrm{d}s^2$ is called the *metric* or *line element* of the spacetime. In a curved 4-dimensional spacetime, $g_{\mu\nu}$ can be seen as components of a 4×4 matrix

$$g_{\mu\nu} = \begin{pmatrix} g_{00} & g_{01} & g_{02} & g_{03} \\ g_{10} & g_{11} & g_{12} & g_{13} \\ g_{20} & g_{21} & g_{22} & g_{23} \\ g_{30} & g_{31} & g_{32} & g_{33} \end{pmatrix}. \tag{2.12}$$

The object $g_{\mu\nu}$ is called the *metric tensor*.

Example 2.2.1. Examples of metric tensors.

(a) We already know the Minkowski metric, whose line element is

$$\mathrm{d}s^2 = \eta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = -\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2.$$

Its metric tensor is

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

(b) The line element of a sphere of radius a is

$$\mathrm{d}s^2 = a^2 \left(\mathrm{d}\theta^2 + \sin^2\theta \, \mathrm{d}\phi^2\right).$$

Its metric tensor is

$$g_{\mu\nu} = \begin{pmatrix} a^2 & 0 \\ 0 & a^2 \sin^2\theta \end{pmatrix}$$

Properties of the metric tensor:

- **Symmetry.** Generally the metric tensor is symmetric, $g_{\mu\nu} = g_{\nu\mu}$. As a specific example in components, $g_{12} = g_{21}$. For an $n \times n$ symmetric metric tensor, this means it has $\frac{1}{2}n(n+1)$ independent components. In particular, the metric tensor of a $n = 4$-dimensional spacetime has 10 independent components.
- **Coordinate dependence.** As can be seen from the line element $\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu$, the form of the metric tensor depends on the choice of coordinate system. Suppose we have another coordinate system y^μ. Recall that generally y^μ can be expressed

as functions of the previous coordinates x^μ. Therefore we have

$$\mathrm{d}y^\mu = \frac{\partial y^\mu}{\partial x^\nu}\mathrm{d}x^\nu. \tag{2.13}$$

Suppose the line element is given in the y-coordinate as $\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}y^\mu \mathrm{d}y^\nu$, we then have

$$\begin{aligned}\mathrm{d}s^2 &= g_{\mu\nu}\left(\frac{\partial y^\mu}{\partial x^\alpha}\mathrm{d}x^\alpha\right)\left(\frac{\partial y^\nu}{\partial x^\beta}\mathrm{d}x^\beta\right) = g_{\mu\nu}\frac{\partial y^\mu}{\partial x^\alpha}\frac{\partial y^\nu}{\partial x^\beta}\,\mathrm{d}x^\alpha \mathrm{d}x^\beta \\ &= \gamma_{\alpha\beta}\mathrm{d}x^\alpha \mathrm{d}x^\beta. \end{aligned}\tag{2.14}$$

Therefore, the equation $\gamma_{\alpha\beta} = g_{\mu\nu}\frac{\partial y^\mu}{\partial x^\alpha}\frac{\partial y^\nu}{\partial x^\beta}$ tells us how the metric transforms under coordinate transformations.

Example 2.2.2. Consider again $\mathcal{M} = \mathbb{R}^2$. In Cartesian coordinates, its line element is

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2, \tag{2.15}$$

with the domain $x, y \in (-\infty, \infty)$. Transforming to a polar coordinates,

$$x = r\cos\phi, \quad y = r\sin\phi,$$

where the domain of the new coordinates are $r \in [0, \infty)$ and $\phi \in [0, 2\pi)$. Its differentials are

$$\begin{aligned}\mathrm{d}x &= \frac{\partial x}{\partial r}\mathrm{d}r + \frac{\partial x}{\partial \phi}\mathrm{d}\phi = \cos\phi\,\mathrm{d}r - r\sin\phi\,\mathrm{d}\phi,\\ \mathrm{d}y &= \frac{\partial y}{\partial r}\mathrm{d}r + \frac{\partial y}{\partial \phi}\mathrm{d}\phi = \sin\phi\,\mathrm{d}r + r\cos\phi\mathrm{d}\phi.\end{aligned}$$

Substituting this into Eq. (2.15),

$$\begin{aligned}\mathrm{d}s^2 &= (\cos\phi\,\mathrm{d}r - r\sin\phi\,\mathrm{d}\phi)^2 + (\sin\phi\,\mathrm{d}r + r\cos\phi\mathrm{d}\phi)^2\\ &= \cos^2\phi\mathrm{d}r^2 - \cancel{2r\sin\phi\cos\phi\mathrm{d}r\mathrm{d}\phi} + r^2\sin^2\phi\,\mathrm{d}\phi^2\\ &\quad + \sin^2\phi\mathrm{d}r^2 + \cancel{2r\cos\phi\sin\phi\mathrm{d}r\mathrm{d}\phi} + r^2\cos^2\phi\mathrm{d}\phi^2\\ &= \mathrm{d}r^2 + r^2\mathrm{d}\phi^2.\end{aligned}$$

Therefore the line element of $\mathbb{R}^2$ in polar coordinates is $\mathrm{d}s^2 = \mathrm{d}r^2 + r^2\mathrm{d}\phi^2$. In terms of the metric tensor, the (same) metric tensor in the two coordinate systems are

$$\textbf{Cartesian:} \quad g_{\mu\nu} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$
$$\textbf{Polar:} \quad \gamma_{\mu\nu} = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}.$$

Example 2.2.3. The metric of a two-sphere. *Consider a sphere of radius a in $\mathbb{R}^3$. The sphere itself is therefore a two-dimensional surface, which we denote by S^2. Determine the metric of S^2 which describes a sphere of radius a centred at the origin.*

Solution. First, we start with the metric on $\mathbb{R}^3$ in standard Cartesian coordinates

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2,$$

where $x, y, z \in (-\infty, \infty)$. Transforming to spherical coordinates,

$$x = r \sin\theta \cos\phi, \quad y = r \sin\theta \sin\phi, \quad z = r\cos\theta,$$

with the domains $r \in [0, \infty)$, $\theta \in [0, \pi]$, and $\phi \in [0, 2\pi)$, the metric becomes

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

This is still the metric for the three-dimensional Euclidean space $\mathbb{R}^3$. The sphere of radius a centred at the origin has points at fixed $r = a =$ constant. Substituting this conidition into the metric,

$$\begin{aligned} \mathrm{d}s^2 &= 0 + a^2 \mathrm{d}\theta^2 + a^2 \sin^2\theta \, \mathrm{d}\phi^2 \\ \mathrm{d}s^2 &= a^2 \left(\mathrm{d}\theta^2 + \sin^2\theta \, \mathrm{d}\phi^2 \right). \end{aligned}$$

2.3 Defining vectors on manifolds and the metric

By now, we should be clear about the distinction between a (spacetime) manifold $\mathcal{M}$ and its coordinates $\mathbb{R}^n$. Generally speaking the manifold $\mathcal{M}$ is where 'physics takes place', and the coordinates $\mathbb{R}^n$ is where 'we do calculations'.

Now we have a problem: How do we define a vector on $\mathcal{M}$? Roughly speaking, the defining feature of a *vector* is typically an object having a *direction* and a *magnitude*. Recall how we originally deal with vectors in Newtonian mechanics. For example, the non-relativistic velocity of a particle is written as

$$\vec{v} = v_x\,\hat{\imath} + v_y\,\hat{\jmath} + v_z\,\hat{k}.$$

It is a vector with three components (v_x, v_y, v_z). Therefore the velocity is an element of $\mathbb{R}^3$.

In the context of differential geometry, the *position is not a vector!* A point on a manifold is simply given as coordinates $(x^1, \ldots, x^n)$; it doesn't really make sense to talk about the 'magnitude/direction' of these coordinates. They are just a set of n numbers! On the other hand, the *velocity* of a particle does make sense as having a magnitude and direction. Same for acceleration.

Defining a tangent vector. Suppose we have a manifold $\mathcal{M}$, and a particle moves along a curve inside $\mathcal{M}$. (This is 'where physics takes place', so there's really a particle travelling inside $\mathcal{M}$.) If $\mathcal{M}$ is a spacetime (for example Minkowski), this curve might be the worldline of a time-like particle. If $\mathcal{M}$ is a surface of a sphere, then the curve describes the trajectory of a particle moving on the sphere's surface.

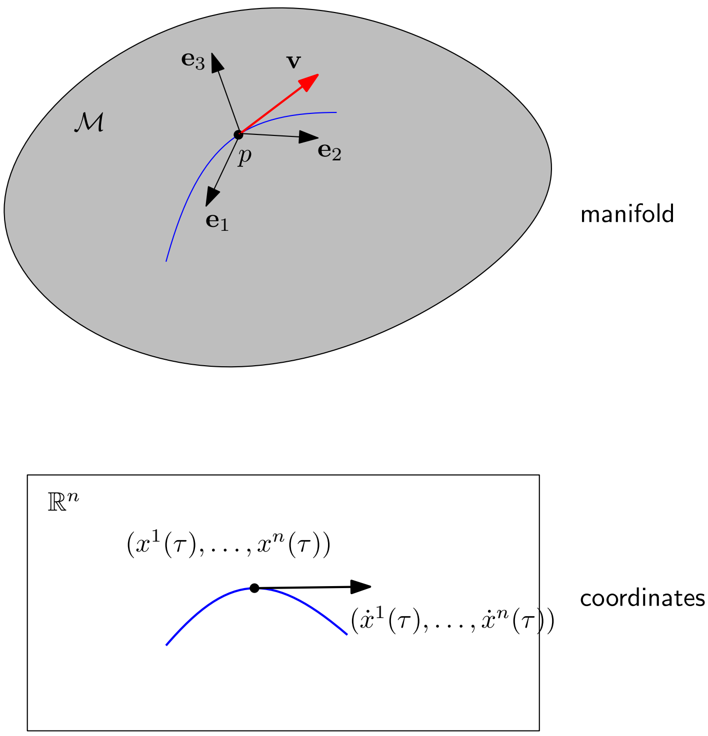


Figure 2.6: Defining a tangent vector in a manifold $\mathcal{M}$.

Consider a curve C in a manifold $\mathcal{M}$, as shown in Fig. 2.6. Suppose when the parameter takes the value τ, the particle is at point p. The coordinates of point p is $(x^1(\tau), \ldots, x^n(\tau))$.

In $\mathbb{R}^n$, we can now easily define the tangent vector as

$$\mathbf{v} = v^\mu \mathbf{e}_\mu, \tag{2.16}$$

whose components are $v^\mu = \dot{x}^\mu = \frac{\mathrm{d}}{\mathrm{d}t} x^\mu(\tau)$ and $\mathbf{e}_\mu$ are the basis vectors constructed at the point p. (See $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ at Fig. 2.6.) Clearly the expression for a vector depends on the coordinate system used. Sometimes we tend to refer v^μ as the vector, even though they are just the *components* of the full vector $\mathbf{v} = v^\mu \mathbf{e}_\mu$. Skipping some mathematical subtleties, *we will identify* $\mathbf{v}$ *as the tangent vector to the curve at point* p.

Coordinate transformation of vectors. Suppose we have a tangent vector $\mathbf{v} = v^\mu \mathbf{e}_\mu$. Its components are $v^\mu = \dot{x}^\mu$, which clearly depends on the coordinate system x^μ. If we change to new coordinate y^μ, what are the components of $\mathbf{v}$?

Assuming the Jacobian is non-zero so we can express each x^μ as functions of the new coordinates y^μ, we apply the chain rule:

$$\begin{aligned}
\dot{x}^\mu &= \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} = \frac{\partial x^\mu}{\partial y^\nu} \dot{y}^\nu \\
\frac{\partial y^\alpha}{\partial x^\mu} \dot{x}^\mu &= \underbrace{\frac{\partial y^\alpha}{\partial x^\mu} \frac{\partial x^\mu}{\partial y^\nu}}_{\delta^\alpha_\nu} \dot{y}^\nu = \delta^\alpha_\nu \dot{y}^\nu \\
\dot{y}^\alpha &= \frac{\partial y^\alpha}{\partial x^\mu} \dot{x}^\mu
\end{aligned}$$

Here, $v^\mu = \dot{x}^\mu$ are the components of the vector in x-coordinates. So we would expect $u^\mu = \dot{y}^\mu$ to be the corresponding components in y-coordinates. Therefore under the coordinate change $x^\mu \to y^\mu$, the vector components transform according to

$$v^\mu \to u^\mu = \frac{\partial y^\mu}{\partial x^\nu} v^\nu. \tag{2.17}$$

Eq. (2.17) gives the transformation law for the vector components. What about the basis vectors? Suppose that in y-coordinates, we denote the basis as $\mathbf{f}_\mu$.

Generally a vector components and its basis depend on the choice of coordinates. But the whole vector $\mathbf{v}$ itself should not depend on coordinates. Therefore we expect

$$\mathbf{v} = v^\mu \mathbf{e}_\mu = u^\nu \mathbf{f}_\nu.$$

We know from Eq. (2.17) how the components transform, therefore

$$
\begin{aligned}
v^\mu \mathbf{e}_\mu &= \frac{\partial y^\nu}{\partial x^\alpha} v^\alpha \mathbf{f}_\nu \\
v^\mu \mathbf{e}_\mu &= v^\alpha \left(\frac{\partial y^\nu}{\partial x^\alpha} \mathbf{f}_\nu \right).
\end{aligned}
$$

Comparing both sides of the equation, we conclude that

$$
\mathbf{e}_\mu = \frac{\partial y^\nu}{\partial x^\mu} \mathbf{f}_\nu \quad \leftrightarrow \quad \mathbf{f}_\mu = \frac{\partial x^\nu}{\partial y^\mu} \mathbf{e}_\nu. \tag{2.18}
$$

This establishes the transformation law for the bases.

Example 2.3.1. Velocities in $\mathbb{R}^2$. On the familiar two-dimensional plane, we can express the (non-relativistic) velocity of a particle in Cartesian coordinates as

$$
\vec{v} = v^x \mathbf{e}_x + v^y \mathbf{e}_y = \dot{x}\,\hat{\imath} + \dot{y}\,\hat{\jmath}. \tag{2.19}
$$

We can also transform to polar coordinate as

$$
\{x = r\cos\phi, \quad y = r\sin\phi\} \quad \leftrightarrow \quad \left\{ r = \sqrt{x^2+y^2}, \quad \tan\phi = \frac{y}{x} \right\}.
$$

(The inverse transformation is also written for later reference.) In our tensor, notation, write the old and new coordinates as

$$
x^\mu = (x, y), \quad y^\mu = (r, \phi)
$$

In the new polar coordinates, the same velocity takes the form

$$
\vec{v} = u^r \mathbf{f}_r + u^\phi \mathbf{f}_\phi = \dot{r}\,\hat{r} + \dot{\phi}\,\hat{\phi}. \tag{2.20}
$$

It is important to emphasise that the **actual** velocity has not changed; we are simply writing the same quantity $\vec{v}$ in different coordinate system. So Eq. (2.19) must be the same as (2.20).

To see that this is the case, note that by the chain rule,

$$
\dot{x} = \frac{\partial x}{\partial r}\dot{r} + \frac{\partial x}{\partial \phi}\dot{\phi}, \quad \dot{y} = \frac{\partial y}{\partial r}\dot{r} + \frac{\partial y}{\partial \phi}\dot{\phi}
$$

These are the components of the transformation equation $v^\mu = \frac{\partial x^\mu}{\partial y^\nu} u^\nu$. The right hand side of Eq. (2.19) is

$$\vec{v} = \left(\frac{\partial x}{\partial r}\dot{r} + \frac{\partial x}{\partial \phi}\dot{\phi}\right)\hat{\imath} + \left(\frac{\partial y}{\partial r}\dot{r} + \frac{\partial y}{\partial \phi}\dot{\phi}\right)\hat{\jmath} = \dot{r}\underbrace{\left(\frac{\partial x}{\partial r}\hat{\imath} + \frac{\partial y}{\partial r}\hat{\jmath}\right)}_{\hat{r}=\mathbf{f}_r} + \dot{\phi}\underbrace{\left(\frac{\partial x}{\partial \phi}\hat{\imath} + \frac{\partial y}{\partial \phi}\hat{\jmath}\right)}_{\hat{\phi}=\mathbf{f}_\phi}$$

Therefore we have shown that

$$\mathbf{f}_r = \frac{\partial x}{\partial r}\mathbf{e}_x + \frac{\partial y}{\partial r}\mathbf{e}_y, \quad \mathbf{f}_\phi = \frac{\partial x}{\partial \phi}\mathbf{e}_x + \frac{\partial y}{\partial \phi}\mathbf{e}_y,$$

which are the components of the equation $\mathbf{f}_\mu = \frac{\partial x^\nu}{\partial y^\mu}\mathbf{e}_\nu$.
To complete this problem, we work out the partial derivatives explicitly:

$$\begin{aligned} \frac{\partial x}{\partial r} &= \cos\phi, \quad \frac{\partial y}{\partial r} = \sin\phi, \\ \frac{\partial x}{\partial \phi} &= -r\sin\phi, \quad \frac{\partial y}{\partial \phi} = r\cos\phi. \end{aligned}$$

So the basis vectors transform as

$$\mathbf{f}_r = \cos\phi\,\mathbf{e}_x + \sin\phi\,\mathbf{e}_y, \quad \mathbf{f}_\phi = -r\sin\phi\,\mathbf{e}_x + r\cos\phi\,\mathbf{e}_y.$$

Inner product of vectors. Our next task is to define the inner product of tangent vectors in manifolds. In other words, we want to generalise the 'dot product' we know from $\mathbb{R}^3$.

Suppose we have an n-dimensional manifold $\mathcal{M}$ with line element

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu. \tag{2.21}$$

At some point p on the manifold, we have two vectors,

$$\mathbf{u} = u^\mu\mathbf{e}_\mu, \quad \mathbf{v} = v^\mu\mathbf{e}_\mu. \tag{2.22}$$

We define the *inner product* between $\mathbf{u}$ and $\mathbf{v}$ as

$$\langle\mathbf{u},\mathbf{v}\rangle = g_{\mu\nu}u^\mu v^\nu. \tag{2.23}$$

Substituting Eq. (2.22) into the left-hand-side of the above equation, we find

$$\langle u^\mu\mathbf{e}_\mu, v^\nu\mathbf{e}_\nu\rangle = g_{\mu\nu}u^\mu v^\nu.$$

With the inner product, we can now define the *norm* of a vector as

$$\|\mathbf{v}\| = \sqrt{|\langle \mathbf{v}, \mathbf{v} \rangle|} = \sqrt{|g_{\mu\nu} v^\mu v^\nu|}. \tag{2.24}$$

Example 2.3.2. Speed in $\mathbb{R}^2$. In the previous examples, we have computed the velocity in Cartesian and polar coordinate systems. In the current language of manifolds, *speed* is defined as the norm of the velocity as

$$v = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = \sqrt{|g_{\mu\nu} v^\mu v^\nu|}$$

In Cartesian coordinates, the line element and velocity is

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2, \quad \mathbf{v} = \dot{x}\mathbf{e}_x + \dot{y}\mathbf{e}_y.$$

Therefore the speed is

$$v = \sqrt{\dot{x}^2 + \dot{y}^2}.$$

On the other hand, in polar coordinates, the line element and velocity is

$$\mathrm{d}s^2 = \mathrm{d}r^2 + r^2 \mathrm{d}\phi^2, \quad \mathbf{v} = \dot{r}\mathbf{e}_r + \dot{\phi}\mathbf{e}_\phi.$$

Therefore the speed is

$$v = \sqrt{\dot{r}^2 + r^2\dot{\phi}^2}.$$

Example 2.3.3. In Minkowski spacetime, the line element is

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2.$$

In Chapter 1, we considered the four-velocity of a particle to have components

$$\dot{x}^\mu = \frac{\mathrm{d}}{\mathrm{d}\tau}\left(t(\tau), x(\tau), y(\tau), z(\tau)\right) = \left(\dot{t}, \dot{x}, \dot{y}, \dot{z}\right).$$

In the notation of this chapter, the four-velocity is $\mathbf{u} = u^\mu \mathbf{e}_\mu = \dot{t}\mathbf{e}_t + \dot{x}\mathbf{e}_x + \dot{y}\mathbf{e}_y + \dot{z}\mathbf{e}_z$. The inner product of the four velocity with itself is

$$\langle \mathbf{u}, \mathbf{u} \rangle = \eta_{\mu\nu}\dot{x}^\mu \dot{x}^\nu = -\dot{t}^2 + \dot{x}^2 + \dot{y}^2 + \dot{z}^2.$$

In Chapter 1, we have shown that $\eta_{\mu\nu}\dot{x}^\mu\dot{x}^\nu = -1$ (in units where $c = 1$ now.) Therefore the four-velocity of a time-like particle in Minkowski spacetime has negative inner product,

$$\langle \mathbf{u}, \mathbf{u} \rangle = -1.$$

Riemannian vs pseudo-Riemannian metric. From the previous example, we note that the inner product $\langle \mathbf{u}, \mathbf{u} \rangle = g_{\mu\nu}u^\mu u^\nu$ is negative, which is an unusual property of inner products. (In $\mathbb{R}^3$, $\vec{u} \cdot \vec{v} = \delta_{ij}u^i v^j$ is always positive, or at worst zero.) This is because the metric $g_{\mu\nu}$ might have some negative component in it. From this we can classify two types of manifolds:

- *Riemannian manifolds* are those where $g_{\mu\nu}u^\mu u^\nu \geq 0$. The familiar $\mathbb{R}^3$ and $\mathbb{R}^2$ are examples of Riemannian manifolds, as all dot products are non-negative. Sometimes, these are called *Euclidean*, because standard Euclidean geometry falls under this category.
- *Pseudo-Riemannian manifolds* are those where $g_{\mu\nu}u^\mu v^\nu$ might be positive, zero, or negative. The Minkowski spacetime is a pseudo-Riemannian manifold.

Contravariant and covariant vectors. Consider again the expression

$$\langle \mathbf{u}, \mathbf{v} \rangle = g_{\mu\nu}u^\mu v^\nu = (g_{\mu\nu}u^\mu)\, v^\nu = (\text{something})_\nu\, v^\nu,$$

where this $(\text{something})_\nu$ is a quantity with a lower index. According to the Einstein summation convention, quantities with a lower index is contracted with another quantity with an upper index to give a scalar. Hence the inner product is indeed a scalar.

Generally we see that we have two categories of vectors. Those with upper indices and those with lower indices. We shall give them the following names:

$$\textbf{Contravariant vectors:} \quad v^\mu,$$
$$\textbf{Covariant vectors:} \quad v_\mu.$$

The conversion from a contravariant vector to a covariant one is done by the metric tensor:

$$v_\mu = g_{\mu\nu}v^\nu. \tag{2.25}$$

Raising and lowering indices. Given a contravariant vector v^μ, we can convert to covariant form by $v_\mu = g_{\mu\nu}v^\nu$. Can we do this in the opposite direction? We start with

$$v_\mu = g_{\mu\nu}v^\nu. \tag{2.26}$$

We can view the above equation as a multiplication of an $n \times n$ matrix $g_{\mu\nu}$ by a vector v^μ, which results in a vector v_μ. So, if we multiply both sides of the equation by the inverse matrix $g^{\alpha\mu}$. We write the inverse as

$$\boxed{g^{\mu\nu} = \text{inverse of } g_{\mu\nu} \quad \leftrightarrow \quad g^{\mu\lambda}g_{\lambda\nu} = \delta^\mu_\nu = \begin{cases} 1 & \text{if } \mu = \nu, \\ 0 & \text{if } \mu \neq \nu, \end{cases}} \tag{2.27}$$

where δ^μ_ν is the Kroenecker delta. If we multiply both sides of Eq. (2.26) with the inverse metric, we get

$$\begin{aligned} g^{\alpha\mu}v_\mu &= g^{\alpha\mu}g_{\mu\nu}v^\nu \\ g^{\alpha\mu}v_\mu &= \delta^\alpha_\nu v^\nu \\ g^{\alpha\mu}v_\mu &= v^\alpha. \end{aligned}$$

Therefore, given a covariant vector v_μ, we convert to contravariant form by contraction with the inverse metric.

$$v^\mu = g^{\mu\nu}v_\nu \tag{2.28}$$

In summary, the metric tensor $g_{\mu\nu}$ can be used to 'lower indices' by $v_\mu = g_{\mu\nu}v^\nu$, and the inverse metric $g^{\mu\nu}$ is used to 'raise indices' by $v^\mu = g^{\mu\nu}v_\nu$. This can be applied to any (r, s)-tensor.

Example 2.3.4. *Consider $\mathbb{R}^3$ in spherical coordinates,*

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

Find the unit radial vector in covariant and contravariant form.

Solution. Its metric tensor is

$$g_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2\sin^2\theta \end{pmatrix}.$$

We start with the contravariant form,

$$n^\mu = (n^r, n^\theta, n^\phi). \tag{2.29}$$

Since n^μ is supposed to be a radial vector, the angular components should be zero. Therefore

$$n^\mu = (n^r, 0, 0).$$

This vector must be a unit vector, therefore the norm must be equal to 1:

$$g_{\mu\nu} n^\mu n^\nu = g_{rr} n^r n^r + 0 + \ldots = 1,$$

all the other terms are zero because $n^\theta = n^\phi = 0$. We therefore have

$$\begin{aligned} g_{rr}(n^r)^2 &= 1 \\ (n^r)^2 &= 1 \quad \rightarrow \quad n^r = 1. \end{aligned}$$

Therefore $n^\mu = (1, 0, 0)$.
To get the covariant form,

$$n_\mu = g_{\mu\nu} n^\nu = (n_r, n_\theta, n_\phi).$$

Since the only non-zero component of n^ν is when $\nu = r$, we have

$$n_\mu = g_{\mu r} n^r = g_{\mu r}.$$

Furthermore, the only non-zero component of $g_{\mu r}$ is when $\mu = r$. Therefore

$$n_r = g_{rr} n^r = 1, \quad n_\theta = 0, \quad n_\phi = 0.$$

The covariant form of the unit radial vector is

$$n_\mu = (n_r, n_\theta, n_\phi) = (1, 0, 0).$$

Example 2.3.5. *In Alice's coordinate system, the Minkowski spacetime has the metric*

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2.$$

A particle is observed to be moving at constant velocity in the x-direction with respect to Alice. Determine the particle's four-velocity in Alice's coordinate system in contravariant and covariant form.

Solution. Let τ be the proper time of the particle. Therefore its worldline is given by coordinates

$$(t(\tau), x(\tau), 0, 0) .$$

The (contravariant) four-velocity is

$$u^\mu = \frac{\mathrm{d}}{\mathrm{d}\tau}(t(\tau), x(\tau), 0, 0) = \left(\dot{t}, \dot{x}, 0, 0\right)$$

The four-velocity must obey

$$-1 = \eta_{\mu\nu}\dot{x}^\mu\dot{x}^\nu = -\dot{t}^2 + \dot{x}^2$$

Therefore we find that

$$\begin{aligned} \dot{t}^2 &= 1 + \dot{x}^2 \\ \dot{t} &= \sqrt{1 + \dot{x}^2}. \end{aligned}$$

Therefore the contravariant four-velocity is

$$u^\mu = \left(u^t, u^x, u^y, u^z\right) = \left(\sqrt{1 + \dot{x}^2}, \dot{x}, 0, 0\right) .$$

To get the covariant form,

$$u_\mu = \eta_{\mu\nu}\dot{x}^\nu = \eta_{tt}(u^t)^2 + \eta_{xx}(u^x)^2 + 0 + \ldots ,$$

where all the other terms are zero. Therefore

$$u_\mu = \left(-\sqrt{1 + \dot{x}^2}, \dot{x}^2, 0, 0\right) .$$

2.4 Tensors

Simply speaking, a tensor is a generalisation of vectors. Actually, let us first recall that a vector is a generalisation of a scalar (number). In Newtonian mechanics in $\mathbb{R}^3$, we take three scalars to form a vector. So in previous sections we now learned how to form a vector in n-dimensional manifolds (spacetimes) by taking n numbers to form a contravariant vector $v^\mu = (v^1, \ldots, v^n)$ or its contravariant vector $v_\mu = (v_1, \ldots, v_n)$.

In order to discuss tensors, let us reinterpret the concept of contravariant and covariant vectors. Start with a contravariant vector V^μ. As discussed previously, we can take the inner product with a covariant vector ω_μ, so that $V^\mu \omega_\mu$ is a (scalar) number. That means, V^μ has *eaten a covariant vector ω_μ to produce a number.*

Equivalently, we can also think of the covariant ω_μ as *eating a contravariant V^μ to produce a number.*

Basic idea of tensors

With this mode of thinking, we see them as mathematical objects that eats contravariant or covariant vectors to produce a number. Then we define **a (r,s)-tensor as an object that *eats* r *covariant vectors and* s *contravariant vectors to produce a number.*** They are written in the form

$$T^{\mu_1 \cdots \mu_r}{}_{\nu_1 \cdots \nu_r}. \tag{2.30}$$

For example, a $(1,1)$-tensor shall be written in the form

$$T^\mu{}_\nu.$$

It eats 1 covariant vector ω_μ and 1 contravariant vector V^ν, so that

$$T^\mu{}_\nu \omega_\mu V^\nu = \text{ a number.}$$

The metric tensor $g_{\mu\nu}$ is a $(0,2)$-tensor. Since it eats two contravariant vectors giving a number (the inner product)

$$\langle \mathbf{u}, \mathbf{v} \rangle = g_{\mu\nu} u^\mu v^\nu = \text{ a number.}$$

A $(1,0)$ tensor is something that only has one upper index – a contravariant vector,

$$v^\mu \quad \text{a (1,0)-tensor.}$$

Similarly, a $(0,1)$ tensor is something that only has one lower index – a covariant tensor,

$$v_\mu \quad \text{a (0,1)-tensor.}$$

A $(0,0)$ tensor has no indices. It's just a single number, or a scalar:

$$\varphi: \quad \text{a (0,0)-tensor}$$

Raising and lowering indices. We can convert a (r,s) tensor into a $(r-1,s+1)$ tensor, or $(r+1,s-1)$ tensor. This is done using the metric and its inverse. For example, given a $(1,1)$-tensor $F^{\mu}{}_{\nu}$, we can convert it into a $(0,2)$-tensor by

$$F_{\mu\nu} = g_{\mu\lambda}F^{\lambda}{}_{\nu}.$$

As another example, suppose we have a $(0,2)$-tensor $A_{\mu\nu}$. We can raise the first index by

$$A^{\mu}{}_{\nu} = g^{\mu\alpha}A_{\alpha\nu}.$$

Instead of the first, we can instead raise the second index as

$$A_{\mu}{}^{\nu} = g^{\nu\alpha}A_{\mu\alpha}.$$

Example 2.4.1. Consider $\mathbb{R}^2$ in polar coordinates, where the line element is

$$\mathrm{d}s^2 = \mathrm{d}r^2 + r^2\mathrm{d}\phi^2.$$

Suppose that we have a $(0,2)$-tensor given by

$$A_{ab} = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}.$$

In other words, $A_{11} = \alpha$, $A_{12} = A_{21} = 0$, and $A_{22} = \beta$. Compute the components of $A^{a}{}_{b} = g^{ac}A_{cb}$.

Solution. *For this we require the inverse metric*

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

The components of $A^a{}_b$ are computed directly:

$$\begin{aligned}
A^1{}_1 &= g^{1c}A_{c1} = g^{11}A_{11} + g^{12}A_{21} = (1)\alpha + 0 = \alpha, \\
A^1{}_2 &= g^{1c}A_{c2} = g^{11}A_{12} + g^{12}A_{22} = 0, \\
A^2{}_1 &= g^{2c}A_{c1} = g^{21}A_{11} + g^{22}A_{21} = 0, \\
A^2{}_2 &= g^{2c}A_{c2} = g^{21}A_{12} + g^{22}A_{22} = 0 + \frac{1}{r^2}\beta = \beta/r^2.
\end{aligned}$$

If we wish to write in matrix form, the result is

$$A^a{}_b = \begin{pmatrix} \alpha & 0 \\ 0 & \frac{\beta}{r^2} \end{pmatrix}.$$

Clearly, in general $A^a{}_b \neq A_{ab}$ hence the locations of the indices (up or down) give different components.

Contraction of tensors. Given two tensors, we can *contract* one lower index of a tensor with an upper index with the other tensor. For example, we are already familiar with the contraction of a contravariant and a covariant vector:

$$\begin{gathered}
u^\mu \text{ a } (1,0)\text{-tensor}, \quad v_\mu \text{ a } (0,1)\text{-tensor}, \\
\downarrow \\
u^\mu v_\mu = \text{a number, or } (0,0)\text{-tensor}.
\end{gathered}$$

We can do this with higher tensors:

$$\begin{gathered}
u^\mu \text{ a } (1,0)\text{-tensor}, \quad B_{\mu\nu} \text{ a } (0,2)\text{-tensor} \\
\downarrow \\
u^\mu B_{\mu\nu} = (\text{something})_\nu \quad \text{a } (0,1)\text{-tensor}
\end{gathered}$$

We can also have multiple contractions:

$$\begin{gathered}
A_{\mu\nu} \text{ a } (0,2)\text{-tensor}, \quad B^{\mu\nu} \text{ a } (2,0)\text{-tensor} \\
\downarrow \\
A_{\mu\nu}B^{\mu\nu} = \text{a number, or } (0,0)\text{-tensor}.
\end{gathered}$$

Even if we have two tensor of, say, all lower indices, we can use the metric tensor to raise one index and then do a contraction. We have done this implicitly when we defined the inner product of contravariant tensors:

$$u^\mu, \quad v^\mu : \quad \rightarrow \quad g_{\mu\nu} u^\mu v^\nu = u_\mu v^\mu.$$

This can be extended to higher tensors:

$$A_{\mu\nu}, \quad B_{\mu\nu} : \quad \rightarrow \quad A_{\mu\lambda} B^\lambda{}_\nu = g^{\lambda\sigma} A_{\mu\lambda} B_{\sigma\nu}.$$

Tensors after coordinate transformation.

Suppose in a n-dimensional manifold $\mathcal{M}$, we first have a coordinate system $x^\mu = (x^1, \ldots, x^n)$. Then suppose we have a second coordinate system $y^\mu = (y^1, \ldots, y^n)$. By the nature of coordinate transformations, the coordinate x^μ can be expressed as a functions of the other coordinates y^μ and vice-versa. That is, each x^μ coordiate is an n-variable function[3]

$$x^\mu(y^1, y^2, \ldots, y^n),$$

and conversely in the other coordinate system, each y^μ is an n-variable function

$$y^\mu(x^1, x^2, \ldots, x^n).$$

As shown in previous section, the transformation matrix between the two coordinate system is given by

$$\frac{\partial y^\mu}{\partial x^\nu} \quad \leftrightarrow \quad \frac{\partial x^\mu}{\partial y^\nu},$$

and these two matrices are inverses of each other.

The central idea for tensors is the following two statements:

- ***All mathematical operations should be the same for any coordinate system.***
- ***Values of scalars (non-vectors, just a single-component objects without indices) should be independent of coordinate systems.***

By mathematical operations, we mean for example raising or lowering indices, $g^{\mu\nu} A_\nu =$

[3]A familiar example is $x = r\cos\phi$ and $y = r\sin\phi$. Each Cartesian coordinate (x, y) is a function of polar coordinates (r, ϕ).

A^μ. This should be true regardless of coordinate system. Otherwise, we need to establish a new rule for each coordinate system!

Transformation of a $(1,0)$-tensor (a contravariant vector). We have seen that the vector (components) must transform under coordinate transformations. Let us show this again. Recall that we define a contravariant vector as a tangent to some curve such that $u^\mu = \dot{x}^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau}$. By applying the chain rule,

$$u^\mu = \frac{\partial x^\mu}{\partial y^\nu}\frac{\mathrm{d}y^\nu}{\mathrm{d}\tau}.$$

But $\frac{\mathrm{d}y^\nu}{\mathrm{d}\tau} = \dot{y}^\nu = u'^\nu$ is just the vector as defined using the y^μ-coordinates. Therefore we have

$$\boxed{u^\mu = \frac{\partial x^\mu}{\partial y^\nu}u'^\nu.} \tag{2.31}$$

This is the coordinate transformation rule for a contravariant vector, or a $(1,0)$-tensor.

Transformation of the metric. The metric is a $(0,2)$-tensor which gives the interval $\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu$. Here $\mathrm{d}s^2$ is a physical (scalar) quantity that is independent of coordinate. So it should have the same formula in y^μ coordinates,

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu = g'_{\alpha\beta}\mathrm{d}y^\alpha\mathrm{d}y^\beta.$$

We apply the chain rule $\mathrm{d}y^\alpha = \frac{\partial y^\alpha}{\partial x^\mu}$ to the differentials on the right hand side,

$$\begin{aligned} g_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu &= g'_{\alpha\beta}\frac{\partial y^\alpha}{\partial x^\mu}\mathrm{d}x^\mu\,\frac{\partial y^\beta}{\partial x^\nu}\mathrm{d}x^\nu \\ g_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu &= \left(\frac{\partial y^\alpha}{\partial x^\mu}\frac{\partial y^\beta}{\partial x^\nu}g'_{\alpha\beta}\right)\mathrm{d}x^\mu\mathrm{d}x^\nu. \end{aligned}$$

By comparing both sides of the equation, we find

$$\boxed{g_{\mu\nu} = \frac{\partial y^\alpha}{\partial x^\mu}\frac{\partial y^\beta}{\partial x^\nu}g'_{\alpha\beta}.} \tag{2.32}$$

This is how a metric transforms under different coordinate systems.

Example 2.4.2. Consider the transformation of the metric in $\mathbb{R}^2$ in spherical and polar coordinates. Starting in Cartesian, the coordinates are $x^\mu = (x, y)$ and the metric is $\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2$. In other words,

$$g_{\mu\nu} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Transforming to polar coordinates,

$$x = r\cos\phi, \quad y = r\sin\phi,$$

where $y^\mu = (r, \phi)$ are the second coordinate system. The transformation matrix is

$$\frac{\partial x^\mu}{\partial y^\nu} = \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \phi} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \phi} \end{pmatrix} = \begin{pmatrix} \cos\phi & -r\sin\phi \\ \sin\phi & r\cos\phi \end{pmatrix}.$$

By Eq. (2.32), the metric in polar coordinates takes the form

$$g'_{\mu\nu} = \frac{\partial x^\sigma}{\partial y^\mu}\frac{\partial x^\lambda}{\partial y^\nu} g_{\sigma\lambda}.$$

Using this we can calculate each component. For instance,

$$\begin{aligned} g_{rr} &= \frac{\partial x^\sigma}{\partial r}\frac{\partial x^\lambda}{\partial r} g_{\sigma\lambda} = \frac{\partial x}{\partial r}\frac{\partial x}{\partial r} g_{xx} + \frac{\partial y}{\partial r}\frac{\partial y}{\partial r} g_{yy} \\ &= \cos^2\phi 1 + \sin^2\phi 1 = 1. \end{aligned}$$

Proceeding this way for all the other components, we find

$$g'_{\mu\nu} = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}.$$

This indeed reproduces the fact that the line element in polar coordinates is $\mathrm{d}s^2 = \mathrm{d}r^2 + r^2\mathrm{d}\phi^2$.

Transformation of a $(0,1)$-tensor (a covariant vector). The covariant vector is obtained by lowering the index of a contravariant vector, $u_\mu = g_{\mu\nu}u^\nu$. But we now know the transformation rule for $g_{\mu\nu}$ and u^ν from the above discussions. Therefore

$$
\begin{aligned}
u_\mu = g_{\mu\nu}u^\nu &= \frac{\partial y^\alpha}{\partial x^\mu}\frac{\partial y^\beta}{\partial x^\nu}g'_{\alpha\beta}\ \frac{\partial x^\nu}{\partial y^\lambda}u'^\lambda \\
&= \frac{\partial y^\alpha}{\partial x^\mu}\underbrace{\frac{\partial y^\beta}{\partial x^\nu}\frac{\partial x^\nu}{\partial y^\lambda}}_{=\delta^\beta_\lambda}g'_{\alpha\beta}u'^\lambda \\
&= \frac{\partial y^\alpha}{\partial x^\mu}g'_{\alpha\lambda}u'^\lambda.
\end{aligned}
$$

But $u'_\alpha = g'_{\alpha\lambda}u'^\lambda$ is just the covariant vector defined in y^μ-coordinates. Therefore we have the transformation rule

$$
\boxed{u_\mu = \frac{\partial y^\alpha}{\partial x^\mu}u'_\alpha.} \tag{2.33}
$$

Transformation of a $(1,1)$-tensor. Given a $(1,1)$-tensor in x^μ coordinates, $A^\mu{}_\nu$, we invoke the fact that contracting all its indices with vectors give a scalar. And scalars must be independent of coordinates. Therefore the same scalar is obtained in the y^μ-coordinates,

$$
A^\mu{}_\nu u_\mu w^\nu = \text{a scalar} = A'^\alpha{}_\beta u'_\alpha w'^\beta.
$$

At the left hand side, u_μ and w^ν are covariant and contravariant vectors respectively. We know that they transforma according to (2.33) and (2.31). Therefore

$$
\begin{aligned}
A^\mu{}_\nu\frac{\partial x^\alpha}{\partial y^\mu}u'_\alpha\ \frac{\partial y^\nu}{\partial x^\beta}w'^\beta &= A'^\alpha{}_\beta u'_\alpha w'^\beta \\
\left(\frac{\partial x^\alpha}{\partial y^\mu}\frac{\partial y^\nu}{\partial x^\beta}A^\mu{}_\nu\right)u'_\alpha w'^\beta &= A'^\alpha{}_\beta u'_\alpha w'^\beta.
\end{aligned}
$$

Comparing both sides of this equation, we find

$$
A'^\alpha{}_\beta = \frac{\partial x^\alpha}{\partial y^\mu}\frac{\partial y^\nu}{\partial x^\beta}A^\mu{}_\nu.
$$

Solving this equation for $A^\mu{}_\nu$,

$$
\boxed{A^\sigma{}_\lambda = \frac{\partial y^\sigma}{\partial x^\alpha}\frac{\partial x^\beta}{\partial y^\lambda}A'^\alpha{}_\beta.} \tag{2.34}
$$

Transformation of a (r,s)-tensor. The previous discussion can be repeatedly generalised to any (r, s)-tensor. So the transformation law should be

$$\boxed{A^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s} = \frac{\partial x^{\mu_1}}{\partial y^{\alpha_1}}\cdots\frac{\partial x^{\mu_r}}{\partial y^{\alpha_r}}\frac{\partial y^{\beta_1}}{\partial x^{\nu_1}}\cdots\frac{\partial y^{\beta_s}}{\partial x^{\nu_s}}A'^{\alpha_1\cdots\alpha_r}{}_{\beta_1\cdots\beta_s}.} \tag{2.35}$$

Eq. (2.35) is used in some textbooks to define the tensor, this is equivalent how we defined them earlier. The important point is that if a tensor obeys (2.35), then any equation[4] takes the same form in any choice of coordinates. The simplest example is Newton's second law $\vec{F} = m\vec{a}$ relates a force $\vec{F}$ to an acceleration $\vec{a}$, and we do not need to write a different 2nd Law if we use a different coordiante system.

Conversely, any quantity that doesn't obey the transformation (2.35) is NOT a tensor.

The partial derivative of a function is a $(0, 1)$-tensor. First, consider a function f, which is a scalar and should be independent of any coordinate choice,

$$f(x^1, \ldots, x^n) = f(y^1, \ldots, y^n)$$

In x^μ-coordinates, we write the partial derivative as $\partial_\mu f = \frac{\partial f}{\partial x^\mu}$. By the chain rule,

$$\partial_\mu f = \frac{\partial y^\lambda}{\partial x^\mu}\frac{\partial f}{\partial y^\lambda}.$$

But $\frac{\partial f}{\partial y^\lambda} = \partial'_\lambda f$ is just the partial derivative in the y^μ-coordinates. Therefore

$$\partial_\mu f = \frac{\partial y^\lambda}{\partial x^\mu}\partial'_\lambda f.$$

This is precisely the transformation (2.33) of a $(0, 1)$-tensor.

The partial derivative of a vector is NOT a tensor! In x^μ-coordinates, consider the partial derivative of a contravariant vector u^ν. By the chain rule,

$$\partial_\mu u^\nu = \frac{\partial}{\partial x^\mu}u^\nu = \frac{\partial y^\lambda}{\partial x^\mu}\frac{\partial}{\partial y^\lambda}u^\nu$$

now, we apply Eq. (2.31) for the vector u^ν,

$$\begin{aligned}\partial_\mu u^\nu = \frac{\partial}{\partial x^\mu}u^\nu &= \frac{\partial y^\lambda}{\partial x^\mu}\frac{\partial}{\partial y^\lambda}\left(\frac{\partial x^\nu}{\partial y^\sigma}u'^\sigma\right) \\ &= \frac{\partial y^\lambda}{\partial x^\mu}\left(\frac{\partial^2 x^\nu}{\partial y^\lambda \partial y^\sigma}u'^\sigma + \frac{\partial x^\nu}{\partial y^\sigma}\frac{\partial u'^\sigma}{\partial y^\lambda}\right).\end{aligned}$$

[4] Lagrangians, gravitational equations, any equation of importance.

In the last term of the brackets, we recognise $\frac{\partial u'^\sigma}{\partial y^\lambda} = \partial'_\lambda u'^\sigma$ as the partial derivative of the vector in y^μ-coordinates. Opening the brackets, we have

$$\partial_\mu u^\nu = \frac{\partial y^\lambda}{\partial x^\mu}\frac{\partial x^\nu}{\partial y^\sigma}\partial'_\lambda u'^\sigma + \frac{\partial y^\lambda}{\partial x^\mu}\frac{\partial^2 x^\nu}{\partial y^\lambda \partial y^\sigma} u'^\sigma .$$

The presence of the second (ugly) term above spoils the transformation law, and it no longer obeys (2.35).

This is a serious problem, because if they are not tensors, the equations of physical laws take different forms in different coordinate systems. And physical laws do indeed involve various derivatives of vector quantities. (Think about the Maxwell equations.)

For physics, derivatives of vectors are important because we often want to know the rate of change of vectors. And the laws of physics involve equations for these rates of changes of vectors. If the laws of physics is to be true for any coordinate system, they must be written in terms of tensors. Therefore we need to have a tensor quantity that somehow describes the rate of change of a vector. To fix this issue, we are led to the concept of covariant derivatives, which is explore in the next section.

2.5 Covariant derivatives

We recall that differentiation involves taking the difference of a quantity at two different points, where the two points are very close to each other. Our goal is to determine the rate of change of some vector v^μ. The first thought is to consider two points p and q on the manifold $\mathcal{M}$. Then the derivative of a vector is expected to take the form

$$\lim_{q \to p} \frac{\mathbf{v}(q) - \mathbf{v}(p)}{q - p}.$$

Immediately we encounter a problem because this expression doesn't make sense! First, q and p are two points on a manifold, they are not numbers which we can subtract. Secondly, $\mathbf{v}(p)$ and $\mathbf{v}(q)$ are vectors attached at points p and q. They belong to different vector spaces. (See Fig. 2.7.) Since p and q are two different points on the some curved manifold $\mathcal{M}$, we can't expect the bases $\{\mathbf{e}_\mu(p)\}$ at p and $\{\mathbf{e}_\mu(q)\}$ at q to match up. Therefore it is not clear how to subtract two vectors attached to p and q.

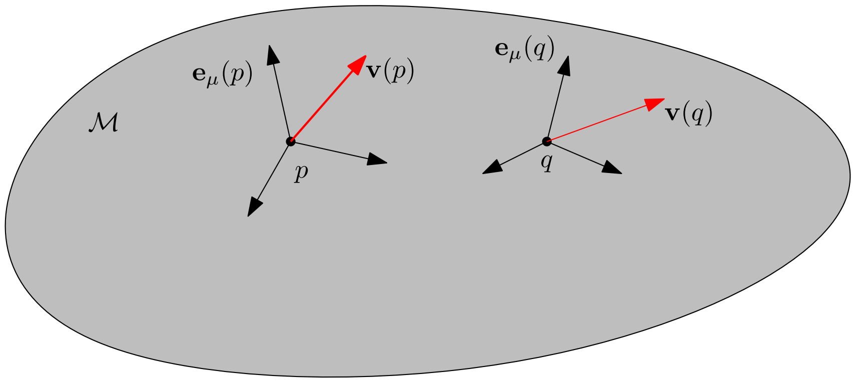


Figure 2.7: Vectors $\mathbf{v}(p)$ and $\mathbf{v}(q)$ at two different points, p and q, respectively. The basis vectors at p are $\{\mathbf{e}_\mu(p)\}$ and the basis vectors at q are $\{\mathbf{e}_\mu(q)\}$.

Nevertheless, we just want to take the derivative. This means we only consider cases where p and q are very close to each other. So even if $\mathbf{e}_\mu(p)$ and $\mathbf{e}_\mu(q)$ are different, they only differ by some small amount:

$$\mathbf{e}_\mu(q) = \mathbf{e}_\mu(p) + \delta\mathbf{e}_\mu, \tag{2.36}$$

where $\delta\mathbf{e}_\mu$ represents this small difference. Correspondingly, the coordinates at p and q also differ by a small amount,

$$x^\mu(q) = x^\mu(p) + \delta x^\mu. \tag{2.37}$$

For each component μ, we can define the change in basis divided by the coordinate interval

$$\frac{\delta\mathbf{e}_\nu}{\delta x^\mu}.$$

And hence we define the partial derivative of the basis by

$$\partial_\mu\mathbf{e}_\nu = \frac{\partial\mathbf{e}_\nu}{\partial x^\mu} = \lim_{\delta x^\nu\to 0}\frac{\delta\mathbf{e}_\nu}{\delta x^\mu}. \tag{2.38}$$

This is the limit where $q \to p$. Therefore this is a quantity that is defined at point p. So the above is a vector quantity that is well-defined as a vector at p. In particular, we can use the basis at p to write

$$\partial_\mu\mathbf{e}_\nu = \frac{\partial\mathbf{e}_\nu}{\partial x^\mu} = \left(\text{components}_{\mu\nu}\right)^\lambda \mathbf{e}_\lambda = \Gamma^\lambda_{\mu\nu}\mathbf{e}_\lambda. \tag{2.39}$$

This quantity tells us how the μ-th basis vector $\mathbf{e}_\mu$ changes if we move away from p infinitesimally along the ν-th direction x^ν. The quantity $\Gamma^\lambda_{\mu\nu}$ is called the *affine connection*, or the *Christoffel symbol*.

At this stage, we can assign two possible categories for the manifold $\mathcal{M}$:

- **Torsionless manifolds:** $\partial_\mu \mathbf{e}_\nu = \partial_\nu \mathbf{e}_\mu$, or $\Gamma^\lambda_{\mu\nu} = \Gamma^\lambda_{\nu\mu}$. That is, the Christoffel symbols of torsionless manifolds are symmetric in the bottom indices.
- **Manifolds with torsion:** $\partial_\mu \mathbf{e}_\nu \neq \partial_\nu \mathbf{e}_\mu$, or $\Gamma^\lambda_{\mu\nu} \neq \Gamma^\lambda_{\nu\mu}$. That is, the Christoffel symbols of manifolds with torsion have no symmetry in its bottom indices.

All spacetimes in the standard theory of General Relativity are torsionless. Therefore in the following we shall assume

$$\Gamma^\lambda_{\mu\nu} = \Gamma^\lambda_{\nu\mu}. \tag{2.40}$$

We now derive an expression for the Christoffel symbols. We take the partial derivative

$$\begin{aligned}
\langle \mathbf{u}, \mathbf{v} \rangle &= g_{\mu\nu} u^\mu v^\nu \\
\partial_\lambda \langle \mathbf{u}, \mathbf{v} \rangle &= (\partial_\lambda g_{\mu\nu}) u^\mu v^\nu + g_{\mu\nu} (\partial_\lambda u^\mu) v^\nu + g_{\mu\nu} u^\mu (\partial_\lambda v^\nu) \\
\langle \partial_\lambda \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \partial_\lambda \mathbf{v} \rangle &= (\partial_\lambda g_{\mu\nu}) u^\mu v^\nu + \langle (\partial_\lambda u^\mu) \mathbf{e}_\mu, v^\nu \mathbf{e}_\nu \rangle + \langle u^\mu \mathbf{e}_\mu, (\partial_\lambda v^\nu) \mathbf{e}_\nu \rangle \\
\langle \partial_\lambda \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \partial_\lambda \mathbf{v} \rangle &= (\partial_\lambda g_{\mu\nu}) u^\mu v^\nu + \langle \partial_\lambda (u^\mu \mathbf{e}_\mu) - u^\mu \partial_\lambda \mathbf{e}_\mu, v^\nu \mathbf{e}_\nu \rangle \\
&\quad + \langle u^\mu \mathbf{e}_\mu, \partial_\lambda (v^\nu \mathbf{e}_\nu) - v^\nu \partial_\lambda \mathbf{e}_\nu \rangle \\
\cancel{\langle \partial_\lambda \mathbf{u}, \mathbf{v} \rangle} + \cancel{\langle \mathbf{u}, \partial_\lambda \mathbf{v} \rangle} &= (\partial_\lambda g_{\mu\nu}) u^\mu v^\nu + \cancel{\langle \partial_\lambda \mathbf{u}, \mathbf{v} \rangle} - u^\mu v^\nu \langle \partial_\lambda \mathbf{e}_\mu, \mathbf{e}_\nu \rangle \\
&\quad + \cancel{\langle \mathbf{u}, \partial_\lambda \mathbf{v} \rangle} - u^\mu v^\nu \langle \mathbf{e}_\mu, \partial_\lambda \mathbf{e}_\nu \rangle \\
u^\mu v^\nu \left[\langle \partial_\lambda \mathbf{e}_\mu, \mathbf{e}_\nu \rangle + \langle \mathbf{e}_\mu, \partial_\lambda \mathbf{e}_\nu \rangle \right] &= g_{\mu\nu} u^\mu v^\nu.
\end{aligned}$$

Since u^μ and v^ν are arbitrary,

$$\partial_\lambda g_{\mu\nu} = \langle \partial_\lambda \mathbf{e}_\mu, \mathbf{e}_\nu \rangle + \langle \mathbf{e}_\mu, \partial_\lambda \mathbf{e}_\nu \rangle.$$

Or, using Eq. (2.39),

$$\begin{aligned}
\partial_\lambda g_{\mu\nu} &= \langle \Gamma^\alpha_{\lambda\mu} \mathbf{e}_\alpha, \mathbf{e}_\nu \rangle + \langle \mathbf{e}_\mu, \Gamma^\beta_{\lambda\nu} \mathbf{e}_\nu \rangle \\
\partial_\lambda g_{\mu\nu} &= \Gamma^\alpha_{\lambda\mu} \langle \mathbf{e}_\alpha, \mathbf{e}_\nu \rangle + \Gamma^\beta_{\lambda\nu} \langle \mathbf{e}_\mu, \mathbf{e}_\beta \rangle \\
\partial_\lambda g_{\mu\nu} &= \Gamma^\alpha_{\lambda\mu} g_{\alpha\nu} + \Gamma^\beta_{\lambda\nu} g_{\mu\beta}.
\end{aligned} \tag{2.41}$$

If we cyclicly permute the indices, we obtain two more versions of the same equation,

$$\partial_\mu g_{\nu\lambda} = \Gamma^\alpha_{\mu\nu} g_{\alpha\lambda} + \Gamma^\alpha_{\mu\lambda} g_{\nu\alpha}, \tag{2.42}$$

$$\partial_\nu g_{\lambda\mu} = \Gamma^\alpha_{\nu\lambda} g_{\alpha\mu} + \Gamma^\alpha_{\nu\mu} g_{\lambda\alpha}. \tag{2.43}$$

Now we take (2.42)+(2.43)−(2.41) to obtain

$$\partial_\mu g_{\nu\lambda} + \partial_\nu g_{\lambda\mu} - \partial_\lambda g_{\mu\nu} = \Gamma^\alpha_{\mu\nu} g_{\alpha\lambda} + \cancel{\Gamma^\alpha_{\mu\lambda} g_{\nu\alpha}} + \cancel{\Gamma^\alpha_{\nu\lambda} g_{\alpha\mu}} + \Gamma^\alpha_{\nu\mu} g_{\lambda\alpha} - \cancel{\Gamma^\alpha_{\lambda\mu} g_{\alpha\nu}} - \cancel{\Gamma^\beta_{\lambda\nu} g_{\mu\beta}},$$

where the cancellations can happen because of the symmetries $\Gamma^\alpha_{\mu\nu} = \Gamma^\alpha_{\nu\mu}$ and $g_{\mu\nu} = g_{\nu\mu}$. Further rearranging, we have

$$g_{\alpha\lambda}\Gamma^\alpha_{\mu\nu} = \frac{1}{2}\left(\partial_\mu g_{\lambda\nu} + \partial_\nu g_{\lambda\mu} - \partial_\lambda g_{\mu\nu}\right).$$

Finally, contracting both sides with $g^{\kappa\alpha}$, we get

$$\begin{aligned} g^{\kappa\lambda} g_{\alpha\lambda}\Gamma^\alpha_{\mu\nu} &= \frac{1}{2} g^{\kappa\lambda}\left(\partial_\mu g_{\lambda\nu} + \partial_\nu g_{\lambda\mu} - \partial_\lambda g_{\mu\nu}\right) \\ \delta^\kappa_\alpha \Gamma^\alpha_{\mu\nu} &= \frac{1}{2} g^{\kappa\lambda}\left(\partial_\mu g_{\lambda\nu} + \partial_\nu g_{\lambda\mu} - \partial_\lambda g_{\mu\nu}\right). \end{aligned}$$

By the property of the Kroenecker delta, $\delta^\kappa_\alpha \Gamma^\alpha_{\mu\nu} = \Gamma^\kappa_{\mu\nu}$. Therefore we have the following formula for the Christoffel symbol:

$$\boxed{\Gamma^\kappa_{\mu\nu} = \frac{1}{2} g^{\kappa\lambda}\left(\partial_\mu g_{\lambda\nu} + \partial_\nu g_{\lambda\mu} - \partial_\lambda g_{\mu\nu}\right).} \tag{2.44}$$

Example 2.5.1. Christoffel symbols of S^2. On sphere S^2 of radius 1, the line element is

$$\mathrm{d}s^2 = \mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2.$$

The metric tensor and its inverse are

$$g_{\mu\nu} = \begin{pmatrix} 1 & 0 \\ 0 & \sin^2\theta \end{pmatrix}, \quad g^{\mu\nu} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\sin^2\theta} \end{pmatrix}.$$

We can directly compute the Christoffel symbols of this manifold:

$$\begin{aligned}
\Gamma^{\theta}_{\theta\theta} &= \frac{1}{2}g^{\theta\lambda}\left(\partial_\theta g_{\lambda\theta} + \partial_\theta g_{\lambda\theta} - \partial_\lambda g_{\theta\theta}\right) = 0,\\
\Gamma^{\phi}_{\theta\theta} &= \frac{1}{2}g^{\phi\lambda}\left(\partial_\theta g_{\lambda\theta} + \partial_\theta g_{\lambda\theta} - \partial_\lambda g_{\theta\theta}\right) = 0,\\
\Gamma^{\theta}_{\theta\phi} &= \frac{1}{2}g^{\theta\lambda}\left(\partial_\theta g_{\lambda\phi} + \partial_\phi g_{\lambda\theta} - \partial_\lambda g_{\theta\phi}\right) = 0,\\
\Gamma^{\phi}_{\theta\phi} &= \frac{1}{2}g^{\phi\lambda}\left(\partial_\theta g_{\lambda\phi} + \partial_\phi g_{\lambda\theta} - \partial_\lambda g_{\theta\phi}\right) = \frac{\cos\theta}{\sin\theta},\\
\Gamma^{\theta}_{\phi\phi} &= \frac{1}{2}g^{\theta\lambda}\left(\partial_\phi g_{\lambda\phi} + \partial_\phi g_{\lambda\phi} - \partial_\lambda g_{\phi\phi}\right) = -\sin\theta\cos\theta,\\
\Gamma^{\phi}_{\phi\phi} &= \frac{1}{2}g^{\phi\lambda}\left(\partial_\phi g_{\lambda\phi} + \partial_\phi g_{\lambda\phi} - \partial_\lambda g_{\phi\phi}\right) = 0.
\end{aligned} \tag{2.45}$$

The covariant derivative. Suppose that we have a vector field on a manifold $\mathcal{M}$. Meaning, that at every point p of the manifold there's a corresponding vector $\mathbf{v}(p)$ attached there. Using the Christoffel symbols we just developed, we will now determine the rate of change of this vector $\mathbf{v}$ as we change the point p slightly.

Ultimately, we wish to find a derivative operator, ∇_λ that can act on any tensor on $\mathcal{M}$. This includes action on scalars, vectors (contravariant/covariant), and arbitrary (r, s)-tensors. We also require that the result of the derivative should still be a tensor. So acting on an (r, s)-tensor, the result must be an $(r, s+1)$-tensor, because ∇_λ introduces an extra lower index. Remember that to be a tensor, it must obey Eq. (2.35).

Let us write, at some point p, the vector has components v^μ. Therefore it can be expanded in its basis as $\mathbf{v} = v^\mu \mathbf{e}_\mu$. Taking the derivative,

$$\begin{aligned}
\partial_\lambda \mathbf{v} &= (\partial_\lambda v^\mu)\mathbf{e}_\mu + v^\mu \partial_\lambda \mathbf{e}_\mu\\
&= (\partial_\lambda v^\mu)\mathbf{e}_\mu + v^\mu \left(\Gamma^\alpha_{\lambda\mu}\right)\mathbf{e}_\alpha\\
&= \underbrace{\left(\partial_\lambda v^\alpha + \Gamma^\alpha_{\lambda\mu} v^\mu\right)}_{=\nabla_\lambda v^\alpha}\mathbf{e}_\alpha\\
\partial_\lambda \mathbf{v} &= \left(\nabla_\lambda v^\alpha\right)\mathbf{e}_\alpha
\end{aligned}$$

where we have defined the components of the resulting derivative as

$$\boxed{\nabla_\mu v^\nu = \partial_\mu v^\nu + \Gamma^\nu_{\mu\lambda} v^\lambda.}$$

This is the *covariant derivative* of a contravariant vector.

What is the covariant derivative of a scalar? (i.e., a $(0, 0)$-tensor.) As we have seen from covariant derivative of vectors, the second term in the formula $\nabla_\mu v^\nu = \partial_\mu v^\nu + \Gamma^\nu_{\mu\alpha} v^\alpha$ is an extra correction added to the ordinary partial derivative. This came from taking into account how the basis vectors $\mathbf{e}_\mu$ changed as we vary p. However, scalar functions does not depend on any basis vector at all. So the extra basis correction term is not neccessary. Therefore the covariant derivative of a scalar $(0,0)$-tensor φ is just identical to the partial derivative,

$$\boxed{\nabla_\lambda \varphi = \partial_\lambda \varphi} \tag{2.46}$$

Next we determine the covariant derivative of covariant vectors, u_μ. So far, we know how to compute $\nabla_\lambda v^\mu$ and $\nabla_\lambda \varphi$. The latter is a scalar. Now, the contraction $u_\mu v^\mu = \varphi$ is also a scalar. So the covariant derivative of this is expected to be just the partial derivative as well,

$$\partial_\lambda \left(u_\mu v^\mu\right) = \nabla_\lambda \left(u_\mu v^\mu\right). \tag{2.47}$$

Now, if ∇_λ is to be a legitimate derivative operator, it should obey the Leibniz rule for derivatives.[5] Therefore we expect

$$\begin{aligned} \partial_\lambda \left(u_\mu v^\mu\right) &= (\nabla_\lambda u_\mu) v^\mu + u_\mu \nabla_\lambda v^\mu \\ (\partial_\lambda u_\mu) v^\mu + u_\mu \partial_\lambda v^\mu &= (\nabla_\lambda u_\mu) v^\mu + u_\mu \left(\partial_\lambda v^\mu + \Gamma^\mu_{\lambda\alpha} v^\alpha\right). \end{aligned}$$

By analogy to the contravariant vectors, we expect a similar correction form $\nabla_\lambda u_\mu = \partial_\lambda u_\mu + C^\alpha_{\lambda\mu} u_\alpha$. Therefore

$$\begin{aligned} (\partial_\lambda u_\mu) v^\mu + u_\mu \partial_\lambda v^\mu &= (\partial_\lambda u_\mu + C^\alpha_{\lambda\mu} u_\alpha) v^\mu + u_\mu \partial_\lambda v^\mu + u_\mu \Gamma^\mu_{\lambda\alpha} v^\alpha \\ 0 &= v^\mu C^\alpha_{\lambda\mu} u_\alpha + v^\alpha \Gamma^\mu_{\lambda\alpha} u_\mu. \end{aligned}$$

Renaming some dummy indices,

$$\begin{aligned} 0 &= v^\beta C^\alpha_{\lambda\beta} u_\alpha + v^\beta \Gamma^\alpha_{\lambda\beta} u_\alpha \\ 0 &= v^\beta \left(C^\alpha_{\lambda\beta} + \Gamma^\alpha_{\lambda\beta}\right) u_\alpha \end{aligned}$$

Since v^β and u_α are arbitrary, we find $C^\alpha_{\lambda\beta} = -\Gamma^\alpha_{\lambda\beta}$. Therefore the covariant derviative of a covariant vector is

$$\boxed{\nabla_\mu u_\nu = \partial_\mu u_\nu - \Gamma^\lambda_{\mu\nu} u_\lambda.} \tag{2.48}$$

[5] Also known as the 'product rule' in calculus.

Extending the covariant derivative to any arbitrary tensor is now straightforward. The rule is simply: every upper indices will bring a term of $+\Gamma$, and every lower index will bring a term of $-\Gamma$. For example, for a $(0,2)$-tensor,

$$\nabla_\lambda A_{\mu\nu} = \partial_\lambda A_{\mu\nu} - \Gamma^\alpha_{\lambda\mu} A_{\alpha\nu} - \Gamma^\alpha_{\lambda\nu} A_{\mu\alpha}.$$

For a $(1,1)$-tensor,

$$\nabla_\lambda A^\mu{}_\nu = \partial_\lambda A^\mu{}_\nu + \Gamma^\mu_{\lambda\alpha} A^\alpha{}_\nu - \Gamma^\alpha_{\lambda\nu} A^\mu{}_\alpha.$$

If $A^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s}$ is an (r,s)-tensor,

$$\begin{aligned}\nabla_\lambda A^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s} = \partial_\lambda A^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s} &+ \Gamma^{\mu_1}_{\lambda\alpha} A^{\alpha\mu_2\cdots\mu_s}{}_{\nu_1\cdots\nu_s} + \ldots + \Gamma^{\mu_r}_{\lambda\alpha} A^{\mu_1\cdots\mu_{r-1}\alpha}{}_{\nu_1\cdots\nu_s} \\ &- \Gamma^\alpha_{\lambda\nu_1} A^{\mu_1\cdots\mu_r}{}_{\alpha\nu_2\cdots\nu_s} - \ldots - \Gamma^\alpha_{\lambda\nu_s} A^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_{s-1}\alpha}.\end{aligned}$$

Example 2.5.2. Covariant derivative of a $(1,1)$-tensor. If $A^\mu{}_\nu$ is a $(1,1)$ tensor, its covariant derivative is a $(1,2)$ tensor:

$$\nabla_\sigma A^\mu{}_\nu = \partial_\sigma A^\mu{}_\nu + \Gamma^\mu_{\sigma\lambda} A^\lambda{}_\nu - \Gamma^\lambda{}_{\sigma\nu} A^\mu{}_\lambda.$$

Example 2.5.3. Let $A_\mu{}^\nu = \nabla_\mu V^\nu$. This is is a well-defined $(1,1)$-tensor. Therefore this itself has a covariant derivative:

$$\nabla_\sigma \nabla_\mu V^\nu = \nabla_\sigma A_\mu{}^\nu = \partial_\sigma A_\mu{}^\nu - \Gamma^\lambda_{\sigma\mu} A_\lambda{}^\nu + \Gamma^\nu_{\sigma\lambda} A_\mu{}^\lambda$$

But also we have $A_\mu{}^\nu = \nabla_\mu V^\nu = \partial_\mu V^\nu + \Gamma^\nu_{\mu\rho} V^\rho$. Therefore

$$\begin{aligned}\nabla_\sigma \nabla_\mu V^\nu &= \partial_\sigma \left(\partial_\mu V^\nu + \Gamma^\nu_{\mu\rho} V^\rho\right) - \Gamma^\lambda_{\sigma\mu} \left(\partial_\lambda V^\nu + \Gamma^\nu_{\lambda\rho} V^\rho\right) + \Gamma^\nu_{\sigma\lambda} \left(\partial_\mu V^\lambda + \Gamma^\lambda_{\mu\rho} V^\rho\right) \\ &= \partial_\sigma \partial_\mu V^\nu + (\partial_\sigma \Gamma^\nu_{\mu\rho}) V^\rho + \Gamma^\nu_{\mu\rho} \partial_\sigma V^\rho - \Gamma^\lambda_{\sigma\mu} \partial_\lambda V^\nu - \Gamma^\lambda_{\sigma\mu} \Gamma^\nu_{\lambda\rho} V^\rho \\ &\quad + \Gamma^\nu_{\sigma\lambda} \partial_\mu V^\lambda + \Gamma^\nu_{\sigma\lambda} \Gamma^\lambda_{\mu\rho} V^\rho.\end{aligned}$$

This (unfortunately) complicated expression is the second derivative of a vector V^μ.

Example 2.5.4. A *Killing vector* is a vector that satisfies *Killing's equation*,

$$\nabla_\mu \xi_\nu + \nabla_\nu \xi_\mu = 0.$$

Show that on the sphere with metric $\mathrm{d}s^2 = \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2$,

$$\xi^\mu = \left(\xi^\theta, \xi^\phi\right) = \left(-\sin\phi, -\frac{\cos\theta}{\sin\theta}\cos\phi\right)$$

is a Killing vector. The Christoffel symbols have been calculated in an earlier example as

$$\Gamma^\phi_{\theta\phi} = \Gamma^\phi_{\phi\theta} = \frac{\cos\theta}{\sin\theta}, \quad \Gamma^\theta_{\phi\phi} = -\sin\theta\cos\theta.$$

Solution. *The Killing's equation give above is for a covariant vector ξ_μ, and it is*

$$\begin{aligned} 0 &= \nabla_\mu\xi_\nu + \nabla_\nu\xi_\mu = \partial_\mu\xi_\nu - \Gamma^\lambda_{\mu\nu}\xi_\lambda + \partial_\nu\xi_\mu - \Gamma^\lambda_{\nu\mu}\xi_\lambda \\ &= \partial_\mu\xi_\nu + \partial_\nu\xi_\mu - 2\Gamma^\lambda_{\mu\nu}\xi_\lambda, \end{aligned}$$

where we have used the symmetry property of the Christoffel symbol $\Gamma^\lambda_{\mu\nu} = \Gamma^\lambda_{\nu\mu}$. We the vector with lower index,

$$\xi_\mu = g_{\mu\nu}\xi^\nu = (\underbrace{-\sin\phi}_{\xi_\theta}, \underbrace{-\cos\theta\sin\theta\cos\phi}_{\xi_\phi})$$

The Killing equation has two indices (μ,ν). We compute each case. For $(\mu\nu) = (\theta\theta)$ the equation is

$$\partial_\theta\xi_\theta - \partial_\theta\xi_\theta - 2\Gamma^\lambda_{\theta\theta}\xi_\lambda = 0.$$

The left-hand side is zero because ξ_θ doesn't depend on θ, and all Christoffel symbols with two θ in its lower index are zero.
Next we look at $(\mu\nu) = (\theta\phi)$, we have

$$\begin{aligned} &\partial_\theta\xi_\phi + \partial_\phi\xi_\theta - 2\Gamma^\lambda_{\theta\phi}\xi_\lambda \\ &\quad = \partial_\theta\left(-\cos\theta\sin\theta\cos\phi\right) + \partial_\phi(-\sin\phi) - 2\Gamma^\phi_{\theta\phi}\xi_\phi \\ &\quad = \sin^2\theta\cos\phi - \cos^2\theta\cos\phi - \cos\phi - 2\frac{\cos\theta}{\sin\theta}(-\cos\theta\sin\theta\cos\phi) \\ &\quad = 0. \end{aligned}$$

The term $(\mu\nu) = (\phi\theta)$ is the same as above by symmetry of the equation. Finally *is the case $(\mu\nu) = (\phi\phi)$, we have*

$$
\begin{aligned}
&\partial_\phi \xi_\phi + \partial_\phi \xi_\phi - 2\Gamma^\lambda_{\phi\phi}\xi_\phi \\
&\qquad = 2\partial_\phi(-\cos\theta\sin\theta\cos\phi) - 2\Gamma^\theta_{\phi\phi}\xi_\theta \\
&\qquad = 2\cos\theta\sin\phi - 2(-\sin\theta\cos\theta)(-\sin\phi) \\
&\qquad = 0.
\end{aligned}
$$

Indeed, the vector ξ^μ as given satisfies the Killing equation. [Note that there are two more vectors in this space that satisfies the equation, they are

$$\left(\cos\phi, -\frac{\cos\theta}{\sin\theta}\sin\phi\right) \quad \textit{and} \quad (0,1)\,.$$

*These three vector fields represent the three independent direction of **spherical symmetry**, the symmetries of the 2-sphere.*
]

Metric compatibility. An important result is that the metric $g_{\mu\nu}$ has zero covariant derivative.

This is proven by direct calculation,

$$
\begin{aligned}
\nabla_\sigma g_{\mu\nu} &= \partial_\sigma g_{\mu\nu} - \Gamma^\lambda_{\sigma\mu} g_{\lambda\nu} - \Gamma^\lambda_{\sigma\nu} g_{\mu\lambda} \\
&= \partial_\sigma g_{\mu\nu} - g_{\lambda\nu}\Gamma^\lambda_{\sigma\mu} - g_{\mu\lambda}\Gamma^\lambda_{\sigma\nu} \\
&= \partial_\sigma g_{\mu\nu} - g_{\lambda\nu}\frac{1}{2}g^{\lambda\alpha}\left(\partial_\sigma g_{\alpha\mu} + \partial_\mu g_{\alpha\sigma} - \partial_\alpha g_{\sigma\mu}\right) - g_{\lambda\mu}\frac{1}{2}g^{\lambda\alpha}\left(\partial_\sigma g_{\alpha\nu} + \partial_\nu g_{\alpha\sigma} - \partial_\alpha g_{\sigma\nu}\right) \\
&= \partial_\sigma g_{\mu\nu} - \frac{1}{2}\delta^\alpha_\nu\left(\partial_\sigma g_{\alpha\mu} + \partial_\mu g_{\alpha\sigma} - \partial_\alpha g_{\sigma\mu}\right) - \frac{1}{2}\delta^\alpha_\mu\left(\partial_\sigma g_{\alpha\nu} + \partial_\nu g_{\alpha\sigma} - \partial_\alpha g_{\sigma\nu}\right) \\
&= \partial_\sigma g_{\mu\nu} - \frac{1}{2}\left(\partial_\sigma g_{\nu\mu} + \cancel{\partial_\mu g_{\nu\sigma}} - \cancel{\partial_\nu g_{\sigma\mu}}\right) - \frac{1}{2}\left(\partial_\sigma g_{\mu\nu} + \cancel{\partial_\nu g_{\mu\sigma}} - \cancel{\partial_\mu g_{\sigma\nu}}\right) \\
&= \partial_\sigma g_{\mu\nu} - \frac{1}{2}\partial_\sigma g_{\nu\mu} - \frac{1}{2}\partial_\sigma g_{\mu\nu} \\
&= 0.
\end{aligned}
$$

The same is true for inverse metric, $\nabla_\sigma g^{\mu\nu} = 0$. This result is called the *metric compatibility of the covariant derivative*,

$$\boxed{\nabla_\sigma g_{\mu\nu} = 0, \quad \nabla_\sigma g^{\mu\nu} = 0.} \tag{2.49}$$

Because of metric compatibility, we can raise/lower indices of tensors by passing through the covariant derivatives.

$$\begin{aligned}
\nabla_\sigma \left(g^{\mu\rho} F_{\mu\nu}\right) &= \left(\nabla_\sigma g^{\mu\rho}\right) F_{\mu\nu} + g^{\mu\rho} \left(\nabla_\sigma F_{\mu\nu}\right) \\
\nabla_\sigma \left(F^\rho{}_\nu\right) &= 0 + g^{\mu\rho} \left(\nabla_\sigma F_{\mu\nu}\right) \\
\nabla_\sigma F^\rho{}_\nu &= g^{\mu\rho} \nabla_\sigma F_{\mu\nu},
\end{aligned}$$

so we can view the metric as 'raising the index from beyond the covariant derivative.'

A useful identity. We derive a useful identity which involves the Christoffel symbol and the divergence. Suppose that we are in some space or spacetime of dimension n with metric

$$\mathrm{d}s^2 = g_{\mu\nu} \mathrm{d}x^\mu \mathrm{d}x^\nu.$$

Think of $g_{\mu\nu}$ as an $n \times n$ square matrix. The inverse is $g^{\mu\nu}$, and the determinant is written as $\det g$.

In this section, we wish to derive an identity for the trace of the Christoffel symbol, $\Gamma^\mu_{\mu\nu}$. By applying the Christoffel symbol formula, it is

$$\begin{aligned}
\Gamma^\mu_{\mu\nu} &= \frac{1}{2} g^{\mu\lambda} \left(\partial_\mu g_{\lambda\nu} + \partial_\nu g_{\lambda\mu} - \partial_\lambda g_{\mu\nu}\right) \\
&= \frac{1}{2} \left(\cancel{g^{\mu\lambda} \partial_\mu g_{\lambda\nu}} + g^{\mu\lambda} \partial_\nu g_{\lambda\mu} - \cancel{g^{\mu\lambda} \partial_\lambda g_{\mu\nu}}\right) \\
&= \frac{1}{2} g^{\mu\lambda} \partial_\nu g_{\mu\lambda}.
\end{aligned} \tag{2.50}$$

Using an identity from linear algebra, for any $n \times n$ matrix M, we have

$$\ln\left(\det M\right) = \mathrm{tr}\left(\ln M\right),$$

where on the right hand side, $A = \ln M$ is a matrix. We will use $M = g$, our metric as the $n \times n$ matrix. In this equation, each entry of the matrix are values that depend on coordinates x^μ. So we can take the partial derivative with respect to the coordinate,

$$\frac{1}{|\det g|} \partial_\lambda \left(|\det g|\right) = \mathrm{tr}\left(g^{-1} \partial_\lambda g\right),$$

where on the right hand side, we have taken $A = \ln g$ so the partial derivative is $\partial_\lambda A = g^{-1} \partial_\lambda g$. Now, $g^{-1} \partial_\lambda g$ is a matrix multiplication. Using our index notation,

$$\begin{aligned}
\frac{1}{\det g} \partial_\lambda \left(|\det g|\right) &= g^{\mu\nu} \partial_\lambda g_{\mu\nu} \\
\partial_\lambda |\det g| &= |\det g| g^{\mu\nu} \partial_\lambda
\end{aligned}$$

Therefore,

$$\partial_\lambda \sqrt{|\det g|} = \frac{\partial_\lambda |\det g|}{2\sqrt{|\det g|}} = \frac{|\det g| g^{\mu\nu} \partial_\lambda g_{\mu\nu}}{2\sqrt{|\det g|}} = \sqrt{|\det g|} \; \underbrace{\frac{1}{2} g^{\mu\nu} \partial_\lambda g_{\mu\nu}}_{\Gamma^\mu_{\mu\lambda}},$$

where we have recognised the expression for the trace of the Christoffel symbol we got earlier in Eq. (2.50). In other words, our identity for the trace is

$$\boxed{\Gamma^\mu_{\mu\lambda} = \frac{1}{\sqrt{|\det g|}} \partial_\lambda \sqrt{|\det g|}.} \tag{2.51}$$

This formula is useful to calculate the divergence of vectors. For instance,

$$\begin{aligned}
\nabla_\mu V^\mu &= \partial_\mu V^\mu + \Gamma^\mu_{\mu\lambda} V^\lambda \\
&= \partial_\mu V^\mu + \frac{1}{\sqrt{|\det g|}} \left(\partial_\lambda \sqrt{|\det g|} \right) V^\lambda \\
&= \frac{1}{\sqrt{|\det g|}} \partial_\mu \left(\sqrt{|\det g|} V^\mu \right).
\end{aligned}$$

The Christoffel symbol is not a tensor. If you're wondering why it's called the Christoffel *symbol* and not a tensor, it's because it isn't. Consider a coordinate transformation $x^\mu \to y^\mu$. Then the metric $\mathrm{d}s^2 = g_{\mu\nu} \mathrm{d}x^\mu \mathrm{d}x^\nu = g'_{\mu\nu} \mathrm{d}y^\mu \mathrm{d}y^\nu$. The metric tensor is a $(0, 2)$-tensor transforming as

$$g'_{\mu\nu} = \frac{\partial x^\sigma}{\partial y^\mu} \frac{\partial x^\rho}{\partial y^\nu} g_{\sigma\rho}, \qquad g'^{\kappa\lambda} = \frac{\partial y^\kappa}{\partial x^\beta} \frac{\partial y^\lambda}{\partial x^\gamma} g^{\beta\gamma}. \tag{2.52}$$

In the following, we will make frequent use of the identity

$$\frac{\partial y^\kappa}{\partial x^\rho} \frac{\partial x^\rho}{\partial y^\nu} = \delta^\kappa_\nu$$

In the new coordinates, the Christoffel symbols should be

$$\Gamma'^\kappa_{\mu\nu} = \frac{1}{2} g'^{\kappa\lambda} \left(\partial'_\mu g'_{\lambda\nu} + \partial'_\nu g'_{\lambda\mu} - \partial'_\lambda g'_{\mu\nu} \right),$$

where $\partial'_\mu = \frac{\partial}{\partial y^\mu}$ is the partial derivative in the new coordinates. Transforming from the old coordinates, we need the quantities

$$
\begin{aligned}
\partial'_\mu g'_{\lambda\nu} &= \frac{\partial}{\partial y^\mu}\left(\frac{\partial x^\sigma}{\partial y^\lambda}\frac{\partial x^\rho}{\partial y^\nu} g_{\sigma\rho}\right) \\
&= \frac{\partial^2 x^\sigma}{\partial y^\mu \partial y^\lambda}\frac{\partial x^\rho}{\partial y^\nu} g_{\sigma\rho} + \frac{\partial x^\sigma}{\partial y^\lambda}\frac{\partial^2 x^\rho}{\partial y^\mu \partial y^\nu} g_{\sigma\rho} + \frac{\partial x^\sigma}{\partial y^\lambda}\frac{\partial x^\rho}{\partial y^\nu}\underbrace{\frac{\partial}{\partial y^\mu} g_{\sigma\rho}}_{=\frac{\partial x^\alpha}{\partial y^\mu}\frac{\partial}{\partial x^\alpha} g_{\sigma\rho}} \\
&= \frac{\partial^2 x^\sigma}{\partial y^\mu \partial y^\lambda}\frac{\partial x^\rho}{\partial y^\nu} g_{\sigma\rho} + \frac{\partial x^\sigma}{\partial y^\lambda}\frac{\partial^2 x^\rho}{\partial y^\mu \partial y^\nu} g_{\sigma\rho} + \frac{\partial x^\sigma}{\partial y^\lambda}\frac{\partial x^\rho}{\partial y^\nu}\frac{\partial x^\alpha}{\partial y^\mu}\partial_\alpha g_{\sigma\rho}
\end{aligned}
$$

The other terms are simply the permutation of the indices. So

$$
\begin{aligned}
\partial'_\mu g'_{\lambda\nu} + \partial'_\nu g'_{\lambda\mu} - \partial'_\lambda g'_{\mu\nu} &= \cancel{\frac{\partial^2 x^\sigma}{\partial y^\mu \partial y^\lambda}\frac{\partial x^\rho}{\partial y^\nu}} g_{\sigma\rho} + \frac{\partial x^\sigma}{\partial y^\lambda}\frac{\partial^2 x^\rho}{\partial y^\mu \partial y^\nu} g_{\sigma\rho} + \frac{\partial x^\sigma}{\partial y^\lambda}\frac{\partial x^\rho}{\partial y^\nu}\frac{\partial x^\alpha}{\partial y^\mu}\partial_\alpha g_{\sigma\rho} \\
&\quad + \cancel{\frac{\partial^2 x^\sigma}{\partial y^\nu \partial y^\lambda}\frac{\partial x^\rho}{\partial y^\mu} g_{\sigma\rho}} + \frac{\partial x^\sigma}{\partial y^\lambda}\frac{\partial^2 x^\rho}{\partial y^\nu \partial y^\mu} g_{\sigma\rho} + \frac{\partial x^\sigma}{\partial y^\lambda}\frac{\partial x^\rho}{\partial y^\mu}\frac{\partial x^\alpha}{\partial y^\nu}\partial_\alpha g_{\sigma\rho} \\
&\quad - \cancel{\frac{\partial^2 x^\sigma}{\partial y^\lambda \partial y^\mu}\frac{\partial x^\rho}{\partial y^\nu}} g_{\sigma\rho} - \cancel{\frac{\partial x^\sigma}{\partial y^\mu}\frac{\partial^2 x^\rho}{\partial y^\lambda \partial y^\nu} g_{\sigma\rho}} - \frac{\partial x^\sigma}{\partial y^\lambda}\frac{\partial x^\rho}{\partial y^\nu}\frac{\partial x^\alpha}{\partial y^\lambda}\partial_\alpha g_{\sigma\rho} \\
&= 2\frac{\partial x^\sigma}{\partial y^\lambda}\frac{\partial^2 x^\rho}{\partial y^\mu \partial y^\nu} g_{\rho\sigma} \\
&\qquad + \partial_\alpha g_{\sigma\rho}\left(\frac{\partial x^\sigma}{\partial y^\lambda}\frac{\partial x^\rho}{\partial y^\nu}\frac{\partial x^\alpha}{\partial y^\mu} + \frac{\partial x^\sigma}{\partial y^\lambda}\frac{\partial x^\rho}{\partial y^\mu}\frac{\partial x^\alpha}{\partial y^\nu} - \frac{\partial x^\sigma}{\partial y^\mu}\frac{\partial x^\rho}{\partial y^\nu}\frac{\partial x^\alpha}{\partial y^\lambda}\right)
\end{aligned}
$$

Next we contract both sides with $\frac{1}{2}g'^{\kappa\lambda} = \frac{1}{2}\frac{\partial y^\kappa}{\partial x^\beta}\frac{\partial y^\lambda}{\partial x^\gamma} g^{\beta\gamma}$ to make the Christoffel symbol,

$$
\begin{aligned}
\Gamma'^\kappa_{\mu\nu} &= \frac{1}{2} g'^{\kappa\lambda}\left(\partial'_\mu g'_{\lambda\nu} + \partial'_\nu g'_{\lambda\mu} - \partial'_\lambda g'_{\mu\nu}\right) \\
&= \frac{1}{2}\frac{\partial y^\kappa}{\partial x^\beta}\color{red}{\frac{\partial y^\lambda}{\partial x^\gamma}}\color{black} g^{\beta\gamma}\partial_\alpha g_{\sigma\rho}\left(\frac{\partial x^\sigma}{\partial y^\lambda}\frac{\partial x^\rho}{\partial y^\nu}\frac{\partial x^\alpha}{\partial y^\mu} + \frac{\partial x^\sigma}{\partial y^\lambda}\frac{\partial x^\rho}{\partial y^\mu}\frac{\partial x^\alpha}{\partial y^\nu} - \frac{\partial x^\sigma}{\partial y^\mu}\frac{\partial x^\rho}{\partial y^\nu}\frac{\partial x^\alpha}{\partial y^\lambda}\right) \\
&\qquad + \frac{1}{\cancel{2}}\frac{\partial y^\kappa}{\partial x^\beta}\underbrace{\frac{\partial y^\lambda}{\partial x^\gamma}\frac{\partial x^\sigma}{\partial y^\lambda}}_{\delta^\sigma_\gamma}\frac{\partial^2 x^\rho}{\partial y^\mu \partial y^\nu}\cancel{2} g^{\beta\gamma} g_{\rho\sigma}.
\end{aligned}
$$

Multiplying $\color{red}{\frac{\partial y^\lambda}{\partial x^\gamma}}$ into the brackets, we get contractions which turn them into Dirac deltas,

$$
\begin{aligned}
\Gamma'^\kappa_{\mu\nu} &= \frac{\partial y^\kappa}{\partial x^\beta}\frac{1}{2} g^{\beta\gamma}\partial_\alpha g_{\sigma\rho}\left(\color{red}{\delta^\sigma_\gamma}\color{black}\frac{\partial x^\rho}{\partial y^\nu}\frac{\partial x^\alpha}{\partial y^\mu} + \color{red}{\delta^\sigma_\gamma}\color{black}\frac{\partial x^\rho}{\partial y^\mu}\frac{\partial x^\alpha}{\partial y^\nu} - \color{red}{\delta^\alpha_\gamma}\color{black}\frac{\partial x^\sigma}{\partial y^\mu}\frac{\partial x^\rho}{\partial y^\nu}\right) + \frac{\partial y^\kappa}{\partial x^\beta}\frac{\partial^2 x^\rho}{\partial y^\mu \partial y^\nu}\underbrace{\delta^\sigma_\gamma g_{\rho\sigma} g^{\beta\gamma}}_{g_{\rho\sigma}g^{\beta\sigma}=\delta^\beta_\rho} \\
&= \frac{\partial y^\kappa}{\partial x^\beta}\frac{1}{2} g^{\beta\gamma}\left(\partial_\alpha g_{\gamma\rho}\frac{\partial x^\rho}{\partial y^\nu}\frac{\partial x^\alpha}{\partial y^\mu} + \partial_\rho g_{\gamma\alpha}\frac{\partial x^\alpha}{\partial y^\mu}\frac{\partial x^\rho}{\partial y^\nu} - \partial_\gamma g_{\sigma\rho}\frac{\partial x^\rho}{\partial y^\mu}\frac{\partial x^\sigma}{\partial y^\nu}\right) + \frac{\partial y^\kappa}{\partial x^\rho}\frac{\partial^2 x^\rho}{\partial y^\mu \partial y^\nu}
\end{aligned}
$$

Renaming some dummy indices, the three Jacobians can be factorised out of the brackets, and we get

$$
\begin{aligned}
\Gamma'^{\kappa}_{\mu\nu} &= \frac{\partial y^\kappa}{\partial x^\beta}\frac{\partial x^\rho}{\partial y^\nu}\frac{\partial x^\alpha}{\partial y^\mu}\underbrace{\frac{1}{2}g^{\beta\gamma}\left(\partial_\alpha g_{\alpha\rho}+\partial_\rho g_{\gamma\alpha}-\partial_\gamma g_{\alpha\rho}\right)}_{\Gamma^\beta_{\alpha\rho}}+\frac{\partial y^\kappa}{\partial x^\rho}\frac{\partial^2 x^\rho}{\partial y^\mu \partial y^\nu} \\
&= \frac{\partial y^\kappa}{\partial x^\beta}\frac{\partial x^\rho}{\partial y^\nu}\frac{\partial x^\alpha}{\partial y^\mu}\Gamma^\beta_{\alpha\rho}+\frac{\partial y^\kappa}{\partial x^\rho}\frac{\partial^2 x^\rho}{\partial y^\mu \partial y^\nu}. \qquad (2.53)
\end{aligned}
$$

This is the transformation property of the Christoffel symbol; the second term violates the tensor transformation law — the Christoffel symbols are not tensors!

There's another equivalent way to express the right hand side of the above equation. First, observe that

$$
\frac{\partial y^\kappa}{\partial x^\rho}\frac{\partial x^\rho}{\partial y^\nu}=\delta^\kappa_\nu
$$

is a constant. So if we differentiate this equation with $\frac{\partial}{\partial y^\mu}$, the right hand side becomes zero, and the left hand side uses the product rule,

$$
\begin{aligned}
\left(\frac{\partial}{\partial y^\mu}\frac{\partial y^\kappa}{\partial x^\rho}\right)\frac{\partial x^\rho}{\partial y^\nu}+\frac{\partial y^\kappa}{\partial x^\rho}\left(\frac{\partial^2 x^\rho}{\partial y^\mu \partial y^\nu}\right) &= 0 \\
\left(\frac{\partial x^\sigma}{\partial y^\mu}\frac{\partial^2 y^\kappa}{\partial x^\sigma \partial x^\rho}\right)\frac{\partial x^\rho}{\partial y^\nu}+\frac{\partial y^\kappa}{\partial x^\rho}\frac{\partial^2 x^\rho}{\partial y^\mu \partial y^\nu} &= 0 \\
\frac{\partial x^\sigma}{\partial y^\mu}\frac{\partial x^\rho}{\partial y^\nu}\frac{\partial^2 y^\kappa}{\partial x^\sigma \partial x^\rho}+\frac{\partial y^\kappa}{\partial x^\rho}\frac{\partial^2 x^\rho}{\partial y^\mu \partial y^\nu} &= 0 \\
\frac{\partial y^\kappa}{\partial x^\rho}\frac{\partial^2 x^\rho}{\partial y^\mu \partial y^\nu} &= -\frac{\partial x^\sigma}{\partial y^\mu}\frac{\partial x^\rho}{\partial y^\nu}\frac{\partial^2 y^\kappa}{\partial x^\sigma \partial x^\rho}.
\end{aligned}
$$

So Eq. (2.53) can be alternatively written as

$$
\Gamma'^{\kappa}_{\mu\nu} = \frac{\partial y^\kappa}{\partial x^\beta}\frac{\partial x^\rho}{\partial y^\nu}\frac{\partial x^\alpha}{\partial y^\mu}\Gamma^\beta_{\alpha\rho}-\frac{\partial x^\sigma}{\partial y^\mu}\frac{\partial x^\rho}{\partial y^\nu}\frac{\partial^2 y^\kappa}{\partial x^\sigma \partial x^\rho}. \qquad (2.54)
$$

It is actually important that the Christoffel symbol is not a tensor, because the partial derivative $\partial_\mu V^\nu$ is also not a tensor. But the combination

$$
\nabla_\mu V^\nu = \partial_\mu V^\nu + \Gamma^\nu_{\mu\lambda}V^\lambda
$$

will obey the tensor transformation law. Generally, the covariant derivative $\nabla_\alpha T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s}$ is a $(r,s+1)$-tensor obeying the correct tensor transformation law.

2.6 Christoffel symbols of a spherically-symmetric spacetime

A general static, spherically-symmetric spacetime is described by

$$\mathrm{d}s^2 = -f(r)\mathrm{d}t^2 + h(r)\mathrm{d}r^2 + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right),$$

where $f(r)$ and $h(r)$ are arbitrary functions. Here we are using lightspeed units such that $t = ct_{\mathrm{SI}}$ where t_{SI} is the time in usual SI units. Now, observe that the piece

$$\gamma_{ij}\mathrm{d}\theta^i\mathrm{d}\theta^j = \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2,$$

is the metric on unit S^2, for which we already calculated the Christoffel symbol earlier. We rewrite the metric as

$$\mathrm{d}s^2 = -f\mathrm{d}t^2 + h\mathrm{d}r^2 + r^2\gamma_{ij}\mathrm{d}\theta^i\mathrm{d}\theta^j, \quad (\theta^1,\theta^2) = (\theta,\phi).$$

The metric tensor and its inverse is

$$g_{\mu\nu} = \begin{pmatrix} -f & 0 & 0 \\ 0 & h & 0 \\ 0 & 0 & r^2\gamma_{ij} \end{pmatrix}, \quad g^{\mu\nu} = \begin{pmatrix} -\frac{1}{f} & 0 & 0 \\ 0 & \frac{1}{h} & 0 \\ 0 & 0 & \frac{1}{r^2}\gamma^{ij} \end{pmatrix}$$

Since there are four coordinates, there are many components of $\Gamma^\kappa_{\mu\nu}$ to calculate. This daunting task can be approached systematically. In fact, since $g_{\mu\nu}$ is diagonal and all parts are independent of t, the calculation simplifies greatly. In particular any derivatives with t are zero immediately.

Keeping κ arbitrary for now, the Christoffel symbols to calculate are

$$\begin{array}{ccc} \Gamma^\kappa_{tt}, & \Gamma^\kappa_{tr}, & \Gamma^\kappa_{tj}, \\ & \Gamma^\kappa_{rr}, & \Gamma^\kappa_{rj}, \\ & & \Gamma^\kappa_{ij}, \end{array}$$

the others are related by symmetry $\Gamma^\kappa_{\mu\nu} = \Gamma^\kappa_{\nu\mu}$. Let's start with the first one

$$\Gamma^\kappa_{tt} = \frac{1}{2}g^{\kappa\lambda}\left(\partial_t g_{\lambda t} + \partial_t g_{\lambda t} - \partial_\lambda g_{tt}\right)$$

Since none of the metric components depend on t, all derivatives with t vanish immediately. We are left with

$$\Gamma^{\kappa}_{tt} = -\frac{1}{2} g^{\kappa\lambda} \partial_\lambda g_{tt}.$$

Now this is a sum over λ with $g_{tt} = -f$, which depend only on r. So the only non-zero term is when $\lambda = r$. Noticing $g^{\kappa\lambda}$ with $\lambda = r$, this term is only non-zero when κ is also r. So, when the lower indices are tt, the only non-zero Christoffel symbol is

$$\Gamma^{r}_{tt} = -\frac{1}{2} g^{rr} \partial_r g_{tt} = -\frac{1}{2h} \partial_r(-f) = \frac{f'}{2h}.$$

Next we consider the Christoffel symbol when the two lower indices are tr:

$$\Gamma^{\kappa}_{tr} = \frac{1}{2} g^{\kappa\lambda} \left(\partial_t g_{\lambda r} + \partial_r g_{\lambda t} - \partial_\lambda g_{tr}\right).$$

The first term is zero because of the t-derivative. The last term is zero because the metric is diagonal hence $g_{tr} = 0$. We are left with

$$\Gamma^{\kappa}_{tr} = \frac{1}{2} g^{\kappa\lambda} \partial_r g_{\lambda t}.$$

This is a sum over λ. But the only non-zero term would be when $\lambda = t$, then $g_{\lambda t} = g_{tt}$. Furthermore, when $\lambda = t$, the only non-zero component is when κ is also t. So when the lower indices are tr, the only non-zero Christoffel is

$$\Gamma^{t}_{tr} = \frac{1}{2} g^{tt} \partial_r g_{tt} = -\frac{1}{2f} \partial_r(-f) = \frac{f'}{2f}.$$

When the lower indices are tj, the Christoffel symbols are always zero.

Proceeding this way for all the other components, we find that the non-zero components of the Christoffel symbols are

$$\begin{aligned} \Gamma^{r}_{tt} = \frac{f'}{2h}, \quad \Gamma^{t}_{tr} = \frac{f'}{2f}, \quad \Gamma^{r}_{rr} = \frac{h'}{2h}, \quad \Gamma^{r}_{ij} = -\frac{r}{h}\tilde{\gamma}_{ij}, \\ \Gamma^{i}_{rj} = \frac{1}{r}\tilde{\delta}^{i}_{j}, \quad \Gamma^{k}_{ij} = \tilde{\Gamma}^{k}_{ij}, \end{aligned} \tag{2.55}$$

where $\tilde{\Gamma}^{k}_{ij} \in \left\{\tilde{\Gamma}^{\theta}_{\theta\theta}, \tilde{\Gamma}^{\theta}_{\phi\phi}, \ldots\right\}$ are the Christoffel symbols for $\tilde{\gamma}_{ij} \mathrm{d}\theta^i \mathrm{d}\theta^j = \mathrm{d}\theta^2 + \sin^2\theta \, \mathrm{d}\phi^2$, which we have calculated in Eq. (2.45).

Chapter 3 Geodesics and curvature

Going back to the weak equivalence principle, GR is based on Einstein's bold claim that the presence of gravity is equivalent to an accelerated frame. A consequence of this is that gravity is not a force.

Now, if gravity is not a 'force', the presence of gravity should not change a particle's momentum. In relativity, the momentum is represented by a four-vector $p^\mu = mu^\mu$, where $u^\mu = \dot{x}^\mu$ is the particle's four velocity. Therefore, if only gravity is present, we expect the four velocity u^μ to be somehow 'constant'. The constancy of this four-vector is represented by the notion of *parallel-transport*.

In the following we shall consider vectors on a general n-dimensional manifold $\mathcal{M}$. The concept of parallel transport can be applied equally well to spacetimes or simply Euclidean spaces like $\mathbb{R}^3$, spheres, etc.

The idea of parallel-transport, as the name suggests, is how to move a vector across different points on a manifold while 'keeping it parallel'. On flat space like $\mathbb{R}^2$, this is straightforward — just keep the angle of the vector constant while sliding its base according to any path we define, as shown in the left side of Fig. 3.1.

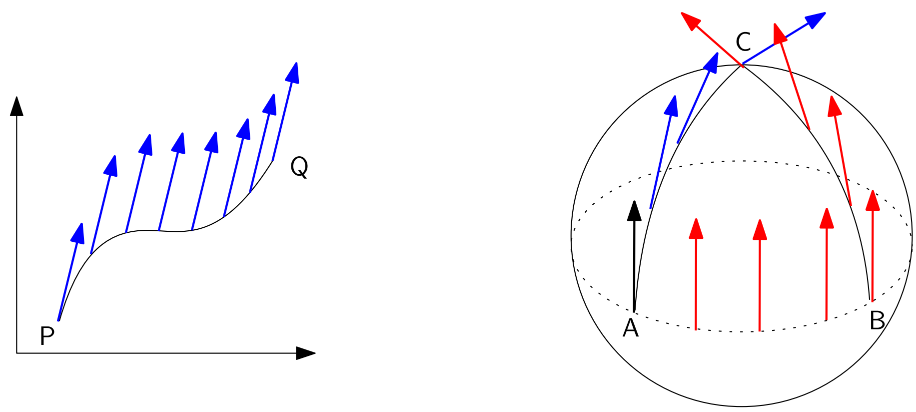


Figure 3.1: Parallel transport of a vector on flat space (left) and on a curved surface (right).

On a curved manifold, the result of the parallel transport depends on the path taken. Consider the right side of Fig. 3.1, where we want to parallel-transport black, initially north-pointing vector on a sphere S^2. We first do this along the curve AC, where point C is the North pole. As expected, to keep the vector parallel means we keep the vector pointing north at all times. So if we parallel-transport it, we end with the blue vector at point C.

But if we were to transport the vector along path ABC instead, we still keep the vector pointing north at all times. However at the final piont, we end with the red vector at point C. Clearly this vector points in a different direction than the vector that took the first path AC! Generally, on curved surfaces, the result of parallel-transport depends on the path taken.

In this chapter, there are two problems to address:

1. Given a vector field on a manifold $\mathcal{M}$, how to find a path such that each vector along the path is parallel? — Answering this will lead us to the concept of *geodesics*. The *geodesic equation* will give us a path where the four-velocity is parallel-transported.

2. A curved manifold is associated with paralle transport being path-dependent. How do we quantify this? — Answering this will lead us to the concept of *spacetime curvature*. In particular, spacetime curvature will be quantified by the *curvature tensors*.

3.1 Geodesics

Consider an n-dimensional manifold $\mathcal{M}$, with some coordinate system where its various points can be written as $x^\mu = (x^1, \ldots, x^n)$. Then, a parametrised curve on $\mathcal{M}$ can be written as $x^\mu(\sigma)$, where σ is the parameter of the curve. In particular, the initial point of the curve is $x^\mu(0)$.

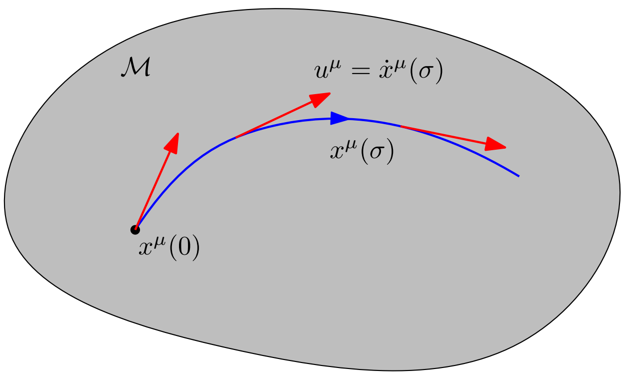


Figure 3.2: A curve on $\mathcal{M}$, described in coordinates by $x^\mu(\sigma)$. The tangent vector at each point along the curve is given by $u^\mu = \dot{x}^\mu(\sigma) = \frac{\mathrm{d}}{\mathrm{d}\sigma}x^\mu(\sigma)$.

Parallel transport and the geodesic equation

At an arbitrary point on the curve $x^\mu(\sigma)$, the tangent vector is $u^\mu = \frac{\mathrm{d}}{\mathrm{d}\sigma}x^\mu(\sigma) = \dot{x}^\mu(\sigma)$. If $\mathcal{M}$ is a spacetime, this might be the four-velocity. Since u^μ is a vector, we can take its covariant derivative

$$\nabla_\nu u^\mu = \partial_\nu u^\mu + \Gamma^\mu_{\nu\alpha} u^\alpha.$$

However, we are focused on the problem of parallel-transport. That means, we want to know the rate of change of u^μ as we go along the curve $x^\mu(\sigma)$. At the instant σ, the direction of motion is also $u^\mu = \dot{x}^\mu(\sigma)$. Therefore, to get the rate of change along this direction, we project the covariant derivative along the direction of u^μ:

$$u^\nu \nabla_\nu u^\mu = \text{rate of change of } u^\mu \text{ along the direction of motion.}$$

If we wish this vector to be parallel-transported, we shall require this rate of change to be *zero*,

$$\boxed{u^\nu \nabla_\nu u^\mu = 0.} \tag{3.1}$$

This is called the *geodesic equation*. Next we shall derive a more practical form of Eq. (3.1). First we apply the formula for covariant derivatives

$$\begin{aligned} u^\nu \nabla_\nu u^\mu = u^\nu \left(\partial_\nu u^\mu + \Gamma^\mu_{\nu\alpha} u^\alpha\right) &= 0 \\ u^\nu \partial_\nu u^\mu + \Gamma^\mu_{\nu\alpha} u^\nu u^\alpha &= 0. \end{aligned}$$

Next we recall that $u^\nu = \dot{x}^\nu$. Therefore, by the chain rule, $u^\nu \partial_\nu u^\mu = \dot{x}^\nu \partial_\nu \dot{x}^\mu =$

$$\begin{aligned} u^\nu \partial_\nu u^\mu = \dot{x}^\nu \partial_\nu \dot{x}^\mu &= \underbrace{\frac{\mathrm{d}x^\nu}{\mathrm{d}\sigma}\frac{\partial}{\partial x^\nu}}_{\frac{\mathrm{d}}{\mathrm{d}\sigma}} \left(\frac{\mathrm{d}x^\mu}{\mathrm{d}\sigma}\right) \\ &= \frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\sigma^2} = \ddot{x}^\mu. \end{aligned}$$

Therefore the geodesic equation becomes

$$\boxed{\ddot{x}^\mu + \Gamma^\mu_{\alpha\beta} \dot{x}^\alpha \dot{x}^\beta = 0.} \tag{3.2}$$

This form of the geodesic equation makes it clear that it's a second-order ordinary differential equation in each coordinate of x^μ. By the usual theory of differential equations, setting an initial position $x^\mu(0)$ and initial velocity $\dot{x}^\mu(0)$ will uniquely determine the resulting trajectory.

Example 3.1.1. *Take the case* $\mathcal{M} = \mathbb{R}^2$*, whose metric in standard Cartesian coordinates is*

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

(a) *Find the geodesics whose initial condition is* $(x(0), y(0)) = (a, b)$ *and initial velocity is* $(\dot{x}(0), \dot{y}) = (1, \frac{1}{2})$.

(b) *Show that a circular motion of radius* a *centred at the origin is not a geodesic.*

Solution.

(a) In Cartesian coordinates, the Christoffel symbols of $\mathbb{R}^2$ are all zero. Therefore the geodesic equation $\ddot{x}^\mu + \Gamma^\mu_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta = 0$ is simply

$$\ddot{x} = 0, \quad \ddot{y} = 0.$$

Solving the differential equation with the given initial conditions, we simply have

$$x(\sigma) = \sigma + a, \quad y(\sigma) = \frac{1}{2}\sigma + b.$$

Therefore, the geodesic in $\mathbb{R}^2$ is a straight line.

(b) Circular motion of radius a centred at the origin is given by the equation

$$x(\sigma) = a\cos\omega\sigma, \quad y(\sigma) = a\sin\omega\sigma, \tag{3.3}$$

where ω is a constant. We have shown in (a) that the geodesic equation reduces to $\ddot{x} = 0$ and $\ddot{y} = 0$. But for Eq. (3.3), we have

$$\ddot{x} = -\omega^2 a\cos\omega\sigma \neq 0,$$
$$\ddot{y} = -\omega^2 a\sin\omega\sigma \neq 0.$$

The geodesic equation is not satisfied. Therefore a circle is not a geodesic in $\mathbb{R}^2$.

Time-like, null, and space-like curves

As discussed in Chapter 1, the worldline of a particle describe its motion in space, as well as its changes in time. We start by generalising the concepts from Chapter 1 to any manifold $\mathcal{M}$ with metric $\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu$, that is not neccessarily Minkowski. The worldline is some parametrised curve $x^\mu(\tau)$.

Time-like trajectories. Consider the scenario in Fig. 3.3, where Alice (defining the metric and coordinate system) is observing the motion of Bob (defining the test particle)

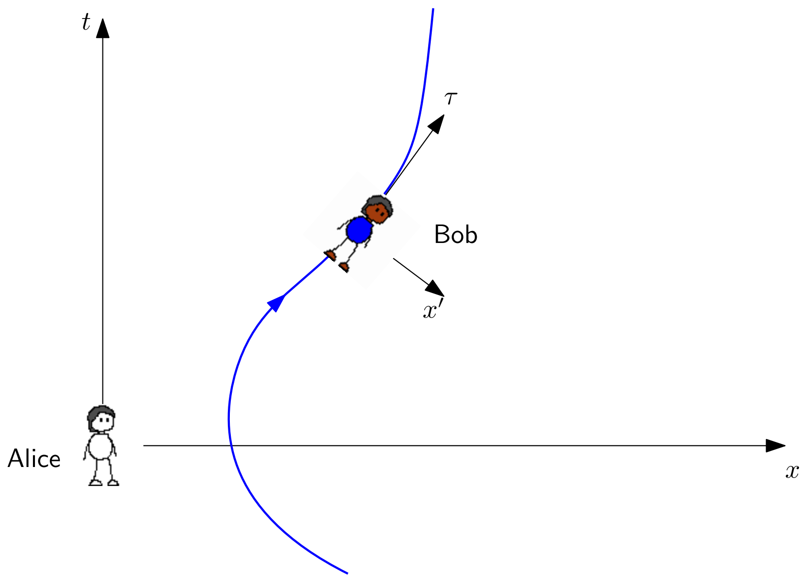


Figure 3.3: Motion of a test particle (Bob), as observed according to coordinate system used by Alice.

Consider an infinitesimal segment of the curve. Generally, a small segment has small displacements in coordinates $\mathrm{d}x^\mu$. The spacetime interval observed by Alice is

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu \tag{3.4}$$

as usual. But suppose from Bob's perspective, he is using a coordinate system attached to himself. So in terms of position, Bob does not move. Therefore if Bob measures his own interval,

$$\mathrm{d}s^2 = -c^2\mathrm{d}\tau, \tag{3.5}$$

where τ is Bob's time coordinate, the proper time. Note that here we are using time in length units, $\tau = c\tau_{\mathrm{SI}}$ where τ_{SI} is the time in SI units measured in seconds. We require that both observers should obtain the same invariant spacetime interval. Hence Eqs. (3.4) and (3.5) are equal to each other. That is,

$$
\begin{aligned}
g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu &= -\mathrm{d}\tau^2 \\
g_{\mu\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}\tau}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau} &= -c^2.
\end{aligned} \tag{3.6}
$$

Observed by Alice, we define the particle's *four velocity* as the derivative with respect to the particle's proper time,

$$
u^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} = \dot{x}^\mu. \tag{3.7}
$$

Here, $u^\mu = \dot{x}^\mu$ is a tangent vector to the worldline. If the tangent vector is taken as the derivative with respect to proper time τ, we get Eq. (3.6), or

$$
g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu = -c^2. \tag{3.8}
$$

That is, the norm of the four velocity $u^\mu = \dot{x}^\mu$ is constant. We call Eq. (3.8) the *proper-time normalisation of the four-velocity.*

Null trajectories. For null trajectories there is no concept of proper time, because no 'Bob' can travel at the speed of light. But the trajectories can still be parametrised, by some quantity λ so that $x^\mu(\lambda)$ is the trajectory. For null paths, the spacetime interval must be zero in any coordinate system

$$
\begin{aligned}
\mathrm{d}s^2 = 0 &= g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu \\
0 &= g_{\mu\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda}.
\end{aligned}
$$

If we write $\dot{x}^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}$, then we have

$$
g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu = 0. \tag{3.9}
$$

Space-like trajectories. No physical particle can travel faster than the speed of light. So they cannot travel on space-like trajectories. However, we can carry an analogous argument by a mathematical displacement in a space direction where 'Bob' displaces by

$$
\mathrm{d}s^2 = +\mathrm{d}\sigma^2, \tag{3.10}
$$

where σ is some spatial direction. This invariant must be the same in Alice's coordinates,

$$
\mathrm{d}s^2 = \mathrm{d}\sigma^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu \quad g_{\mu\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}\sigma}\frac{\mathrm{d}x^\nu}{\mathrm{d}\sigma} = +1.
$$

If we write $\dot{x}^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\sigma}$, we have

$$g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu = +1. \tag{3.11}$$

For compactness, Eqs. (3.8), (3.9), and (3.11), can be written together as

$$g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu = \epsilon = \begin{cases} -c^2, & \text{time-like,} \\ 0, & \text{null/light-like,} \\ +1, & \text{space-like} \end{cases} \tag{3.12}$$

where, depending on ϵ, $\dot{x}^\mu$ is understood to be derivatives with respect to τ, λ, or σ, depending on the problem we are solving. For the physics of particle motion, particles with non-zero mass travel along curves where $\epsilon = -c^2$. Massless particles like photons take $\epsilon = 0$. It is physically impossible for any particle to take $\epsilon = +1$. But they are relevant in solving mathematical problems of finding geodesic curves on curved surfaces.

Given a spacetime of metric $g_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu$, what determines the curve of a time-like/null particle? According to relativistic theory, they move along *geodesics*. We will show that this is true according to Lagrangian mechanics.

3.2 Geodesics from variational principles

Lagrangian formulation of time-like geodesics

In the previous section, we derived the geodesic equation $\ddot{x}^\kappa + \Gamma^\kappa_{\mu\nu}\dot{x}^\mu\dot{x}^\nu = 0$ from the mathematical problem of parallel transport, but we claimed that physical particle travel this curve. Now, usually physical behaviours are determined from physical principles, perhaps involving a Lagrangian or Hamiltonian. We now show that there is indeed an action principle that reproduces the geodesic equation, so physical particles do really follow a geodesic curve.

Let m be the mass of a test particle[1] and τ be its proper time. Its four velocity is $\dot{x}^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau}$. The motion of this particle in spacetime is governed by the action

$$S[x] = \int_{\tau_1}^{\tau_2} \mathrm{d}\tau\, \frac{1}{2} m g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu, \tag{3.13}$$

where $\mathcal{L} = \frac{1}{2} m g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu$ is the Lagrangian and τ_1 and τ_2 are the initial and final proper times of the particle, respectively.

[1] Here, 'test particle' means m is small such that its own gravitational field is negligible.

The trajectory is written as a curve parametrised in proper time. Therefore performing a variation means we perturb

$$x^\mu \to x^\mu + \delta x^\mu, \tag{3.14}$$

where δx^μ is a small variation from some arbitrary trajectory $x^\mu(\lambda)$. The action under this variation is

$$\begin{aligned} S[x+\delta x] &= S[x] + \delta \int \mathrm{d}\tau\, \frac{1}{2} m g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu \\ \delta S = S[x+\delta x] - S[x] &= \frac{m}{2}\int \mathrm{d}\tau \delta\left(g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu\right). \end{aligned} \tag{3.15}$$

We are seeking the conditions for which $\delta S = 0$, yielding the extremum of the action. To do this we have to evaluate

$$\delta\left(g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu\right) = (\delta g_{\mu\nu})\dot{x}^\mu\dot{x}^\nu + g_{\mu\nu}\delta\dot{x}^\mu\dot{x}^\nu + g_{\mu\nu}\dot{x}^\mu\delta\dot{x}^\nu.$$

Now, since $g_{\mu\nu}$ is a function of x^μ, varying x^μ will cause a corresponding change by $\delta x^\lambda \partial_\lambda g_{\mu\nu}$. (Think of this as a sort of 'chain rule'.) Furthermore, $g_{\mu\nu} = g_{\nu\mu}$ is symmetric. So the last two terms of the above equation are equal and we have

$$\begin{aligned} \delta\left(g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu\right) &= \delta x^\sigma(\partial_\sigma g_{\mu\nu})\dot{x}^\mu\dot{x}^\nu + 2g_{\mu\nu}\dot{x}^\mu\delta\dot{x}^\nu \\ &= \delta x^\sigma\partial_\sigma g_{\mu\nu} - 2\delta x^\nu\frac{\mathrm{d}}{\mathrm{d}\tau}(g_{\mu\nu}\dot{x}^\mu) + \frac{\mathrm{d}}{\mathrm{d}\tau}\left(2g_{\mu\nu}\dot{x}^\mu\delta x^\nu\right) \end{aligned}$$

where in the last equality we applied the usual integration-by-parts typical of any variational calculation. Working out the derivatives in the middle term,

$$\begin{aligned} \delta\left(g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu\right) &= \delta x^\sigma(\partial_\sigma g_{\mu\nu})\dot{x}^\mu\dot{x}^\nu - 2\delta x^\nu\left(\dot{x}^\lambda(\partial_\lambda g_{\mu\nu})\dot{x}^\mu + g_{\mu\nu}\ddot{x}^\mu\right) + \frac{\mathrm{d}}{\mathrm{d}\tau}\left(2g_{\mu\nu}\dot{x}^\mu\delta x^\nu\right) \\ &= -\delta x^\sigma\left[g_{\sigma\mu}\ddot{x}^\mu + (\partial_\lambda g_{\sigma\mu})\dot{x}^\mu\dot{x}^\lambda\right] - \delta x^\sigma\left[-\frac{1}{2}(\partial_\sigma g_{\mu\nu})\dot{x}^\mu\dot{x}^\nu\right] + \frac{\mathrm{d}}{\mathrm{d}\tau}\left(2g_{\mu\nu}\dot{x}^\mu\delta x^\nu\right) \\ &- \delta x^\sigma\left[g_{\sigma\kappa}\ddot{x}^\kappa + \frac{1}{2}(\partial_\mu g_{\sigma\nu})\dot{x}^\mu\dot{x}^\nu + \frac{1}{2}(\partial_\nu g_{\sigma\mu})\dot{x}^\mu\dot{x}^\nu - \frac{1}{2}(\partial_\sigma g_{\mu\nu})\dot{x}^\mu\dot{x}^\nu\right] \\ &\qquad + \frac{\mathrm{d}}{\mathrm{d}\tau}\left(2g_{\mu\nu}\dot{x}^\mu\delta x^\nu\right) \\ &- \delta x^\sigma\left[g_{\sigma\kappa}\ddot{x}^\kappa + \frac{1}{2}\delta^\lambda_\sigma\left(\partial_\mu g_{\lambda\nu} + \partial_\nu g_{\lambda\mu} - \partial_\lambda g_{\mu\nu}\right)\dot{x}^\mu\dot{x}^\nu\right] + \frac{\mathrm{d}}{\mathrm{d}\tau}\left(2g_{\mu\nu}\dot{x}^\mu\delta x^\nu\right). \end{aligned}$$

Next we use $g_{\sigma\kappa}g^{\kappa\lambda} = \delta^\lambda_\sigma$ so $g_{\sigma\kappa}$ can be factorised out of the square bracket

$$\delta\left(g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu\right) = -\delta x^\sigma g_{\sigma\kappa}\left[\ddot{x}^\kappa + \frac{1}{2}g^{\kappa\lambda}\left(\partial_\mu g_{\lambda\nu} + \partial_\nu g_{\lambda\mu} - \partial_\lambda g_{\mu\nu}\right)\dot{x}^\mu\dot{x}^\nu\right] + \frac{\mathrm{d}}{\mathrm{d}\tau}\left(2g_{\mu\nu}\dot{x}^\mu\delta x^\nu\right). \tag{3.16}$$

We recognise the formula for the Christoffel symbol, so we obtain

$$\delta\left(g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu\right) = -\delta x^\sigma g_{\sigma\kappa}\left[\ddot{x}^\kappa + \Gamma^\kappa_{\mu\nu}\dot{x}^\mu\dot{x}^\nu\right] + \frac{\mathrm{d}}{\mathrm{d}\tau}\left(2g_{\mu\nu}\dot{x}^\mu\delta x^\nu\right). \tag{3.17}$$

Therefore the variation of the action is

$$\begin{aligned}\delta S &= \frac{m}{2}\int_{\tau_1}^{\tau_2}\mathrm{d}\tau\left[-\delta x^\sigma g_{\sigma\kappa}\left(\ddot{x}^\kappa + \Gamma^\kappa_{\mu\nu}\dot{x}^\mu\dot{x}^\nu\right) + \frac{\mathrm{d}}{\mathrm{d}\tau}\left(2g_{\mu\nu}\dot{x}^\mu\delta x^\nu\right).\right] \\ &= -\frac{m}{2}\int_{\tau_1}^{\tau_2}\mathrm{d}\tau\delta x^\sigma g_{\sigma\kappa}\left(\ddot{x}^\kappa + \Gamma^\kappa_{\mu\nu}\dot{x}^\mu\dot{x}^\nu\right) + m\left[g_{\mu\nu}\dot{x}^\mu\delta x^\nu\right]^{\tau_2}_{\tau_1}.\end{aligned}$$

We assume $\delta x^\nu = 0$ at the endpoints, so the boundary term is zero, $\left[g_{\mu\nu}\dot{x}^\mu\delta x^\nu\right]^{\tau_2}_{\tau_1} = 0$ and we end up with

$$\delta S = -\frac{m}{2}\int_{\tau_1}^{\tau_2}\mathrm{d}\tau\delta x^\sigma g_{\sigma\kappa}\,\left(\ddot{x}^\kappa + \Gamma^\kappa_{\mu\nu}\dot{x}^\mu\dot{x}^\nu\right) \tag{3.18}$$

To have the extremum $\delta S = 0$ for arbitrary δx^σ, the integrand must be zero and therefore the trajectory satisfies

$$\ddot{x}^\kappa + \Gamma^\kappa_{\mu\nu}\dot{x}^\mu\dot{x}^\nu = 0. \tag{3.19}$$

We simply have re-derived the geodesic equation using an action principle.

Momenta and energy of particles

Here we are interested in calculating physical quantities of particles. Therefore we consider time-like and null geodesics only.

Time-like particles. In Chapter 1, we (operationally) defined the four-momentum of a time-like particle of mass m to be $p^\mu = mu^\mu = m\dot{x}^\mu$. With the Lagrangian

$$\mathcal{L} = \frac{1}{2}mg_{\mu\nu}\dot{x}^\mu\dot{x}^\nu, \tag{3.20}$$

we can do this more rigorously, as the *canonical momentum* in any Lagrangian theory is

$$p_\mu = \frac{\partial \mathcal{L}}{\partial \dot{x}^\mu} = m g_{\mu\nu} \dot{x}^\mu. \tag{3.21}$$

How does an observer measure the momentum and energy of this particle?

First, an observer is also a participant in spacetime, having their own 4-velocity V^μ. We define the energy measured by this observer to be

$$E = -V^\mu p_\mu. \tag{3.22}$$

This is a scalar quantity that is independent of choice of coordinates, which is a suitable quantity for energy.

Photons. Photons or other massless particles travel along null geodesics, where $g_{\mu\nu} u^\mu u^\nu = 0$. In this case no notion of proper time can be defined, and $u^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda} = \dot{x}^\mu$ for some arbitrary parameter λ. Then, the trajectory of the photon is given by the Lagrangian

$$\mathcal{L} = \frac{1}{2} \alpha g_{\mu\nu} u^\mu u^\nu, \tag{3.23}$$

where α is an arbitrary constant chosen to ensure that the Lagrangian has the correct units of (energy) = (mass) (length)^2 (time)$^{-2}$. Clearly α is not the mass, since photons are massless. It's just an arbitrary constant. The canonical momentum is

$$p_\mu = \alpha g_{\mu\nu} u^\mu. \tag{3.24}$$

Note that

$$g^{\mu\nu} p_\mu p_\nu = \alpha^2 g_{\mu\nu} u^\mu u^\nu = 0, \tag{3.25}$$

since the trajectory has null tangent vector.

The energy of a photon measured by an observer with 4-velocity V^μ is again defined to be

$$E = -V^\mu p_\mu. \tag{3.26}$$

Example 3.2.1. Photons in Minkowski space. In Minkowski space, the Lagrangian for a photon is

$$\mathcal{L} = \frac{1}{2}\alpha\left(-c^2\dot{t}^2 + \dot{x}^2 + \dot{y}^2 + \dot{z}^2\right),$$

where dots here are derivatives with respect to λ. The canonical momenta are

$$p_t = -\alpha c\dot{t}, \quad p_x = \alpha\dot{x}, \quad p_y = \alpha\dot{y}, \quad p_z = \alpha\dot{z}.$$

Consider again Alice, who in her own coordinate system has velocity $V^\mu = (1,0,0,0)$. She measures the energy of the photon to be

$$E = -V^\mu p_\mu = -c\left(-\alpha c\dot{t}\right) = \alpha c^2\dot{t} = -\frac{p_t}{c}.$$

Therefore we find $p_t = -\frac{E}{c}$. Raising the index, $p^t = \frac{E}{c}$. (Without minus sign.) What are the interpretations of p_x, p_y, and p_z? These are the momentum of the photon *particle* that is consistent with quantum mechanics. To see this, note that

$$\begin{aligned} g^{\mu\nu}p_\mu p_\nu = -p_t^2 + p_x^2 + p_y^2 + p_z^2 = 0 \\ -\frac{E^2}{c^2} + p_x^2 + p_y^2 + p_z^2 = 0, \end{aligned}$$

or

$$\frac{E^2}{c^2} = p_x^2 + p_y^2 + p_z^2.$$

Now, according to quantum mechanics, $E = \hbar\omega$ (Planck's formula), and $p_i = \hbar k_i$, where k_i is the wave-vector component in the i-th direction (de Broglie's wave-particle duality). The above equation is now

$$\frac{\omega^2}{c^2} = k_x^2 + k_y^2 + k_z^2.$$

The magnitude of the wave-vector is related to the wavelength by $k = \frac{2\pi}{\lambda}$. Therefore taking the square root of the above equation, we get

$$\begin{aligned} \frac{\omega}{c} = \frac{2\pi f}{c} = \pm\frac{2\pi}{\lambda}. \\ c = f\lambda, \end{aligned}$$

recovering the familiar formula for electromagnetic waves. Therefore, with the identification $E = -V^\mu p_\mu$ and $p_i = \hbar k_i$, we recover results that are simultaneously consistent with quantum mechanics and electrodynamics.

Geodesics as the 'path of shortest/longest length'

We now show that geodesics on some manifold $\mathcal{M}$ can be interpreted as 'a path of least/longest distance'. We already know that geodesics on $\mathbb{R}^2$ and $\mathbb{R}^3$ are straight lines, and indeed those are paths connecting any two points with the shortest distance.

For a manifold M, the distance measure is the integral

$$I = \beta \int \mathrm{d}s = \beta \int \sqrt{|g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu|} = \beta \int \mathrm{d}\tau \sqrt{|g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu|}, \tag{3.27}$$

where β is some arbitrary constant and $\dot{x}^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau}$. Note that in pseudo-Riemannian manifolds, $g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu$ may possibly be negative. In order to avoid complex numbers for the interval, we take the absolute value

$$|g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu| = \epsilon g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu, \quad \epsilon = \begin{cases} -1 & \text{if } g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu < 0 \\ +1 & \text{if } g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu > 0. \end{cases}$$

To extremise[2] the curve, we apply the variational principle

$$x^\mu(\tau) \to x^\mu(\tau) + \delta x^\mu,$$

the integral varies as

$$\delta I = \beta\delta \int \mathrm{d}\tau \sqrt{g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu} = \beta \int \mathrm{d}\tau \, \frac{\epsilon\delta\left(g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu\right)}{2\sqrt{|g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu|}}.$$

Now, if we consider geodesics with normalised tangent vectors, $|g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu| = 1$ along the varied paths. Therefore,

$$\delta I = \frac{1}{2}\epsilon\beta \int \mathrm{d}\tau \; \delta\left(g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu\right)$$

This is the same as Eq. (3.15), up to an overall constant. Therefore $\delta S = 0$ in Eq. (3.15) leads to the same requirements to $\delta I = 0$. And the geodesic equation is again recovered.

[2]It's either *maximising* or *minimising*, the distinction doesn't matter because β is arbitrary. Whether β is positive or negative changes the distinction.

3.3 The curvature tensors

In Chapter 2 we have defined the covariant derivative of a vector V^μ as

$$\nabla_\mu V^\nu = \partial_\mu V^\nu + \Gamma^\nu_{\mu\lambda} V^\lambda,$$

where $\Gamma^\nu_{\mu\lambda} = \frac{1}{2} g^{\nu\sigma} \left(\partial_\mu g_{\sigma\lambda} + \partial_\lambda g_{\sigma\mu} - \partial_\sigma g_{\mu\lambda} \right)$. Loosely speaking, we can imagine this quantity the 'rate of change' of the vector V^ν in the direction μ, like in Fig. 3.4.

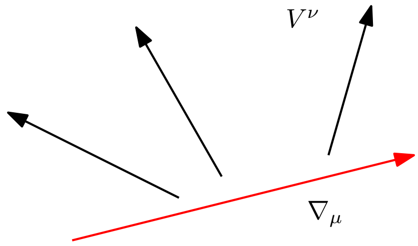


Figure 3.4: A crude 'visualisation' of $\nabla_\mu V^\mu$ as the change of a vector V^ν in an infinitesimal displacement ∇_μ.

Now, Fig. 3.4 can be seen as an infinitesimal version of parallel displacement in the direction specified by ∇_μ. At the start of this chapter, we argued that on curved manifolds, parallel transporting a vector over a closed loop may result in a final vector different from the initial one. (The right panel of Fig. 3.1.)

The Riemann tensor

Let us now take this concept as a defining feature of curvature. In particular, consider a modified version of the parallel transport procedure depicted in Fig. 3.5. Where the closed loop takes the shape of an infinitesimla trapezium, whose sides are directions defined by ∇_μ and ∇_ν.

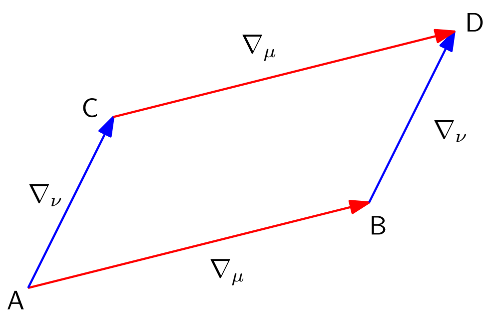


Figure 3.5: 'Parallel transporting a vector along an infinitesimal trapezium.

Consider parallel transporting an arbitrary vector V^σ along two paths. First is by path ACD. The 'infinitesimal change' of the vector is

$$\begin{aligned}\nabla_\mu\nabla_\nu V^\rho &= \partial_\mu \nabla_\nu V^\rho + \Gamma^\rho_{\mu\lambda}\nabla_\nu V^\lambda - \Gamma^\lambda_{\mu\nu}\nabla_\lambda V^\rho \\ &= \partial_\mu\left(\partial_\nu V^\rho + \Gamma^\rho_{\nu\sigma}V^\sigma\right) + \Gamma^\rho_{\mu\lambda}\left(\partial_\nu V^\lambda + \Gamma^\lambda_{\nu\sigma}V^\sigma\right) - \Gamma^\lambda_{\mu\nu}\nabla_\lambda V^\rho \\ &= \partial_\mu\partial_\nu V^\rho + \left(\partial_\mu \Gamma^\rho_{\nu\sigma}\right)V^\sigma + \Gamma^\rho_{\nu\sigma}\partial_\mu V^\sigma + \Gamma^\rho_{\mu\lambda}\partial_\nu V^\lambda + \Gamma^\rho_{\mu\lambda}\Gamma^\lambda_{\nu\sigma}V^\sigma - \Gamma^\lambda_{\mu\nu}\nabla_\lambda V^\rho.\end{aligned}$$

Alternatively, one could also take the path ABD, and the change in vector can be calculated similarly as above, just by interchanging $\mu \leftrightarrow \nu$:

$$\nabla_\nu\nabla_\mu V^\rho = \partial_\nu\partial_\mu V^\rho + \left(\partial_\nu\Gamma^\rho_{\mu\sigma}\right)V^\sigma + \Gamma^\rho_{\mu\sigma}\partial_\nu V^\sigma + \Gamma^\rho_{\nu\lambda}\partial_\mu V^\lambda + \Gamma^\rho_{\nu\lambda}\Gamma^\lambda_{\mu\rho}V^\rho.$$

Now, what is the difference in the final results? In particular, we calculate

$$\begin{aligned}\nabla_\mu\nabla_\nu V^\rho - \nabla_\nu\nabla_\mu V^\rho &= \left(\partial_\mu\Gamma^\rho_{\nu\sigma} - \partial_\nu\Gamma^\rho_{\mu\sigma} + \Gamma^\rho_{\mu\lambda}\Gamma^\lambda_{\nu\sigma} - \Gamma^\rho_{\nu\lambda}\Gamma^\lambda_{\mu\sigma}\right)V^\sigma \\ &= R^\rho{}_{\sigma\mu\nu}V^\sigma,\end{aligned}$$

where we have defined the *Riemann tensor*

$$\boxed{R^\rho{}_{\sigma\mu\nu} = \partial_\mu\Gamma^\rho_{\nu\sigma} - \partial_\nu\Gamma^\rho_{\mu\sigma} + \Gamma^\rho_{\mu\lambda}\Gamma^\lambda_{\nu\sigma} - \Gamma^\rho_{\nu\lambda}\Gamma^\lambda_{\mu\sigma}.} \tag{3.28}$$

This is a $(1,3)$-tensor that quantifies the curvature of a given manifold. For flat spacetimes, all Γ's are zero, hence all components of the Riemann tensor are zero so $\nabla_\mu\nabla_\nu V^\rho - \nabla_\nu\nabla_\mu V^\rho = 0$, which is indeed true for flat spacetime.

Example 3.3.1. Riemann tensor of S^2. On S^2 where the metric is

$$\mathrm{d}s^2 = \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2,$$

we have calculated the Christoffel symbols in Chapter 2. The only non-zero ones are

$$\Gamma^\phi_{\theta\phi} = \frac{\cos\theta}{\sin\theta} = \Gamma^\phi_{\phi\theta}, \quad \Gamma^\theta_{\phi\phi} = -\sin\theta\cos\theta.$$

We can calculate the Riemann tensor systematically as follows. Not that $R^\rho{}_{\sigma\mu\nu}$ is anti-symmetric in μ and ν. Therefore, keeping ρ and σ arbitrary for now, we only need to compute

$$R^\rho{}_{\sigma\theta\phi} = -R^\rho{}_{\sigma\phi\theta}.$$

Applying the formula,

$$R^\rho{}_{\sigma\theta\phi} = \partial_\theta \Gamma^\rho_{\phi\sigma} - \underbrace{\partial_\phi \Gamma^\rho_{\theta\sigma}}_{=0} + \Gamma^\rho_{\theta\lambda}\Gamma^\lambda{}_{\phi\sigma} - \Gamma^\rho_{\phi\lambda}\Gamma^\lambda{}_{\theta\sigma}.$$

The term $\partial_\phi \Gamma^\rho_{\theta\sigma}$ vanishes that none of the Christoffel symbols depend on ϕ. The last two terms involving a sum over $\lambda = \theta, \phi$:

$$R^\rho{}_{\sigma\theta\phi} = \partial_\theta \Gamma^\rho_{\phi\sigma} + \underbrace{\Gamma^\rho_{\theta\theta}\Gamma^\theta{}_{\phi\sigma}}_{=0} + \Gamma^\rho_{\theta\phi}\Gamma^\phi{}_{\phi\sigma} - \Gamma^\rho_{\phi\theta}\Gamma^\theta{}_{\theta\sigma} - \Gamma^\rho_{\phi\phi}\Gamma^\phi{}_{\theta\sigma}.$$

The term $\Gamma^\rho_{\theta\theta}\Gamma^\theta{}_{\phi\sigma}$ equals zero because there are no non-zero Christoffels where both lower indices are θ.
We now calculate explicit entries for ρ and σ:

$$R^\theta{}_{\theta\theta\phi} = \partial_\theta \Gamma^\theta_{\phi\theta} + \Gamma^\theta_{\theta\phi}\Gamma^\phi_{\phi\theta} + \Gamma^\rho_{\theta\phi}\Gamma^\phi_{\phi\sigma} - \Gamma^\theta_{\phi\theta}\Gamma^\theta_{\theta\theta} - \Gamma^\theta_{\phi\phi}\Gamma^\phi_{\theta\theta} = 0$$

because each term has at least a zero-Christoffel symbol. Similarly, one can proceed to get $R^\phi{}_{\phi\theta\phi} = 0$ in the same way. As for

$$\begin{aligned} R^\theta{}_{\phi\theta\phi} &= \partial_\theta \Gamma^\theta_{\phi\phi} + \Gamma^\theta_{\theta\phi}\Gamma^\phi_{\phi\phi} - \Gamma^\theta_{\phi\phi}\Gamma^\phi_{\theta\phi} \\ &= \partial_\theta \Gamma^\theta_{\phi\phi} - \Gamma^\theta_{\phi\phi}\Gamma^\phi_{\theta\phi} \\ &= \partial_\theta\left(-\sin\theta\cos\theta\right) - \left(-\sin\theta\cos\theta\right)\frac{\cos\theta}{\sin\theta} \\ &= \sin^2\theta = -R^\theta{}_{\phi\phi\theta}. \end{aligned}$$

Similarly one can find that

$$R^\phi{}_{\theta\theta\phi} = \partial_\theta \Gamma^\phi_{\phi\theta} + \Gamma^\phi_{\theta\phi}\Gamma^\phi_{\phi\theta} = \partial_\theta\left(\frac{\cos\theta}{\sin\theta}\right) + \frac{\cos^2\theta}{\sin^2\theta} = -1 = -R^\phi_{\theta\phi\theta}.$$

Collecting our results, the non-zero Riemann tensor components for unit S^2 are

$$R^\theta{}_{\phi\theta\phi} = \sin^2\theta = -R^\theta{}_{\phi\phi\theta} \quad R^\phi{}_{\theta\phi\theta} = 1 = -R^\phi{}_{\theta\theta\phi}. \tag{3.29}$$

Properties of the Riemann tensor. First, it is 'nicer' to have a Riemann tensor with all indices lowered,

$$R_{\rho\sigma\mu\nu} = g_{\rho\lambda}R^\lambda{}_{\sigma\mu\nu}. \tag{3.30}$$

This all lower index version has the following symmetry properties.

- Antisymmetric in its first two and also last two indices (separately)

$$R_{\rho\sigma\mu\nu} = -R_{\sigma\rho\mu\nu}, \quad R_{\rho\sigma\mu\nu} = -R_{\rho\sigma\nu\mu}. \tag{3.31}$$

- Symmetric upon interchange of the first pair with the second pair (while keeping the order within each pair fixed):

$$R_{\rho\sigma\mu\nu} = R_{\mu\nu\rho\sigma}. \tag{3.32}$$

- Sum of the cyclic permutation of the last three indices equal zero:

$$R_{\rho\sigma\mu\nu} + R_{\rho\mu\nu\sigma} + R_{\rho\nu\sigma\mu} = 0. \tag{3.33}$$

- *Bianchi identity*,

$$\nabla_\lambda R_{\rho\sigma\mu\nu} + \nabla_\rho R_{\sigma\lambda\mu\nu} + \nabla_\sigma R_{\lambda\rho\mu\nu} = 0. \tag{3.34}$$

The Ricci tensor and scalar

The *Ricci tensor* is simply a contraction of the Riemann tensor,

$$\boxed{R_{\sigma\nu} = R^\mu{}_{\sigma\mu\nu} = \partial_\mu \Gamma^\mu_{\nu\sigma} - \partial_\nu \Gamma^\mu_{\mu\sigma} + \Gamma^\mu_{\mu\lambda}\Gamma^\lambda_{\nu\sigma} - \Gamma^\rho_{\nu\lambda}\Gamma^\lambda_{\mu\sigma}.} \tag{3.35}$$

On the right hand side of the formula, there is a double summation involving μ and λ.

One useful property of the Ricci tensor is that it is symmetric,

$$R_{\mu\nu} = R_{\nu\mu}. \tag{3.36}$$

From the Ricci tensor, one can construct the *Ricci scalar*.

$$\boxed{R = g^{\mu\nu} R_{\mu\nu}} \tag{3.37}$$

Example 3.3.2. Ricci tensor and scalar of S^2. For the two-sphere $\mathrm{d}s^2 = \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2$, we have calculated the Riemann tensor components in an earlier example. The result was Eq. (3.29). We now contract to obtain the Ricci tensor:

$$R_{\theta\theta} = R^\lambda{}_{\theta\lambda\theta} = R^\theta{}_{\theta\theta\theta} + R^\phi{}_{\theta\phi\theta} = 0 + 1 = 1,$$
$$R_{\phi\phi} = R^\lambda{}_{\phi\lambda\phi} = R^\theta{}_{\phi\theta\phi} + R^\phi{}_{\phi\phi\phi} = \sin^2\theta + 0 = \sin^2\theta.$$

Similarly, we find $R_{\theta\phi} = R_{\phi\theta} = 0$. The non-zero components are

$$R_{\theta\theta} = 1, \quad R_{\phi\phi} = \sin^2\theta.$$

Further contracting to get the Ricci scalar,

$$\begin{aligned} R = g^{\mu\nu}R_{\mu\nu} &= g^{\theta\theta}R_{\theta\theta} + g^{\phi\phi}R_{\phi\phi} + (\text{other terms zero}) \\ &= 2. \end{aligned}$$

Maximally-symmetric spaces

A *maximally-symmetric* space or spacetime is defined as a manifold with the largest possible number of *symmetries*. We have not discussed the concept of symmetries in these lectures yet. But for many useful purposes we need to the following formula. If a d-dimensional manifold with metric $g_{\mu\nu}$ is maximally symmetric, then its Riemann tensor obeys

$$R_{\rho\sigma\mu\nu} = K\left(g_{\rho\mu}g_{\sigma\nu} - g_{\rho\nu}g_{\sigma\mu}\right), \tag{3.38}$$

where K is some constant. We further contract to get the Ricci tensor

$$R_{\sigma\nu} = g^{\rho\mu}R_{\rho\sigma\mu\nu} = Kg^{\rho\mu}\left(g_{\rho\mu}g_{\sigma\nu} - g_{\rho\nu}g_{\sigma\mu}\right) = K\left(dg_{\sigma\nu} - \delta^\mu_\nu g_{\sigma\mu}\right) = K(d-1)g_{\sigma\nu}. \tag{3.39}$$

A space(time) where its Ricci tensor is proportional to the metric,

$$R_{\mu\nu} = (\text{const.})g_{\mu\nu} = K(d-1)g_{\mu\nu} \tag{3.40}$$

is called an *Einstein space* or *Einstein spacetime*. If $K = 0$, then

$$R_{\mu\nu} = 0. \tag{3.41}$$

This is called a *Ricci-flat* space(time).

When K is non-zero, we define $|K| = \frac{1}{\ell^2}$, where ℓ is the *curvature radius/curvature scale* of the Einstein space. Therefore

$$R_{\mu\nu} = \pm\frac{d-1}{\ell^2}g_{\mu\nu} = \begin{cases} \frac{d-1}{\ell^2}g_{\mu\nu}, & \text{positive curvature,} \\ -\frac{d-1}{\ell^2}g_{\mu\nu}, & \text{negative curvature.} \end{cases} \tag{3.42}$$

In the limit of infinite radius, $\ell \to \infty$, we recover the Ricci-flat space. Important spacetimes that have this property is the *de Sitter and anti-de Sitter spacetimes*

Example 3.3.3. The 2-sphere S^2 is an Einstein space. In the previous example, calculated the Ricci tensor for S^2

$$\mathrm{d}s^2 = \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2.$$

The non-zero comonents of the Ricci tensor are $R_{\theta\theta} = 1$ and $R_{\phi\phi} = \sin^2\theta$. Which are identical to the metric tensor $g_{\theta\theta} = 1$ and $g_{\phi\phi} = \sin^2\theta$. Therefore a 2-sphere is a maximally-symmetric Einstein space with

$$R_{ab} = g_{ab}.$$

Therefore $K = 1$ for the unit S^2.

Example 3.3.4. The 2-sphere of radius a. Consider the sphere where the radius is a, not neccessarily equal to 1. The metric is

$$\mathrm{d}s^2 = a^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right).$$

The non-zero Christoffel symbols can be calculated accordingly as

$$\Gamma^{\phi}_{\theta\phi} = \frac{\cos\theta}{\sin\theta}, \quad \Gamma^{\theta}_{\phi\phi} = -\sin\theta\cos\theta.$$

It is often the case where we just need the Ricci tensor and not the Riemann tensor. So we can immediately compute the contraction

$$R_{\sigma\nu} = R^{\mu}{}_{\sigma\mu\nu} = \partial_\mu\Gamma^{\mu}_{\nu\sigma} - \partial_\nu\Gamma^{\mu}_{\mu\sigma} + \Gamma^{\mu}_{\mu\lambda}\Gamma^{\lambda}_{\nu\sigma} - \Gamma^{\mu}_{\nu\lambda}\Gamma^{\lambda}_{\mu\sigma}.$$

The right hand side is double summation over λ and μ. First consider the component

$$R_{\theta\theta} = \partial_\mu\Gamma^{\mu}_{\theta\theta} - \partial_\theta\Gamma^{\mu}_{\mu\theta} + \Gamma^{\mu}_{\mu\lambda}\Gamma^{\lambda}_{\theta\theta} - \Gamma^{\mu}_{\theta\lambda}\Gamma^{\lambda}_{\mu\theta}.$$

Writing out the summation, the only non-zero Christoffel symbols are

$$R_{\theta\theta} = -\partial_\theta\Gamma^{\phi}_{\phi\theta} - \Gamma^{\phi}_{\theta\phi}\Gamma^{\phi}_{\theta\phi} = -\partial_\theta\left(\frac{\cos\theta}{\sin\theta}\right) - \frac{\cos^2\theta}{\sin^2\theta} = 1.$$

Proceeding similarly, we find

$$R_{\phi\phi} = \sin^2\theta.$$

Here, note that

$$R_{\mu\nu} = \frac{1}{a^2} g_{\mu\nu}.$$

Comparing with the equation $R_{\mu\nu} = K(d-1)g_{\mu\nu}$ for general Einstein spaces of $d = 2$, we see that

$$K = \frac{1}{a^2}.$$

This gives the interpretation of the constant K as the inverse curvature squared. One can also check that the Riemann tensor for this case is

$$R^{\rho}{}_{\sigma\mu\nu} = \frac{1}{a^2}\left(\delta^{\rho}_{\mu} g_{\sigma\nu} - \delta^{\rho}_{\nu} g_{\sigma\mu}\right). \tag{3.43}$$

Other curvature tensors

In General Relativity, the Riemann and Ricci tensor will be the most relevant curvature quantities. However, sometimes other types of tensors are considered. We will not use these very often in this course, but we should at least mention them as they are used in certain areas of GR research.

Weyl tensor.[3] Loosely speaking, this is the 'trace-free' part of the Riemann tensor. It is defined by

$$\begin{aligned} C_{\rho\sigma\mu\nu} = R_{\rho\sigma\mu\nu} &+ \frac{1}{n-2}\left(R_{\rho\nu} g_{\sigma\mu} - R_{\rho\mu} g_{\sigma\nu} + R_{\sigma\mu} g_{\rho\nu} - R_{\sigma\nu} g_{\rho\mu}\right) \\ &+ \frac{1}{(n-1)(n-2)}\left(g_{\rho\mu} g_{\sigma\nu} - g_{\rho\nu} g_{\sigma\mu}\right) R. \end{aligned}$$

Schouten tensor. Defined by

$$L_{\mu\nu} = \frac{1}{n-2}\left(R_{\mu\nu} - \frac{1}{2(n-1)} R g_{\mu\nu}\right).$$

[3]Pronounced like 'Wyle', as in rhymes with 'Wild'.

Bach tensor.[4] Defined by

$$B_{\mu\nu} = L_{\rho\sigma}C_{\mu}{}^{\rho}{}_{\nu}{}^{\sigma} + \nabla^{\lambda}\nabla_{\lambda}L_{\mu\nu} - \nabla^{\lambda}\nabla_{\mu}L_{\nu\lambda}.$$

Cotton tensor. Defined by

$$\mathcal{C}_{\mu\nu\rho} = \nabla_{\rho}R_{\mu\nu} - \nabla_{\nu}R_{\mu\rho} + \frac{1}{2(n-1)}\left(\nabla_{\nu}Rg_{\mu\rho} - \nabla_{\rho}Rg_{\mu\nu}\right).$$

Kretschmann invariant. Defined by

$$\mathcal{K} = R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma}.$$

This is called an 'invariant' because all indices have been contracted, hence it is a scalar. Its value will be independent of choice of coordinate system.

3.4 Rindler spacetime

The first postulate of Special Relativity states that *all laws of physics are the same in all inertial frames.* The idea of inertial frames is important. In particular it distinguishes the correct coordinate system to use for the Twin Paradox. In this section, we will show that even in the absence of gravity, an accelerating observer will not percieve Minkowski spacetime.

Uniformly accelerated observers

Recall that the Minkowski metric is

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2.$$

In this spacetime, suppose that we have a particle that is moving along a curve

$$x^{\mu}(\tau) = \left(\frac{1}{\beta}\sinh\beta c\tau, \frac{1}{\beta}\cosh\beta c\tau, 0, 0\right), \tag{3.44}$$

[4]Pronounced like Arnold Schwarzernegger saying 'I'll be back'.

where τ is the particle's proper time. The 4-velocity and 4-acceleration are

$$u^\mu = \dot{x}^\mu = (c\cosh\beta\tau, c\sinh\beta\tau, 0, 0)\,,$$
$$a^\mu = \ddot{x}^\mu = \left(\beta c^2\sinh\beta c\tau, \beta c^2\cosh\beta c\tau, 0, 0\right). \tag{3.45}$$

We can check that

$$u^\mu u_\mu = \eta_{\mu\nu}u^\mu u^\nu = -c^2, \quad a_\mu a^\mu = \eta_{\mu\nu}a^\mu a^\nu = \beta^2 c^4.$$

In particular, this particle has constant 4-acceleration $\sqrt{a_\mu a^\mu} = \beta c^2$. In the Minkowski Cartesian coordinates (x, t), the particle is moving along a hyperbola $x^2 - c^2t^2 = \frac{1}{\beta^2}$, as shown in Fig. 3.6.

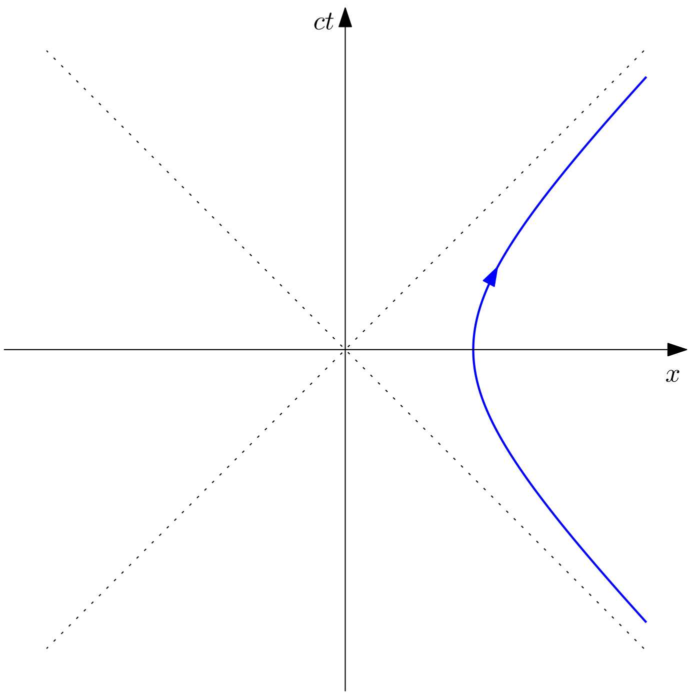


Figure 3.6: Worldline of a particle in uniform acceleration $1/\beta$.

Suppose we define coordinates adapted to uniform observers. In these new coordinates, we want a uniformly accelerating observer to have zero displacement. This suggests that we define coordinates which are constant along hyperbola,

$$x = \rho\cosh\eta, \quad ct = \rho\sinh\eta. \tag{3.46}$$

Note that this transformation is only valid in cases where $|t| < x$ and $x > 0$. Taking the differentials,

$$\mathrm{d}x = \mathrm{d}\rho\cosh\eta + \rho\sinh\eta\mathrm{d}\eta, \quad \mathrm{d}(ct) = c\mathrm{d}t = \mathrm{d}\rho\sinh\eta + \rho\cosh\eta\mathrm{d}\eta.$$

Substituting into the Minkowski metric, we get

$$\begin{aligned} \mathrm{d}s^2 &= -c^2\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 \\ &= -\rho^2\mathrm{d}\eta^2 + \mathrm{d}\rho^2 + \mathrm{d}y^2 + \mathrm{d}z^2. \end{aligned}$$

This is the *Rindler metric*. This is not a curved spacetime, as it is simply a coordinate transformation of the (flat) Minkowski metric. The new coordinate are (η, ρ, y, z). The grid of constant η and ρ are shown in Fig. 3.7.

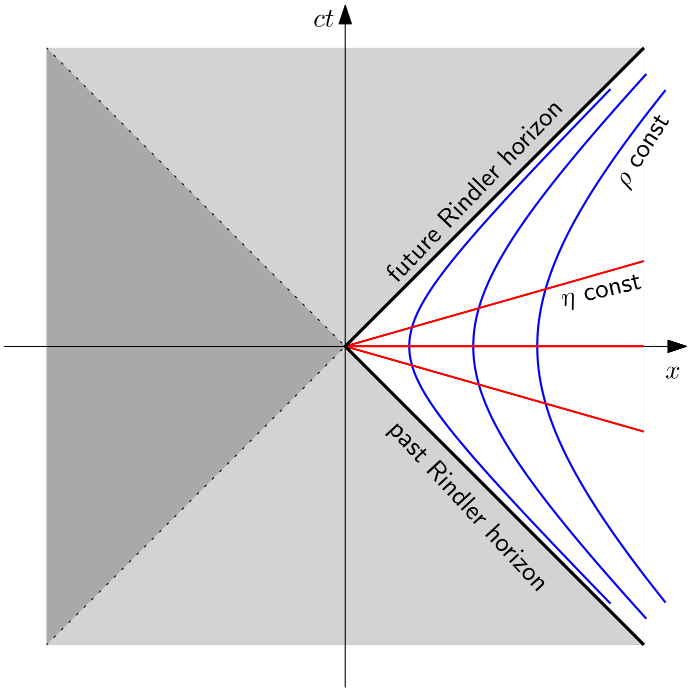


Figure 3.7

Observers lying on ρ =const are called *uniformly accelerating observers*. In the limit $\rho \to 0$, the hyperbola converges to a triangular shape that is a null path. Also in this limit we have $-\rho^2\mathrm{d}\eta^2 \to 0$. This is a *singularity* of the metric, because apparently the evolution of the time coordinate becomes undefined. This location is called the *future and past Rindler horizons*, as indicated in Fig. 3.7. The white, unshaded region where ρ and η are defined is called the *right Rindler wedge*.

Roughly speaking, the significance of Rindler horizons is related to the fact that no signals can travel faster than light. Any observer in the Rindler can receive information from beyond the future Rindler horizon. And no observer in the Rindler wedge can send information beyond the past Rindler horizon.

We will speak more of horizons when we discuss black holes. The lesson here is that, even in flat spacetime, there can be horizons. In this case, there are events in Minkowski space that are not accessible to an accelerating observer. This is because if an observer is accelerating, signals from far away travelling at maximum (light) speed cannot catch up to it.

Geodesics of the Rindler spacetime

Let us study the motion of particles in Rindler spacetime. As usual there are two ways to do this. The first way is by using the Lagrangian

$$\mathcal{L} = \frac{1}{2}k\left(-\rho^2\dot{\eta}^2 + \dot{\rho}^2 + \dot{y}^2 + \dot{z}^2\right),$$

where $k = m$ for time-like particles of mass m, or arbitrary $k = \alpha$ for massless particles (photons) moving along null geodesics. The Euler–Lagrange equations give

$$\ddot{\eta} = -\frac{2\dot{\rho}\dot{\eta}}{\rho}, \quad \ddot{\rho} = -\rho\dot{\eta}^2, \quad \ddot{y} = 0, \quad \ddot{z} = 0. \tag{3.47}$$

Note that for time-like or null particles, the normalisation condition $g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu = \epsilon = -1, 0$ means that $\dot{\eta}$ cannot be zero. (Otherwise we cannot ensure the RHS of this equation to be equal to -1 or 0.) The second equation of (3.47) means that ρ cannot be a constant.

Now, recall that the Rindler spacetime is simply a coordinate transformation of Minkowski space. We do know that geodesics in Minkowski space are straight lines, obeying

$$\ddot{t} = 0, \quad \ddot{x} = 0, \quad \ddot{y} = 0, \quad \ddot{z} = 0.$$

We need to check if this is consistent with what we got previously. First, note that differentiating Eq. (3.46) with respect to τ gives

$$\begin{aligned}\dot{x} &= \dot{\rho}\cosh\eta + \rho\sinh\eta\dot{\eta},\\ \ddot{x} &= \left(\ddot{\rho} + \rho\dot{\eta}^2\right)\cosh\eta + (2\dot{\rho}\dot{\eta} + \rho\ddot{\eta})\sinh\eta,\end{aligned}$$

and

$$\begin{aligned}\dot{t} &= \dot{\rho}\sinh\eta + \rho\cosh\eta\dot{\eta},\\ \ddot{t} &= \left(\ddot{\rho} + \rho\dot{\eta}^2\right)\sinh\eta + (2\dot{\rho}\dot{\eta} + \rho\ddot{\eta})\cosh\eta.\end{aligned}$$

For Minkowski geodesics, $\ddot{x} = 0$ and $\ddot{t} = 0$. Therefore the above equations lead to

$$\left(\ddot{\rho} + \rho\dot{\eta}^2\right)\cosh\eta = -\left(2\dot{\rho}\dot{\eta} + \rho\ddot{\eta}\right)\sinh\eta, \tag{3.48a}$$
$$\left(\ddot{\rho} + \rho\dot{\eta}^2\right)\sinh\eta = -\left(2\dot{\eta}\dot{\rho} + \rho\ddot{\eta}\right)\cosh\eta. \tag{3.48b}$$

These are actually equivalent to Eq. (3.47). To see this, rearrange Eq. (3.48b),

$$2\dot{\eta}\dot{\rho} + \rho\ddot{\eta} = -\left(\ddot{\rho} + \rho\dot{\eta}^2\right)\frac{\sinh\eta}{\cosh\eta}.$$

Then substitute Eq. (3.49) into (3.48a),

$$\begin{aligned}
\left(\ddot{\rho} + \rho\dot{\eta}^2\right)\cosh\eta &= -\left[-\left(\ddot{\rho} + \rho\dot{\eta}^2\right)\frac{\sinh\eta}{\cosh\eta}\right]\sinh\eta \\
\left(\ddot{\rho} + \rho\dot{\eta}^2\right)\cosh^2\eta &= \left(\ddot{\rho} + \rho\dot{\eta}^2\right)\sinh^2\eta \\
\left(\ddot{\rho} + \rho\dot{\eta}^2\right)\underbrace{\left(\cosh^2\eta - \sinh^2\eta\right)}_{=1} &= 0 \\
\ddot{\rho} + \rho\dot{\eta}^2 &= 0
\end{aligned}$$

The second equation $\rho\ddot{\eta} + 2\dot{\eta}\dot{\rho} = 0$ is derived in a similar way.

The same equations can be derived using $\ddot{x}^\kappa + \Gamma^\kappa_{\mu\nu}\dot{x}^\mu\dot{x}^\nu = 0$. One can check that the non-zero Christoffel symbols for the Rindler metric is

$$\Gamma^\rho_{\eta\eta} = \rho, \quad \Gamma^\eta_{\eta\rho} = \frac{1}{\rho}. \tag{3.49}$$

3.5 Curvatures of the spherically-symmetric spacetime

In the previous chapter, we have computed the Christoffel symbol for the static, spherically-symmetric metric

$$\mathrm{d}s^2 = -f\mathrm{d}t^2 + h\mathrm{d}r^2 + r^2\tilde{\gamma}_{ij}\mathrm{d}\theta^i\mathrm{d}\theta^j,$$

where $\tilde{\gamma}_{ij}\mathrm{d}\theta^i\mathrm{d}\theta^j = \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2$ is the metric on the 2-sphere S^2. For convenient reference, they are

$$\begin{gathered}
\Gamma^r_{tt} = \frac{f'}{2h}, \quad \Gamma^t_{tr} = \frac{f'}{2f}, \quad \Gamma^r_{rr} = \frac{h'}{2h}, \quad \Gamma^r_{ij} = -\frac{r}{h}\tilde{\gamma}_{ij}, \\
\Gamma^i_{rj} = \frac{1}{r}\tilde{\delta}^i_j, \quad \Gamma^k_{ij} = \tilde{\Gamma}^k_{ij},
\end{gathered} \tag{3.50}$$

Here, when we use lowercase Latin indices $i, j, k, l, m, n, \ldots$, it refers to either θ or ϕ. We now proceed to compute its Riemann and Ricci scalars.

Riemann tensor. The Riemann tensor $R_{\rho\sigma\mu\nu} = g_{\rho\lambda}R^{\lambda}{}_{\sigma\mu\nu}$ has many symmetry properties, which means that we only need to calculate 4 of them,

$$R_{trtr} = g_{tt}R^{t}{}_{rtr}, \quad R_{titj} = g_{tt}R^{t}{}_{itj}, \quad R_{rirj} = g_{rr}R^{r}{}_{irj}, \quad R_{klij} = g_{kn}R^{n}{}_{lij}.$$

The rest are either zero, or connected to the above by symmetry properties.[5]

Let us start with

$$R^{t}{}_{rtr} = \partial_r\Gamma^{t}_{rr} - \partial_r\Gamma^{t}_{tr} + \Gamma^{t}_{t\lambda}\Gamma^{\lambda}_{rr} - \Gamma^{t}_{r\lambda}\Gamma^{\lambda}_{tr}.$$

The first term is zero because $\Gamma^{t}_{rr} = 0$. The last two terms involve summations over λ. But for each summation, there is only one non-zero term:

$$\begin{aligned}
\Gamma^{t}_{t\lambda}\Gamma^{\lambda}_{rr} &= \Gamma^{t}_{tt}\Gamma^{t}_{rr} + \Gamma^{t}_{tr}\Gamma^{r}_{rr} + \Gamma^{t}_{ti}\Gamma^{i}_{rr} = 0 + \Gamma^{t}_{tr}\Gamma^{r}_{rr} + 0,\\
\Gamma^{t}_{r\lambda}\Gamma^{\lambda}_{tr} &= \Gamma^{t}_{rt}\Gamma^{t}_{tr} + \Gamma^{t}_{rr}\Gamma^{r}_{tr} + \Gamma^{t}_{ri}\Gamma^{i}_{tr} = \Gamma^{t}_{rt}\Gamma^{t}_{tr} + 0 + 0.
\end{aligned}$$

So there are only three terms,

$$\begin{aligned}
R^{t}{}_{rtr} &= 0 - \partial_r\Gamma^{t}_{tr} + \Gamma^{t}_{tr}\Gamma^{r}_{rr} - \Gamma^{t}_{rt}\Gamma^{t}_{tr}\\
&= -\partial_r\left(\frac{f'}{2f}\right) + \frac{f'}{2f}\frac{h'}{2h} - \frac{f'}{2f}\frac{f'}{2f}\\
&= -\frac{f''}{2f} + \frac{f'^2}{2f^2} + \frac{f'h'}{4fh} - \frac{f'^2}{4f^2}\\
&= -\frac{f''}{2f} + \frac{f'^2}{4f^2} + \frac{f'h'}{4fh}.
\end{aligned}$$

We can use the symmetries of the Riemann tensor to obtain $R^{r}{}_{trt}$. First, note that

$$\begin{aligned}
R_{trtr} = g_{tt}R^{t}{}_{rtr} &= -f\left(-\frac{f''}{2f} + \frac{f'^2}{4f^2} + \frac{f'h'}{4fh}\right)\\
&= \frac{f''}{2} - \frac{f'^2}{4f} - \frac{f'h'}{4h}
\end{aligned}$$

By the anti-symmetry of the first and second pair of indices,

$$R_{trtr} = -R_{rttr} = -\left(-R_{rtrt}\right) = R_{rtrt} = \frac{f''}{2} - \frac{f'^2}{4f} - \frac{f'h'}{4h}.$$

[5]For example, $R_{rtrt} = -R_{trrt} = -(-R_{trtr}) = R_{trtr}$.

Raising the first index,

$$
\begin{aligned}
R^r{}_{trt} = g^{rr} R_{rtrt} &= \frac{1}{h}\left(\frac{f''}{2} - \frac{f'^2}{4f} - \frac{f'h'}{4h}\right) \\
&= \frac{f''}{2h} - \frac{f'^2}{4fh} - \frac{f'h'}{4h^2}.
\end{aligned}
$$

Next we proceed to

$$R^t{}_{itj} = \partial_t \Gamma^t_{ji} - \partial_j \Gamma^t_{ti} + \Gamma^t_{t\lambda}\Gamma^\lambda_{ji} - \Gamma^t_{j\lambda}\Gamma^\lambda_{ti}.$$

The first two and final term is zero, because all those Christoffel symbols are zero. There is only

$$R^t{}_{itj} = \Gamma^t_{tr}\Gamma^r_{ji} = \frac{f'}{2f}\left(-\frac{r}{h}\tilde{\gamma}_{ji}\right) = -\frac{rf'}{2fh}\tilde{\gamma}_{ij}.$$

The third one is computed in the similar method. (Present the details as a tutorial.)

$$R^r{}_{irj} = \frac{rh'}{2h^2}\tilde{\gamma}_{ij},$$

The last one is slightly different. All the indices belong to either θ or ϕ. Writing them out we find

$$R^m{}_{nij} = \partial_i \Gamma^m_{jn} - \partial_j \Gamma^m_{in} + \Gamma^m_{i\lambda}\Gamma^\lambda_{jn} - \Gamma^m_{j\lambda}\Gamma^\lambda_{in}.$$

Observe that there is a sum over $\lambda = t, r, l$. When $\lambda = t$, the Christoffels there are zero. So

$$R^m{}_{nij} = \partial_i \Gamma^m_{jn} - \partial_j \Gamma^m_{in} + \Gamma^m_{ir}\Gamma^r_{jn} + \Gamma^m_{il}\Gamma^l_{jn} - \Gamma^m_{jr}\Gamma^r_{in} - \Gamma^m_{jl}\Gamma^l_{in}.$$

We put the terms involving r at the back,

$$R^m{}_{nij} = \underbrace{\partial_i \Gamma^m_{jn} - \partial_j \Gamma^m_{in} + \Gamma^m_{il}\Gamma^l_{jn} - \Gamma^m_{jl}\Gamma^l_{in}}_{\tilde{R^m}{}_{nij}} + \Gamma^m_{ir}\Gamma^r_{jn} - \Gamma^m_{jr}\Gamma^r_{in}.$$

Note that $\Gamma^k_{ij} = \tilde{\Gamma}^k_{ij}$. This means when all the indices belong to the angles θ and ϕ, the Christoffel symbols are just the Christoffel symbols for S^2, which we calculated previously. With these Christoffels, we also recognise the formula for $\tilde{R}^m_{nij}$, the Riemann tensor for the unit S^2 which was given in Eq. (3.43). Therefore

$$\begin{aligned} R^m{}_{nij} &= \delta^m_i \tilde{\gamma}_{nj} - \delta^m_j \tilde{\gamma}_{ni} + \frac{1}{r}\delta^m_i \left(-\frac{r}{h}\tilde{\gamma}_{jn}\right) - \frac{1}{r}\delta^m_j \left(-\frac{r}{h}\tilde{\gamma}_{in}\right) \\ &= \left(1 - \frac{1}{h}\right)\left(\delta^m_i \tilde{\gamma}_{nj} - \delta^m_j \tilde{\gamma}_{nj}\right). \end{aligned}$$

Collecting our results, the relevant Riemann tensor components are

$$\begin{aligned} R^t{}_{rtr} &= -\frac{f''}{2f} + \frac{f'^2}{4f^2} + \frac{f'h'}{4fh}, \\ R^r{}_{trt} &= \frac{f''}{2h} - \frac{f'^2}{4fh} - \frac{f'h'}{4h^2} \\ R^t{}_{itj} &= -\frac{rf'}{2fh}\tilde{\gamma}_{ij}, \\ R^i{}_{tjt} &= \frac{f'}{2rh}\tilde{\delta}^i_j \\ R^r{}_{irj} &= \frac{rh'}{2h^2}\tilde{\gamma}_{ij}, \\ R^i{}_{rjr} &= \frac{h'}{2rh}\tilde{\delta}^i_j, \\ R^m{}_{nij} &= \left(1 - \frac{1}{h}\right)\left(\delta^m_i \tilde{\gamma}_{nj} - \delta^m_j \tilde{\gamma}_{nj}\right). \end{aligned}$$

Our next task is to find the Ricci tensor. It is the contraction

$$R^\mu{}_{\sigma\mu\nu} = R^t{}_{\sigma t\nu} + R^r{}_{\sigma r\nu} + R^i{}_{\sigma i\nu}.$$

So, the components are

$$\begin{aligned} R_{tt} &= \frac{f''}{2h} - \frac{f'^2}{4fh} - \frac{f'h'}{4h^2} + \frac{f'}{rh}, \\ R_{rr} &= -\frac{f''}{2f} + \frac{f'^2}{4f^2} + \frac{f'h'}{4fh} + \frac{h'}{rh}, \\ R_{ij} &= \tilde{\gamma}_{ij}\left[1 - \frac{1}{h} + \frac{r}{2h}\left(\frac{h'}{h} - \frac{f'}{f}\right)\right]. \end{aligned}$$

Finally, the Kretschmann invariant of this spacetime is (taking general d dimensions)

$$\mathrm{d}s^2 = -f\mathrm{d}t^2 + h\mathrm{d}r^2 + r^2\tilde{\gamma}_{ij}\mathrm{d}\theta^i\mathrm{d}\theta^j.$$

(taking arbitrary d dimensions.) The Kretschmann invariant for this spacetime is

$$\begin{aligned}\mathcal{K} = R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma} = \frac{1}{f^2h^2}&\left[f''^2 - f''f'\left(\frac{f'}{f}+\frac{h'}{h}\right)+\frac{1}{4}f'^2\left(\frac{f'}{f}+\frac{h'}{h}\right)^2\right]\\ &+\frac{d-2}{r^2h^2}\left(\frac{f'^2}{f^2}+\frac{h'^2}{h^2}\right)+\frac{2(d-2)(d-3)}{r^4}\left(k-\frac{1}{h}\right)^2.\end{aligned}$$

For the four-dimensional case, we insert $d=4$.

Chapter 4 Matter and the Einstein equation

The basic idea for General Relativity is '*matter causes spacetime curvature*' which then manifests as gravity. We have spent the last two chapters learning how to describe curvature. Now we turn our attention to matter. The central concept is the idea of the *stress-energy tensor*. In any classical field theory, the main equation often takes the form

$$\nabla\,(\mathsf{field}) = (\mathsf{sources,\ or,\ current})\,.$$

Here, sources or current mean the substance that creates the fields in question. For example, the current density $\vec{\jmath}$ is the source of the magnetic field $\vec{B}$ in Amperé's law. We also know that a planet's mass M is the source of the gravitational field in Newton's law, which we are now trying to generalise to the relativistic version.

4.1 The energy-momentum tensor

Current density in electromagnetism

Let us consider how to write down a current density for non-relativistic electromagnetism. In this case we are seeking the amount of current flowing through an area $\mathrm{d}A$, as shown in Fig. 4.1. Suppose that the the flow consists of particles each with charge q, so that the charge density is $\rho = \frac{\mathrm{d}q}{\mathrm{d}V}$.

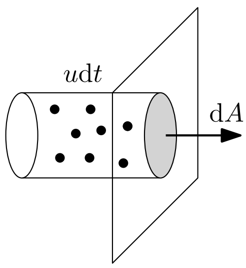


Figure 4.1: Current flowing through a surface element $\mathrm{d}A$.

Let the particles be moving at speed v. According to Fig. 4.1, within a time interval $\mathrm{d}t$, all particles within the cylinder of volume $\mathcal{V} = u\mathrm{d}t\mathrm{d}A$ would have passed through the shaded area element $\mathrm{d}A$. So the total charge of $\mathrm{d}q = \rho u\mathrm{d}t\mathrm{d}A$ has passed through this area. The current is then defined as

$$\mathsf{current} = \frac{\mathrm{d}q}{\mathrm{d}t} = \rho u \mathrm{d}A,$$

and subsequently the current density is

$$j = \frac{\mathsf{current}}{\mathsf{area}} = \frac{\mathrm{d}q}{\mathrm{d}t\mathrm{d}A} = \rho u.$$

In an arbitrary direction in $\mathbb{R}^3$, we have

$$\vec{\jmath} = \rho \vec{u} \quad \leftrightarrow \quad j^i = \rho u^i, \quad i = 1, 2, 3.$$

In other words, we have a rough definition of 'current' as

$$\mathsf{current\ density} = (\mathsf{source\ density}) \times (\mathsf{velocity}) \,. \tag{4.1}$$

Concept of current in relativity: Perfect dust

We know from Newtonian gravity that the source of gravity are masses. But since then we have also learned from SR that $E = mc^2$, so energy and mass are equivalent, so energy can be viewed as a source of gravity. Furthermore in Chapter 1, we actually learned the full equation

$$E = \sqrt{(mc^2)^2 + (pc)^2} \quad \rightarrow \quad mc^2 = \sqrt{E^2 - (pc)^2}$$

so both energy and momentum are equivalent to m, and both contribute to the gravitational field.

Under this (very crude) reasoning, we conclude that the sources of gravitational field are energy and momentum. So we seek the corresponding equation for their currents.

We shall now take the current to be due to a flow of particles of mass m. We first assume that these particles do not interact with each other. This model is called *perfect dust.* Suppose that there are many particles contained within some volume $\mathcal{V}$. Then the mass density is

$$\rho = \frac{\mathsf{mass}}{\mathsf{volume}} = \frac{\mathrm{d}M}{\mathrm{d}V} = m\frac{\mathrm{d}N}{\mathrm{d}V},$$

where N is the number of particles.[1]

[1] In some cases it is convenient to define the *number density* $n = \frac{\mathrm{d}N}{\mathrm{d}V}$.

For each individual particle, the Lagrangian of a particle is

$$\mathcal{L} = \frac{1}{2}m\left(-c^2\dot{t}^2 + \dot{x}^2 + \dot{y}^2 + \dot{z}^2\right).$$

We now treat $x^0 = ct$ as the fourth variable, hence we have four momenta $p_\mu = \frac{\partial \mathcal{L}}{\partial \dot{x}^\mu}$,

$$p_0 = -mc\dot{t}, \quad p_1 = m\dot{x}, \quad p_2 = m\dot{y}, \quad p_3 = m\dot{z}.$$

Recall also from Chapter 1 that the four-momentum takes the form

$$p_\mu = \left(-\frac{E}{c}, p_1, p_2, p_3\right) = \left(mc\dot{t}, \dot{x}, \dot{y}, \dot{z}\right),$$

where p_1, p_2, and p_3 are the traditional Newtonian momenta. In particular, observe that

$$\textsf{energy},\ E \propto \dot{t}, \quad \textsf{momentum},\ p_i \propto \dot{x}^i$$

and $p_0 = \frac{E}{c}$ so the energy E is simply the 'momentum in the time direction'. The rate of change of E with respect to (proper) time tells us the 0-th component of the four-force.

Using Eq. (4.1) as a guide, we try to identify the 'source density' and the 'velocity'. First, the velocity we already know as follows

$$u^\mu = \begin{pmatrix} c\dot{t} \\ \dot{x}^i \end{pmatrix} = \begin{pmatrix} \textsf{'speed of time'} \\ \textsf{normal velocity} \end{pmatrix} \tag{4.2}$$

So there are two conceptually distinct components for velocity.

Next for the source densities, we have the energy and momentum:

$$\begin{aligned} \textsf{energy density} &\propto \left(\frac{\textsf{mass}}{\textsf{volume}} \times \dot{t}\right), \\ \textsf{momentum density} &\propto \left(\frac{\textsf{mass}}{\textsf{volume}} \times \dot{x}^i\right). \end{aligned}$$

So our 'velocity' has 2 types, and 'source density' has 2 types. Using Eq. (4.1) as the guide, there are four types of currents we can write!

$$\begin{aligned}
&& \text{current density} = (\text{source density}) \times (\text{velocity}) \\
\text{energy current density in the } t\text{-direction}: && \frac{\text{mass}}{\text{volume}} \times c\dot{t} \times c\dot{t} \\
\text{energy current density in the } x^i\text{-direction}: && \frac{\text{mass}}{\text{volume}} \times c\dot{t} \times \dot{x}^i \\
\text{momentum current density in the } t\text{-direction}: && \frac{\text{mass}}{\text{volume}} \times \dot{x}^j \times c\dot{t} \\
\text{momentum current density in the } x^i\text{-direction}: && \frac{\text{mass}}{\text{volume}} \times \dot{x}^j \times \dot{x}^i,
\end{aligned}$$

where we have included c to ensure that all terms have the same SI unit. Notice that all four quantities above can viewed as different components of a single $(2,0)$-tensor

$$\boxed{T^{\mu\nu} = \rho u^\mu u^\nu.} \tag{4.3}$$

This is the *stress-energy tensor of dust*, which is a flow of non-interacting particles with 4-velocity $u^\mu = (c\dot{t}, \dot{x}, \dot{y}, \dot{z})$. To re-emphasise, the parts of the stress energy tensor are

$$\begin{aligned}
T^{00} &= \text{energy current density in the } x^0 = ct \text{ direction}, \\
T^{0j} &= \text{energy current density in the } x^j \text{ direction}, \\
T^{i0} &= \text{momentum current density in the } x^0 = ct \text{ direction}, \\
T^{ij} &= \text{momentum current density in the } x^j \text{ direction},
\end{aligned}$$

Since $u^\mu = (c\dot{t}, \dot{x}, \dot{y}, \dot{z})$, we have in particlar $T^{00} = \rho c^2$. So this is the energy density. (Using $\text{energy} = (\text{mass})\, c^2$). Furthermore, we can clearly see that the stress-energy tensor is symmetric, $T^{\mu\nu} = T^{\nu\mu}$. In particular, $T^{0j} = T^{j0}$. The energy current density in the x^j direction is equal to the momentum current density in the t direction.

In a frame in which the particles are at rest, $u^\mu = (c\dot{t}, 0, 0, 0)$. Then, the energy-momentum tensor for dust is simply

$$T^{\mu\nu} = \begin{pmatrix} \rho c^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Inclusion of stress

Dust was the simplest case where the particles do not interact with each other. So in a frame where the particles are at rest, we have the simple form of the energy-momentum tensor given above.

In more general cases, suppose that there is exchange of energy among the particles. For example, heat flow and thermal motion may occur within the particles. The energy-momentum tensor will have a new contribution:

$$T^{\mu\nu} = \rho u^\mu u^\nu + \mathcal{T}^{\mu\nu}.$$

We write this new contribution as

$$\mathcal{T}^{\mu\nu} = \begin{pmatrix} 0 & \mathcal{T}^{0j} \\ \mathcal{T}^{i0} & \mathcal{T}^{ij} \end{pmatrix},$$

where

$$\begin{aligned} \mathcal{T}^{0j} &= \textsf{additional energy current density in the } x^j \textsf{ direction.} \\ \mathcal{T}^{i0} &= \textsf{additional momentum current density in the } x^0 \textsf{ direction.} \\ \mathcal{T}^{ij} &= \textsf{stresses.} \end{aligned}$$

These additional energy and momentum current density may come from heat flow within the fluid. This heat flow constituets moving energy and momentum which contributes to these terms.

Now, the last part $\mathcal{T}^{ij}$ comes from the stresses within the fluid, which we now explain in some detail.

In a non-relativistic theory, one often seeks the net force of some volume $\mathcal{V}$ of a fluid, as in Fig. 4.2.

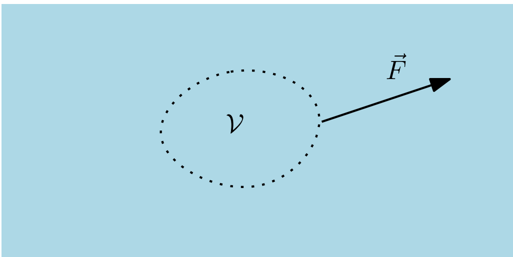


Figure 4.2: Net force on a volume $\mathcal{V}$ of a fluid.

The tricky question for fluid is: how to determine $\vec{F}$? To start, let us draw the volume $\mathcal{V}$ as a cube whose sides are aligned along the Cartesian axis, like in Fig. 4.2.

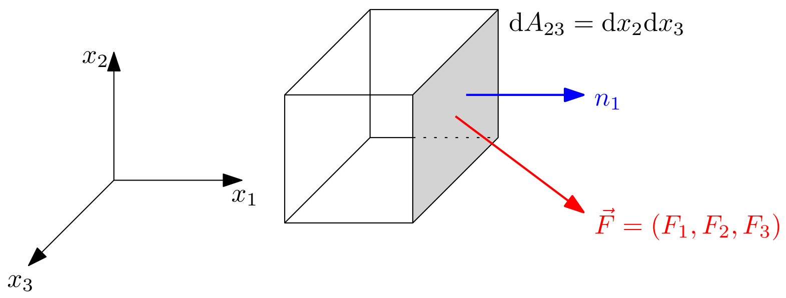


Figure 4.3

Let us consider the net force on each surface. Consider the shaded surface in Fig. 4.3. This is a surface perpendicular to the x_1-direction. We let $\vec{n} = (1, 0, 0)$ be the unit vector along the x_1-direction. Now, the force per unit area $\vec{\mathcal{T}}_{(1)}$ itself has three components which we write as $\vec{\mathcal{T}}_{(1)} = (\mathcal{T}_{11}, \mathcal{T}_{12}, \mathcal{T}_{13})$. The net force over the shaded region is the integral

$$\vec{F} = \int_{A_{23}} \vec{\mathcal{T}}_{(1)} dx_2 dx_3.$$

The three components of this force vector $\vec{F} = (F_1, F_2, F_3)$ are given explicitly by the calculation

$$\begin{aligned}
F_1 &= \int \mathcal{T}^{11} dx_2 dx_3 \quad \rightarrow \quad \mathcal{T}^{11} = \frac{\partial^2 F_1}{\partial x_2 \partial x_3}, \\
F_2 &= \int \mathcal{T}^{12} dx_2 dx_3 \quad \rightarrow \quad \mathcal{T}^{12} = \frac{\partial^2 F_2}{\partial x_2 \partial x_3} \\
F_3 &= \int \mathcal{T}^{13} dx_2 dx_3 \quad \rightarrow \quad \mathcal{T}^{13} = \frac{\partial^2 F_3}{\partial x_2 \partial x_3}.
\end{aligned}$$

This can be applied to the other surfaces as well. There are three possible surfaces, perpendicular to x ($dydz$), perpendicular to y ($dxdz$), and perpendicular to z ($dxdy$.)

So, generally, we have

$$\begin{aligned}
\mathcal{T}^{ij} &= \text{force per area in the } j\text{-th direction, on the surface element} \\
&\qquad \text{perpendicular to the } i\text{-th direction.} \\
&= \epsilon_{jkl} \frac{\partial^2 F_i}{\partial x_k \partial x_l},
\end{aligned}$$

where ϵ_{jkl} is the *totally anti-symmetric tensor* defined by

$$\epsilon_{123} = +1,$$

and it flips sign after every exchange of indices; $\epsilon_{132} = -1$, $\epsilon_{312} = +1$, etc.

Observe that the dimension of T_{ij} is force/area, which is a *pressure* whose SI unit is the Pascal. The diagonal elements T_{ii} correspond to the component of force that is perpendicular to the surface, and is indeed our usual understanding of pressure. The off-diagonal elements T_{ij}, $i \neq j$ are the forces parallel to the surface. These are the *shear* forces.

Therefore, if we assume the fluid is shear-free, the off-diagonal terms are zero, and only the diagonal pressures are present:

$$\mathcal{T}^{ij} = \begin{pmatrix} p_x & 0 & 0 \\ 0 & p_y & 0 \\ 0 & 0 & p_z \end{pmatrix}.$$

Isotropic perfect fluids. We define the isotropic perfect fluid as **fluids** which have no heat conduction, shear-free, and isotropic pressures. The latter condition means

$$p_x = p_y = p_z \equiv p, \quad \text{(isotropy)}. \tag{4.4}$$

This results in the energy-momentum-stress tensor given by

$$T^{\mu\nu} = \begin{pmatrix} \rho c^2 & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}. \tag{4.5}$$

Using the Minkowski metric, this can be written as

$$T^{\mu\nu} = \left(\rho + \frac{p}{c^2}\right) u^\mu u^\nu + p\eta^{\mu\nu}. \tag{4.6}$$

Substituting static fluids $u^\mu = (c, 0, 0, 0)$ into Eq. (4.6) reproduces Eq. (4.5).

The generalisation to curved spacetime is straightforward. It simply done by taking Eq. (4.6) and replace the Minkowski metric $\eta_{\mu\nu}$ with a general metric $g_{\mu\nu}$ so that the isotropic perfect fluid on curved spacetime is

$$\boxed{T^{\mu\nu} = \left(\rho + \frac{p}{c^2}\right) u^\mu u^\nu + pg^{\mu\nu}.} \quad \text{(Perfect fluids on any spacetime } g_{\mu\nu}.) \tag{4.7}$$

4.2 Conservation of the stress-energy tensor

As a warm-up, let us consider a Newtonian, non-relativistic fluid. Its mass density is ρ, and the mass current density is $\vec{\jmath} = \rho\vec{u}$. We consider a volume $\mathcal{V}$ of the fluid. The rate of mass exiting from this volume is the total flux of mass current density flowing through the boundary $\partial\mathcal{V}$ of the volume,

$$\oint_{\partial\mathcal{V}} \vec{\jmath} \cdot \mathrm{d}\vec{\Sigma} = \int_{\mathcal{V}} \vec{\nabla} \cdot \vec{\jmath}\, \mathrm{d}\mathcal{V}, \tag{4.8}$$

where the divergence theorem have been applied. On the other hand, the rate of mass being lost from $\mathcal{V}$ can also be calculated as

$$-\frac{\partial}{\partial t} \int_{\mathcal{V}} \rho\, \mathrm{d}\mathcal{V} = \int_{\mathcal{V}} \frac{\partial \rho}{\partial t} \mathrm{d}\mathcal{V}. \tag{4.9}$$

Equating (4.8) and (4.9), we find

$$\begin{aligned} \int_{\mathcal{V}} \frac{\partial \rho}{\partial t} \mathrm{d}\mathcal{V} + \int_{\mathcal{V}} \vec{\nabla} \cdot \vec{\jmath}\, \mathrm{d}\mathcal{V} &= 0 \\ \int_{\mathcal{V}} \mathrm{d}\mathcal{V} \left(\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot \vec{\jmath} \right) &= 0. \end{aligned}$$

To ensure this equation is true regardless of any volume shape $\mathcal{V}$, we get

$$\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot \vec{\jmath} = 0. \tag{4.10}$$

This is the non-relativistic *continuity equation*. However, if we were to apply our tensor index notation with $x^0 = ct$, we see that

$$\begin{aligned} \partial_0(\rho c) + \partial_i j^i &= 0 \\ \partial_\mu J^\mu &= 0, \end{aligned}$$

if we define the four-vector as $J^\mu = (c\rho, j^1, j^2, j^3)$.

Conservation of the stress-energy tensor. We now want to show a similar conservation behaviour for the relativistic stress-energy tensor $T^{\mu\nu}$. Let us first consider the case in Special Relativity. Note that $T^{\mu\nu}$ consists of two types of current. Let us show the explicit steps for the energy. The steps for momentum are similar and is left for the Tutorial.

Recall that T^{00} is the energy density and T^{0j} is the energy current density in the x^j direction. Consider an arbitrary 3-dimensional space volume $\mathcal{V}$ (that doesn't involve the

t direction). The rate of energy flowing out through the boundary of $\mathcal{V}$ is

$$\oint_{\partial\mathcal{V}} T^{0j}\,\mathrm{d}\Sigma_j = \int_{\mathcal{V}} \partial_j T^{0j}\,\mathrm{d}\mathcal{V}, \tag{4.11}$$

where $\mathrm{d}\Sigma_j = n_j \mathrm{d}A$ is the area element vector, whose direction is n_j and magnitude is the infinitesimal area $\mathrm{d}A$. The divergence theorem have been applied to convert the surface integral into a volume integral of the divergence.

On the other hand, the rate of energy loss from within volume $\mathcal{V}$ is

$$-\frac{1}{c}\frac{\partial}{\partial t}\int_{\mathcal{V}} T^{00}\,\mathrm{d}\mathcal{V} = -\int_{\mathcal{V}} \partial_0 T^{00}\,\mathrm{d}\mathcal{V}. \tag{4.12}$$

Equating the two quantities and rearranging we find

$$\begin{aligned}\int_{\mathcal{V}} \mathrm{d}\mathcal{V}\left(\partial_0 T^{00} + \partial_j T^{0j}\right) &= 0\\ \int_{\mathcal{V}} \mathrm{d}\mathcal{V}\;\partial_\mu T^{\mu 0} &= 0\end{aligned}$$

To ensure that this equation is satisfied regardless of volume shape $\mathcal{V}$, we obtain

$$\partial_\mu T^{\mu 0} = 0, \tag{4.13}$$

where the anti-symmetry of $T^{\mu\nu}$ have been applied. By applying the same steps for the conservation of momentum, we also show $\partial_\mu T^{\mu i} = 0$. Both results together imply

$$\boxed{\partial_\mu T^{\mu\nu} = 0.} \tag{4.14}$$

This is the equation for *conservation of the stress-energy tensor.*

Let us work out the consequences when the stress-energy tensor comes from a perfect fluid, $T^{\mu\nu} = \left(\rho + \frac{p}{c^2}\right)u^\mu u^\nu + p\eta^{\mu\nu}$. Taking the divergence we find

$$\begin{aligned}
0 &= \partial_\mu T^{\mu\nu}\\
0 &= \left(\partial_\mu \rho + \frac{1}{c^2}\partial_\mu p\right)u^\mu u^\nu + \left(\rho + \frac{p}{c^2}\right)\left[(\partial_\mu u^\mu)u^\nu + u^\mu(\partial_\mu u^\nu)\right] + \eta^{\mu\nu}\partial_\mu p\\
0 &= \left(\frac{u^\mu u^\nu}{c^2} + \eta^{\mu\nu}\right)\partial_\mu p + \underbrace{u^\mu u^\nu \partial_\mu \rho + \rho(\partial_\mu u^\mu)u^\nu}_{=u^\nu\partial_\mu(\rho u^\mu)} + \frac{p}{c^2}(\partial_\mu u^\mu)u^\nu + \left(\rho + \frac{p}{c^2}\right)u^\mu \partial_\mu u^\nu\\
&= \left(\frac{u^\mu u^\nu}{c^2} + \eta^{\mu\nu}\right)\partial_\mu p + u^\nu \partial_\mu\left(\rho u^\mu\right) + \frac{p}{c^2}u^\nu \partial_\mu u^\mu + \left(\rho + \frac{p}{c^2}\right)u^\mu \partial_\mu u^\nu
\end{aligned} \tag{4.15}$$

Now, contract this equation with u_ν. Using $u_\nu u^\nu = -c^2$, we find that

$$0 = \underbrace{\left(u^\mu + \frac{-\cancel{c^2} u^\mu}{\cancel{c^2}}\right)}_{=0} \partial_\mu p - c^2 \partial_\mu \left(\rho u^\mu\right) - p \partial_\mu u^\mu + \left(\rho + \frac{p}{c^2}\right) u^\mu u_\nu \partial_\mu u^\nu. \tag{4.16}$$

The last term is zero because $u_\nu \partial_\mu u^\nu = 0$. (Tutorial.) Hence we end up with

$$\partial_\mu \left(\rho u^\mu\right) + \frac{p}{c^2} \partial_\mu u^\mu = 0. \tag{4.17}$$

This is called the *continuity equation*. Using this equation to substitute the blue term in Eq. (4.15), we have

$$0 = \left(\frac{u^\mu u^\nu}{c^2} + \eta^{\mu\nu}\right) \partial_\mu p - u^\nu \frac{p}{c^2} \partial_\mu u^\mu + \frac{p}{c^2} u^\nu \partial_\mu + \left(\rho + \frac{p}{c^2}\right) u^\mu \partial_\mu u^\nu.$$

The middle two terms cancel each other. Therefore

$$\left(\rho + \frac{p}{c^2}\right) u^\mu \partial_\mu u^\nu = -\left(\eta^{\mu\nu} + \frac{u^\mu u^\nu}{c^2}\right) \partial_\mu p.$$

This is called the *equation of motion* for the perfect fluid.

Non-relativistic limit. We now attempt to recover the non-relativistic limit of Eqs. (4.17) and (4.18). In this limit, the pressure is small compared to the speed of light, so p/c^2 is negligible and Eq. (4.17) becomes

$$\begin{gathered} \partial_\mu \left(\rho u^\mu\right) = 0 \\ \partial_0 \left(\rho c \dot{t}\right) + \partial_i \left(\rho u^i\right) = 0. \end{gathered}$$

In the non-relativistic limit, $\dot{t} = 0$ (no time dilation) and $\partial_0 = \frac{\partial}{\partial(ct)} = \frac{1}{c}\frac{\partial}{\partial t}$. So in 3-vector notation,

$$\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot (\rho \vec{u}) = 0,$$

which exactly the non-relativistic continuity equation (4.10) for $\vec{\jmath} = \rho \vec{u}$.

On curved spacetime, we simply replace ∂_μ with ∇_μ. The corresponding continuity equation is

$$\boxed{\nabla_\mu T^{\mu\nu} = 0.} \tag{4.18}$$

For the perfect fluid, $T^{\mu\nu} = \left(\rho + \frac{p}{c^2}\right) u^\mu u^\nu + p g^{\mu\nu}$. It is straightforward to show that the continuity equation and equation of motion is

$$\boxed{\nabla_\mu (\rho u^\mu) + \frac{p}{c^2}\nabla_\mu u^\mu = 0,} \tag{4.19}$$

and the equation of motion is

$$\boxed{\left(\rho + \frac{p}{c^2}\right) u^\mu \nabla_\mu u^\nu = -\left(g^{\mu\nu} + \frac{u^\mu u^\nu}{c^2}\right)\nabla_\mu p.} \tag{4.20}$$

Observe that if the pressure is constant, $\nabla_\mu p = 0$, then the equation of motion reduces to

$$u^\mu \nabla_\mu u^\nu = 0.$$

This is just the geodesic equation. So in fluids of constant pressure, the fluid flow lines follow geodesic curves.

4.3 Energy conditions

The stress-energy tensor $T^{\mu\nu}$ quantifies the energy and momentum content of the space-time. We have argued that the conservation of energy and momentum requires $\nabla_\mu T^{\mu\nu} = 0$. So we should have physically reasonable distribution of energy and momentum.

But conservation alone is not enough to guarantee that it is 'physically reasonable'. Note that the energy and momentum of a particle depends on the relative motion of the observer. So we should try to check that the measurement made by an observer gives 'reasonable' outcomes. (For example, does not give complex numbers for energy.) Furthermore, the observer themselves must be 'physically reasonable'. That is, any relevant observer should not travel faster than light.

Therefore an important quantity is the four-velocity of the observer. Let V^μ be the observer's four-velocity. As usual, we classify the observers as

$$V^\mu V_\mu \begin{cases} < 0 & \text{time-like observers,} \\ = 0 & \text{null observers.} \end{cases} \tag{4.21}$$

Given a stress-energy tensor $T^{\mu\nu}$, the energy of the system measured by this observer is

$$\mathcal{E} = T_{\mu\nu} V^\mu V^\nu. \tag{4.22}$$

Because there are different kinds of observers, and that there is no clear rules of what energy *should* be (aside from previous experience in Newtonian or quantum mechanics), there is no unique condition that must be satisfied by $T^{\mu\nu}$. In fact there are *several* different energy conditions. We list the main ones here:

1. **Weak Energy Condition (WEC).** $T_{\mu\nu}V^\mu V^\nu \geq 0$ for any time-like observer, $V^\mu V_\mu < 0$.

2. **Null Energy Condition (NEC).** $T_{\mu\nu}k^\mu k^\nu \geq 0$ for any null observer, $k^\mu k_\mu = 0$.

3. **Dominant Energy Condition (DEC).** For any time-like observer, $V_\mu V^\mu < 0$, the stress tensor must satify $T_{\mu\nu}V^\mu V^\nu \geq 0$ (WEC) and also $T_{\mu\nu}T^\nu{}_\lambda V^\mu V^\lambda \leq 0$. This second condition means $T^{\mu\nu}V_\mu$ is non-space-like.

4. **Strong Energy Condition (SEC).** $T_{\mu\nu}V^\mu V^\nu \geq \frac{1}{2}T^\lambda{}_\lambda V^\sigma V_\sigma$ for any time-like observer $V^\mu V_\mu < 0$.

Energy conditions for the isotropic perfect fluid

Let us see what the energy conditions mean for the case

$$T_{\mu\nu} = \left(\rho + \frac{p}{c^2}\right) u_\mu u_\nu + p g_{\mu\nu}.$$

Recall that u^μ is a normalised time-like vector $u^\mu u_\mu = -c^2$. First, to check the weak energy condition, we take V^μ as another time-like vector and compute

$$\begin{aligned} T_{\mu\nu}V^\mu V^\nu &= \left(\rho + \frac{p}{c^2}\right) u_\mu V^\mu u_\nu V^\nu + p V_\mu V^\mu \\ &= \left(\rho + \frac{p}{c^2}\right) (u_\mu V^\mu)^2 - pc^2. \end{aligned}$$

Next we have to compute the inner product $u_\mu V^\mu$. In general, we may decompose V^μ into a two parts, one parallel and one orthogonal to u^μ:

$$V^\mu = a u^\mu + b n^\mu,$$

where n^μ is a unit vector orthogonal to u^μ, hence $n^\mu u_\mu = 0$. But both u^μ and V^μ are time-like, so they both satisfy $u_\mu u^\mu = -c^2$ and $V_\mu V^\mu = -c^2$. Therefore

$$\begin{aligned} V^\mu V_\mu = -c^2 &= (au^\mu bn^\mu)\,(au_\mu + bn_\mu) \\ -c^2 &= a^2 u^\mu u_\mu + 2ab\,\underbrace{u^\mu n_\mu}_{0} + b^2 n^\mu n_\mu \\ -c^2 &= -a^2c^2 + b^2 \\ c^2 &= a^2c^2 - b^2. \end{aligned}$$

We can thus parametrise a and b by

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

and without loss of generality, any arbitrary unit time-like observer V^μ can be expressed in terms of u^μ as

$$V^\mu = \cosh\beta u^\mu + c\sinh\beta\, n^\mu.$$

This will be useful because the product with u^μ is

$$u_\mu V^\mu = \cosh\beta u_\mu u^\mu = -c^2\cosh\beta.$$

We therefore have

$$\begin{aligned} T_{\mu\nu}V^\mu V^\nu &= \left(\rho + \frac{p}{c^2}\right)c^4 - pc^2 \\ &= c^2\left[\left(\rho c^2 + p\right)\cosh^2\beta - p\right] \end{aligned}$$

For the WEC, $T_{\mu\nu}V^\mu V^\nu \geq 0$, and this must be positive in the above expression for any (real) β. The domain for $\cosh^2\beta$ is $1 \leq \cosh^2\beta < \infty$. Let us check $\beta = 0$, then $\cosh^2\beta = 1$ and we have $T_{\mu\nu}V^\mu V^\nu = c^2\left[\rho c^2 + p - p\right] = \rho c^4$. Therefore we require $\rho \geq 0$. For large β, then $(\rho c^2 + p)\cosh^2\beta - p \simeq (\rho c^2 + p)\cosh^2\beta$. (The second term becomes negligible). To make sure this is still positive we require $\rho c^2 + p \geq 0$. Therefore the WEC for the perfect fluid is

$$\rho^2 \geq 0, \quad \rho c^2 + p \geq 0. \quad \text{(WEC for perfect fluids.)} \tag{4.23}$$

We turn next to the Null energy condition (NEC), we take null vectors k^μ, where $k_\mu k^\mu = 0$.

Computing the contraction with the stress tensor,

$$\begin{aligned} T_{\mu\nu}k^\mu k^\nu &= \left(\rho + \frac{p}{c^2}\right) u_\mu k^\mu \, u_\nu k^\nu + p g_{\mu\nu} k^\mu k^\nu \\ &= \left(\rho + \frac{p}{c^2}\right) (u_\mu k^\mu)^2 + 0. \end{aligned}$$

Clearly this is non-negative as long as

$$\rho c^2 + p \geq 0. \quad \text{(NEC for perfect fluids.)} \tag{4.24}$$

We also see that any fluid that satisfies the WEC will satisfy the NEC.

By similar methods, we obtain the other energy conditions for the perfect fluid:

$$\rho c^2 + p \geq 0, \quad \text{(NEC for perfect fluids)}, \tag{4.25}$$

$$\rho c^2 \geq p, \quad \text{(DEC for perfect fluids)}, \tag{4.26}$$

$$\rho c^2 + p \geq 0, \quad \rho c^2 + 3p \geq 0, \quad \text{(SEC for perfect fluids)}. \tag{4.27}$$

4.4 Philosophical principles leading to GR

The equivalence principles

As we discussed during the introduction of this course, it is an experimental fact that inertial mass is identical to ‘gravitational charge’. The consequences of this is the *universality of gravitation*, meaning all masses have the same acceleration in the same gravitational field. This is why we always use $g = 9.80\ \mathrm{m\,s^{-2}}$ on Earth’s surface. Since this is an experimental fact, we can only be sure of this up to experimental accuracy. Therefore, we shall assume that this is true as a *principle*. More precisely, this is the *weak equivalence principle (WEP)*.

Another situation where every object expriences the same acceleration is when they are in an accelerating frame. For example everybody gets thrown forwards when a car suddenly brakes. The WEP is a statement about *free-falling* particles. Meaning particles that experience no other forces besides gravity.

Consider a situation where an observer is in a sealed box with no windows. When she releases a particle in front of her, she notices that it accelerates towards one side of the box. Is this box accelerating in space, or is the box sitting on a planet with some gravity? (See Fig. 4.4.)

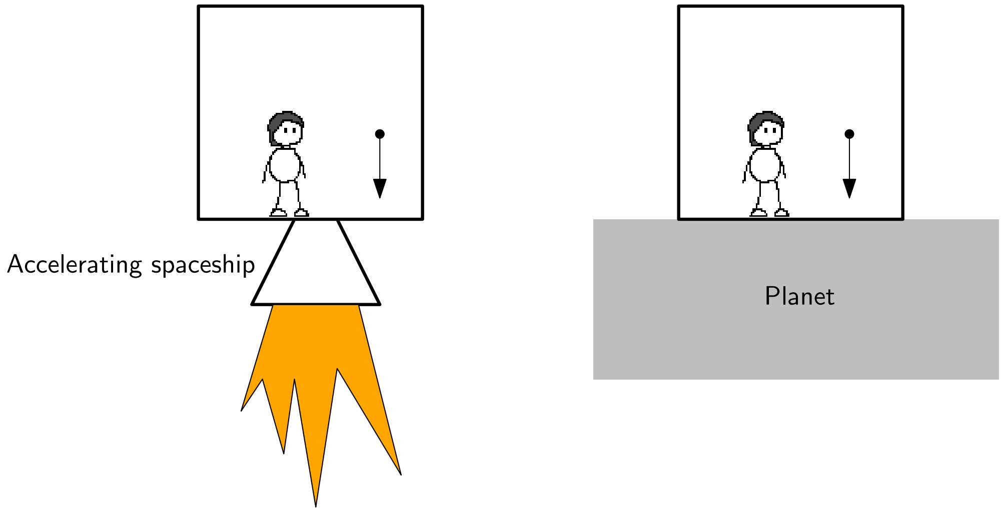


Figure 4.4: Two situations indistinguishible by an observer observing a free-falling particle inside a sealed box.

The precise formulation of the WEP is

> **Weak Equivalence Principle.** *The motion of a free-falling particle in a uniform gravitational field is the same as in a uniformly accelerated frame, for small enough regions of spacetime.*

The condition *'small enough regions of spacetime'* is the requirement that the gravitational field is **uniform**. This is because we can obviously distinguish the gravitational field when it's not uniform, for example on a spherical planet shown in Fig. 4.5.

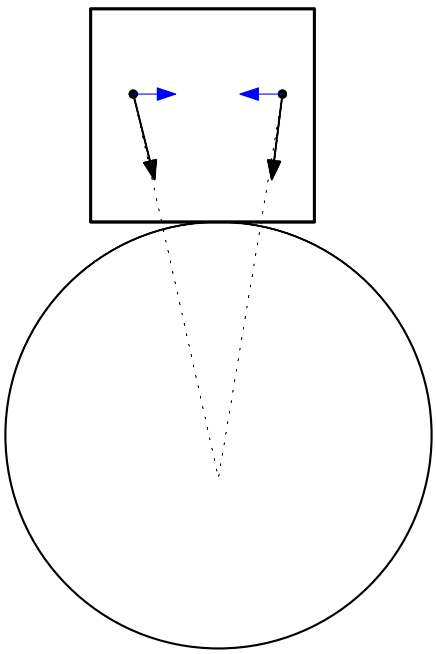


Figure 4.5: Tidal forces due to a non-uniform gravitational field. In this case, the field is radial.

As it currently stands, the WEP is a statement about the equivalence between inertial mass and gravitational mass (the 'gravitational charge'). This is not unique to relativity, as Newtonian gravity and Galilean transformations are alrady sufficient ingredients to identify WEP. But after the introduction of Special Relativity and the discussion of stress-energy tensor in the previous chapter, we know that energy and momentum are also equivalent to mass. So the WEP should be generalised not only as a statement about inertial mass, but also includes energy and momentum as well.

Therefore General Relativity actually requires a stronger principle than the WEP. Einstein generalises the condition to WEP to include not just free-falling particles. But also *any* non-gravitational interactions in the experiments. Today we call this the Einstein Equivalence Principle,

> **Einstein Equivalence Principle (EEP).** *All non-gravitational laws of physics in a uniform gravitational field are the same as in a uniformly accelerated frame.*

For example, experiments involving charged particles (which uses Coulomb's law), done in an accelerating frame or in a uniform gravitational field.

The EEP has a testable prediction. Consider two spaceships accelerating uniformly with the same acceleration a as in Fig. 4.6. Then both will be static in a frame that is co-accelerating with both of them.

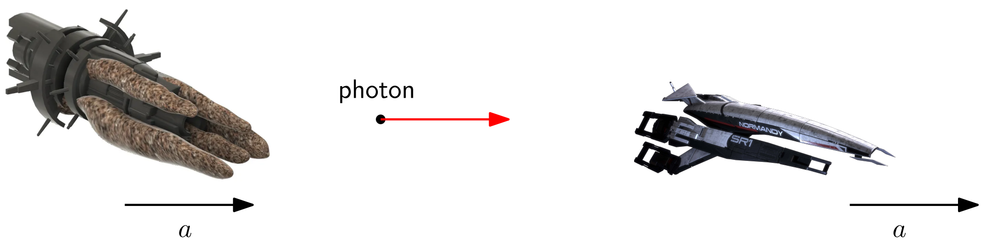


Figure 4.6: A Collector Ship (left) and SSV Normandy (right) are moving under the same acceleration a. The Collector Ship emits a photon of wavelength λ_0. [Normandy image from [17]. Collector ship image from [18].]

The ship on the left (Collector) emits a photon towards the ship on the right (Normandy). During emission the photon has wavelength λ_0 and frequency $f_0 = c/\lambda_0$ according to the emitting ship.

In an inertial background coordinate system, the two ships accelerate at a, and the photon travels some distance L over a time $\Delta t = L/c$. So when the photon reaches the Normandy,

Normandy's speed has increased by approximately $\Delta v = a\Delta t = aL/c$.

$$\Delta v = a\Delta t = \frac{aL}{c}.$$

What is the wavelength measured by the Normandy? This can calculated from the measured frequency. The Normandy has relative velocity Δv **relative to the Collector earlier when it was emitting the photon**. The observed wavelength will have a difference due to the Doppler effect

$$\frac{\Delta\lambda}{\lambda_0} = \frac{\Delta v}{c^2} = \frac{aL}{c^2}. \tag{4.28}$$

This is the result occuring within a system undergoing uniform acceleration a. According to the EEP, the same result will occur in a non-accelerating system, but in an external gravitational field. Let $g = a$ be the acceleration due to this gravity, and consider a photon emitted from a tower of height L. During emission its wavelength was λ_0, as shown in Fig. 4.7.

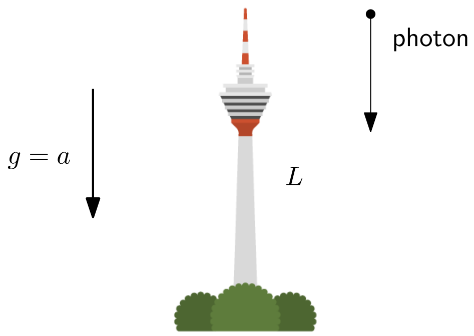


Figure 4.7: KL Tower illustration by tabisozai https://tabisozai.net/en/kl-tower/.

When the photon is measured by an observer at the ground floor, the EEP state that the measured wavelength is the same as Eq. (4.28), so

$$\frac{\Delta\lambda}{\lambda_0} = \frac{aL}{c^2} = \frac{gL}{c^2}. \tag{4.29}$$

That is, the wavelength of the photon has been changed by the gravitational field! This phenomena is called the *gravitational redshift*. This is because $\Delta\lambda$ is positive, so the measured photon has longer wavelength than when it was first emitted. So the observe light becomes *redder*. For the KL Tower whose height is $L = 421$ m, the redshift is

$$\frac{\Delta\lambda}{\lambda_0} = \frac{(9.80\ \mathrm{m\,s^{-2}})(421\ \mathrm{m})}{(2.998\times10^{8}\ \mathrm{m\,s^{-1}})^2} = 4.584\times10^{-14}\ \mathrm{m^{-1}},$$

which is hardly noticeable with the naked eye. However, the gravitational redshift has been successfully observed with precision experiments Pound et al. [19, 20].

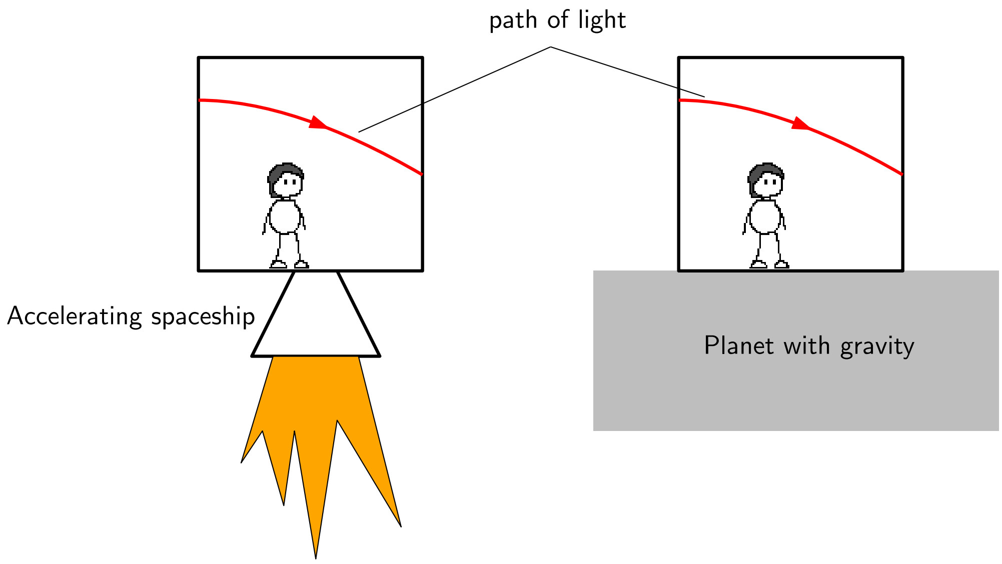


Figure 4.8: By the equivalence principle, the curving of photon trajectory by a gravitational field.

There is another observable phenomena that comes from EEP. Consider again an observer in a sealed box. All objects fall to one side of the box, and according to EEP, it is distinguish the difference between an accelerating box, or a box sitting in an external gravitational field (Fig. 4.8).

The WEP applies only to particles with inertial/gravitational masses, while the EEP extends to consider all particles, including massless photons. So if the box is accelerating past a beam of light, the observer inside the box will see that it's moving on a curve trajectory relative to the box. (The left panel of Fig. 4.8.)

By the EEP, this situation is equivalent to the box sitting in a gravitational field. So an observer in gravity will also observe a curve trajectory of light. This means that the gravity will also curve the trajectory of light!

Finally, one can continue extending EEP to include *all* laws of physics, gravitational and non-gravitational alike. This is the Strong Equivalence Principle,

> **Strong Equivalence Principle (SEP).** *All laws of physics in a uniform gravitational field are the same as in a uniformly-accelerated frame.*

This means that the experiments inside the box includes also gravity. For example, two masses m_1 and m_2 inside a box, which will gravitationally attract each other. The behaviour of these two particles in an external gravitational field will be the same as a uniformly accelerating box.

Geodesics as trajectories of free particles

How do we turn these philosophical observations into equations that describe the gravitational field? In the Minkowski space of Special Relativity, for the metric

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2,$$

free particles move in straight lines,

$$\ddot{x}^0 = c\ddot{t} = 0, \quad \ddot{x} = 0, \quad \ddot{y} = 0, \quad \ddot{z} = 0. \tag{4.30}$$

In Chapter 3, we have seen that an example of an *accelerating observer* uses the Rindler metric

$$\mathrm{d}s^2 = -\rho^2\mathrm{d}\eta^2 + \mathrm{d}\rho^2 + \mathrm{d}y^2 + \mathrm{d}z^2,$$

then, Eq. (4.30) becomes the geodesic equations for Rindler spacetime. By the EEP, this is equivalent to a non-accelerating observer, but in a gravitational field. Therefore, the EEP requires the gravitational field to make the particle move according to a geodesic equation as well:

$$\ddot{x}^\kappa + \Gamma^\kappa_{\mu\nu}\dot{x}^\mu\dot{x}^\nu = 0. \tag{4.31}$$

Of course, in the presence of gravity, even for static observers, particles do not move in straight lines. So in general $\Gamma^\kappa_{\mu\nu}$ should be non-zero. Since the Christoffel symbols are calculated from the metric, we conclude that the gravitational field has to be described by the metric $g_{\mu\nu}$. In general, having non-trivial $\Gamma^\kappa_{\mu\nu}$ requires a curved manifold. With this we arrive at our central statement of General Relativity,

The gravitational field is described by a curved manifold with metric $g_{\mu\nu}$.

This is a drastic statement because it represents a big leap from a previously well-established theory: Newtonian mechanics. In Newtonian gravity, the gravitational potential is given by some scalar function Φ, and the gravitational force is $\vec{F} = m\vec{\nabla}\Phi$. Then by the second law, $\vec{F} = m\vec{a}$,

$$\frac{\mathrm{d}^2\vec{r}}{\mathrm{d}t^2} = -\vec{\nabla}\Phi \quad \leftrightarrow \quad \frac{\mathrm{d}^2x^i}{\mathrm{d}t^2} = -\delta^{ij}\partial_j\Phi, \tag{4.32}$$

where the particle's mass m has cancelled out (this is the WEP already discussed.) We have written both the vector calculus notation as well as our tensor-index notation. For the new theory using (4.31) to be correct, it must recover Eq. (4.32) in the non-relativistic limit, when the velocity is small compared to c. Furthermore, the relativistic effects like gravitational redshift was not noticeable for weak gravitational fields.

This means that, in this limit, the metric is a small perturbation from Minkowski,

$$\begin{aligned} g_{\mu\nu} &= (\text{Minkowski}) + (\text{small perturbation}) \\ &= \eta_{\mu\nu} + h_{\mu\nu}, \end{aligned}$$

where $h_{\mu\nu}$ represents the small perturbation. By 'small', the components have small values, $|h_{\mu\nu}| \ll 1$. Later, we will need an expression for the inverse metric

$$g^{\mu\nu} = \eta^{\mu\nu} + h^{\mu\nu}.$$

To determine the expression for $h^{\mu\nu}$, we take the identity

$$\begin{aligned} g^{\mu\lambda} g_{\lambda\nu} &= \delta^\mu_\nu \\ \left(\eta^{\mu\lambda} + h^{\mu\lambda}\right)\left(\eta_{\lambda\nu} + h_{\lambda\nu}\right) &= \delta^\mu_\nu \\ \underbrace{\eta^{\mu\lambda}\eta_{\lambda\nu}}_{\delta^\mu_\nu} + \eta^{\mu\lambda} h_{\lambda\nu} + h^{\mu\lambda}\eta_{\lambda\nu} + h^{\mu\lambda} h_{\lambda\nu} &= \delta^\mu_\nu \end{aligned}$$

Since $h_{\mu\nu}$ is small, we consider $h^{\mu\lambda} h_{\lambda\nu} \simeq h^2$ to be negligible. Cancelling the Kronecker delta on both sides,

$$h^{\mu\lambda}\eta_{\lambda\nu} = -\eta^{\mu\lambda} h_{\lambda\nu}.$$

Further contracting both sides with $\eta^{\nu\sigma}$, we have

$$h^{\mu\sigma} = -\eta^{\mu\lambda}\eta^{\nu\sigma} h_{\lambda\nu}. \tag{4.33}$$

Therefore, $g^{\mu\nu} = \eta^{\mu\nu} - \eta^{\mu\rho}\eta^{\nu\sigma} h_{\rho\sigma}$.

Turning now to the geodesic equation,

$$\begin{aligned} \ddot{x}^\kappa + \Gamma^\kappa_{\mu\nu}\dot{x}^\mu \dot{x}^\nu &= 0 \\ \ddot{x}^\kappa + \Gamma^\kappa_{00}\dot{x}^0\dot{x}^0 + 2\Gamma^\kappa_{0j}\dot{x}^0\dot{x}^j + \Gamma^\kappa_{ij}\dot{x}^i\dot{x}^j &= 0. \end{aligned}$$

In the low velocity limit, the particle must be moving slow compared to light. To compare, we divide both sides by c^2,

$$\frac{\ddot{x}^\kappa}{c^2} + \Gamma^\kappa_{00}\frac{\dot{x}^0\dot{x}^0}{c^2} + 2\Gamma^\kappa_{0j}\frac{\dot{x}^0\dot{x}^j}{c^2} + \Gamma^\kappa_{ij}\frac{\dot{x}^i\dot{x}^j}{c^2} = 0. \tag{4.34}$$

The conditions for non-relativistic limit is that the particle moves slower than c, and also the time dilation becomes negligible. This means

$$\frac{\dot{x}^0}{c} = \frac{\mathrm{d}t}{\mathrm{d}\tau} \simeq 1, \quad \frac{\dot{x}^i}{c} \ll 1$$

Therefore the last two terms of (4.34) is negligible and we have

$$\begin{aligned} \frac{\ddot{x}^\kappa}{c^2} + \Gamma^\kappa_{00} &= 0 \\ \frac{\ddot{x}^\kappa}{c^2} + \frac{1}{2}g^{\kappa\lambda}\left(\partial_0 g_{\lambda 0} + \partial_0 g_{\lambda 0} - \partial_\lambda g_{00}\right) &= 0. \end{aligned}$$

We assume the spacetime is static (doesn't change with time). So $\partial_0 g_{\mu\nu} = 0$, and we now have

$$\begin{aligned} \frac{\ddot{x}^\kappa}{c^2} - \frac{1}{2}g^{\kappa\lambda}\partial_\lambda g_{00} &= 0 \\ \frac{\ddot{x}^\kappa}{c^2} - \frac{1}{2}g^{\kappa j}\partial_j g_{00} &= 0 \\ \frac{\ddot{x}^\kappa}{c^2} - \frac{1}{2}\left(\eta^{\kappa j} - \eta^{\kappa\sigma}\eta^{jk}h_{\sigma k}\right)\partial_j\left(\eta_{00} + h_{00}\right) &= 0 \\ \frac{\ddot{x}^\kappa}{c^2} - \frac{1}{2}\eta^{\kappa j}\partial_j h_{00} &= 0 \end{aligned}$$

For $\kappa = 0$ and $\kappa = i = 1, 2, 3$, respectively, we have

$$\frac{\ddot{x}^0}{c^2} = \frac{\mathrm{d}^2 t}{\mathrm{d}\tau^2} = 0, \quad \frac{\ddot{x}^i}{c^2} = \frac{1}{2}\delta^{ij}\partial_j h_{00} \tag{4.35}$$

The first equation is consistent with the absence of time dilation, $\dot{t} = \frac{\mathrm{d}t}{\mathrm{d}\tau} = 1$, so we can take $t = \tau$. For the second equation,

$$\frac{\mathrm{d}^2 x^i}{\mathrm{d}t^2} = \frac{1}{2}\partial_j\left(c^2 h_{00}\right). \tag{4.36}$$

By comparing Eqs. (4.36) with (4.32), we have identified that

$$\frac{1}{2}c^2 h_{00} = -\Phi \quad \leftrightarrow \quad h_{00} = -\frac{2\Phi}{c^2} \tag{4.37}$$

is the gravitational potential in the non-relativistic, weak gravity limit.

4.5 The Einstein equation

The question of '*what determines the gravitational field?*' is now equivalent to '*what determines $g_{\mu\nu}$?*' In Newtonian gravity, we know that masses determines the strength of the gravitational field. In Chapter 4, we have seen that relativity says mass is equivalent to energy and momentum. So we expect that the stress-energy-momentum tensor $T^{\mu\nu}$ should determine the gravitational field.

In the previous section, we have seen that g_{00} contains the gravitational potential in the Newtonian limit. Newtonian gravity states that

$$\nabla^2\Phi = 4\pi G\rho.$$

Furthermore, the mass density ρ is contained in the component T^{00} of the stress-energy tensor. The appearance of ∇^2 suggests that the field equation should include second derivatives in $g_{\mu\nu}$.

We do know that the Ricci tensor $R_{\mu\nu}$ contains second derivatives in the metric, so one could guess that the law of gravity should take the form

$$R_{\mu\nu} \propto T_{\mu\nu}.$$

But this is not possible, because the stress tensor must obey conservation of energy-momentum, $\nabla_\mu T^{\mu\nu} = 0$. The divergence of the Ricci tensor is

$$\nabla_\mu R^\mu{}_\nu = \frac{1}{2}\nabla_\nu R. \tag{4.38}$$

(Tutorial 3, Question 9.) So the field equation must have zero divergence. At this stage we observe that a rearrangement of (4.38) gives

$$\nabla_\mu \left(R^{\mu\nu} - \frac{1}{2} R g^{\mu\nu} \right) = 0.$$

So we define the *Einstein tensor*

$$\boxed{\mathcal{G}_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu},} \tag{4.39}$$

which has zero divergence.

With this insight, it is natural to propose the field equation

$$\mathcal{G}_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = \kappa T_{\mu\nu}. \tag{4.40}$$

We finally encounter the *Einstein's field equations*, the main subject of this course. But it is not complete yet because κ is a constant of proportionality whose value remains to be fixed. Both sides of the equation now has zero divergence, satisfying conservation of energy-momentum. It is often convenient to remove the Ricci scalar from the equation. First we contract both sides with with $g^{\mu\nu}$, giving

$$\begin{aligned} R - \frac{4}{2} R = -R &= \kappa T^{\lambda}{}_{\lambda} \\ R &= -\kappa T^{\lambda}{}_{\lambda}. \end{aligned}$$

Substituting this expression for R back into (4.40), we get

$$\begin{aligned} R_{\mu\nu} - \frac{1}{2} \left(-\kappa T^{\lambda}{}_{\lambda} \right) g_{\mu\nu} &= \kappa T_{\mu\nu} \\ R_{\mu\nu} &= \kappa \left(T_{\mu\nu} - \frac{1}{2} T^{\lambda}{}_{\lambda} g_{\mu\nu} \right). \end{aligned}$$

This is called the *traced-reversed* form of the Einstein equation.

To determine the constant κ, we should require that our field equation should recover Newtonian gravity in the non-relativistic, weak gravity limit,

$$\nabla^2 \Phi = 4\pi G \rho. \tag{4.41}$$

In the previous chapter we have seen that $T_{00} = \rho c^2$ so let us focus on the 00 components of this equation,

$$R_{00} = \kappa \left(T_{00} - \frac{1}{2} T^{\lambda}{}_{\lambda} g_{00} \right).$$

As discussed earlier, we represent 'weak' gravity as a perturbation of Minkowski spacetime,

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

where $h_{\mu\nu}$ is small such that terms like h^2 are negligible. Now since $\Gamma \simeq \partial g \simeq \partial h$, this also means that two Christoffel symbols multiplied together are also negligible. Also, we assume that our metric components are static, so $\partial_0 \Gamma^{\kappa}_{\mu\nu} = 0$. The 00-Ricci tensor is therefore

$$\begin{aligned}
R_{00} &= \partial_\mu \Gamma^\mu_{00} - \underbrace{\partial_0 \Gamma^\mu_{\mu 0}}_{\simeq 0} + \underbrace{\Gamma^\mu_{\mu\lambda}\Gamma^\lambda_{00} - \Gamma^\mu_{0\lambda}\Gamma^\lambda_{\mu 0}}_{\simeq \text{negligible.}} \\
&= \partial_\mu \Gamma^\mu_{00} \\
&= \partial_\mu \frac{1}{2} g^{\mu\lambda} \left(\partial_0 g_{\lambda 0} + \partial_0 g_{\lambda 0} - \partial_\lambda g_{00}\right) \\
&= \frac{1}{2}\partial_\mu \left[g^{\mu\lambda}\left(-\partial_\lambda g_{00}\right)\right] \\
&= -\frac{1}{2}\eta^{\mu\lambda}\partial_\mu \partial_\lambda h_{00} \\
&= -\frac{1}{2}\eta^{ij}\partial_i \partial_j h_{00}.
\end{aligned}$$

For the spatial components of the inverse Minkowski metric is simply $\eta^{ij} = \delta^{ij}$. Therefore

$$R_{00} = -\frac{1}{2}\delta^{ij}\partial_i \partial_j h_{00}. \tag{4.42}$$

Turning to the right hand side, we assume

$$T_{\mu\nu} \simeq \begin{pmatrix} \rho c^2 & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}.$$

In Chapter 4, we discussed the non-relativistic limit of the stress tensor to be $p \ll \rho c^2$. So the pressures are negligible and it becomes the dust $T^{\mu\nu} = \rho u^\mu u^\nu$. Then the trace is ${T^\lambda}_\lambda = g^{\mu\nu} T_{\mu\nu} = -\rho c^2$. This leads to

$$T_{00} - \frac{1}{2}{T^\lambda}_\lambda g_{00} = \rho c^2 - \frac{1}{2}(-\rho c^2)\left(\eta_{00} + h_{00}\right) \tag{4.43}$$

Since $\eta_{00} = -1$ and $h_{00} \ll 1$, the second term above is also negligible and we have $T_{00} - \frac{1}{2}{T^\lambda}_\lambda g_{00} \simeq \frac{1}{2}\rho c^2$. We now have both sides of the field equation, giving

$$-\frac{1}{2}\delta^{ij}\partial_i \partial_j h_{00} = \frac{1}{2}\kappa \rho c^2. \tag{4.44}$$

In Eq. (4.37), we have identified $h_{00} = -\frac{2\Phi}{c^2}$ to be related to the Newtonian gravitational potential, so

$$\begin{aligned}
-\frac{1}{2}\delta^{ij}\partial_i\partial_j\left(-\frac{2\Phi}{c^2}\right) &= \frac{1}{2}\kappa\rho c^2 \\
\delta^{ij}\partial_i\partial_j\Phi &= \frac{\kappa c^4}{2}\rho \\
\nabla^2\Phi &= \frac{1}{2}\kappa c^4\rho.
\end{aligned}$$

Comparing this with Eq. (4.41), we conclude that $\kappa = \frac{8\pi G}{c^4}$. Therefore the Einstein equation, along with its its traced-reversed form, is

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu}, \quad R_{\mu\nu} = \frac{8\pi G}{c^4}\left(T_{\mu\nu} - \frac{1}{2}T^{\lambda}{}_{\lambda}g_{\mu\nu}\right). \tag{4.45}$$

If we are measuring time in length units, $c = 1$ and the Einstein equation appears simply as

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi G T_{\mu\nu}, \quad R_{\mu\nu} = 8\pi G\left(T_{\mu\nu} - \frac{1}{2}T^{\lambda}{}_{\lambda}g_{\mu\nu}\right). \tag{4.46}$$

The cosmological constant. One of the main reasons we arrived at this form of Einstein equation is we required $\nabla_\mu T^{\mu\nu} = 0$. The left hand side is satisfied because $\nabla_\mu\left(R^{\mu\nu} - \frac{1}{2}Rg^{\mu\nu}\right) = 0$. Now, we notice that the metric itself will satisfy $\nabla_\mu g^{\mu\nu} = 0$, due to metric compatibility.[2] So, we can actually add a term $(\mathsf{constant})\, g_{\mu\nu}$, and all our conditions will still be satisfied.

So, our requirements are still satisfied if we take the Einstein equation to be

$$\boxed{R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu},} \tag{4.47}$$

where Λ is called the *cosmological constant*.

The traced-reversed form of Einstein equation with cosmological constant (in four dimensions) is

$$\boxed{R_{\mu\nu} = \Lambda g_{\mu\nu} + \frac{8\pi G}{c^4}\left(T_{\mu\nu} - \frac{1}{2}T^{\lambda}{}_{\lambda}g_{\mu\nu}\right).} \tag{4.48}$$

As a final remark, we note that the equation governing a gravitational system, Eq. (4.47), depends on three physical constants G, c, and Λ. The first two are known from various

[2] Recall that metric compatibility means that the covariant derivative of the metric is always zero, $\nabla_\sigma g_{\mu\nu} = 0$.

experiments to be

$$G = 6.6743 \times 10^{-11}\ \mathrm{kg}^{-1}\,\mathrm{m}^3\,\mathrm{s}^{-2}, \quad c = 2.99\,792\,458 \times 10^8\ \mathrm{m\,s}^{-1}.$$

4.6 On units

Reading books and papers on General Relativity may be slightly confusing because different authors use different conventions for units. This is usually not a big problem if we remember the interpretations of the physical quantities and perform simple dimensional analysis.

Recall that our fundamental dimensions are

$$[\mathsf{mass}] = M, \quad [\mathsf{length}] = L, \quad [\mathsf{time}] = T. \tag{4.49}$$

First, consider the line element

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu.$$

This is a square of an infinitesimal length. So $\mathrm{d}s^2$ should have dimension L^2. Let us now take the convention that x^μ carries the length dimension and hence $g_{\mu\nu}$ is dimensionless,[3]

$$[g_{\mu\nu}] = 1. \tag{4.50}$$

This convention is convenient because raising or lowering indices of tensors does not change the dimension. For example the stress tensor $T^{\mu\nu}$ and $T_{\mu\nu}$ will have the same dimension.

The physical constants carry the following dimensions:

$$[G] = M^{-1}L^3T^{-2}, \quad [c] = LT^{-1}.$$

Consider the perfect fluid stress tensor. The mass density ρ and pressure[4] p respectively has dimension

$$[\rho] = ML^{-3}, \quad [p] = \frac{[\mathsf{force}]}{[\mathsf{area}]} = \frac{MLT^{-2}}{L^2} = ML^{-1}T^{-2}.$$

[3]This is sometimes violated if x^μ is an angular coordinate where $\mathrm{d}\theta$ is dimensionless.

[4]The unit for pressure is abbreviated as the *Pascal*, where $1\ \mathrm{Pa} = 1\ \mathrm{kg\,m}^{-1}\,\mathrm{s}^{-2}$.

Therefore, ρ and p/c^2 has the same dimension, as

$$\left[\frac{p}{c^2}\right] = \frac{ML^{-1}T^{-2}}{L^2T^{-2}} = ML^{-3}.$$

Each component of the 4-velocity has the dimension of velocity, $[u^\mu] = LT^{-2}$. This is easy to remember since the time-like condition of 4-velocities is

$$u^\mu u_\mu = g_{\mu\nu}u^\mu u^\nu = -c^2,$$

which is the square of a velocity.

Therefore the stress tensor carries the dimension

$$[T_{\mu\nu}] = \left[\left(\rho + \frac{p}{c^2}\right) u_\mu u_\nu + pg_{\mu\nu}\right] = \frac{M}{L^3} \times (\mathsf{velocity})^2 = ML^{-1}T^{-2}.$$

The right hand side of the Einstein equation is $\frac{8\pi G}{c^4}T_{\mu\nu}$. With 8π being a dimensionless number, we have need to check the dimensions of the physical constants,

$$\left[\frac{G}{c^4}\right] = \frac{M^{-1}L^3T^{-2}}{L^4T^{-4}} = M^{-1}L^{-1}T^2.$$

Therefore we have

$$\left[\frac{8\pi G}{c^4}T_{\mu\nu}\right] = M^{-1}L^{-1}T^2 \cdot ML^{-1}T^{-2} = L^{-2}. \tag{4.51}$$

The connection and curvature components are defined in terms of derivatives of $g_{\mu\nu}$. Since $[\partial_\mu] = L^{-1}$, we have

$$\begin{aligned}
\text{Christoffel symbols}: \quad & \left[\Gamma^\kappa_{\mu\nu}\right] \sim \ [\partial g] = L^{-1}, \\
\text{Riemann tensor}: \quad & [R^\rho{}_{\sigma\mu\nu}] \ \sim [\partial\Gamma + \ldots + \Gamma\Gamma] = L^{-2}, \\
\text{Ricci tensor}: \quad & [R_{\mu\nu}] \ \sim \left[R^\lambda_{\mu\lambda\nu}\right] = L^{-2}, \\
\text{Ricci scalar}: \quad & [R] \ \sim [g^{\mu\nu}R_{\mu\nu}] = L^{-2}
\end{aligned}$$

In particular, the Einstein tensor $R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}$ also has dimension L^{-2}. Therefore we have confirmed that both sides of the Einstein equation have the same dimension of L^{-2}.

In the following, we describe two unit systems. Henceforth quantities in SI units will be labelled with 'SI'.

Lightspeed units

In relativistic calculations, we have been frequently using $x^0 = t = ct_{\mathrm{SI}}$. That is, we measure time in length units. Therefore the time dimension no longer explicitly appears. This means we absorb the numerical constant

$$c = 2.998 \times 10^8 \ \mathrm{m\,s^{-1}} \tag{4.52}$$

into physical quantities and constants. Previously, we already noted that ct has the dimension of length,

$$[ct] = L.$$

The ratio

$$\frac{G}{c^2} = 7.426 \times 10^{-28} \ \mathrm{kg^{-1}\ m} \tag{4.53}$$

has dimension $M^{-1}L$. We also note that

$$\left[\frac{1}{c^2}T_{\mu\nu}\right] = T^2L^{-2} \cdot ML^{-1}T^{-2} = ML^{-3}.$$

In particular,

$$[\rho] = ML^{-3} = \left[\frac{p}{c^2}\right], \tag{4.54}$$

and the 4-velocity

$$\left[\frac{u^\mu}{c}\right] = 1$$

is dimensionless.

Let us rearrange the perfect fluid stress tensor to isolate the quantities we got above:

$$\begin{aligned}\frac{8\pi G}{c^4}T_{\mu\nu} &= 8\pi\left(\frac{G}{c^2}\right)\frac{1}{c^2}\left[\left(\rho + \frac{p}{c^2}\right)u_\mu u_\nu + pg_{\mu\nu}\right]\\ &= 8\pi\left(\frac{G}{c^2}\right)\left[\left(\rho + \frac{p}{c^2}\right)\frac{u_\mu u_\nu}{c^2} + \frac{p}{c^2}g_{\mu\nu}\right].\end{aligned}$$

Therefore, if we redefine our quantities by aborbing the factors of c,

$$\frac{G}{c^2} \to G, \quad \frac{p}{c^2} \to p, \quad \frac{u^\mu}{c} \to u^\mu, \quad ct \to t, \tag{4.55}$$

and the Einstein equation now becomes

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = 8\pi G T_{\mu\nu} = 8\pi G \left[(\rho + p) \, u_\mu u_\nu + p g_{\mu\nu} \right], \tag{4.56}$$

and the time-like condition of the dimensionless 4-velocity becomes

$$u^\mu u_\mu = g_{\mu\nu} u^\mu u^\nu = -1.$$

This looks like "substituting $c = 1$" into the Einstein equation of SI units. Hence this is why lightspeed units is often called 'units where $c = 1$' in the literature. Though personally I don't like this terminology, as I would like think of c as always $c = 2.998 \times 10^8 \ \mathrm{m\,s^{-1}}$, an explicit conversion factor between units. Personal taste aside, $c = 1$ is still correct, because that is the (dimensionless) speed of light in this unit system. Nevertheless we shall call this *lightspeed units*.

To restore SI units, we simply pull the factors of c back out:

$$G \to \frac{G}{c^2}, \quad p \to \frac{p}{c^2}, \quad u^\mu \to \frac{u^\mu}{c}, \quad t \to ct. \tag{4.57}$$

The same can be done for all other physical quantities like energy, momentum, etc. This is done by essentially multiplying/dividing c in various ways to find a combination that no longer has the T dimension. Table 4.1 shows the conversion of various physical quantities.

Quantity	SI dim.	Lightpeed. dim.	Conversion
Length	L	L	$x = \tilde{x}$
Time	T	L	$t = c\tilde{t}$
Velocity	LT^{-1}	1	$v = \frac{\tilde{v}}{c}$
Acceleration	LT^{-2}	L^{-1}	$a = \frac{\tilde{a}}{c^2}$
Mass	M	M	$m = \tilde{m}$
Mass density	ML^{-3}	ML^{-3}	$\rho = \tilde{\rho}$
Energy	ML^2T^{-2}	M	$E = \frac{1}{c^2}\tilde{E}$
Momentum	MLT^{-1}	M	$p_\mu = \frac{1}{c}\tilde{p}_\mu$
Pressure	$ML^{-1}T^{-2}$	L^{-2}	$p = \frac{1}{c^2}\tilde{p}$
Gravitational constant	$M^{-1}L^3T^{-2}$	$M^{-1}L$	$G = \frac{1}{c^2}\tilde{G}$

Table 4.1: Conversion between SI and lightspeed units. Quantities with a tilde '$\sim$' denote quantities in SI units.

Geometric units

Geometric units is the convention where not only time, but *all* physical quantities are mesured in length dimensions. As we seen above, multiplying/dividing by factors of c removes T from physical quantities. We then multiply/divide with G to further remove M.

To start, using $[G/c^2] = M^{-1}L$, we see that

$$\left[\frac{G}{c^2}\rho\right] = M^{-1}L \cdot ML^{-3} = L^{-2},$$

purely has a dimension of length. Since p/c^2 has the same SI units as ρ, we also have

$$\left[\frac{G}{c^2}\frac{p}{c^2}\right] = \left[\frac{G}{c^4}p\right] = L^{-2}.$$

From earlier, we already have that u^μ/c is dimensionless. Therefore in geometric units we absorb

$$\frac{G}{c^2}\rho \to \rho, \quad \frac{G}{c^4}p \to p, \quad \frac{u^\mu}{c} \to u^\mu, \quad ct \to t.$$

In geometric units the Einstein equation is

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi T_{\mu\nu} = 8\pi\left[(\rho + p)\,u_\mu u_\nu + pg_{\mu\nu}\right],$$

and the time-like condition of the 4-velocity is

$$u^\mu u_\mu = g_{\mu\nu}u^\mu u^\nu = -1.$$

Hence, G and c do not appear explicitly when equations are written in geometric units. In the literature, this is sometimes stated as '$G = c = 1$'. To restore SI units from geometric units, we pull the factors back out:

$$\rho \to \frac{G}{c^2}\rho, \quad p \to \frac{G}{c^4}p, \quad u^\mu \to \frac{u^\mu}{c}, \quad t \to ct.$$

Table 4.2 shows the conversions between geometric and SI quantities.

Quantity	SI dim.	Geom. dim.	Conversion
Length	L	L	$x = \tilde{x}$
Time	T	L	$t = c\tilde{t}$
Velocity	LT^{-1}	1	$v = \frac{\tilde{v}}{c}$
Acceleration	LT^{-2}	L^{-1}	$a = \frac{\tilde{a}}{c^2}$
Mass	M	L	$m = \frac{G\tilde{m}}{c^2}$
Mass density	ML^{-3}	L^{-2}	$\rho = \frac{G}{c^2}\tilde{\rho}$
Energy	ML^2T^{-2}	L	$E = \frac{G}{c^4}\tilde{E}$
Pressure	$ML^{-1}T^{-2}$	L^{-2}	$p = \frac{G}{c^4}\tilde{p}$

Table 4.2: Conversion between SI and geometric units. Quantities with a tilde '$\sim$' denote quantities in SI units.

Part II

Gravitational physics

Chapter 5 The Schwarzschild solution

In this chapter we will spend some time and effort in studying a particular solution to Einstein's equation, which bears the name of Karl Schwarzschild. (Fig. 5.1.)

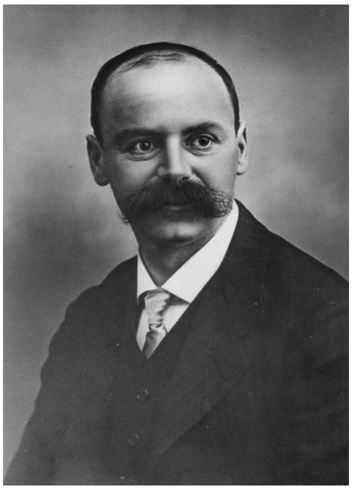


Figure 5.1: Karl Schwarzschild.

Schwarzschild's solution is perhaps the first known exact solution to Einstein's equation in the vacuum case: $T_{\mu\nu} = 0$. It was found just a few months after Einstein published his field equations in 1915, while Schwarzschild was serving in the military during World War I [21]. Tragically he died very shortly afterwards in 1916 from a skin disease at the age of 42.

5.1 Derivation of the Schwarzschild solution

In this section we will work in lightspeed units ($c = 1$). As a reminder, we absorb c into time and gravitational constant,

$$ct \to t, \quad \frac{G}{c^2} \to G.$$

Therefore Einstein's equations, in the trace-reversed form, reads as

$$R_{\mu\nu} = 8\pi G \left(T_{\mu\nu} - \frac{1}{2} T^{\lambda}{}_{\lambda} g_{\mu\nu} \right).$$

We are seeking *vacuum* solutions, meaning spacetimes where there is no mass or matter present. This means setting $T_{\mu\nu} = 0$ and the equation we wish solve is

$$R_{\mu\nu} = 0. \tag{5.1}$$

In essence, we are looking for spacetimes with a vanishing Ricci tensor. These are called Ricci-flat.

Of course, for an arbitrary metric $\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu$, Eq. (5.1) still becomes an extremely complicated set of coupled partial differential equations. In particular, a 4×4 symmetric matrix $g_{\mu\nu}$ has 10 independent components. Therefore the Einstin equation will lead to 10 second-order, coupled partial differential equations!

Therefore to make progress, we shall assume some symmetries to be carried by the spacetime.

Static, spherically-symmetric spacetimes

We will seek spacetimes which are:

- *Static.* Meaning that they are unchanging with time.
- *Spherically-symmetric.* They carry spherical symmetry, so the metric should have the same features if we change the angles of the spherical coordinate θ and ϕ.

In fact, these are reasonable assumptions for most astrophysical bodies. Our sun is approximately static and spherically-symmetric, as are most planets and stars in the galaxy.

A trivial example of a static, spherically-symmetric spacetime is Minkowski, written in spherical coordinates as

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + \mathrm{d}r^2 + r^2 \left(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\phi^2\right). \tag{5.2}$$

Note that surfaces of constant time and radius are spheres. We see this by setting $\mathrm{d}t = \mathrm{d}r = 0$ at some $r = a$ and $t = b$

$$\mathrm{d}l^2 = a^2 \left(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\phi^2\right).$$

This is precisely the metric of a sphere which we studied in previous chapters.

Now, the metric (5.2) clearly has $\Gamma^\kappa_{\mu\nu} = 0$, leading to $R_{\mu\nu} = 0$ since all the metric coefficients are constants. To find a less trivial solution, let us modify the Minkowski metric by assuming the metric coefficients are non-constant functions,

$$\mathrm{d}s^2 = -A(r)\mathrm{d}t^2 + B(r)\mathrm{d}r^2 + C(r)\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right). \tag{5.3}$$

This is still spherically-symmetric since surfaces of constant $\mathrm{d}t = 0$ and $\mathrm{d}r = 0$ are still spheres. At the moment, $A(r)$, $B(r)$, and $C(r)$ are arbitrary functions of r. They do not depend on t because of the static condition; the spacetime should not change with time.

We can reduce the number of arbitrary functions by one if we let

$$C(r) = \bar{r}^2, \quad \to \quad C'(r)\mathrm{d}r = 2\bar{r}\mathrm{d}\bar{r} = 2\sqrt{C(r)}\mathrm{d}\bar{r}.$$

The inverse transformation is

$$r = C^{-1}(\bar{r}^2) = r(\bar{r}),$$

where C^{-1} is the inverse function of C, and we can simply view r as a(n arbitrary) function of $\bar{r}$. In terms of $\bar{r}$, the metric (5.3) is

$$\mathrm{d}s^2 = -A\left(r(\bar{r})\right)\mathrm{d}t^2 + \frac{B\left(r(\bar{r})\right)C'^2}{4C(r(\bar{r}))}\mathrm{d}\bar{r}^2 + \bar{r}^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right), \tag{5.4}$$

where $A(r(\bar{r}))$ and $B(r(\bar{r}))$ are functions of $\bar{r}$, due to the composition of functions. Since all these are arbitrary functions of $\bar{r}$, we let

$$f(\bar{r}) = A(r(\bar{r})), \quad h(\bar{r}) = \frac{B\left(r(\bar{r})\right)C'^2}{4C(r(\bar{r}))}, \tag{5.5}$$

then (5.4) now becomes

$$\mathrm{d}s^2 = -f(\bar{r})\mathrm{d}t^2 + h(\bar{r})\mathrm{d}\bar{r}^2 + \bar{r}^2\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\right).$$

Finally, we rename the dummy variable $\bar{r}$ as r, so our static, spherically-symmetric ansatz is

$$\begin{aligned} \mathrm{d}s^2 &= -f(r)\mathrm{d}t^2 + h(r)\mathrm{d}r^2 + r^2\tilde{\gamma}_{ij}\mathrm{d}\theta^i\mathrm{d}\theta^j, \\ \tilde{\gamma}_{ij}\mathrm{d}\theta^i\mathrm{d}\theta^j &= \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2 \end{aligned} \tag{5.6}$$

Therefore, instead of 10 unknown components of $g_{\mu\nu}$, the (reasonable) assumptions that the spacetime is static and spherically-symmetric reduces to just unknown functions $f(r)$ and $h(r)$.

In the previous chapters, we have worked out the Ricci tensor of this metric, which we reproduce here:

$$\begin{aligned} R_{tt} &= \frac{f''}{2h} - \frac{f'^2}{4fh} - \frac{f'h'}{4h^2} + \frac{f'}{rh}, \\ R_{rr} &= -\frac{f''}{2f} + \frac{f'^2}{4f^2} + \frac{f'h'}{4fh} + \frac{h'}{rh}, \\ R_{ij} &= \tilde{\gamma}_{ij}\left[1 - \frac{1}{h} + \frac{r}{2h}\left(\frac{h'}{h} - \frac{f'}{f}\right)\right]. \end{aligned}$$

Therefore Einstein's equation $R_{\mu\nu} = 0$ simply leads to

$$\frac{f''}{2h} - \frac{f'^2}{4fh} - \frac{f'h'}{4h^2} + \frac{f'}{rh} = 0, \tag{5.7a}$$

$$-\frac{f''}{2f} + \frac{f'^2}{4f^2} + \frac{f'h'}{4fh} + \frac{h'}{rh} = 0, \tag{5.7b}$$

$$1 - \frac{1}{h} + \frac{r}{2h}\left(\frac{h'}{h} - \frac{f'}{f}\right) = 0. \tag{5.7c}$$

The first two are coupled second-order ODEs, while the third consists of only first derivatives, which is regarded as a *constraint.*

To go about solving this, we notice that Eqs. (5.7a) and (5.7b) look very similar, if it weren't for the different denominators. We can create similar terms by multiplying (5.7a) with $\frac{1}{f}$, and (5.7b) with $\frac{1}{h}$,

$$\frac{f''}{2fh} - \frac{f'^2}{4f^2h} - \frac{f'h'}{4fh^2} + \frac{f'}{rfh} = 0, \tag{5.8}$$

$$-\frac{f''}{2fh} + \frac{f'^2}{4f^2h} + \frac{f'h'}{4fh^2} + \frac{h'}{rh^2} = 0. \tag{5.9}$$

Now the first three terms of the two equations have opposite signs, but the last has the same sign. So if we take (5.8) + (5.9), we get

$$\frac{1}{rh}\left(\frac{f'}{f} + \frac{h'}{h}\right) = 0 \quad \to \quad \frac{f'}{f} + \frac{h'}{h} = 0. \tag{5.10}$$

This equation is satisfied if $h = \frac{1}{f}$, since

$$\begin{aligned} h' &= -\frac{f'}{f^2} = -f\frac{f'}{f} = -\frac{1}{h}\frac{f'}{f} \\ \frac{h'}{h} &= -\frac{f'}{f}. \end{aligned}$$

With this expression for h, Eq. (5.7) now becomes

$$\frac{1}{2}\left(f'' + \frac{2}{r}f'\right) = 0, \tag{5.11a}$$

$$-\frac{1}{2}\left(f'' + \frac{2}{r}f'\right) = 0, \tag{5.11b}$$

$$1 - f - rf' = 0. \tag{5.11c}$$

Note that (5.11a) is identical to (5.11b). So we are now down to a single differential equation for a single function $f(r)$ and a first-order constraint,

$$f'' + \frac{2}{r}f' = 0, \tag{5.12a}$$

$$1 - f - rf' = 0. \tag{5.12b}$$

First, we consider Eq. (5.12a). It looks like an outcome of a product rule. The coefficient 2 suggests that it came out of differentiating $r^2 f'$. We can check that (5.12a) is equivalent to

$$\frac{1}{r^2}\left(r^2 f'\right)' = 0. \tag{5.13}$$

This means $r^2 f'$ is a constant,

$$\begin{aligned} r^2 f' &= C \\ f' &= \frac{C}{r^2} \\ f &= \int \frac{C}{r^2}\mathrm{d}r = -\frac{C}{r} + B, \end{aligned}$$

where B and C are integration constants. We have now fixed $f(r) = B - C/r$ and $h(r) = 1/f = 1/\left(B - C/r\right)$, but it should be checked that Eq. (5.12b) is satisfied. Substituting the solution into (5.12b), we see that it reduces to

$$1 - B = 0.$$

This constraint fixes the value of B. So there is only one arbitrary constant left which parametrises the Schwarzschild solution,

$$\mathrm{d}s^2 = -f(r)\mathrm{d}t^2 + \frac{\mathrm{d}r^2}{f(r)} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right), \tag{5.14a}$$

$$f(r) = 1 - \frac{C}{r}. \tag{5.14b}$$

To determine the physical meaning of the constant C, we recall that in Chapter 4, in the weak gravity case, the metric is a perturbation of the Minkowski metric,

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

and that the $tt = 00$ component is $h_{tt} = h_{00} = -2\Phi$[1] where Φ is the Newtonian potential. Therefore we see that

$$\frac{C}{r} = -2\Phi.$$

In this regime for a spherically-symmetric configuration, the Newtonian potential is $\Phi = -GM/r$. Therefore we conclude that

$$C = 2GM.$$

We have now derived the Schwarzschild solution, in lightspeed units

$$\boxed{\begin{aligned} \mathrm{d}s^2 &= -f(r)\mathrm{d}t^2 + f(r)^{-1}\mathrm{d}r^2 + r^2\mathrm{d}\Omega^2, && (5.15\text{a}) \\ f(r) &= 1 - \frac{2GM}{r}, \quad \mathrm{d}\Omega^2 = \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2 && (5.15\text{b}) \end{aligned}}$$

Restoring to SI units, we replace $t \to ct$ and $G \to G/c^2$,

$$\begin{aligned} \mathrm{d}s^2 &= -f(r)c^2\mathrm{d}t^2 + f(r)^{-1}\mathrm{d}r^2 + r^2\mathrm{d}\Omega^2, \\ f(r) &= 1 - \frac{2GM}{c^2 r}. \end{aligned}$$

5.2 Properties of the Schwarzschild spacetime

Asymptotic flatness

The Schwarzschild spacetime describes the gravitational field due to a spherically-symmetric mass distribution of total mass M. Indeed, we observe that for $r \gg 2GM$, the function $f(r)$ tends to 1 and the metric becomes approximately Minkowski,

$$\mathrm{d}s^2 \simeq -\mathrm{d}t^2 + \mathrm{d}r^2 + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right), \quad r \gg 2GM.$$

Far from the mass, the gravitational field becomes negligible. Roughly speaking, spacetimes with these properties are called *asymptotically flat*.

[1] Recall that we are using lightspeed units, $c = 1$.

Singularities

How about small r? Notice that $f(r)$ vanishes at $r = 2GM$. The component g_{tt} vanishes and g_{rr} diverges. This is a *singularity* of the spacetime, and this is an important location called the *Schwarzschild radius*,

$$R_s = 2GM. \tag{5.16}$$

In SI units, the Schwarzschild radius is $\frac{2GM}{c^2}$. The larger the mass of the system, the larger its Schwarzschild radius.

There is also another singularity at $r = 0$. Where now g_{tt} diverges and g_{rr} vanishes.

When a metric becomes singular, it is important to check whether this singularity is merely a bad choice of coordinates (coordinate singularity), or if the singularity is physical. An example of a coordinate singularity occurs for the sphere S^2,

$$\mathrm{d}s^2 = \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2.$$

At $\theta = 0$ or $\theta = \pi$, the component $g_{\phi\phi}$ vanishes. This is simply due to the coordinate system used, as we know those two points are the north and south poles of the sphere, and the whole sphere is finite and smooth.

The metric components $g_{\mu\nu}$ depend on the choice of coordinates. So to check if a singularity is physical, we should look at scalar quantities, which do not depend on coordinates. Scalars constructed out of the metric are

$$R, \quad R_{\mu\nu}R^{\mu\nu}, \quad R_{\rho\sigma\mu\nu}R^{\rho\sigma\mu\nu}, \quad R_{\rho\sigma\mu\nu}R^{\rho\sigma}R^{\mu\nu}, \ldots$$

These are called *curvature invariants*. Since the Schwarzschild solution satisfies $R_{\mu\nu} = 0$, anything with the Ricci scalar and Ricci tensor are immediately zero. However, the Kretschmann invariant $R_{\rho\sigma\mu\nu}R^{\rho\sigma\mu\nu}$ may not be zero. Computing this for the Schwarzschild metric,

$$R_{\rho\sigma\mu\nu}R^{\rho\sigma\mu\nu} = \frac{48G^2M^2}{r^6}. \tag{5.17}$$

We see that the Kretschmann invariant is perfectly smooth and finite at the Schwarzschild radius $r = 2GM$. But blows up at $r = 0$. Therefore the origin $r = 0$ is a genuine *curvature singularity* of the spacetime.

We will discuss black holes in later sections. But here we comment that the Schwarzschild radius marks the event horizon of the Schwarzschild black hole. And here we see that the event horizon encloses the curvature singularity.

Continuing into $0 < r < 2GM$, the function $f(r)$ becomes negative. So g_{tt} becomes positive and g_{rr} negative. The metric is now

$$\mathrm{d}s^2 = \left(\frac{2GM}{r} - 1\right)\mathrm{d}t^2 - \left(\frac{2GM}{r} - 1\right)^{-1}\mathrm{d}r^2 + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right). \tag{5.18}$$

Since g_{rr} is the negative component, this implies that r is now the 'time' coordinate, and t is a 'space' coordinate! Since the metric components depend on r, this spacetime is non-static inside the Schwarzschild radius $0 < r < 2GM$.

Clearly strange things happen if we cross the Schwarzschild radius. These are phenomena related to *black holes*, and we will discuss black holes in more detail in later sections.

As an exterior model of a star and Birkhoff's Theorem

The Schwarzschild spacetime can be used to describe the gravitational field outside a star. Inside a star, $T_{\mu\nu}$ is non-zero and the metric is no longer described by the Schwarzschild metric. But *outside*, $T_{\mu\nu} = 0$ and therefore the Schwarzschild spacetime describes the spacetime outside a star, or any non-zero spherically-symmetric mass distribution.

We can model the stellar material as a perfect fluid. Letting R_0 be the radius of the star, we may write the stress tensor as

$$T_{\mu\nu} = \begin{cases} (\rho + p)\, u_\mu u_\nu + p g_{\mu\nu}, & 0 < r \leq R_0, \\ 0, & r > R. \end{cases} \tag{5.19}$$

Therefore the spacetime can also be split into two part,

$$\mathrm{d}s^2 = \begin{cases} -\left(1 - \frac{2GM}{r}\right)\mathrm{d}t^2 + \left(1 - \frac{2GM}{r}\right)^{-1}\mathrm{d}r^2 + r^2\mathrm{d}\Omega^2, & r > R_0, \\ -f(r)\mathrm{d}t^2 + h(r)\mathrm{d}r^2 + r^2\mathrm{d}\Omega^2, & 0 < r < R_0. \end{cases} \tag{5.20}$$

where $f(r)$ and $h(r)$ needs to be determined from the Einstein equation. (We will not do this here, but see Sec. 10.5 of [3].)

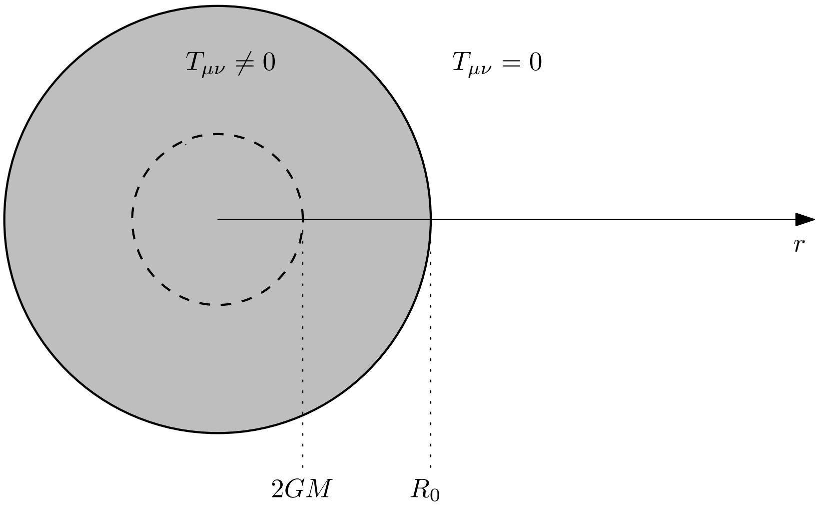


Figure 5.2: Sketch of the model of a spherically-symmetric star of mass M and radius $R_0 > 2GM$ (in lightspeed units). For $r > R_0$, the spacetime is described by the Schwarzschild metric.

Since a star is clearly not a black hole, the outside Schwarzschild metric is valid as long as the radius of the star is larger than it Schwarzschild radius, as sketched in Fig. 5.2.

In fact, there is a theorem which states that the metric outside *any* spherically-symmetric mass distribution is the Schwarzschild metric. This is Birkhoff's theorem.

Theorem 5.2.1 (Birkhoff's Theorem)**.** The metric outside any spherically-symmetric matter distribution is given by the Schwarzschild metric.

This theorem is very useful, as most astrophysical bodies are spherically symmetric, and so their gravitational fields are well-approximated by the Schwarzschild metric. This holds true for the gravity outside the surfaces of the Sun, Earth, Moon, and all the planets. As well as most stars.

An important condition is that the radius of the star is larger than its Schwarzschild radius $2GM$, because if the mass is concentrated in a region less than $2GM$, we will have the coordinate singularity and the metric is no longer static inside. (Actually it becomes a black hole, which we will discuss later.) As a rough check, let us consider the Sun, which has a mass

$$M_\odot = 1.989 \times 10^{30} \text{ kg}. \tag{5.21}$$

Recall that in lightspeed units, $G \to G_{\text{SI}}/c_{\text{SI}}^2 = 7.426 \times 10^{-28} \text{ kg}^{-1}\,\text{m}$. Therefore the Schwarzschild radius of the sun is

$$R_S = 2GM\odot = 2953 \text{ m} \simeq 3 \text{ km}. \tag{5.22}$$

The radius of the sun is $R_\odot = 696\,340$ km which is way bigger than the Schwarzschild radius. Therefore the spacetime outside the sun is regular and free of singularities. (Fortunately for us.)

5.3 Time-like geodesics

Equations of motion

In Chapter 3, we established that test particles[2] of mass m follow time-like geodesics. Suppose that the trajectory of the particle $x^\mu(\tau) = (t(\tau), r(\tau), \theta(\tau), \phi(\tau))$, parametrised by the particle's proper time τ. Its 4-velocity is therefore

$$u^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} = \dot{x}^\mu = \left(\dot{t}, \dot{r}, \dot{\theta}, \dot{\phi}\right),$$

with the normalisation condition (in lightspeed units)

$$u_\mu u^\mu = -f\dot{t}^2 + \frac{\dot{r}^2}{f} + r^2\dot{\theta}^2 + r^2\sin^2\theta\dot{\phi}^2 = -1. \tag{5.23}$$

We may start with the Lagrangian

$$\mathcal{L} = \frac{1}{2}m\left(-f\dot{t}^2 + \frac{\dot{r}^2}{f} + r^2\dot{\theta}^2 + r^2\sin^2\theta\dot{\phi}^2\right), \tag{5.24}$$

where m is the mass of the particle. The canonical momenta are computed from $p_\mu = \frac{\partial\mathcal{L}}{\partial\dot{x}^\mu}$, they are

$$p_t = -mf\dot{t}, \quad p_r = \frac{m\dot{r}}{f}, \quad p_\theta = mr^2\dot{\theta}, \quad p_\phi = mr^2\sin^2\theta\dot{\phi}.$$

The equations of motion (i.e., the geodesic equations) are determined from the Euler–Lagrange equation $\frac{\mathrm{d}}{\mathrm{d}\tau}\frac{\partial\mathcal{L}}{\partial\dot{x}^\mu} = \frac{\partial\mathcal{L}}{\partial x^\mu}$. First, we observe that the metric, and hence the Lagrangian, is completely independent of t and ϕ. Therefore $\frac{\partial\mathcal{L}}{\partial t} = 0$ and $\frac{\partial\mathcal{L}}{\partial\phi} = 0$. The Euler–Lagrange equation for these two are

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\tau}\frac{\partial\mathcal{L}}{\partial\dot{t}} = 0 \quad &\rightarrow \quad p_t = \frac{\partial\mathcal{L}}{\partial\dot{t}} = -mf\dot{t} = \text{constant},\\
\frac{\mathrm{d}}{\mathrm{d}\tau}\frac{\partial\mathcal{L}}{\partial\dot{\phi}} = 0 \quad &\rightarrow \quad p_\phi = \frac{\partial\mathcal{L}}{\partial\dot{\phi}} = mr^2\sin^2\theta\dot{\phi} = \text{constant},
\end{aligned}$$

[2]Note that *test* particles mean that their mass is small enough that it doesn't exert its own gravitational field.

We let these constants be equal to $-\mathcal{E}$ and J respectively, so that

$$\dot{t} = \frac{\mathcal{E}}{mf}, \quad \dot{\phi} = \frac{J}{mr^2 \sin^2\theta}.$$

We will see later that $\mathcal{E}$ and J are related to the energy and angular momentum of the particle.

In PHY201 *Theoretical Mechanics*, we have learned that for a spherically-symmetric Lagrangian, the motion of the particle is confined on a plane. We can also see this by considering the Euler–Lagrange equation for θ,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\tau}\frac{\partial\mathcal{L}}{\partial\dot{\theta}} &= \frac{\partial\mathcal{L}}{\partial\theta} \\ \frac{\mathrm{d}}{\mathrm{d}\tau}\left(mr^2\dot{\theta}\right) &= mr^2\sin\theta\cos\theta\dot{\phi}^2. \end{aligned}$$

This equation is always satisfied if $\theta = \frac{\pi}{2} =$ constant. We can always choose a coordinate system such that the particle moves on the plane $\theta = \frac{\pi}{2}$.

The Euler–Lagrange equation for r is

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\tau}\frac{\partial\mathcal{L}}{\partial\dot{r}} &= \frac{\partial\mathcal{L}}{\partial r} \\ \frac{\mathrm{d}}{\mathrm{d}\tau}\left(\frac{\dot{r}}{f}\right) &= -\frac{1}{2}f'\dot{t}^2 - \frac{f'\dot{r}^2}{2f^2} + r\dot{\theta}^2 + r\sin^2\theta\dot{\phi}^2 \\ \frac{\ddot{r}}{f} - \frac{f'\dot{r}^2}{f^2} &= -\frac{f'\dot{r}^2}{2f} + r\dot{\theta}^2 - \frac{f'\mathcal{E}^2}{2m^2f} + \frac{fJ^2}{m^2r^3\sin^2\theta} \\ \ddot{r} &= \frac{f'\dot{r}}{2f} + r\dot{\theta}^2 - \frac{f'\mathcal{E}^2}{2m^2f} + \frac{fJ^2}{m^2r^3\sin^2\theta}. \end{aligned}$$

Since we have shown the polar angle to be constant at $\theta = \frac{\pi}{2}$, the equation for r is

$$\ddot{r} = \frac{f'\dot{r}}{2f} - \frac{f'\mathcal{E}^2}{2m^2f} + \frac{fJ^2}{m^2r^3\sin^2\theta}.$$

This form is suitable for numerical integration. However, to analyse the properties of the solution, we consider the time-like normalisation condition (5.23). Substituting $\dot{t}$ and $\dot{\phi}$ in terms of E and L, along with $\theta = \frac{\pi}{2}$ and rearranging, we find

$$\frac{\dot{r}^2}{f} - \frac{\mathcal{E}^2}{m^2f} + \frac{J^2}{m^2r^2} = -1. \tag{5.25}$$

Further rearranging, we find

$$\begin{aligned}
\dot{r}^2 &= \frac{\mathcal{E}^2}{m^2} - \left(\frac{J^2}{m^2 r^2} + 1\right)\left(1 - \frac{2GM}{r}\right) \\
\dot{r}^2 &= \frac{\mathcal{E}^2}{m^2} - \underbrace{\left(1 - \frac{2GM}{r} + \frac{J^2}{m^2 r^2} - \frac{2GMJ^2}{m^2 r^3}\right)}_{U_{\text{eff}}}.
\end{aligned}$$

This looks like an equation of a Newtonian particle of mass 2 with total energy $\mathcal{E}^2/m^2$ moving in a potential

$$U_{\text{eff}} = 1 - \frac{2GM}{r} + \frac{J^2}{m^2 r^2} - \frac{2GMJ^2}{m^2 r^3}. \tag{5.26}$$

This is called the *effective potential* of the motion.

Throughout our calculations, we notice that $\mathcal{E}/m$ and J/m always appear together. Therefore it is convenient to write

$$E = \frac{\mathcal{E}}{m}, \quad L = \frac{J}{m}. \tag{5.27}$$

Where E is dimensionless and L has dimensions of length.

Collecting our results, the equations of motion of a particle in Schwarzschild spacetime are

$$\boxed{\begin{aligned}
\dot{t} &= \frac{E}{f}, \quad \dot{\phi} = \frac{L}{r^2}, \quad \theta = \frac{\pi}{2} = \text{constant}, && \text{(5.28a)} \\
\ddot{r} &= \frac{f'\dot{r}}{2f} - \frac{f'E^2}{2f} + \frac{fL^2}{r^3}, && \text{(5.28b)} \\
\dot{r}^2 &= E^2 - \underbrace{\left(1 - \frac{2GM}{r} + \frac{L^2}{r^2} - \frac{2GML^2}{r^3}\right)}_{U_{\text{eff}}}. && \text{(5.28c)}
\end{aligned}}$$

Analysis of time-like particle motion

To begin analysing the problem of time-like geodesics, let us take stock of the parameters of the system. The geodesics are parametrised by its constants of motion E and L. Here E is dimensionless and L has dimension of length. The Schwarzschild spacetime itself is governed by the mass M. Multiplying with the gravitational constant, the quantity GM (half the Schwarzschild radius) has a dimension of length, and it sets length scale of the spacetime. So all parameters can be measured in length units of GM.

The effective potential equation (5.28c) is very useful in analysing the behaviour of the particle. First, we see that $U_{\text{eff}} \to 1$ as $r \to \infty$. So particles with energy $E^2 > 1$ can escape to infinity, and particles with $0 < E^2 < 1$ are bounded.

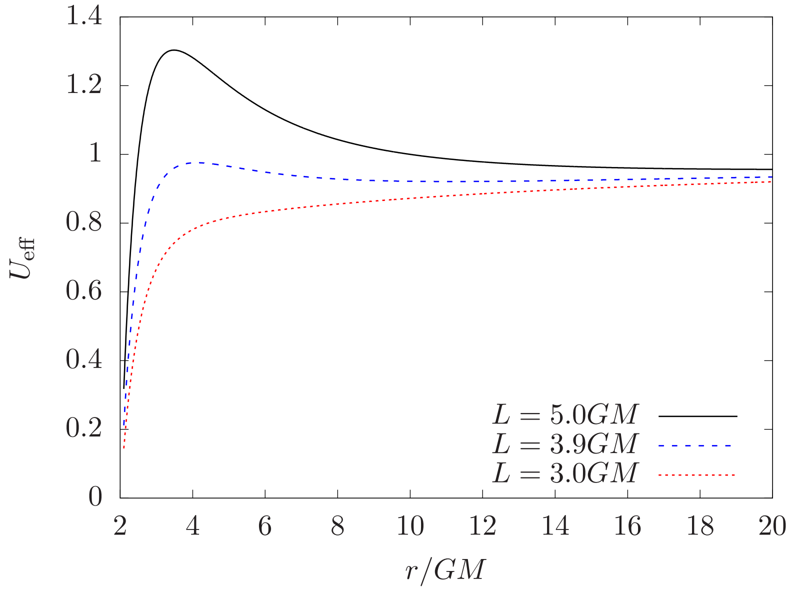


Figure 5.3: Plots of U_{eff} vs r for various values of angular momenta.

In Fig. 5.3, the effective potentials for various values of angular momenta L are plotted. We see that as L is increased, there is a growing potential barrier. Particles of energy E less than the height of the barrier will be prevented from falling into the black hole.

To obtain an orbit explicitly, we choose values for the constants E and L. The turning points of the motion occur when $E = U_{\text{eff}}$, the location where $\dot{r} = 0$. The number of turning points determine whether the orbit is bounded, escaping, or falling into the black hole.

Let us run through an explicit example. Suppose we choose

$$E = 0.9685, \quad L = 3.9GM.$$

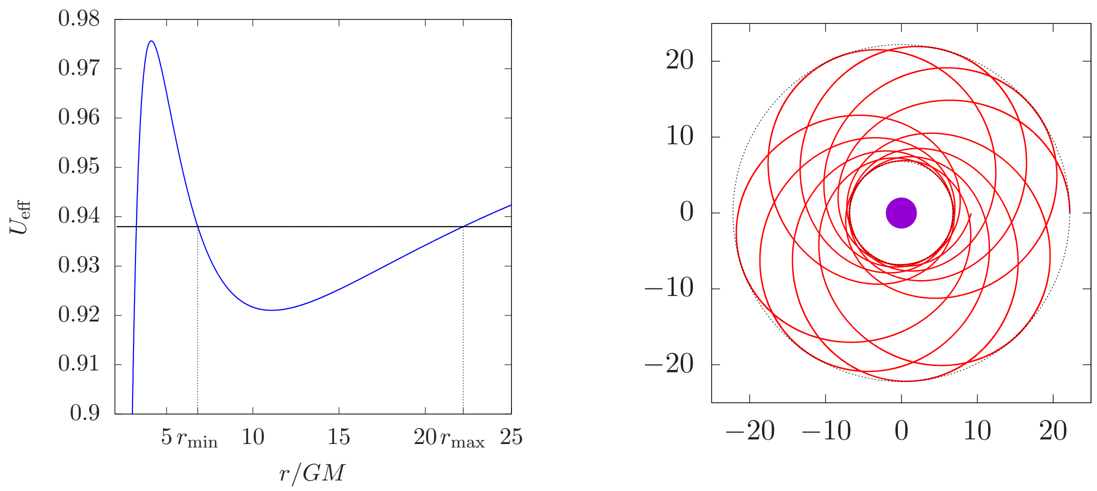
Substituting this and $\dot{r} = 0$ into Eq. (5.28c) gives three roots, $3.246146212GM$, $6.807431572GM$, and $22.20045498GM$. By sketching the graph, we see that $E - U_{\text{eff}} = \dot{r}^2 \geq 0$ between the two larger roots, as seen in Fig. 5.4a. $\dot{r}$ is also positive for r less than the smaller root, but this region contains the Schwarzschild radius, so particles there will fall into the black hole. Hence we have found that the particle is bounded between the radii

$$r_{\min} \leq r \leq r_{\max}; \quad r_{\min} = 6.807431572GM, \quad r_{\max} = 22.20045498GM.$$

To plot the orbit explicitly, we need to choose initial conditions to solve the differential equation. Let's suppose that the particle starts at $r = r_{\max}$, then we know $\dot{r} = 0$. This fixes the initial conditions. With these initial conditions, one can solve Eq. (5.28b) numerically (for example, by applying the fourth-order Runge–Kutta algorithm) and the results are plotted in Cartesian coordinates with

$$X = r\cos\phi, \quad Y = r\sin\phi.$$

Fig. 5.4b shows the results of the numerical integration.



(a) U_{eff} vs r.

(b) Orbit.

Figure 5.4: Orbit of a particle with $E = 0.9685$ and $L = 3.9GM$. In this case $r_{\min} = 6.8074GM$ and $r_{\max} = 22.200GM$. Fig. 5.4a shows the effective potential graph, it intersects the energy at $U_{\text{eff}} = E^2 = 0.9380$. The orbit is plotted in Fig. 5.4b; we can see that the trajectory is bounded between the circles $r_{\min} \le r \le r_{\max}$. The shaded circle marks the location inside the Schwarzschild radius.

Circular orbits

An important type of relativistic orbits is the circular orbit. They are important, for example, in the study of accretion disks around a black hole.

By observing Fig. 5.3 or Fig. 5.4a, we notice that U_{eff} sometimes have a local minimum. Take a particle in bounded orbit. If we lower the energy, the domain $r_{\min} \le r \le r_{\max}$ gets narrower and narrower. Until the energy is lowered to a local minimum of U_{eff}, the domain shrinks to a point, and the particle gets 'stuck' at constant radius.

Therefore, the locations of circular orbits are determined by the local minima of U_{eff}.

Differentiating,

$$U'_{\text{eff}} = \frac{2GM}{r^2} - \frac{2L^2}{r^3} + \frac{6GML^2}{r^4} = \frac{2}{r^4}\left(2GMr^2 - 2L^2 r + 6GML^2\right).$$

Let r_0 be the radius of the circular orbit. Then its value is determined by the roots of $2GMr^2 - 2L^2 r + 6GML^2 = 0$. This is a quadratic equation. A choice of L determines the roots. But since L is an undetermined parameter anyway, it is easier to solve for L^2,

$$L_0 = r_0\sqrt{\frac{GM}{r_0 - 3GM}}. \tag{5.29}$$

Without loss of generality, we assume L is positive.[3] Substituting this into $E^2 = U_{\text{eff}}$, we can also determine the energy of the circular orbit to be

$$E_0 = \frac{r_0 - 2GM}{\sqrt{r_0(r_0 - 3GM)}}. \tag{5.30}$$

From Classical Mechanics, refer to $U'_{\text{eff}} = 0$ as an equilibrium point of the system. The equilibrium here is in the radial displacement of the particle. The *stability* of the equilibrium depends on whether U_{eff} is a minimum or maximum. Therefore let us check the second derivative

$$\begin{aligned} U''_{\text{eff}} = -\frac{4GM}{r^3} + \frac{6L^2}{r^4} - \frac{24GML^2}{r^5} &= \frac{2}{r^5}\left(-2GMr^2 + 3L^2 r - 12GML^2\right) \\ &= \frac{2}{r^5}\left[3L^2(r - 4GM) - 2GMr^2\right] \end{aligned}$$

We evaluate this at the location $U'_{\text{eff}} = 0$. There, we know that L is given by (5.29). We then have

$$\begin{aligned} U''_{\text{eff}} &= \frac{2}{r_0^5}\left[3r_0^2\frac{GM}{r_0 - 3GM}(r_0 - 4GM) - 2GMr_0^2\right] \\ &= \frac{2}{r_0^5}\left[\frac{3GMr_0^2(r_0 - 4GM) - 2GMr_0^2(r_0 - 3GM)}{r_0 - 3GM}\right] \\ &= \frac{2}{r_0^5}\left[\frac{r_0^2\,(r_0 - 6GM)}{r_0 - 3GM}\right]. \end{aligned}$$

The condition for stable equilibrum is $U_{\text{eff}} > 0$. Therefore circular orbits of radii

$$r_{\text{ISCO}} < r_0 < \infty; \quad r_{\text{ISCO}} = 6GM$$

are stable, where $r_{\text{ISCO}} = 6GM$ is the *innermost stable circular orbit* (ISCO). Circular orbits of radii $3GM < r_0 < r_{\text{ISCO}} = 6GM$ are unstable. Orbits with $r_0 < 3GM$ do not

[3] The case $L < 0$ is equivalent to reflecting the coordinate system by $\phi \to -\phi$.

exist because L_0 and E_0 becomes complex.

The values of E and L of circular orbits are shown in Fig. 5.5.

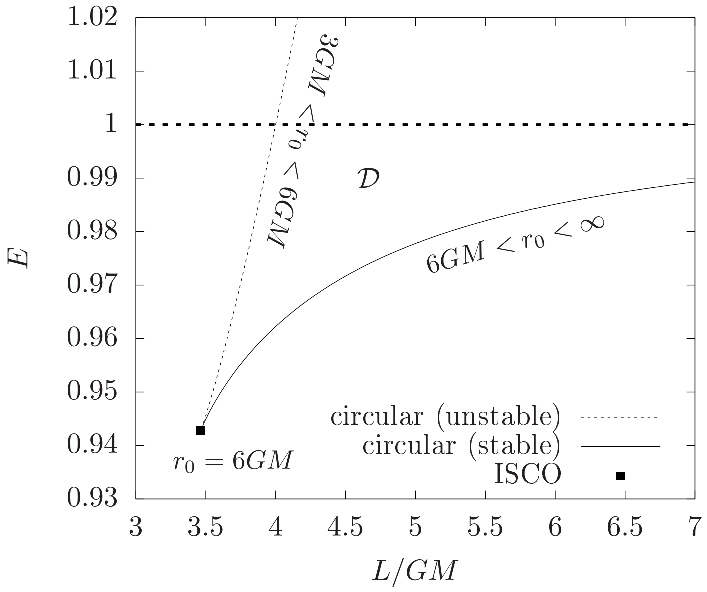


Figure 5.5: Energies E and angular momenta L of circular orbits. The solid and dotted curves indicate stable and unstable orbits, respectively. The region $\mathcal{D}$ bounded by curves of circular orbit and the line $E = 1$ are values for bounded, non-escaping orbits.

Exact solution

Taking Eq. (5.28a) and (5.28c) applying the chain rule, we have

$$\begin{aligned}\frac{\dot{r}}{\dot{\phi}} = \frac{\mathrm{d}r}{\mathrm{d}\phi} &= \pm\frac{r^2}{L}\sqrt{E^2 - 1 + \frac{2GM}{r} - \frac{L^2}{r^2} + \frac{2GML^2}{r^3}} \\ &= \pm r^2\sqrt{\frac{-(1-E^2)}{L^2} + \frac{2GM}{L^2 r} - \frac{1}{r^2} + \frac{2GM}{r^3}}.\end{aligned} \tag{5.31}$$

By introducing the substitution $r = 1/u$, the equation becomes

$$\frac{\mathrm{d}u}{\mathrm{d}\phi} = \mp\sqrt{P(u)},$$

where $P(u)$ is a degree-3 polynomial

$$P(u) = -\frac{1-E^2}{L^2} + \frac{2GM}{L^2}u - u^2 + 2Mu^3. \tag{5.32}$$

Note that the roots of $P(u) = 0$ is equivalent to the turning points where $E^2 = U_{\text{eff}}$, and the motion of the particle lies in the domain where $P(u) \geq 0$. As a degree-3 polynomial, its roots can always, in principle, be found. Let us consider the case when all 3 roots are real,

$$u_1 \leq u_2 \leq u_3.$$

These occur when E and L takes values in region $\mathcal{D}$ of Fig. 5.5. Furthermore the leading coefficient of $P(u)$ is positive, so the graph of $P(u)$ looks like Fig. 5.6.

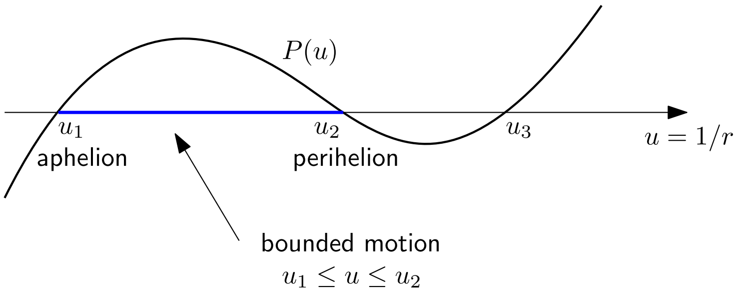


Figure 5.6: Sketch of $P(u)$ defined in Eq. (5.32).

Bounded motion occurs in $u_1 \leq u \leq u_2$. The polynomial is factorised as

$$P(u) = 2GM(u_3 - u)(u_2 - u)(u - u_1). \tag{5.33}$$

Since $r = 1/u$, the smaller root corresponds to $r_{\max} = 1/u_1$ and the larger root is $r_{\min} = 1/u_2$. Let us choose initial conditions such that the particle starts at $r_{\max}$, and take the lower sign of the square root.

Separating the variables and integrating, we find

$$\int_0^\phi \mathrm{d}\phi' = \frac{1}{\sqrt{2GM}} \int_{u_1}^{u} \frac{\mathrm{d}u'}{\sqrt{(u_3 - u')(u_2 - u')(u' - u_1)}}$$

This integral is listed in (pg 254, 3.131-3) of [22]. Therefore the solution is

$$\phi(u) = \frac{2}{\sqrt{2GM(u_3 - u_1)}} \mathrm{F}\left(\arcsin\sqrt{\frac{u - u_1}{u_2 - u_1}}, \sqrt{\frac{u_2 - u_1}{u_3 - u_1}}\right),$$

where $\mathrm{F}(\varphi, k)$ is the incomplete elliptic integral of the first kind.

Inverting gives $1/u(\phi) = r(\phi)$, a parametric solution in polar coordinates,

$$r(\phi) = \frac{1}{u_1 + (u_2 - u_1)\mathrm{sn}^2\left(\sqrt{2GM(u_3 - u_1)}\,\frac{\phi}{2}, \sqrt{\frac{u_2 - u_1}{u_3 - u_1}}\right)}. \tag{5.34}$$

where $\mathrm{sn}(\vartheta, k)$ is the Jacobi sine function of elliptic modulus k. This gives the exact solution describing time-like Schwarzschild geodesics. For details utilising the exact solution, see [8, 23].

Non-relativistic limit

We have argued with the support of Birkhoff's theorem that the spacetime outside the Sun or the Earth is approximated by the Schwarzschild metric. Though for the solar system, it is well known that planetary orbits are described to good accuracy by non-relativistic classical mechanics.

The equations of Schwarzschild geodesic should contain this non-relativistic limit. To find it, we look at Eq. (5.28c). We have been calculating using lightspeed units where $c = 1$. To recover the non-relativistic limit we should restore SI units since velocities should be slower than light.

First, restore

$$E = \frac{\mathcal{E}}{m}, \quad L = \frac{J}{m},$$

where $\mathcal{E}$ and J were our energy and angular momentum in lightspeed units. To restore SI units, we pull the factors of c back out:

$$\mathcal{E} \to \frac{\mathcal{E}}{c^2}, \quad J \to \frac{J}{c}, \quad G \to \frac{G}{c^2} \quad \tau \to c\tau, \quad t \to ct.$$

Then Eq. (5.28c) restored to SI units is

$$\begin{aligned}
\dot{r}^2 &= E^2 - \left(1 - \frac{2GM}{r} + \frac{L^2}{r^2} - \frac{2GML^2}{r^3}\right) \\
&\downarrow \\
\frac{1}{c^2}\dot{r}^2 &= \frac{\mathcal{E}^2}{m^2c^4} - \left(1 - \frac{2GM}{c^2 r} + \frac{J^2}{m^2c^2r^2} - \frac{2GMJ^2}{m^2c^4}\right) \\
\dot{r}^2 &= \frac{\mathcal{E}^2}{m^2c^2} - \left(c^2 - \frac{2GM}{r} + \frac{J^2}{m^2r^2} - \frac{2GMJ^2}{m^2c^2r^3}\right)
\end{aligned}$$

Furthermore, we have

$$\mathcal{E} = e + mc^2,$$

where e is the Newtonian energy of the particle.[4] In the non-relativistic limit, e is small

[4] We are running out of symbols to denote the various energies.

compared to mc^2. Hence we have

$$\begin{aligned}
\dot{r}^2 &= \frac{(e+mc^2)^2}{m^2c^2} - c^2 - \left(-\frac{2GM}{r} + \frac{J^2}{m^2r^2} - \frac{2GMJ^2}{m^2c^2r^3}\right) \\
&= \frac{e^2 + 2emc^2 + m^2c^4}{m^2c^2} - c^2 - \left(-\frac{2GM}{r} + \frac{J^2}{m^2r^2} - \frac{2GMJ^2}{m^2c^2r^3}\right) \\
&= \frac{e^2}{m^2c^2} + \frac{2e}{m} + \cancel{c^2} - \cancel{c^2} - \left(-\frac{2GM}{r} + \frac{J^2}{m^2r^2} - \frac{2GMJ^2}{m^2c^2r^3}\right) \\
\frac{1}{2}m\dot{r}^2 &= e + \frac{e^2}{2mc^2} - \left(-\frac{GM}{mr} + \frac{J^2}{2mr^2} - \frac{GMJ^2}{mc^2r^3}\right)
\end{aligned}$$

In the non-relativistic limit, $\frac{e^2}{2mc^2}$ and $\frac{GMJ^2}{mc^2r^3}$ becomes negligible. Furthermore in the non-relativistic limit the proper time becomes equal to coordinate time,

$$\dot{r} = \frac{\mathrm{d}r}{\mathrm{d}\tau} = \frac{\mathrm{d}r}{\mathrm{d}t}.$$

Then this equation reduces to

$$\frac{1}{2}m\dot{r}^2 = e - \left(-\frac{GMm}{r} + \frac{J^2}{2mr^2}\right)$$

This is exactly the equation for the central force problem

$$\frac{1}{2}m\dot{r}^2 = e - V_{\text{eff}}, \quad V_{\text{eff}} = -\frac{GMm}{r} + \frac{J^2}{2mr^2},$$

as one can check, for example, in Eq. (5.17) of *PHY201 Theoretical Mechanics*. It is a remarkable sigh of relief that General Relativity, a theory foundationally vastly different from Newtonian gravitation, still reduces to a limit that agrees with Newton. It is all to easy to come up with a radical new theory. (Many people have tried to become 'The Next Einstein'.) But any new theory must be discarded if it doesn't agree in the Newtonian limit, since this limit has been observationally and experimentally confirmed.

Perihelion shift

The explanation of Mercury's perihelion shift is one of the early observational confirmation of General Relativity. In the following treatment we follow [6]. We rewrite Eq. (5.31) as

$$\left(\frac{\mathrm{d}u}{\mathrm{d}\phi}\right)^2 = P(u) = \frac{-(1-E^2)}{L^2} + \frac{2GM}{L^2}u - u^2 + 2GMu^3.$$

When we take the non-relativistic limit, it is the last term $2GMu^3$ that vanishes to recover the Kepler equation. Therefore the remaining terms are the ones that reduces to Newton.

So we keep track of this term.

Differentiating both sides with respect to ϕ,

$$
\begin{aligned}
2\frac{\cancel{\mathrm{d}u}}{\cancel{\mathrm{d}\phi}}\frac{\mathrm{d}^2u}{\mathrm{d}\phi^2} &= P'(u)\frac{\cancel{\mathrm{d}u}}{\cancel{\mathrm{d}\phi}} \\
2\frac{\mathrm{d}^2u}{\mathrm{d}\phi^2} &= \frac{2GM}{L^2} - 2u + 6GMu^2 \\
\frac{\mathrm{d}^2u}{\mathrm{d}\phi^2} &= \frac{GM}{L^2} - u + 3GMu^2.
\end{aligned}
$$

This is a second-order differential equation for u.

As mentioned earlier, if the last term $3GMu^2$ is absent, we would get the non-relativistic case, for which the differential equation is

$$
\frac{\mathrm{d}^2u}{\mathrm{d}\phi^2} = \frac{GM}{L^2} - u \quad \text{(non-relativistic case)}
$$

The solution to the non-relativistic case is simply

$$
u_0 = \frac{GM}{L^2}\left(1 + e\cos\phi\right).
$$

This is the equation of the Kepler ellipse with the constant e being the eccentricity.

To get relativistic corrections, we expand about the non-relativistic solution by writing

$$
u = u_0 + u_1 = \frac{GM}{L^2}\left(1 + e\cos\phi\right) + u_1.
$$

Substituting into the differential equation,

$$
\begin{aligned}
\frac{\mathrm{d}^2u_0}{\mathrm{d}\phi^2} + \frac{\mathrm{d}^2u_1}{\mathrm{d}\phi^2} &= \frac{GM}{L^2} - u_0 - u_1 + 3GM(u_0+u_1)^2 \\
\frac{\mathrm{d}^2u_1}{\mathrm{d}\phi^2} &= -u_1 + 3GM\left(u_0^2 + 2u_0u_1 + u_1^2\right).
\end{aligned}
$$

We are considering u_1 to be very small such that GMu_1 and u_1^2 are negligible. Then we have

$$\begin{aligned}\frac{\mathrm{d}^2u_1}{\mathrm{d}\phi^2} &= -u_1 + 3GMu_0^2 = -u_1 + \frac{3G^3M^3}{L^4}\left(1+e\cos\phi\right)^2 \\ &= -u_1 + \frac{3G^3M^3}{L^4}\left(1+2e\cos\phi+e^2\cos^2\phi\right) \\ &= -u_1 + \frac{3G^3M^3}{L^4}\left(1+\frac{e^2}{2}+2e\cos\phi+\frac{e^2}{2}\cos 2\phi\right) \\ &= -u_1 + \frac{3G^3M^3}{L^4}\left(1+\frac{e^2}{2}\right) + \frac{3G^3M^3}{L^4}\left(2e\cos\phi+\frac{e^2}{2}\cos 2\phi\right)\end{aligned}$$

We can take care of the constant term by writing

$$u_1 = \frac{3G^3M^3}{L^4}\left(1+\frac{e^2}{2}\right) + f(\phi),$$

where $f(\phi)$ is an unknown function to be determined. Substituting into the differential equation, we get

$$\frac{\mathrm{d}^2f}{\mathrm{d}\phi^2} = -f + \frac{3G^3M^3}{L^4}\left(2e\cos\phi+\frac{e^2}{2}\cos 2\phi\right). \tag{5.35}$$

To solve this equation, we observe that

$$\frac{\mathrm{d}^2}{\mathrm{d}\phi^2}\left(\phi\sin\phi\right) = 2\cos\phi - \phi\sin\phi, \qquad \frac{\mathrm{d}^2}{\mathrm{d}\phi^2}\cos 2\phi = -4\cos 2\phi.$$

This suggests that we write f in the form

$$f(\phi) = A\phi\sin\phi + B\cos 2\phi,$$

where A and B are constants. The second derivative of f is

$$\begin{aligned}\frac{\mathrm{d}^2f}{\mathrm{d}\phi^2} &= 2A\cos\phi - A\phi\sin\phi - 4B\cos 2\phi \\ &= \underbrace{-A\phi\sin\phi - B\cos 2\phi}_{-f} + 2A\cos\phi - 3B\cos 2\phi \\ \frac{\mathrm{d}^2f}{\mathrm{d}\phi^2} &= -f + 2A\cos\phi - 3B\cos 2\phi.\end{aligned}$$

Comparing this with Eq. (5.35), we fix the constants

$$A = \frac{3G^3M^3}{L^4}e, \quad B = -\frac{3G^3M^3}{L^4}\frac{e^2}{6}.$$

This gives the full solution for $f(\phi)$. Therefore the solution to the differential equation is

$$u_1 = \frac{3G^3M^3}{L^4}\left(1+\frac{e^2}{2}\right) + f(\phi) = \frac{3G^3M^3}{L^4}\left(1+\frac{e^2}{2} + e\phi\sin\phi - \frac{e^2}{6}\cos 2\phi\right).$$

Even as a first-order perturbation, this solution is still somewhat complicated sum of three terms. But the first term is only a constant shift, while the last term oscillates about zero. It is the middle term that is important as it grows as the orbit evolves in ϕ.

Keeping only this term, we install this to our perturbed expression for u,

$$\begin{aligned} u &= u_0 + u_1 \\ &\simeq \frac{GM}{L^2}(1+e\cos\phi) + \frac{3G^3M^3}{L^4} e\phi\sin\phi \\ &\simeq \frac{GM}{L^2}\left(1 + e\left(\cos\phi + \frac{3G^2M^2}{L^2}\phi\sin\phi\right)\right) \end{aligned}$$

To simplify this expression even further, we note that

$$\cos\left[(1-\alpha)\phi\right] \simeq \cos\phi + \alpha\phi\sin\phi + \mathcal{O}(\alpha^6),$$

so if α is small, the solution can be written as

$$u = \frac{GM}{L^2}\left(1 + e\cos\left[(1-\alpha)\phi\right]\right), \tag{5.36}$$

which is valid when

$$\alpha = \frac{3G^2M^2}{L^2}$$

is small.

Now, the Newtonian solution $u_0 = \frac{GM}{L^2}(1+e\cos\phi)$ describes a closed ellipse. Suppose for a Newtonian orbit starting from the *perihelion* (maximum distance), if ϕ evolves after a period 2π, it returns exactly to the perihelion again, thereby forming a closed ellipse.

On the other hand, the solution (5.36) has a small modification at the angular frequency of ϕ. The planet returns still hasn't returned to the perihelion after when ϕ evolves by 2π. The discrepancy is

$$(1-\alpha)2\pi = 2\pi - \delta\phi, \tag{5.37}$$

so the particle would need to evolve a further amount of

$$\delta\phi = 2\pi\alpha = \frac{6\pi G^2 M^2}{L^2}.$$

Now the angular momentum L is not a conveniently measurable quantity. Since we are taking values close to the Newtonian orbits, the Newtonian angular momentum is

$$L^2 \simeq GM(1-e^2)a,$$

where a is the semi-major axis of the Kepler ellipse. Restoring SI units, the perihelion shift is

$$\delta\phi = \frac{6\pi GM}{c^2(1-e^2)a} \tag{5.38}$$

So the orbits do not form a perfectly closed ellipse, as the planet hits the perihelion always slightly later. The consequence is that the orbit looks like a precessing (rotating) ellipse, such as in Fig. 5.7.

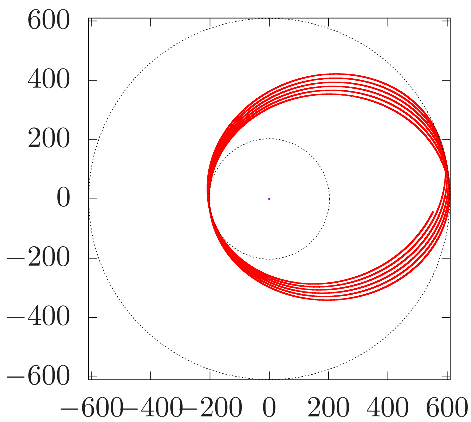


Figure 5.7: Time-like Schwarzschild orbits near the Newtonian limit is a precessing ellipse. This case is solved with $E = 0.9987710703$, $L = 17.54583311GM$. This gives the perihelion distance at $r_{\max} = 609.1424293GM$ and aphelion $r_{\min} = 203.0474683GM$.

Perihelion shift of Mercury. Most of our solar system is usually well-described by Newtonian mechanics. Only planet Mercury is close enough to the sun where the gravitational field is strong enough that relativistic effects start to occur. In fact, when Einstein was still formulating GR, it was already observed by astronomers that Mercury's ellipse seem to be precessing, and it was not known why.[5]

[5]A plausible explanation would be an unseen planet exerting an additional gravitational force to

Putting in the orbital parameters for Mercury,

$$\frac{GM_\odot}{c^2} = 1.48 \times 10^3 \text{ m}, \quad a = 5.79 \times 10^{10} \text{ m}, \quad e = 0.2056,$$

we get

$$\delta\phi_{\text{Mercury}} = 5.01 \times 10^{-7} \text{ rad/orbit} = 0.103 \text{ as/orbit}.$$

This means the perihelion of Mercury's 'ellipse' shifts by 0.013 arcseconds each orbit.[6] This seems like a very small quantity. But the effects accumulate over many years. Mercury completes an orbit once every 88 Earth days. So over a century, the perihelion shifts by

$$\delta\phi_{\text{Mercury}} = 43.0 \text{ as/century},$$

this was the discrepancy that was unexplained before the advent of GR!

5.4 Null geodesics

We now consider the trajectory of light in the Schwarzschild spacetime (outside $r = 2GM$.) Recall that there is no concept of proper time for photons. Therefore the trajectory $x^\mu(\lambda) = (t(\lambda), r(\lambda), \theta(\lambda), \phi(\lambda))$ is parametrised by some arbitrary parameter λ. The 4-velocity of the photon is

$$u^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda} = \left(\dot{t}, \dot{r}, \dot{\theta}, \dot{\phi}\right).$$

Equations of motion

The Lagrangian for null geodesics is

$$\mathcal{L} = \frac{1}{2}\alpha\left(-f\dot{t}^2 + \frac{\dot{r}}{f} + r^2\dot{\theta}^2 + r^2\sin^2\theta\,\dot{\phi}^2\right).$$

As expected, this looks almost similar to the Lagrangian for time-like particles. The difference is that the constant α no longer represent mass, as the photon is massless.

displace the ellipse.

[6] 1 as $= 4.84814 \times 10^{-6}$ rad.

Another difference is $u^\mu u_\mu$ is now equal to zero and not -1,

$$u^\mu u_\mu = 0 = -f\dot{t}^2 + \frac{\dot{r}}{f} + r^2\dot{\theta}^2 + r^2\sin^2\theta\,\dot{\phi}^2.$$

Following the same steps as before (Tutorial) the equations of motion for photons are derived as follows.

$$\dot{t} = \frac{E}{f}, \qquad \dot{\phi} = \frac{L}{r^2}, \qquad \theta = \frac{\pi}{2} = \text{constant}, \tag{5.39a}$$

$$\ddot{r} = \frac{f'\dot{r}}{2f} - \frac{f'E^2}{2f} + \frac{fL^2}{r^3}, \tag{5.39b}$$

$$\dot{r}^2 = E^2 - \underbrace{\frac{L^2}{r^2}\left(1 - \frac{2GM}{r}\right)}_{U_{\text{eff}}}. \tag{5.39c}$$

Photon geodesics can also be solved exactly [8, 24], but as before, it is simpler to analyse the effective potential. Unlike the potential of time-like geodesics, there are fewer terms in the photon case. This means that there are fewer turning points. From Fig. 5.8, we see that there is a potential barrier, but no local minima. So there is no potential well for a photon to be in a bounded orbit. A typical photon orbit starts from infinity, reaches minimum distance, and goes back off to infinity.

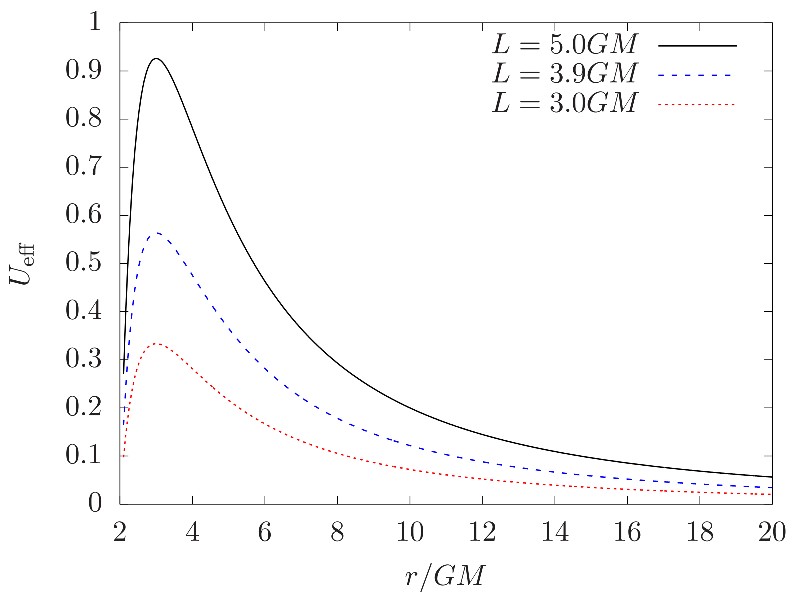
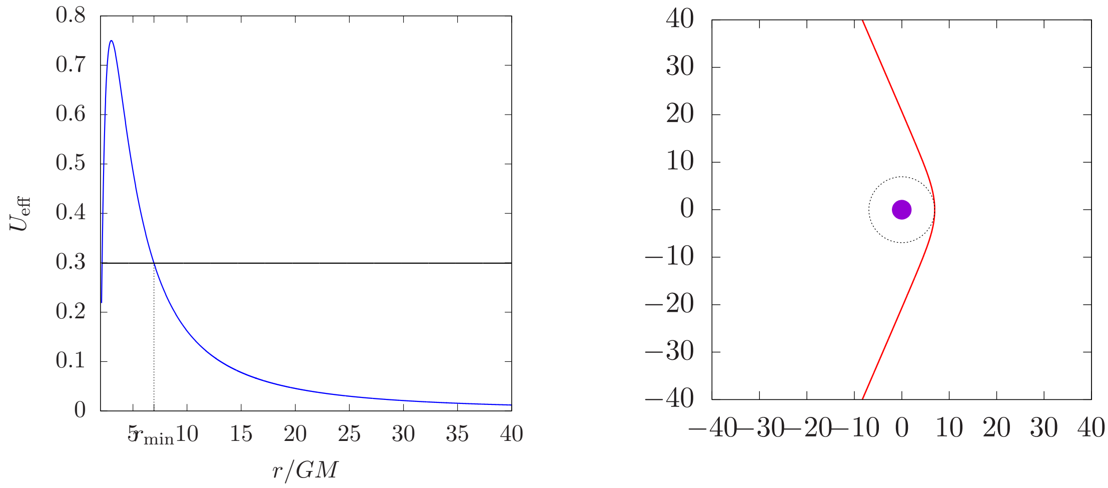


Figure 5.8: Effective potential for null geodesics for various angular momenta L.

The photon trajectories are slightly less interesting than time-like ones, due to the absence of bound orbits. As mentioned earlier, a typical case is of a photon coming from infinity. Then reaches a turning point at r_{min}, and gets deflected off to infinity again. An example of this is shown in Fig. 5.9.



(a) U_{eff} vs r.

(b) Orbit.

Figure 5.9: Orbit of a photon with $E = 0.547$ and $L = 4.5GM$. In this case $r_{\text{min}} = 6.940990940GM$. Fig. 5.9a shows the effective potential graph, it intersects the energy at $U_{\text{eff}} = E^2 = 0.299209$. The orbit is plotted in Fig. 5.9b; we can see that the trajectory comes from infinity, reaches r_{min}, and goes to infinity again. The radial domain is $r_{\text{min}} \leq r \leq \infty$. The shaded circle marks the location inside the Schwarzschild radius.

Circular orbits

We now seek circular orbits of photons. To this end we require the derivatives of the potential

$$U'_{\text{eff}} = -\frac{2L^2}{r^3} - \frac{6GML^2}{r^4}, \quad U''_{\text{eff}} = \frac{6L^2}{r^4} - \frac{24GML^2}{r^5}. \tag{5.40}$$

This time, $U'_{\text{eff}} = 0$ is solved by

$$r_0 = 3GM, \tag{5.41}$$

regardless of the value of L. So there is only one circular orbit for photons and it lies on radius $3GM$. Substituting $r = r_0 = 3GM$ into the second derivative, we find

$$U''_{\text{eff}} = \frac{6L^2}{r_0^4}\left(1 - \frac{4GM}{r}\right) = \frac{6L^2}{r_0^4}\left(-\frac{1}{3}\right) < 0.$$

Therefore this circular orbit is unstable.

We have found that the gravitational field of the Schwarzchild spacetime is strong enough to bend light into a circular orbit. The radius is $r = 3GM > 2GM$, which is outside the Schwarzschild radius! So light is in circular motion, but has not yet fallen into the black hole. Since this orbit is unstable, a small perturbation will either make it fall into the black hole or escape to infinity.

This circular light orbit is sometimes called a *photon/light ring*. Furthermore, since $\theta = \frac{\pi}{2}$ is an arbitrary choice of coordinates, the surface of the sphere of radius $3GM$ is where light goes in circular orbit in any orientation. Therefore it is sometimes called a *photon sphere* in the literature. In spacetimes without spherical symmetry, they are called *photon surfaces* [25], or *photon cones* [26].

Gravitational lensing and exact solutions

When the photon is infinitely far away, assuming no other gravitating bodies are present, then it's in Minkowski space and it travels in a straight line. The as it gets closer to the black hole, the path becomes curved. This can be seen clearly in Fig. 5.9. This phenomena is called *gravitational lensing*. Because the black hole effectively focuses the path of light, just like normal lenses.

This has been observed in space. A collection of galaxy clusters have masses large enough to bend light around it, and light from other galaxies behind it gets bent into an arc (sort of similar to normal lenses). Fig. 5.10 shows an example of a gravitationally-lensed image.

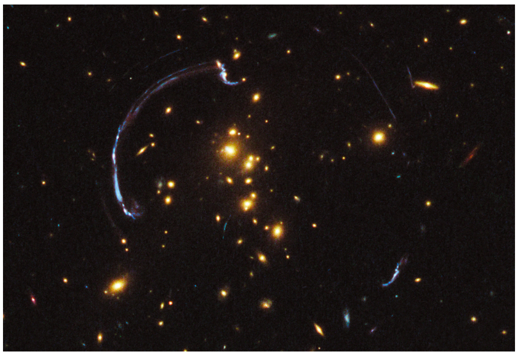


Figure 5.10: Gravitational lens cluster RCS2 032727-132623, which is about 5×10^9 light years away. The blue arcs is the distorted image of a galaxy from further behind the cluster. Credit: NASA, ESA, J. Rigby (NASA GSFC), K. Sharon (KICP, U Chicago), and M. Gladders and E. Wuyts (U Chicago)

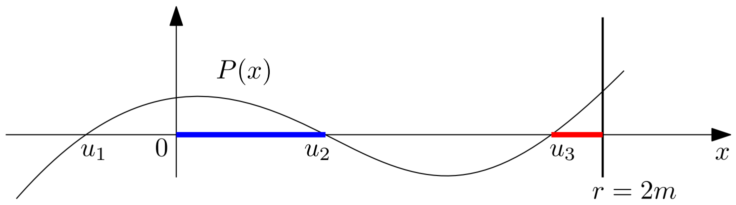


Figure 5.11: Sketch of the graph of $P(u)$, where $u = 1/r$. For a given value of η, the photon can access the radial coordinates for which $P \geq 0$. For the thick blue domain, photons can move in radial coordinates $\frac{1}{u_2} \leq r < \infty$, while for the thick red domain is for $2m < r < \frac{1}{u_3}$.

We can derive exact solutions for gravitational lensing by the following procedure. Using

$$\begin{aligned}\frac{\dot{r}}{\dot{\phi}} = \frac{\mathrm{d}r}{\mathrm{d}\phi} &= \pm\frac{r^2}{L}\sqrt{E^2 - \frac{L^2}{r^2}\left(1 - \frac{2GM}{r}\right)} \\ &= \pm r^2\sqrt{\frac{E^2}{L^2} - \frac{1}{r^2} + \frac{2GM}{r^3}}.\end{aligned}$$

Letting $\eta = E^2/L^2$ and making the substitution $u = 1/r$, we have $\frac{\mathrm{d}u}{\mathrm{d}r} = -\frac{1}{u^2}$. Therefore

$$\begin{aligned}\frac{\mathrm{d}u}{\mathrm{d}\phi} &= \frac{\mathrm{d}u}{\mathrm{d}r}\frac{\mathrm{d}r}{\mathrm{d}\phi} = \mp\frac{1}{r^2}r^2\sqrt{\eta - \frac{1}{r^2} + \frac{2GM}{r^3}} \\ \frac{\mathrm{d}u}{\mathrm{d}\phi} &= \mp\sqrt{\eta - \frac{1}{r^2} + \frac{2GM}{r^3}} \\ \frac{\mathrm{d}u}{\mathrm{d}\phi} &= \mp\sqrt{P(u)}, \quad P(u) = \eta - u^2 + 2GMu^3. \end{aligned} \tag{5.42}$$

The polynomial $P(u)$ is of degree 3. From the equations of motion, it is clear that photons can only exist in domains of u (or equivalently, r) where $P \geq 0$. The boundaries of these domains are the roots of P, which are the turning points $\frac{\mathrm{d}u}{\mathrm{d}\phi} = 0$, or equivalently $\dot{r} = 0$. Using the Descartes rule of signs, one can show that P has one negative root and two real roots. The graph of $y = P(u)$ is sketched in Fig. 5.11. Let us write these roots as $u_1 < 0 \leq u_2 \leq u_3$.

For the interest of gravitational lensing, our primary domain of interest is $0 < u \leq u_2$. Therefore we shall write $P(u)$ as

$$P(u) = 2GM(u - u_1)(u_2 - u)(u_3 - u). \tag{5.43}$$

Consider a photon coming in from infinity $u = 0$ with some value of η. Such a photon can reach a distance of closest approach $r_b = \frac{1}{u_2}$. We infer that

$$\eta = \frac{1}{r_b^2}\left(1 - \frac{2GM}{r_b}\right) = u_2^2\left(1 - 2GMu_2\right). \tag{5.44}$$

An important quantity is the *impact parameter* which we define as follows. Suppose that the photon emitted from a source S, passes the black hole at coordinate distance of closest approach r_b, and incident at an observer O at large r, as sketched in Fig. 5.12. By 'large r', we mean that the observer's radial coordinate is $r \gg M$. Therefore the equation of motion for a photon in the vicinity of the observer is

$$\frac{\mathrm{d}\phi}{\mathrm{d}x} \simeq \pm\frac{1}{\sqrt{\eta - u^2}},$$

which has a solution $u = \frac{1}{r} = \sqrt{\eta}\,\sin\phi$. From the perspective of the observer, the angular position of the image relative to the lens depends on $b = r\sin\phi_\infty$. We define this distance b to be the impact parameter, and is related to η and u_2 by

$$b = \frac{1}{\sqrt{\eta}} = \frac{1}{u_2\sqrt{1 - 2GMu_2}}. \tag{5.45}$$

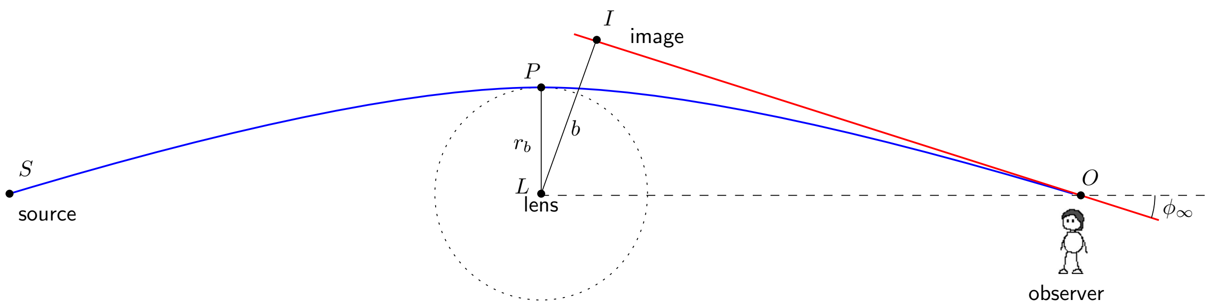


Figure 5.12: Gravitational lensing of a photon emitted from a source S by a lens L, which is the Schwarzschild black hole, and arriving at the observer O. The tangent of the null geodesic at O makes an angle ϕ_∞ with the optic axis $\overrightarrow{LO}$. Assuming the observer is at large distance r far away from the lens, the impact parameter is approximately $b = r\sin\phi_\infty$.

The differential equation (5.42) can now be integrated as

$$\begin{aligned}\phi(u) &= \frac{1}{\sqrt{2GM}}\int_u^{u_2}\frac{\mathrm{d}u}{\sqrt{(u-u_1)(u_2-u)(u_3-u)}}\\ &= \frac{2}{\sqrt{2GM(u_3-u_1)}}\mathrm{F}\left(\arcsin\sqrt{\frac{(u_3-u_1)(u_2-u)}{(u_2-u_1)(u_3-u)}}, \sqrt{\frac{u_2-u_1}{u_3-u_1}}\right),\end{aligned} \tag{5.46}$$

where $\mathrm{F}\left(\Phi, k\right)$ is the incomplete elliptical integral of the first kind. Solving for $r = 1/u$, we get

$$r(\phi) = \frac{(u_2 - u_1)\mathrm{sn}^2\,(\omega\phi, k) - (u_3 - u_1)}{(u_2 - u_1)u_3\mathrm{sn}^2\,(\omega\phi, k) - (u_3 - u_1)u_2}, \tag{5.47a}$$

$$\omega = \frac{1}{2}\sqrt{2GM(u_3 - u_1)}, \quad k = \sqrt{\frac{u_2 - u_1}{u_3 - u_1}}. \tag{5.47b}$$

It is convenient to parametrise the roots by

$$u_1 = -\frac{(1-\sigma^2)}{2GM(3+\sigma^2)}, \quad u_2 = \frac{1-\sigma}{GM(3+\sigma^2)}, \quad u_3 = \frac{1+\sigma}{GM(3+\sigma^2)}, \tag{5.48}$$

where $0 \le \sigma < 1$. By substituting into Eq. (5.43) and comparing coefficients with (5.42), we see that

$$\eta = \frac{(1-\sigma^2)^2}{M^2(3+\sigma^2)^3} = \frac{1}{b^2}. \tag{5.49}$$

In terms of the parameter σ, Eq. (5.46) is

$$\begin{aligned}\phi(u) = {} & 2\sqrt{\frac{3+\sigma^2}{(3-\sigma)(1+\sigma)}} \\ & \times \mathrm{F}\left(\arcsin\sqrt{\frac{(1+\sigma)(3-\sigma)\left[1-\sigma-(3+\sigma^2)Mu\right]}{(1-\sigma)(3+\sigma)\left[1+\sigma-(3+\sigma^2)Mu\right]}}, \sqrt{\frac{(3+\sigma)(1-\sigma)}{(3-\sigma)(1+\sigma)}}\right).\end{aligned} \tag{5.50}$$

The trajectory is symmetric about the turning point u_2. Therefore, for a particle coming in from infinity, reaching the coordinate distance of closest approach, then proceeding off to infinity again, the total angular evolution is $\Delta\phi = 2\phi(0)$. The Einstein bending angle is

$$\begin{aligned}\hat{\alpha} &= -\pi + 2\phi(0) \\ &= -\pi + 4\sqrt{\frac{3+\sigma^2}{(3-\sigma)(1+\sigma)}}\mathrm{F}\left(\arcsin\sqrt{\frac{3-\sigma}{3+\sigma}}, \sqrt{\frac{(3+\sigma)(1-\sigma)}{(3-\sigma)(1+\sigma)}}\right).\end{aligned} \tag{5.51}$$

By this procedure, the problem of gravitational lensing is completely determined by the parameter $\sigma \in (0,1)$. A choice of σ determines the coordinate distance of closest approach with $r_b = 1/u_2 = 2m(3+\sigma^2)/(1-\sigma)$, and the impact parameter b with Eq. (5.49), and the bending angle $\hat{\alpha}$ with (5.51).

The case $\sigma = 0$ is where u_2 and u_3 becomes complex, and P is positive all the way down to the horizon $x = 1/2M$. In other words, photons coming from infinity will fall into the black hole. Reversing the direction of geodesics, clearly no photons can be emanated from the horizon[7] and therefore no photons of $\sigma \leq 0$ will reach an observer an infinity. Using Eq. (5.49),

$$b = \frac{(3+\sigma^2)^{3/2}}{1-\sigma^2} GM,$$

we see that an observer will be able to observe photons with $0 < \sigma < \infty$, for which $b_{\rm sh} < b < \infty$, where

$$b_{\rm sh} = 3\sqrt{3} GM. \tag{5.52}$$

An observer will observe a dark spot of radius $b_{\rm sh}$; this is called the *shadow*. The radius of the black spot of the Event Horizon Telescope collaboration (EHT) [27] (Fig. 5.13) is the shadow $b_{\rm sh}$ (not the Schwarzschild radius).

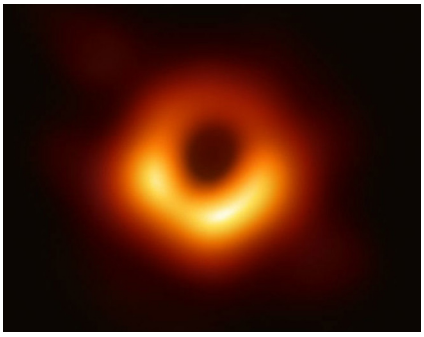


Figure 5.13: Image of the M87 black hole by the EHT collaboration [27].

Most textbooks provided the result $\hat{\alpha} \simeq \frac{4GM}{b}$ under the approximation that the impact parameter b is large. In our present notation, this is where σ is close to 1. We can recover the approximate result by taking $\sigma = 1 + \epsilon$ in Eq. (5.51) and expanding in small ϵ. This gives

$$\hat{\alpha} = \epsilon + \left(\frac{1}{4} + \frac{15\pi}{64}\right)\epsilon^2 + \mathcal{O}(\epsilon^3). \tag{5.53}$$

In terms of η and σ, the impact parameter is

$$b = \frac{1}{\sqrt{\eta}} = \frac{4GM}{\epsilon} - GM + \frac{11}{8} GM\epsilon + \mathcal{O}(\epsilon^2), \tag{5.54}$$

or $\epsilon \simeq \frac{4GM}{b}$. We indeed the approximate bending angle is

$$\hat{\alpha} \simeq \frac{4GM}{b}.$$

[7]We are considering classical General Relativity and therefore ingoring quantum effects such as Hawking radiation, etc.

This is the deflection angle of light that does not approach too close to the black hole. (Large impact parameter.)

Gravitational redshift

In Chapter 4 we briefly discussed the gravitational redshift as a consequence of the equivalence principle. However, back then we didn't have the full Einstein equations yet and our calculations were only approximate. We can now perform a more concrete calculation using the Schwarzschild metric. As discussed above, this metric well-approximates the gravitational field above the surface of a planet, like the Earth.

Consider two observers, Alice and Bob, located at different positions in Schwarzschild spacetime. For example, Alice is on Earth's furface, and Bob is on the top floor of a high tower. Alice and Bob's spatial coordinates are respectively

$$(r_A, \theta_A, \phi_A), \quad (r_B, \theta_B, \phi_B).$$

Let us their differences in the angular coordinates are negligible. So we only consider r_A and r_B. Taking into account the time, their worldlines are depicted in the spacetime diagram in Fig. 5.14.

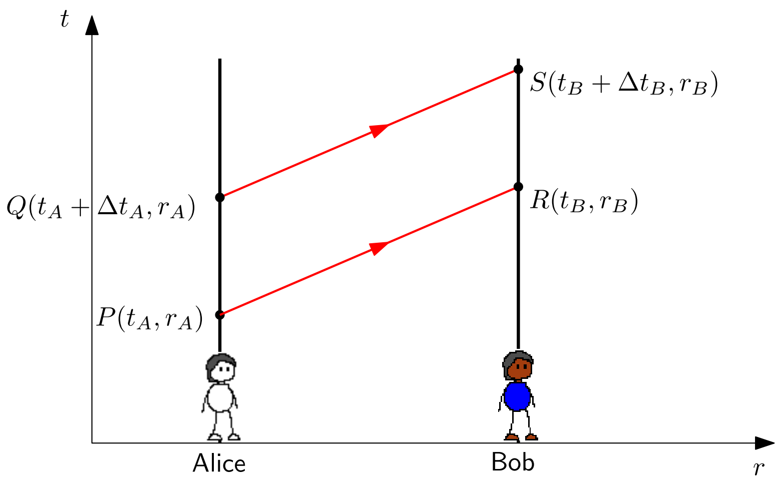


Figure 5.14: Worldlines of Alice and Bob. At time $t = t_A$, Alice emits a photon, which is received by Bob at time $t = t_B$. At a later time $t = t_A + \Delta t_A$ Alice emits another photon which is received by Bob at time $t = t_B + \Delta t_B$.

At time t_A, Alice emits a photon (Event P). This photon is received at Bob's location at time t_B (Event R). A little later, at time $t_A + \Delta t_A$ Alice emits another photon (Event Q), which is received by Bob at $t_B + \Delta t_B$ (Event S).

Let the trajectory of the photon be described by $x^\mu(\lambda) = (t(\lambda), r(\lambda), \theta(\lambda), \phi(\lambda))$, where λ

is an arbitrary parameter. The 4-velocity of the photon is

$$u^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda} = \dot{x}^\mu = \left(\dot{t}, \dot{r}, \dot{\theta}, \dot{\phi}\right). \tag{5.55}$$

Since photon follow null geodesics, we have

$$\begin{aligned} u^\mu u_\mu = g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu = 0 \\ -f\dot{t}^2 + \frac{\dot{r}^2}{f} + r^2\dot{\theta}^2 + r^2 \sin^2\theta \dot{\phi}^2 = 0 \end{aligned}$$

solving for $\dot{t}$, and assuming this to be positive[8]

$$\dot{t} = f^{-1/2}\left[\frac{\dot{r}^2}{f} + r^2\dot{\theta}^2 + r^2\sin^2\theta\dot{\phi}^2\right]^{1/2}.$$

We simplify the problem by considering the case where Alice and Bob have the same angular coordinates. So the photon only travels in the radial direction with $\dot{\theta} = \dot{\phi} = 0$:

$$\begin{aligned} \dot{t} &= \frac{\dot{r}}{f} \\ \frac{\dot{t}}{\dot{r}} &= \frac{\mathrm{d}t}{\mathrm{d}r} = \frac{1}{f} \\ \mathrm{d}t &= \frac{\mathrm{d}r}{f} \\ \int \mathrm{d}t &= \int \frac{\mathrm{d}r}{f}. \end{aligned}$$

We compute this integral for the first photon that was emitted by Alice at P and received by Bob at R. The appropriate limits of integration are

$$\begin{aligned} \int_{t_A}^{t_B} \mathrm{d}t &= \int_{r_A}^{r_B} \frac{\mathrm{d}r}{f} \\ t_B - t_A &= \int_{r_A}^{r_B} \frac{\mathrm{d}r}{f}. \end{aligned} \tag{5.56}$$

For the second photon emitted by Alice at Q and receved by Bob at S, the integration is

$$\begin{aligned} \int_{t_A+\Delta t_A}^{t_B+\Delta t_B} \mathrm{d}t &= \int_{r_A}^{r_B} \frac{\mathrm{d}r}{f} \\ t_B + \Delta t_B - t_A - \Delta t_A &= \int_{r_A}^{r_B} \frac{\mathrm{d}r}{f}. \end{aligned} \tag{5.57}$$

Clearly, the right hand sides of (5.56) and (5.57) are exactly equal. From this we conclude

[8] As in, t increases as λ moves forward.

for the left-hand sides that

$$\Delta t_B = \Delta t_A. \tag{5.58}$$

These are differences in *coordinate* time, which may not be the time according to watches carried by Alice and Bob. To do this, we must determine the *proper time*. In Alice's frame, Alice's proper time identified by equating $\mathrm{d}s^2$ with $-\mathrm{d}\tau_A^2$, where τ_A is Alice's proper time,

$$\begin{aligned}
\mathrm{d}s^2 = -\mathrm{d}\tau_A^2 = -f(r_A)\mathrm{d}t^2 &= -\left(1 - \frac{2GM}{r_A}\right)\mathrm{d}t^2 \\
\frac{\mathrm{d}\tau_A}{\mathrm{d}t} &= \sqrt{1 - \frac{2GM}{r_A}}.
\end{aligned}$$

Integrating both sides,

$$\Delta\tau_A = \int_{t_A}^{t_A+\Delta t_A} \mathrm{d}t\, \sqrt{1 - \frac{2GM}{r_A}} = \Delta t_A \sqrt{1 - \frac{2GM}{r_A}}. \tag{5.59}$$

Performing a similar calculation for Bob,

$$\Delta\tau_B = \Delta t_B \sqrt{1 - \frac{2GM}{r_B}}. \tag{5.60}$$

The ratio between the two is

$$\frac{\Delta\tau_B}{\Delta\tau_A} = \sqrt{\frac{1 - 2GM/r_B}{1 - 2GM/r_A}}. \tag{5.61}$$

This is essentially the time-dilation phenomena in Schwarzschild spacetime. If Bob is at a high tower with Alice on the ground, $r_B > r_A$, therefore $\Delta\tau_B > \Delta\tau_B$.

Now suppose that $\Delta\tau_A$ is the oscillation period of the light wave measured by Alice. Then the frequency of the photon emitted by her is $f_A = 1/\Delta\tau_A$. When received by Bob, the frequency he meaures is $f_B = 1/\Delta\tau_B$. With $c = f\lambda$, we find

$$\begin{aligned}
\frac{f_A}{f_B} &= \frac{c/\lambda_A}{c/\lambda_B} \\
\frac{1/\Delta\tau_A}{1/\Delta\tau_B} &= \frac{\lambda_B}{\lambda_A} \\
\frac{\Delta\tau_B}{\Delta\tau_A} &= \frac{\lambda_B}{\lambda_A} = \sqrt{\frac{1 - 2GM/r_B}{1 - 2GM/r_A}} \\
\lambda_B &= \lambda_A \sqrt{\frac{1 - 2GM/r_B}{1 - 2GM/r_A}}.
\end{aligned}$$

Therefore if $r_B > r_A$, Bob receives a longer wavelength than when it was emitted by Alice. Longer wavelength means it moves towards the red side of the light spectrum. Therefore we say that the light has been *redshifted*. So if Alice is wearing a blue shirt, Bob might see the shirt as red in colour.

The redshift becomes more drastic if r_A is approaches the Schwarzschild radius. At the Schwarzschild radius itself, the redshift becomes infinite.

5.5 The Schwarzschild black hole

We now address what happens at the Schwarzschild radius $r = 2GM$. It is the peculiar phenomena associated with this radius which makes it a *black hole*. In public discourse, black holes are often said to be where '*gravity is so strong that even light cannot escape*'. In this section we will see what does this statement means using the Schwarzschild metric.

Infalling particles

We consider Alice, who is a particle falling radially towards the Schwarzschild radius. Alice carries a watch that emits light signals at periods of $\Delta\tau_A$, which is the proper time for Alice. These light signals are received by Bob, who is very far away at location $r_B \to \infty$.

By the gravitational redshift calculation Eq. (5.61), gives

$$\Delta\tau_B = \Delta\tau_A \sqrt{\frac{1 - \frac{2GM}{r_B}}{1 - \frac{2GM}{r_A}}} = \frac{\Delta\tau_A}{\sqrt{1 - \frac{2GM}{r_A}}}$$

since we are taking $r_B \to \infty$. From this equation we see that as Alice approaches the black hole, r_A decreases and Bob receives the signals at increasingly large periods. Until the point where Alice reaches the Schwarzshild radius, $r_A \to 2GM$, we have $\Delta\tau_B \to \infty$. The time dilation on Alice observed by Bob becomes infinite as she approaches the Schwarzschild radius.

From the perspective of Alice herself, she is able to reach $2GM$ in finite proper time. We can see this by assuming she is free-falling radially, so her motion is described by time-like geodesics with zero angular momentum, $L = 0$. We then have

$$\dot{t} = \frac{E}{f}, \quad \dot{\phi} = 0, \quad \dot{r} = \pm\sqrt{E^2 - 1 + \frac{2GM}{r}}.$$

We take the negative square root, since she is falling in decreasing r. Then,

$$\begin{aligned} \dot{r} = \frac{\mathrm{d}r}{\mathrm{d}\tau} &= -\sqrt{E^2 - 1 + \frac{2GM}{r}} \\ \int_0^{\Delta\tau} \mathrm{d}\tau &= -\int_{r_i}^{2GM} \frac{\mathrm{d}r}{\sqrt{E^2 - 1 + \frac{2GM}{r}}}, \end{aligned}$$

where $r_i > 2GM$ is some initial position of Alice. One can check that for any choice of r_i and E, the result $\Delta\tau$ is finite.

The falling of Alice can also be seen by checking the light cone structure of the Schwarzschild spacetime. Photons travel along null paths, therefore $\mathrm{d}s^2 = 0$. Furthermore let us consider radially moving photons, so $\mathrm{d}\theta = \mathrm{d}\phi = 0$. Then we have

$$\begin{aligned} \mathrm{d}s^2 = 0 &= -f\mathrm{d}t^2 + \frac{\mathrm{d}r^2}{f} \\ \frac{\mathrm{d}t}{\mathrm{d}r} &= \pm\frac{1}{f}. \end{aligned} \tag{5.62}$$

The value of $\frac{\mathrm{d}t}{\mathrm{d}r}$ represens the slope of the light cone when plotted on the r-t axis. This slope depends on f, whose value depends on the position r, as seen in Fig. 5.15.

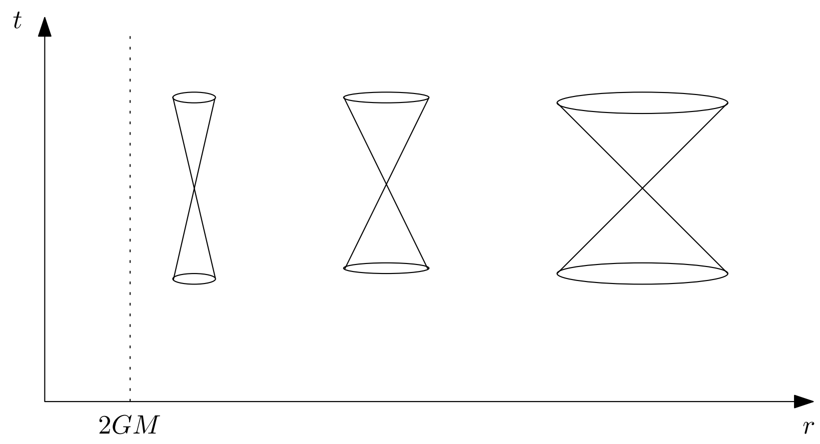


Figure 5.15: Schwarzschild light cones in (r, t)-coordinates.

Far away, at r very large, f becomes close to 1, and we have the usual 45° light cones approaching Minkowski space. As r approaches $2GM$, f approaches zero so the slope of the light cone gets more and more vertical.

Knowing the slope of light cones at each r, we can sketch the path of the light signals emitted from Alice as she falls in, such as in Fig. 5.16.

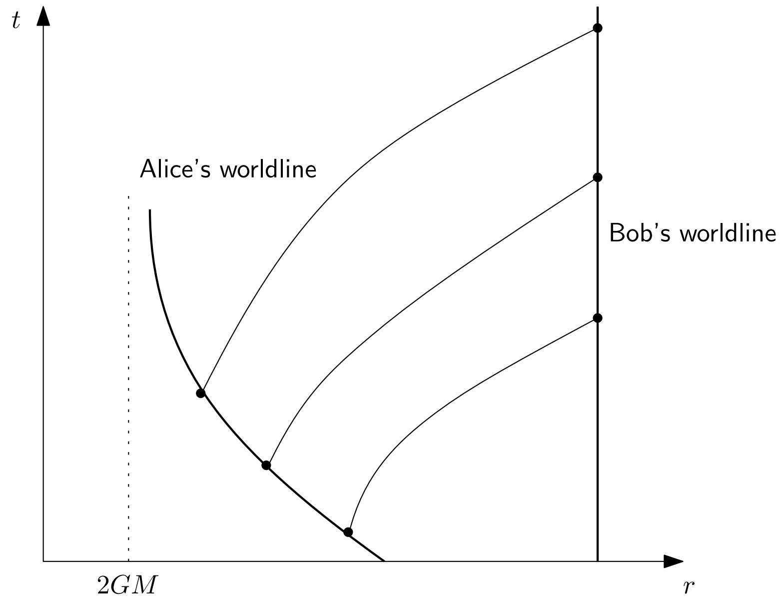


Figure 5.16: Signals emitted by Alice in regular intervals according to her proper time, received by Bob at longer and longer intervals.

Even as Alice emits signals at regular intervals according to her proper time, Bob receives them at longer and longer intervals, approaching infinity as Alice approaches $2GM$.

Now we wish to actually see what happens when we cross the horizon. As we argued earlier, the Schwarzschild metric in standard coordinates is badly behaved at the Schwarzschild radius. So we wish to transform to a more appropriate set of coordinates analyse it.

First we introduce the *tortoise coordinates* in the following way. We manipulate the Schwarzschild metric into the following form,

$$\mathrm{d}s^2 = f\left(-\mathrm{d}t^2 + \frac{\mathrm{d}r^2}{f^2}\right) + r^2\mathrm{d}\Omega^2.$$

If we take

$$\mathrm{d}r_* = \frac{\mathrm{d}r}{f},$$

then, the metric is

$$\mathrm{d}s^2 = f\left(-\mathrm{d}t^2 + \mathrm{d}r_*^2\right) + r^2\mathrm{d}\Omega^2.$$

In this new coordinates, the (t, r_*)-part of the metric is just an overall factor times something like Minkowski. Since $\mathrm{d}s^2 = 0$ for null radial geodesics, all light cones are at 45°. The transformation for r_* is explicitly

$$r_* = \int \frac{\mathrm{d}r}{f} = \int \frac{\mathrm{d}r}{1 - \frac{2GM}{r}} = r + 2GM \ln \left(\frac{r}{2GM} - 1 \right)$$

This is called the *tortoise coordinate*. The domains of each coordinates are mapped as in Fig. 5.17.

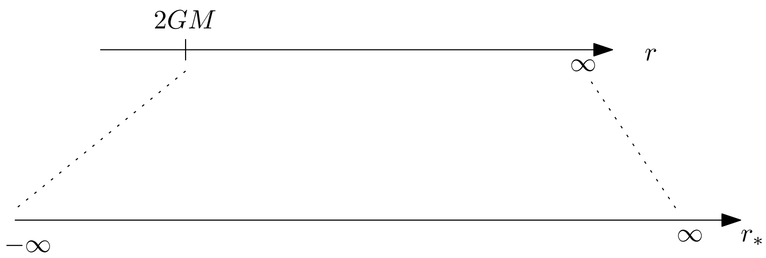


Figure 5.17

We see that $2GM < r < \infty$ is mapped to $-\infty < r_* < \infty$. The tortoise coordinate is undefined inside the Schwarzschild radius.

The event horizon

To investigate inside the Schwarzschild radius we transform into *Eddington–Finkelstein coordinates* by

$$t = v - r_*$$

where r_* is the tortoise coordinate defined previously. Note that

$$\begin{aligned} \mathrm{d}t &= dv - dr_* = \mathrm{d}v - \frac{\mathrm{d}r}{f}, \\ \mathrm{d}t^2 &= \mathrm{d}v^2 - \frac{2\mathrm{d}v\mathrm{d}r}{f} + \frac{\mathrm{d}r^2}{f^2}, \end{aligned}$$

therefore the Schwarzschild metric is transformed as

$$\begin{aligned} \mathrm{d}s^2 = -f\mathrm{d}t^2 + \frac{\mathrm{d}r^2}{f} + r^2\mathrm{d}\Omega^2 &= -f\mathrm{d}v^2 + 2\mathrm{d}v\mathrm{d}r - \frac{\mathrm{d}r^2}{f} + \frac{\mathrm{d}r^2}{f} + r^2\mathrm{d}\Omega^2 \\ &= -f\mathrm{d}v^2 + 2\mathrm{d}v\mathrm{d}r + r^2\mathrm{d}\Omega^2. \end{aligned} \tag{5.63}$$

Null radial geodesics in Eddington–Finkelstein coordinates now follow

$$\begin{aligned} 0 &= -f\mathrm{d}v^2 + 2\mathrm{d}v\mathrm{d}r \\ 0 &= \mathrm{d}v\left(-f\mathrm{d}v + 2\mathrm{d}r\right). \end{aligned}$$

There are two cases,

$$v = \text{const}, \quad \frac{\mathrm{d}v}{\mathrm{d}r} = \frac{1}{2f}.$$

The two solutions describe the slopes of each side of the light cone. Now the light cones does not close up, an improvement over the standard coordinates. If we were to sketch this in (r, v) coordinates, we get Fig. 5.18.

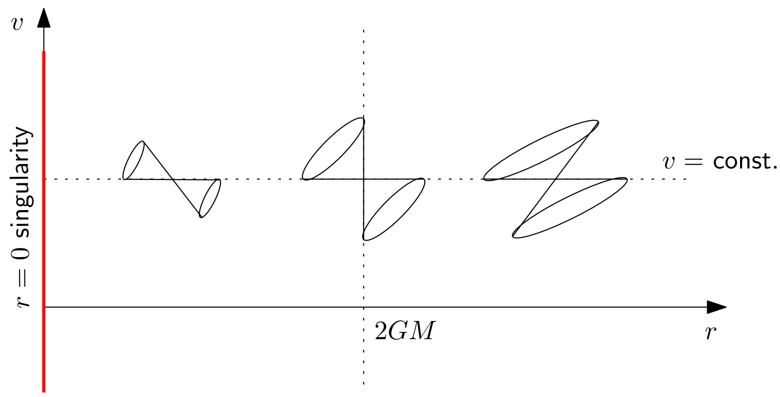


Figure 5.18: Light cones in Eddington–Finkelstein coordinates.

From this figure, one side of the light cone aligns with $v = \text{const}$. The other side of the light cone has slope $1/2f$, which is positive for $r > 2GM$. This side becomes vertical at the Schwarzschild radius $r = 2GM$, and the cone points completely inwards at $r < 2GM$. This means the future of any event always points towards the $r = 0$ singularity. No signals emitted from inside $r < 2GM$ can reach $r > 2GM$. We call the surface $r = 2GM$ the *event horizon*. Since no light can escape from inside the horizon, this is called a *black hole*. Any photon or particle that crosses the event horizon inevitably head towards the $r = 0$ singularity. Escaping this requires a worldline that goes outside the light cone, meaning one has to travel faster than light.

Kruskal extension

Let us return to the Schwarzscchild metric in tortoise coordinates,

$$\mathrm{d}s^2 = f\left(-\mathrm{d}t^2 + \mathrm{d}r_*^2\right) + r^2\mathrm{d}\Omega^2.$$

We now define *null coordinates* by

$$v = t + r_*, \quad u = t - r_*. \tag{5.64}$$

The inverse transformation is

$$t = \frac{1}{2}(v + u), \quad r_* = \frac{1}{2}(v - u).$$

In these coordinates, the metric now becomes

$$\mathrm{d}s^2 = -f\mathrm{d}v\mathrm{d}u + r^2\mathrm{d}\Omega^2. \tag{5.65}$$

At this stage the horizon is at $v = -\infty$ or $u = +\infty$. We make a transformation to bring it to finite values,

$$\begin{aligned}
v' &= \mathrm{e}^{v/4GM}, \quad u' = -\mathrm{e}^{-u/4GM}, \\
\mathrm{d}v' &= \frac{\mathrm{d}v}{4GM}\mathrm{e}^{v/4GM}, \quad \mathrm{d}u' = \frac{\mathrm{d}u}{4GM}\mathrm{e}^{-u/4GM}, \\
\mathrm{d}v &= 4GM\mathrm{e}^{-v/4GM}\mathrm{d}v', \quad \mathrm{d}u = 4GM\mathrm{e}^{u/4GM}\mathrm{d}u'.
\end{aligned}$$

Therefore the metric now becomes

$$\mathrm{d}s^2 = -f16G^2M^2\mathrm{e}^{\frac{u-v}{4GM}}\mathrm{d}u'\mathrm{d}v'r^2\mathrm{d}\Omega^2.$$

Note that $u - v = 2r_* = -2r - 4GM\ln\left(\frac{r}{2GM} - 1\right)$. Therefore

$$\begin{aligned}
\mathrm{d}s^2 &= -16G^2M^2 f\exp\left[\frac{-2r - 4GM\ln\left(\frac{r}{2GM} - 1\right)}{4GM}\right]\mathrm{d}u'\mathrm{d}v'r^2\mathrm{d}\Omega^2 \\
&= -16G^2M^2 f\mathrm{e}^{-r/2GM}\mathrm{e}^{-\ln\left(\frac{r}{2GM}-1\right)}\mathrm{d}u'\mathrm{d}v' + r^2\mathrm{d}\Omega^2 \\
&= -16G^2M^2 f\mathrm{e}^{-r/2GM}\left(\frac{r}{2GM} - 1\right)^{-1}\mathrm{d}u'\mathrm{d}v' + r^2\mathrm{d}\Omega^2 \\
&= -16G^2M^2\cancel{f}\mathrm{e}^{-r/2GM}\left(\frac{r}{2GM}\right)^{-1}\cancel{\left(1 - \frac{2GM}{r}\right)^{-1}}\mathrm{d}u'\mathrm{d}v' + r^2\mathrm{d}\Omega^2.
\end{aligned}$$

We finally brought the metric into the form

$$\mathrm{d}s^2 = -\frac{32G^3M^3}{r}\mathrm{e}^{-r/2GM}\mathrm{d}u'\mathrm{d}v' + r^2\mathrm{d}\Omega^2. \tag{5.66}$$

In this form, the metric is clearly non-singular at $r = 2GM$, and we can cross the event horizon easily in these coordinates.

To investigate light cones, we introduce another transformation

$$T = \frac{1}{2}\left(v' + u'\right), \quad R = \frac{1}{2}\left(v' - u'\right). \tag{5.67}$$

To see the meaning of T, we find that

$$\begin{aligned}
T &= \frac{1}{2}\left(v' + u'\right) = \frac{1}{2}\left(\mathrm{e}^{v/4GM} - \mathrm{e}^{-u/4GM}\right) \\
&= \frac{1}{2}\left[\mathrm{e}^{\frac{1}{4GM}\left[t+r+2GM\ln\left(\frac{r}{2GM}-1\right)\right]} - \mathrm{e}^{-\frac{1}{4GM}\left[t-r-2GM\ln\left(\frac{r}{2GM}-1\right)\right]}\right] \\
&= \frac{1}{2}\left[\mathrm{e}^{t/4GM}\mathrm{e}^{r/4GM}\mathrm{e}^{\frac{1}{2}\ln\left(\frac{r}{2GM}-1\right)} - \mathrm{e}^{-t/4GM}\mathrm{e}^{r/4GM}\mathrm{e}^{\frac{1}{2}\ln\left(\frac{r}{2GM}-1\right)}\right] \\
&= \mathrm{e}^{r/4GM}\mathrm{e}^{\frac{1}{2}\ln\left(\frac{r}{2GM}-1\right)}\frac{1}{2}\left(\mathrm{e}^{t/4GM} - \mathrm{e}^{-t/4GM}\right) \\
&= \left(\frac{r}{2GM} - 1\right)^{1/2}\mathrm{e}^{r/4GM}\sinh\left(\frac{t}{4GM}\right).
\end{aligned}$$

Similarly, one can show

$$R = \frac{1}{2}\left(v' - u'\right) = \left(\frac{r}{2GM} - 1\right)^{1/2}\mathrm{e}^{r/4GM}\cosh\left(\frac{t}{4GM}\right).$$

The result is the metric in *Kruskal coordinates*,

$$\mathrm{d}s^2 = \frac{32G^3M^3}{r}\mathrm{e}^{-r/2GM}\left(-\mathrm{d}T^2 + \mathrm{d}R^2\right) + r^2\mathrm{d}\Omega^2, \tag{5.68a}$$

$$T = \left(\frac{r}{2GM} - 1\right)^{1/2}\mathrm{e}^{r/4GM}\sinh\left(\frac{t}{4GM}\right), \tag{5.68b}$$

$$R = \left(\frac{r}{2GM} - 1\right)^{1/2}\mathrm{e}^{r/4GM}\cosh\left(\frac{t}{4GM}\right). \tag{5.68c}$$

In particular, we notice that

$$R^2 - T^2 = \frac{r}{2GM}\left(1 - \frac{2GM}{r}\right)\mathrm{e}^{r/2GM}. \tag{5.69}$$

So, positions of constant r are hyperbolae in the R-T plane. The orientations of the hyperbolae depend on whether we are outside ($r > 2GM$) or inside ($r < 2GM$) the horizon. Sketching the light cones in (R, T)-coordinates, we get Fig. 5.19. The event horizon is now a diagonal line at 45°. All light cones are also 45°. The region inside the horizon is in the upper part of the diagram. Again we see that all signals emitted from inside the horizon cannot escape to $r > 2GM$.

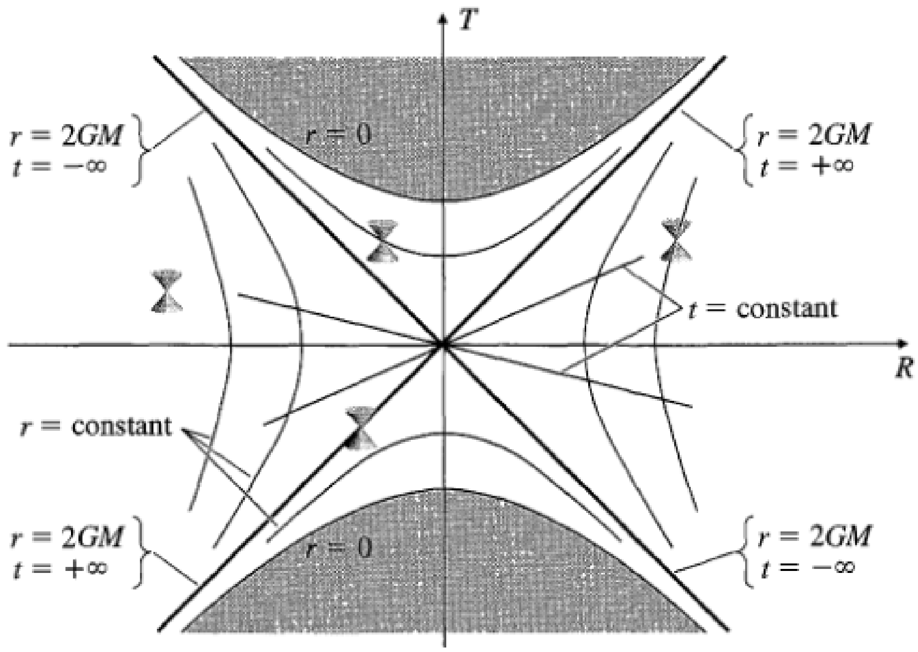


Figure 5.19: Kruskal diagram depicting the causal structure of the Schwarzschild black hole.

Chapter 6 Cosmology

Cosmology is an exceedingly ambitious endeavour – seeking to understand the dynamics and behaviour of the *entire universe*. Indeed, humans have always had curiosity about the origin and future of the universe, and many religions and cultures have their own creation myths about how the universe was created.[1]

The universe is an unfathomably huge place. If we look out into space, we find a collection of stars grouped into *galaxies*. The Milky Way is the particular galaxy that we live in. Further out, there are billions of other galaxies. The galaxies themselves tend to group among each other called *galaxy clusters*. Clusters group with other clusters to form *superclusters*. Fig. 6.1 shows the supercluster containing the Local Group, which itself contains the Milky Way.

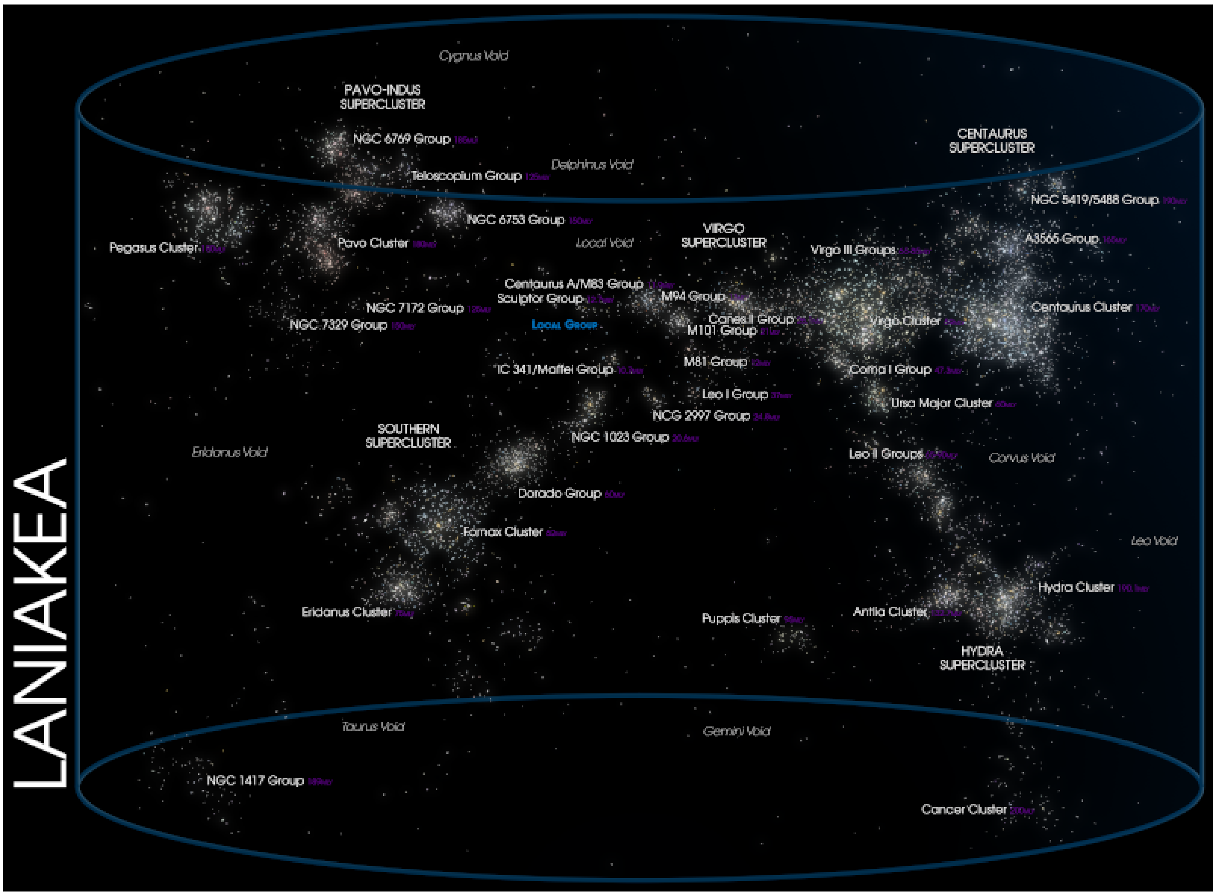


Figure 6.1: The Laniakea Supercluster. Map by Andrew Z. Colvin [CC BY-SA 4.0.]

The entire collection of the universe consists of the mass of all stars and galaxies. By the theory of General Relativity, this huge collection of mass will affect the behaviour

[1]See excellent book by Helge Kragh [28] for the history of cosmology.

of spacetime. The gravity around galaxy clusters are often strong enough to induce gravitational lensing, such as Abell 2744 in Fig. 6.2. Here, we will study the effect of *all* matter in the universe and see the spacetime of the entire universe itself. This is the study of cosmology, where we attempt to determine the behaviour and evolution of the universe.

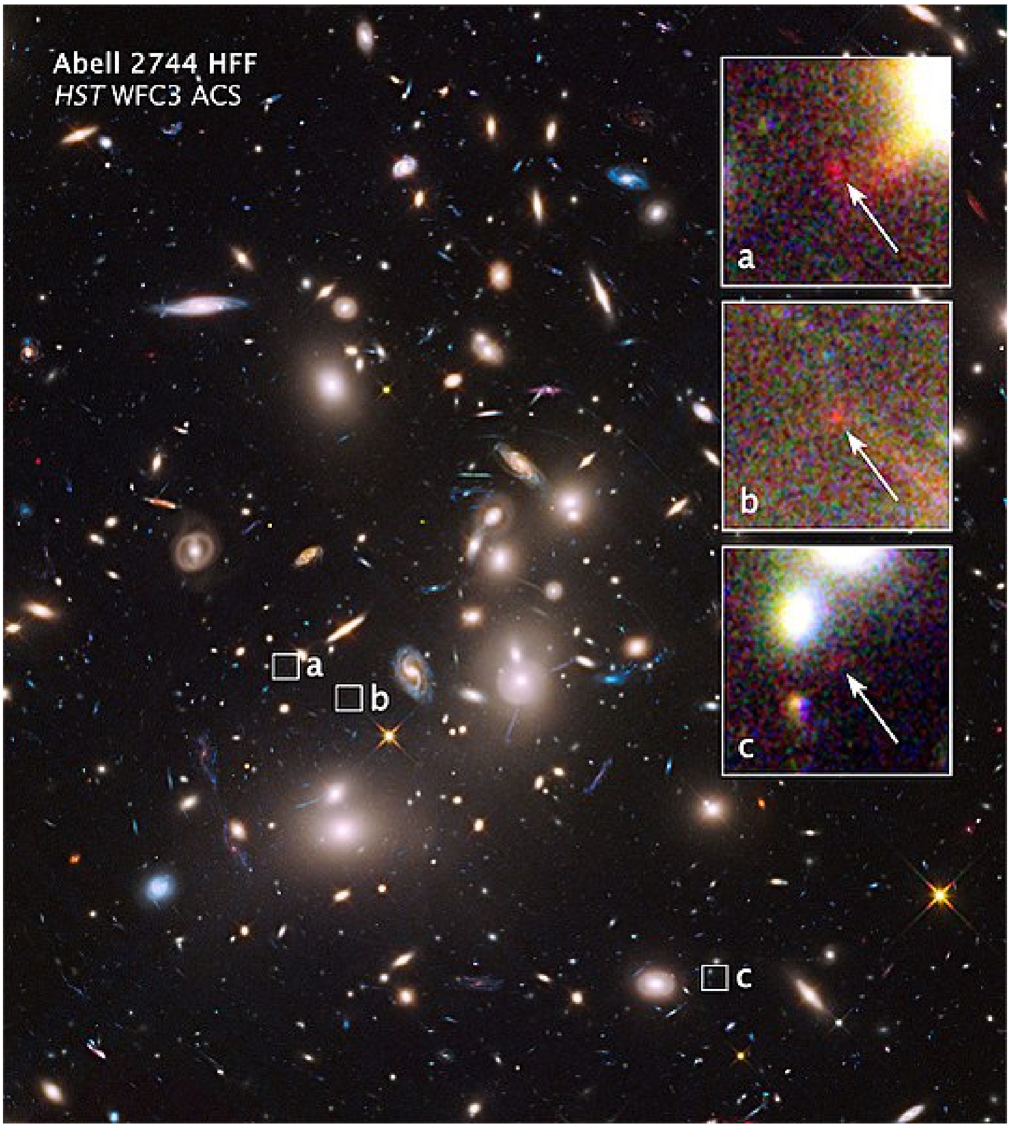


Figure 6.2: Gravitational lensing around galaxy cluster Abell 2744. [NASA, J. Lotz, (STScI)].

Length scales, astrophysics, and cosmology

In considering a gravitational system, if relativistic effects are negligible, we can simply use the Newtonian theory adequately. Assuming spherically-symmetric mass distributions, a

Object	M (kg)	R (m)	$\frac{GM}{c^2R}$
Sun	1.989×10^{30}	6.96×10^{8}	2×10^{-6}
Galaxy	10^{41}	4.628×10^{20}	3.2×10^{-7}
Galaxy cluster	10^{45}	10^{22}	4×10^{-5}

Table 6.1: Orders of magnitude of various objects in cosmology.

reasonable relativistic benchmark is the Schwarzschild spacetime. In SI units,

$$ds^2 = -\left(1 - \frac{2GM}{c^2r}\right)c^2dt^2 + \left(1 - \frac{2GM}{c^2r}\right)^{-1}dr^2 + r^2d\Omega^2.$$

Roughly speaking, if the size of an object is R and the mass is M, then relativistic GR effects are non-negligible when

$$\frac{GM}{c^2R} \sim 1.$$

We can use Newtonian theory if $\frac{GM}{c^2R} \ll 1$. Using

$$G = 6.674 \times 10^{-11}\ \mathrm{kg^{-1}\,m^3\,s^{-2}}, \quad c = 2.998 \times 10^8\ \mathrm{m\,s^{-1}},$$

we find

$$\frac{G}{c^2} = 7.4258 \times 10^{-28}\ \mathrm{kg^{-1}\,m}.$$

So relativistic effects are important when the mass to size ratio $\frac{M}{R}$ is larger than $10^{28}\ \mathrm{kg\,m^{-1}}$. That is, large masses and small sizes.

Let us consider objects of various scales in space. The sun has mass and radius

$$M_\odot = 1.989 \times 10^{30}\ \mathrm{kg}, \quad R_\odot = 6.96 \times 10^8\ \mathrm{m}.$$

In cosmology, the typical unit of length is the *parsec*. A parsec is a length unit

$$1\ \mathrm{pc} = 3.09 \times 10^{22}\ \mathrm{m} \simeq 3.26156\ \text{light years}.$$

Our solar system is a member of the Milky Way galaxy, a collection of about 10^{11} stars. So the mass is $10^{11}M\odot \sim 10^{41}$ kg and the size is about 15 kpc $\sim 4.628 \times 10^{20}$ m.[2] Galaxies themselves are members of a *galaxy cluster*, whose total mass typically lie in the range between $10^{14}M_\odot$ to $10^{15}M_\odot$, and its size is about 1 Mpc to 5 Mpc. From these orders of magnitude, we compute Table 6.1.

[2] The usual unit in cosmologi is the *parsec*, where 1 pc $= 3.0857 \times 10^{16}$ m or 3.26 light years.

We see that all these objects have ratios GM/c^2R of 10^5 or smaller. So Newtonian gravity is enough to describe these phenomena.

Only at length scales *larger* than galaxy clusters that GR effects start to be important. Elaborate surveys an observations concludes that in a region of size $R \sim 6$ Gpc $\sim 3 \times 10^{25}$ m, the average mass density of matter is $\rho \sim 10^{-26}$ kg m^{-3}. So the total mass contained in such a region is around

$$M \sim \frac{4}{3}\pi R^3 \rho \sim 1.23 \times 10^{51}\text{kg}.$$

This gives

$$\frac{GM}{c^2 R} \sim 0.03,$$

so GR is needed in these length scales. Physical phenomena occurring over the length scales of $\sim$ Gpc or larger is called *cosmological length scales*.

6.1 The Robertson–Walker metric

What do we know about our universe?

At the largest scales and in every direction we look at, we see galaxies spread everywhere. It 'looks the same' in any direction. Therefore we make the assumption that the universe is *isotropic* from the point of our observation. A clear picture of an isotropic universe is the *cosmic background radiation* (CMB). The CMB is the background radiation coming from all directions which we can measure. It turns out that the spectrum of this radiation follows a blackbody radiation with a peak giving the temperature of the blackbody. In all directions, the thermal spectrum corresponds to a temperature of $T = 2.725$ K, with small fluctuations of order 10^{-3}. The fluctuations are measured with great precision, such as by the Wilkinson Microwave Anisotropy Probe (WMAP) [29] shown in Fig. 6.3.

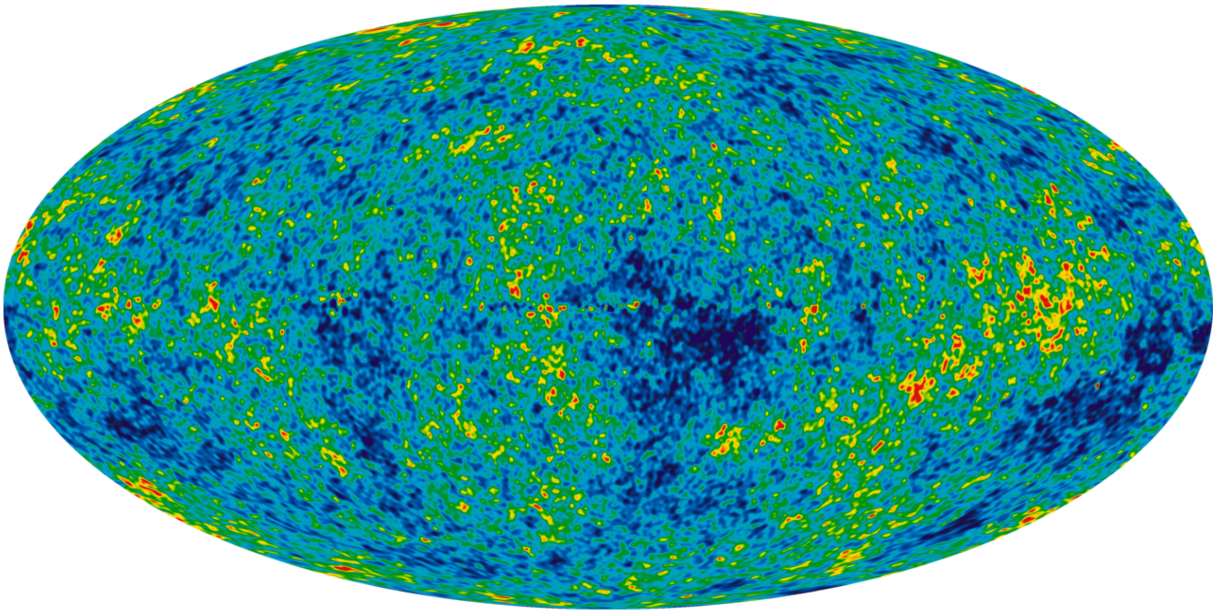


Figure 6.3: The temperature variation of the cosmic microwave background measured by the Wilkinson Microwave Anisotropy Probe (WMAP).

So far, this assumption is based on what we observe from our perspective living on Earth. Will the universe look the same from a different galaxy far away? Let us see the lessons we learned in history. As a whole, humans are quite self-centred. We first assumed that we live in the centre of the universe. Only a few centuries ago did we realise and accept the fact that we are not even the centre of our solar system – that's the Sun. The solar system itself is not at the centre of the Milky Way; we are somewhere near the side. The Milky Way is just one galaxy amongst many other galaxy in the Local Cluster.

This means that it's a wise decision to assume our perspective is not special. This is formulated as the *Copernican principle*, which states

> **Copernican Principle.** *The Earth does not occupy a priviliged position in the universe. Therefore observations from other parts of the universe will be similar to Earth's.*

The Copernican Principle requires that the universe is *homogeneous*. Loosely speaking, this means the universe is the 'same' everywhere. Mathematically, we can use the same metric $g_{\mu\nu}$ to describe observations from Earth or any other galaxy.

Hubble expansion

On the other hand, the universe is not homogenous in time. In the 1920s, Edwin Hubble discovered that our universe is expanding. Hubble's main result is the following: He found that *all* galaxies outside the Milky Way (our home galaxy) is moving away from us! Furthermore, galaxies that are further away are receding away at higher speeds. Let X be the distance of a galaxy, and the speed of the galaxy moving away is $V = \dot{X}$. Hubble's result is that

$$V = HX. \tag{6.1}$$

This is *Hubble's law*. The proportionality factor H is called the *Hubble's parameter* and was found to be about $H = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. The Hubble parameter is not necessarily a constant. The value at present day is called H_0, the *Hubble constant*.

How is it that *every galaxy* is moving away from ours? On the surface it might appear that we are living in the centre of the universe if everything is expanding radially away from us. But, if we apply the Copernican Principle, the explanation is the following: Consider points lying on a uniformly-distributed grid. If this grid expands uniformly as in Fig. 6.4, we will see that from the perspective of each point, all other points are receding away from it. Therefore we are not in a special position in the universe, because all other galaxies will observe the same thing.

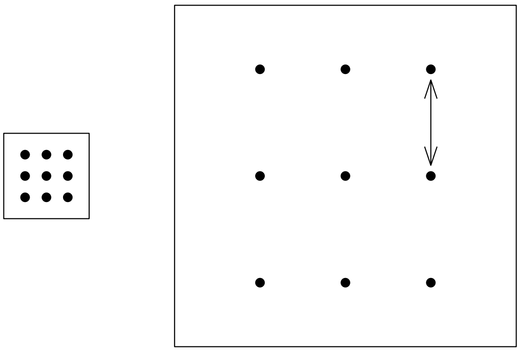


Figure 6.4

This grid explanation leads us to conclude that the galaxies are not actually moving *through* space. Their own proper velocity is zero. Rather, it is the *space itself that expands*, making the distance between two galaxies increasing over time. This can be understood as a coordinate axis that is expanding over time, as in Fig. 6.5. In essence, Hubble's discovery implies that *the universe is expanding*.

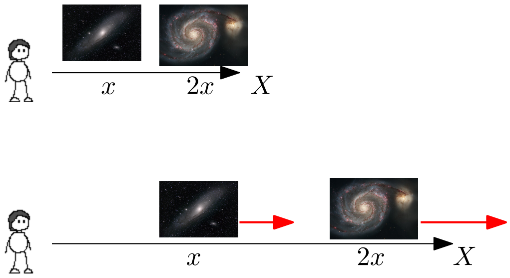


Figure 6.5: A one-dimensional model representing an expanding length scale over time.

In Fig. 6.5, let the observer (Earth) be located at the origin of the X-axis that increases over time. Suppose at time $t = 0$, the position of a galaxy is x. If the axis scale is expanding over time, the distance of this galaxy from the observer is

$$X = a(t)x, \tag{6.2}$$

where $a(t)$ is a *scale factor*, a function increasing with time. The recession speed of this galaxy is obtained by differentiating with respect to t,

$$v = \dot{X} = \dot{a}(t)x. \tag{6.3}$$

Substituting Eqs. (6.2) and (6.3) into Hubble's law gives

$$\begin{aligned} v &= HX \\ \dot{a}x &= Hax \\ H &= \frac{\dot{a}}{a}, \end{aligned}$$

giving an expression of H in terms of scale factor a. The Hubble constant (i.e., Hubble parameter at present day) is therefore

$$H_0 = \frac{\dot{a}(t_0)}{a(t_0)}.$$

With these considerations, we now attempt to write down the metric representing the spacetime of the entire universe.

From observations the expansion rate is the same in all directions of space, so the Copernican principle in the 3 space dimensions continue to hold. Henceforth we shall assume that our universe is *homogeneous and isotropic*. There are precise mathematical definitions of homogeneity and isotropy. But for the present purposes it is easier to think of homogeneity as invariance under spatial translations, and isotropy as invariance under spatial rotations.

At every moment in time, the universe should be homogeneous and isotropic. Therefore the *3-dimensional* part representing the space coordinates is

$$\mathrm{d}\ell^2 = \bar{g}_{ab}\mathrm{d}X^a\mathrm{d}X^b.$$

To be consistent with homogeneity and isotropy, we will take $\bar{g}_{ab}$ to be maximally-symmetric. As discussed above, $X = a(t)x$. Therefore small length intervals will scale as $\mathrm{d}X^a = a(t)\mathrm{d}x^a$. Therefore the three-dimensional spatial metric is (in lightspeed units)

$$\mathrm{d}\ell^2 = a(t)^2\bar{g}_{ab}\mathrm{d}x^a\mathrm{d}x^b.$$

Now we include the time coordinate so that the full spacetime is

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a(t)^2\bar{g}_{ab}\mathrm{d}x^a\mathrm{d}x^b. \tag{6.4}$$

Loosely speaking, $a(t)$ is treated like the 'size' of the universe. We shall fix the convention that t_0 denotes present-day time, and normalise the function such that

$$a(t_0) = 1 \qquad (\text{scale factor of the present day}).$$

Our next task is to determine the form of $\bar{g}_{ab}\mathrm{d}x^a\mathrm{d}x^b$. As mentioned earlier, since we assume our universe is homogeneous and isotropic, we require $\bar{g}_{ab}$ to be maximally-symmetric. By maximal symmetry it should, in particular, contain spherical symmetry. So by similar reasoning we did in Chapter 5, we claim that the metric takes the form

$$\mathrm{d}\sigma^2 = \bar{g}_{ab}\mathrm{d}x^a\mathrm{d}x^b = h(r)\mathrm{d}r^2 + r^2\tilde{\gamma}_{ij}\mathrm{d}\theta^i\mathrm{d}\theta^j,$$

where $\tilde{\gamma}_{ij}\mathrm{d}\theta^i\mathrm{d}\theta^j = \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2 = \mathrm{d}\Omega^2$ is the metric of a unit 2-sphere.

Since it is maximally symmetric, the Ricci tensor of $\bar{g}_{ab}$ should obey

$$\begin{gathered}(\text{Ricci tensor}) = (\text{constant})\,(\text{metric})\\ \bar{R}_{ab} = 2k\bar{g}_{ab}\end{gathered} \tag{6.5}$$

for some constant k. To do this, we need to calculate the Christoffel symbols to get the Ricci tensor. They are

$$\Gamma^r_{rr} = \frac{h'}{2h}, \quad \Gamma^i_{rj} = \frac{1}{r}\tilde{\delta}^i_j, \quad \Gamma^r_{ij} = -\frac{r}{h}\tilde{\gamma}_{ij}, \quad \Gamma^k_{ij} = \tilde{\Gamma}^k_{ij}. \tag{6.6}$$

We just need the Ricci tensor, and not *all* the components of the Riemann tensor. So we can skip the irrelevant parts by immediately computing

$$R_{\sigma\nu} = R^{\mu}{}_{\sigma\mu\nu} = \partial_\mu \Gamma^\mu_{\nu\sigma} - \partial_\nu \Gamma^\mu_{\mu\sigma} + \Gamma^\mu_{\mu\lambda}\Gamma^\lambda_{\nu\sigma} - \Gamma^\mu_{\nu\lambda}\Gamma^\lambda_{\mu\sigma}. \tag{6.7}$$

The result is

$$R_{rr} = \frac{h'}{rh}, \quad R_{ij} = \tilde{\gamma}_{ij}\left(1 - \frac{1}{h} + \frac{rh'}{2h^2}\right). \tag{6.8}$$

Then Eq. (6.5) leads to

$$\frac{h'}{2rh} = kh, \quad 1 - \frac{1}{h} + \frac{rh'}{2h^2} = 2kr^2.$$

A solution that satisfies both equations is

$$h = \frac{1}{1 - kr^2}. \tag{6.9}$$

Therefore the result is $\bar{g}_{ab}\mathrm{d}x^a\mathrm{d}x^b = \frac{\mathrm{d}r^2}{1-kr^2} + r^2\mathrm{d}\Omega^2$, and the metric for the universe is

$$\boxed{\mathrm{d}s^2 = -\mathrm{d}t^2 + a(t)^2\left(\frac{\mathrm{d}r^2}{1 - kr^2} + r^2\mathrm{d}\Omega^2\right).} \tag{6.10}$$

This is the *Robertson–Walker* metric which describes our universe. The form of the scale factor $a(t)$ has to be determined from the Einstein equation. At this stage, the value of the constant k is not yet established. We have three possibilities: $k = 0$ (flat universe), $k > 0$ (closed universe), and $k < 0$ (open universe).

6.2 The Friedmann equation

So far, we have obtained the Robertson–Walker metric based on general observations of homogeneity and isotropy, along with the fact that the universe is expanding. Aside from that, a cosmologist seeks answers to the following questions:

- How did the universe *start*? What are the conditions in the early universe before the expansion reached its current state today?
- What is the future of the universe? Will it continue expanding forever? Or will the expansion slow down and it will contract again.
- What is the spatial curvature of the universe? In other words, what is the value of k in the Robertson–Walker metric?

Answers to most of these questions lies in the function $a(t)$. So we need to determine what is actually this function. Is it monotinically increasing? Oscillating? Does it have a root; which means $a(t) = 0$?

By the theory of General Relativity, this can be determined by the Einstein equation. Whatever $a(t)$ is, the Robertson–Walker metric must satisfy the Einstein equation. This will be our present task.

To get the Einstein equation, we need to compute the Einstein tensor. Which means we need to get the Christoffel symbols and Ricci tensor. To this end, let us write Eq. (6.10) again as

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a(t)^2 \bar{g}_{ab} \mathrm{d}x^a \mathrm{d}x^b.$$

The non-zero Christoffel symbols are

$$\Gamma^c_{tb} = \frac{\dot{a}}{a} \bar{\delta}^c_b, \quad \Gamma^t_{ab} = a\dot{a}\bar{g}_{ab}, \quad \Gamma^c_{ab} = \bar{\Gamma}^c_{ab}, \tag{6.11}$$

where $\dot{a} = \frac{\mathrm{d}a}{\mathrm{d}t}$. The components $\bar{\Gamma}^c_{ab}$ were already calculated in (6.6).

Using the fact that $\bar{R}_{ab} = 2k\bar{g}_{ab}$, the Ricci tensor components are

$$R_{tt} = -3\frac{\ddot{a}}{a}, \quad R_{ab} = \left(\frac{\ddot{a}}{a} + 2\frac{\dot{a}^2}{a^2} + \frac{2k}{a^2}\right) a^2 \bar{g}_{ab}. \tag{6.12}$$

For the right hand side of Einstein's equation, we need the stress tensor. Here we will simply model the matter content of the universe with that of a perfect fluid,

$$T_{\mu\nu} = (\rho + p)\, u_\mu u_\nu + p g_{\mu\nu}. \tag{6.13}$$

This means that, the universe is so large that the galaxies inside it are like particles in a fluid. As mentioned in the beginning of this chapter, the typical length scale for this to hold is about 6 gigaparsecs.

Let assume that the fluid co-moving. In other words, the fluid (stars and galaxies) are generally static in the frame of the Robertson–Walker metric. So the 4-velocity is

$$u^\mu = (1, 0, 0, 0), \quad u_\mu = g_{\mu\nu} u^\nu = (-1, 0, 0, 0).$$

With this 4-velocity, the components of the stress tensor is

$$T_{tt} = \rho, \quad T_{ab} = p a^2 \bar{g}_{ab}.$$

We will also need the components with one index raised, $T^\mu{}_\nu = g^{\mu\lambda}T_{\lambda\nu}$. The non-zero components are

$$\begin{aligned} T^t{}_t &= g^{tt}T_{tt} = -\rho, \\ T^a{}_b &= g^{ac}T_{cb} = \frac{1}{a^2}\bar{g}^{ac} \cdot pa^2\bar{g}_{cb} = p\bar{\delta}^a_b. \end{aligned}$$

We will also require the trace,

$$T^\lambda{}_\lambda = T^t{}_t + T^c{}_c = -\rho + 3p$$

The relation between ρ, p, and scale factor a can be seen with the conservation equation,

$$0 = \nabla_\mu T^\mu{}_\nu = \partial_\mu T^\mu{}_\nu + \Gamma^\mu_{\mu\lambda}T^\lambda{}_\nu - \Gamma^\lambda_{\mu\nu}T^\mu{}_\lambda.$$

The component $\nu = t$ is

$$\begin{aligned} 0 &= \partial_\mu T^\mu{}_t + \Gamma^\mu_{\mu\lambda}T^\lambda{}_t - \Gamma^\lambda_{\mu t}T^\mu{}_\lambda \\ &= \partial_t T^t{}_t + \Gamma^\mu_{\mu t}T^t{}_t - \Gamma^t_{tt}T^t{}_t - \Gamma^c_{dt}T^d{}_c \\ &= -\dot{\rho} + \Gamma^c_{ct}T^t{}_t - 0 - \Gamma^c_{dt}T^d{}_c \\ &= -\dot{\rho} + 3\frac{\dot{a}}{a}(-\rho) - \frac{\dot{a}}{a}\bar{\delta}^c_d p\bar{\delta}^d_c \\ &= -\dot{\rho} - 3\frac{\dot{a}}{a}\rho - 3\frac{\dot{a}}{a}p. \end{aligned}$$

The result is called the *continuity equation*,

$$\boxed{\dot{\rho} + 3\frac{\dot{a}}{a}\left(\rho + p\right) = 0.}$$

The component $\nu = a$ is identically satisfied. (Tutorial.)

Turning to the Einstein equation itself, we use the trace-reversed form

$$R_{\mu\nu} = \Lambda g_{\mu\nu} + 8\pi G\left(T_{\mu\nu} - \frac{1}{2}T^\lambda{}_\lambda g_{\mu\nu}\right). \tag{6.14}$$

The (tt)-component is

$$\begin{aligned}
R_{tt} &= 8\pi G\left(T_{tt}-\frac{1}{2}T^{\lambda}{}_{\lambda}g_{tt}\right)\\
-3\frac{\ddot{a}}{a} &= \Lambda g_{tt}+8\pi G\left[\rho-\frac{1}{2}(3p-\rho)(-1)\right]\\
-3\frac{\ddot{a}}{a} &= -\Lambda+8\pi G\left[\rho+\frac{3}{2}p-\frac{1}{2}\rho\right]\\
\frac{\ddot{a}}{a} &= \frac{\Lambda}{3}-\frac{4\pi}{3}G\left(\rho+3p\right). \qquad (6.15)
\end{aligned}$$

The (ab)-component is

$$\begin{aligned}
R_{ab} &= \Lambda g_{ab}+8\pi G\left(T_{ab}-\frac{1}{2}T^{\lambda}{}_{\lambda}g_{ab}\right)\\
\left(\frac{\ddot{a}}{a}+2\frac{\dot{a}^2}{a^2}+\frac{2k}{a^2}\right)a^2\bar{g}_{ab} &= \Lambda a^2\bar{g}_{ab}+8\pi G\left[pa^2\bar{g}_{ab}-\frac{1}{2}\left(3p-\rho\right)a^2\bar{g}_{ab}\right]\\
\frac{\ddot{a}}{a}+2\frac{\dot{a}^2}{a^2}+\frac{2k}{a^2} &= \Lambda+8\pi G\left[-\frac{1}{2}p+\frac{1}{2}\rho\right]\\
\frac{\ddot{a}}{a}+2\frac{\dot{a}^2}{a^2}+\frac{2k}{a^2} &= \Lambda+4\pi G\left(\rho-p\right). \qquad (6.16)
\end{aligned}$$

We use Eq. (6.15) to substitute for $\frac{\ddot{a}}{a}$,

$$\begin{aligned}
\frac{\Lambda}{3}-\frac{4\pi}{3}G\left(\rho+3p\right)+2\frac{\dot{a}^2}{a^2}+\frac{2k}{a^2} &= \Lambda+4\pi G\left(\rho-p\right)\\
2\frac{\dot{a}^2}{a^2}+\frac{2k}{a^2} &= \frac{2\Lambda}{3}+4\pi G\left(\rho-p+\frac{\rho}{3}+p\right)\\
\frac{\dot{a}^2}{a^2}+\frac{k}{a^2} &= \frac{\Lambda}{3}+\frac{8\pi}{3}G\rho.
\end{aligned}$$

Simplifying and cancelling an overall factor of 2, we get the *Friedmann equation*,

$$\boxed{\frac{\dot{a}^2}{a^2}=\frac{\Lambda}{3}+\frac{8\pi}{3}G\rho-\frac{k}{a^2}.} \qquad (6.17)$$

The continuity equation (6.14) and the Friedmann equation (6.17) determine the scale factor $a(t)$, and therefore the evolution of the universe. The study of the evolution of the universe using these equations is called *Friedmann–Lemaître–Robertson–Walker (FLRW) cosmology.*

From Eq. (6.17), it is the combination $\dot{a}a$ depends on a simple sum of cosmological constant, matter, and curvature terms. We define this convenient combination as the *Hubble* *parameter*

$$H = \frac{\dot{a}}{a}, \tag{6.18}$$

which depends on t. The the present-day value H_0 is an important measurement target in observational cosmology. However, there is no clear agreement as to the value of H_0. Various observational methods seem to converge towards two different values. In SI units, they are

$$H_0 = 73\ \mathrm{km\,s^{-1}\,Mpc^{-1}} \quad \text{vs.} \quad H_0 = 67.7\ \mathrm{km\,s^{-1}\,Mpc^{-1}}, \quad (\textsf{SI units}).$$

In lightspeed units, we take $\frac{H_0}{c} \to H_0$ and they are

$$H_0 = 2.44 \times 10^{-4}\ \mathrm{Mpc^{-1}}, \quad H_0 = 2.26 \times 10^{-4}\ \mathrm{Mpc^{-1}}, \quad (\textsf{lightspeed units}).$$

This problem is called the *Hubble tension* and at the time of writing, still unresolved.

6.3 Equation of state

At this stage, ρ and p are still arbitrary. The *type* of matter may determine a relation between ρ and p. Such an equation is called an *equation of state*. The fluids in cosmology can be assumed have a linear equation of state

$$p = w\rho, \tag{6.19}$$

where w is some constant depending on the type of matter involved. With this equation of state, the continuity equation (6.14) becomes

$$\begin{aligned}
\frac{\dot{\rho}}{\rho} &= -3(1+w)\frac{\dot{a}}{a} \\
\frac{1}{\rho}\frac{\mathrm{d}\rho}{\mathrm{d}t} &= -3(1+w)\frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}t} \\
\int \frac{1}{\rho}\frac{\mathrm{d}\rho}{\mathrm{d}t}\mathrm{d}t &= -3(1+w)\int \frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}t}\mathrm{d}t \\
\int \frac{\mathrm{d}\rho}{\rho} &= -3(1+w)\int \frac{\mathrm{d}a}{a} \\
\ln\rho &= -3(1+w)\ln a + \text{const.}
\end{aligned}$$

Solving for ρ, we have

$$\rho = \rho_0 a^{-3(1+w)}, \tag{6.20}$$

where C is some constant. The matter density of the universe changes with the scale factor a.

We list a few common types of matter considered in cosmology.

- **Matter.** In this case $w = 0$. This means $p = 0$, and we treat the fluid as a pressureless dust. This is a close approximation of present-era universe. In this case we have $\rho \propto a^{-3}$. The matter content gets diluted as the universe expands.
- **Radiation.** In the early universe, before the formation of stars and galaxies, the universe is mostly filled with radiation (photons). The equation of state for radiation is $w = \frac{1}{3}$, for which a^{-4}. This is because the number of photons are like particles, which scale like a^{-3}. But the photons themselves lose energy because they get redshifted. This introduces another factor of a^{-1}.
- **Vacuum.** In this case, $w = -1$ and we get $\rho = \text{const}$. This is essentially the same behaviour as a cosmological constant.

6.4 Evolution of the universe

The scale factor $a(t)$ is seen to obey the differential equation (6.17). If the solution is known, the evolution of the matter density ρ is determined with Eq. (6.14).

In the present day, we assume most of the density ρ comes from the galactic matter as pressureless dust, so we take $w = 0$ in Eq. (6.20). To analyse the Friedmann equation (6.17), it is easier to express the equation in terms of the Hubble parameter $H(t) = \dot{a}(t)/a(t)$,

$$H(t)^2 = \frac{8\pi}{3} G\rho_0 a^{-3} + \frac{\Lambda}{3} - ka^{-2}.$$

Divide this by the present-day Hubble constant,

$$\frac{H(t)^2}{H_0} = \frac{8\pi G\rho_0}{3H_0^2} a^{-3} + \frac{\Lambda}{3H_0^2} + \left(-\frac{k}{H_0^2}\right) a^{-2}.$$

Define the constant parameters

$$\Omega_{\rm m} = \frac{8\pi G\rho_0}{3H_0^2}, \quad \Omega_\Lambda = \frac{\Lambda}{3H_0^2}, \quad \Omega_k = -\frac{k}{H_0^2},$$

the Friedmann equation is now written more compactly as

$$\boxed{\frac{H(t)^2}{H_0^2} = \Omega_{\rm m}a^{-3} + \Omega_\Lambda + \Omega_k a^{-2}.} \tag{6.21}$$

Sometimes cosmologists prefer to express the cosmological constant and curvature terms as effective 'densities',

$$\rho_\Lambda = \frac{3H_0^2}{8\pi G}\Omega_\Lambda, \quad \rho_k = \frac{3H_0^2}{8\pi G}\Omega_k$$

so the above equation is written as $\frac{H(t)^2}{H_0^2} = \frac{8\pi G}{3H_0^2}\left(\rho_0 a^{-3} + \rho_\Lambda + \rho_k a^{-2}\right)$. We now analyse Eq. (6.21) in attempt to draw some conclusions about our universe.

Einstein's epic fail. First, before Hubble's observations, it was assumed that the universe as a whole is flat and static. Therefore $\Omega_k = 0$ and $H = \dot{a}/a = 0$. Then we have

$$\Omega_{\rm m}a^{-3} = -\Omega_\Lambda$$

If the cosmological constant is zero, a flat and static universe requires $\Omega_{\rm m} = 0$. This was the reason why Einstein assumed the cosmological constant is non-zero and included it in the Einstein equation.

If the universe is non-static, $H(t) \neq 0$. Even if it's flat, we have $H(t)^2/H_0^2 = \Omega_{\rm m}a^{-3} + \Omega_\Lambda$ which can contain $\Omega_{\rm m} \neq 0$ even if Ω_Λ. Hubble discovered the expansion of the universe which leads to $H(t) \neq 0$, and Einstein said adding the cosmological constant was the '*biggest blunder of his life*', and he shouldn't have added Λ to his equation in the first place. In another plot twist occuring around 1998, it turns out that it is probably $\Lambda \neq 0$, for a different reason. (We will study this in the next section.) Apparently Einstein's actual blunder was to think he blundered the first time. For now we do not assume anything about Λ yet and see where the analysis takes us.

Is our universe flat? Now consider the present day, $H(t_0) = H_0$ and $a(t_0) = 1$, Eq. (6.21) is

$$1 = \Omega_{\rm m} + \Omega_\Lambda + \Omega_k. \tag{6.22}$$

Is our universe flat, open, or closed? We see that it depends on

$$\Omega_k = 1 - \Omega_\mathrm{m} - \Omega_\Lambda.$$

So it is open if $\Omega_\mathrm{m} + \Omega_\Lambda < 1$, closed if $\Omega_\mathrm{m} + \Omega_\Lambda > 1$ universe is open, and flat if $\Omega_\mathrm{m}+\Omega_\Lambda = 0$. the matter content of the universe. Various cosmological observations can be done independently of Ω_m and Ω_Λ, and they conclude that our universe is approximately flat. So often one concludes $\Omega_k \simeq 0$, which leads to

$$\Omega_\mathrm{m} + \Omega_\Lambda \simeq 1. \tag{6.23}$$

The future of our universe. As observed by Hubble, the universe is expanding. Will it continue to expand forever? Will the expansion stop, or will $a(t)$ reverse direction and the universe recollapses. We can see this by checking whether $\dot{a}$ changes sign, which occurs if $H(t_*) = \dot{a}(t_*)/a(t_*) = 0$. Let the scale factor at this critical time be $a_* = a(t_*)$. Then Eq. (6.21) is

$$\begin{aligned} \Omega_\Lambda a_*^3 + \Omega_k a_* + \Omega_\mathrm{m} &= 0 \\ \Omega_\Lambda a_*^3 + (1 - \Omega_\mathrm{m} - \Omega_\Lambda)\, a_* + \Omega_\mathrm{m} &= 0, \end{aligned} \tag{6.24}$$

where we have used Eq. (6.22) to write $\Omega_k = 1 - \Omega_\mathrm{m} - \Omega_\Lambda$. So the situation depends on two independent quantities Ω_m and Ω_Λ. Now, Eq. (6.24) is a polynomial of degree 3. The existence of a real root implies there exist a point a_* where $H(t)$ changes sign, and the universe recollapses.

To this end, let $a_* = x$ and consider the polynomial

$$P(x) = \Omega_\Lambda x^3 + (1 - \Omega_\mathrm{m} - \Omega_\Lambda)\, x + \Omega_\mathrm{m}.$$

The present-day value is $P(1) = 1$. But if Ω_Λ is negative, we see that $\lim_{x\to\infty} P(x) = -\infty$. So the graph crosses from $P = 1$ to $P < 0$ at some point. In other words, the universe will always recollapse if $\Omega_\Lambda \sim \Lambda$ is negative. Are there positive values of Λ for which the universe collapses? We investigate the discriminant of P. They can be obtained by solving

$$P(x) = P'(x) = 0.$$

Solving the simultaneous equation for Ω_m and Ω_Λ,we get

$$\Omega_\mathrm{mc} = \frac{2x^3}{2x^3 - 3x^2 + 1}, \quad \Omega_{\Lambda\mathrm{c}} = \frac{1}{2x^3 - 3x^2 + 1}.$$

Here $\Omega_{\rm m} > 1$ in the domain $x > 1$ for future values. In particular $\Omega_{\Lambda c} > 0$ for $x > 1$.One can verify that in the domain

$$\Omega_{\rm mc} > 1, \quad \Omega_\Lambda \leq \Omega_{\Lambda c}, \tag{6.25}$$

which contains a positive region for Ω_Λ.

Our reasoning can be summarised in Fig. 6.6. Points $(\Omega_{\rm m}, \Omega_\Lambda)$ lying on the blue curve means $P(x)$ will have a degenerate root. Hence points above the blue curve means there are complex or negative roots, and our universe will expand forever, and points below means the universe will recollapse. The red line is the equation $\Omega_{\rm m} + \Omega_\Lambda = 1$, which means a flat universe. So far, the observational data constrains the values to the shaded region for a flat universe around $(\Omega_{\rm m}, \Omega_\Lambda) = (0.3, 0.7)$. (We will see how these values are obtained in the next section.) Therefore, it is likely that our universe will expand forever.

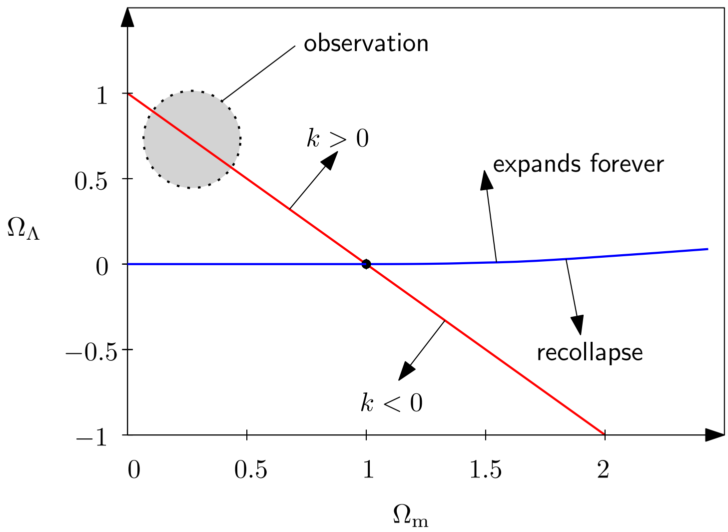


Figure 6.6: Parameters for $(\Omega_{\rm m}, \Omega_\Lambda)$ and possible future evolution of the universe. The shaded region indicates the best estimate of $(\Omega_{\rm m}, \Omega_\Lambda)$ from observations.

6.5 Redshift and measurements

The scale factor $a(t)$ is not easily measurable. To measure anything in the universe, we only have access to the light that reaches the earth. If photons are emitted from a distant galaxy, they take billions of years to finally reach the Earth. So astronomical observations of photons (and today, possibly gravitational waves) provide the only access to information about the early universe. Therefore we should consider null geodesics in the Robertson–Walker metric.

Redshift

For simplicity, let us assume we are in a flat universe so $k = 0$. We also choose our coordinate system so that the photon is travelling radially. We define two events: Event P is the moment and place the photon is emitted, and event Q is the moment the photon is observed,

$$P(t_{\text{em}}, 0, 0, 0), \quad Q(t_{\text{obs}}, r_{\text{obs}}, 0, 0). \tag{6.26}$$

The photon is travelling radially at constant $\theta = \phi = 0$. So $\dot{\theta} = \dot{\phi} = 0$ and the Lagrangian is

$$\mathcal{L} = \frac{1}{2}\alpha\left(-\dot{t} + a^2\dot{r}^2\right).$$

This metric is independent of r. Hence the radial momentum

$$p_r = \frac{\partial \mathcal{L}}{\partial \dot{r}} = \alpha a^2 \dot{r}$$

is constant. The momentum in the t-direction is

$$p_t = \frac{\partial \mathcal{L}}{\partial \dot{t}} = -\alpha \dot{t},$$

which is not constant.[3] Since the photon follows null geodesics,

$$\begin{aligned} p_\mu p^\mu &= 0 \\ -p_t^2 + \frac{1}{a^2}p_r^2 &= 0 \\ p_t &= -\frac{p_r}{a}. \end{aligned}$$

We now consider the energy measured by a static observer, whose 4-velocity is $V^\mu = (1, 0, 0, 0)$. The energy measured is

$$E = -V^\mu p_\mu = -p_t = \frac{p_r}{a(t)}.$$

So this energy depends on the value of the scale factor.

Suppose that this photon is emitted from static star at time t_{em}, and reaches a static

[3] To show this, check the Euler–Lagrange equation.

observer at time $t_{\rm obs}$. Then,

$$E_{\rm em} = \frac{p_r}{a(t_{\rm em})}, \quad E_{\rm obs} = \frac{p_r}{a(t_{\rm obs})}.$$

Since p_r is constant, it is the same on both equations. We can eliminate it by dividing,

$$\frac{E_{\rm obs}}{E_{\rm em}} = \frac{a(t_{\rm em})}{a(t_{\rm obs})}. \tag{6.27}$$

Using Planck's formula $E = \hbar\omega = 2\pi\hbar f = \hbar c/\lambda$ for the energy of a photon with wavelength λ, we have the relations between the frequencies and wavelengths at the emitter and observer,

$$\frac{\lambda_{\rm em}}{\lambda_{\rm obs}} = \frac{a(t_{\rm em})}{a(t_{\rm obs})}, \quad \frac{f_{\rm obs}}{f_{\rm em}} = \frac{a(t_{\rm em})}{a(t_{\rm obs})}$$

In other words, the colour of the photon has changed as it travels across the universe. Furthermore, since $\delta\tau = 1/f$ is the frequency of the electromagnetic oscillations, we have a relation between time intervals between observer and emitter,

$$\frac{\delta\tau_{\rm em}}{\delta\tau_{\rm obs}} = \frac{a(t_{\rm em})}{a(t_{\rm obs})}. \tag{6.28}$$

The *redshift* is defined as the fractional change in wavelength,

$$z = \frac{\lambda_{\rm obs} - \lambda_{\rm em}}{\lambda_{\rm em}}.$$

It is this redshift z that is the measurable quantity obtained by investigating the absorptoin spectrum of the light received by telescopes. Of course, this light is measured by present-day astronomers, so $t_{\rm em} = t_0$. Since we have fixed $a(t_0) = 1$, solving for $a(t_{\rm em})$ gives

$$\boxed{a(t_{\rm em}) = \frac{1}{1+z}.} \tag{6.29}$$

With this relation, Eqs. (6.27) and (6.28) can be expressed in terms of z with

$$E_{\rm obs} = \frac{E_{\rm em}}{1+z}, \quad \delta\tau_{\rm obs} = (1+z)\delta\tau_{\rm em}. \tag{6.30}$$

Luminosity distance of supernovae

Because the spatial part of the metric changes with time, the concept of 'distance' is tricky to define. Since the only observable quantity is the photons from distant galaxies, we attempt to use light measurements to define a concept of distance. Then we are able to discuss the distances of various distant objects.

To start, consider the observation of light made in the present day, $t_{\text{obs}} = t_0$, emitted from a distant source. This light has to be emitted at some earlier time $t_{\text{em}} < t_0$. At this time, $a(t_{\text{em}}) < (a_0 = 1)$ the universe was smaller. When considering light sources in astronomy, we consider the *luminosity* defined by

$$L = (\textsf{energy radiated per unit time}) = \frac{\delta E}{\delta \tau}.$$

Measurement devices detect the *flux* of light it received,

$$F = (\textsf{energy per unit time per area}) = \frac{\delta E}{A\delta \tau},$$

The value of $L = \frac{\delta E_{\text{em}}}{\delta \tau_{\text{em}}}$ depends on the physical property of the emitter. Of course, we are far from the emitter, and can only measure the flux of photons $F = \frac{\delta E_{\text{em}}}{A_{\text{obs}}\delta \tau_{\text{em}}}$ that arrived in Earth, as depicted in Fig. 6.7.

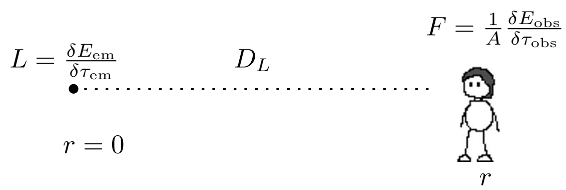


Figure 6.7: Emission of photon at luminosity L from an emitter. The flux F is measured by an observer.

Note that L/F has the units of area. In the non-relativistic case, this is the area of a sphere of radius equal to the distance between emitter and observer. With these considerations we define the *luminosity distance* as

$$4\pi D_L^2 = \frac{L}{F}$$

Note that we are not in Minkowski space, so $D_L \neq r$!

The flux measured by the observer is

$$F = \frac{1}{A}\frac{\delta E_{\text{obs}}}{\delta \tau_{\text{obs}}}$$

Using Eq. (6.30), we express in terms of E_{em} and $\delta\tau_{\text{em}}$,

$$F = \frac{1}{A}\frac{\frac{\delta E_{\text{em}}}{1+z}}{(1+z)\delta\tau_{\text{em}}} = \frac{1}{A(1+z)^2}\frac{\delta E_{\text{em}}}{\delta\tau_{\text{em}}}.$$

But $\frac{\delta E_{\text{em}}}{\delta\tau_{\text{em}}}$ is just the luminosity L, and we have

$$F = \frac{L}{A(1+z)^2} \quad \to \quad \frac{L}{F} = A(1+z)^2.$$

Substituting this result into the definition of luminosity distance,

$$4\pi D_L^2 = A(1+z)^2. \tag{6.31}$$

Now we compute the area. This is the proper area of a sphere at coordinate r,

$$A = \int \sqrt{g_{\theta\theta}g_{\phi\phi}}\mathrm{d}\theta\mathrm{d}\phi = \int_0^{2\pi}\mathrm{d}\phi\int_0^{\pi}\mathrm{d}\theta\; a^2r^2 = 4\pi a_0^2r^2. \tag{6.32}$$

For the present day, $a_0 = 1$ and we get $A = 4\pi r^2$ which agrees with the usual formula. Therefore

$$4\pi D_L^2 = 4\pi r^2(1+z)^2 \quad \to \quad D_L = (1+z)r.$$

Now, r is the total coordinate distance covered by the photon,

$$r = \int_0^r \mathrm{d}r'.$$

Since these photons travel along radial null geodesics,

$$\mathrm{d}s^2 = 0 = -\mathrm{d}t^2 + a^2\mathrm{d}r^2 \quad \to \quad \mathrm{d}r = \frac{\mathrm{d}t}{\mathrm{d}a}.$$

Therefore

$$r = \int_{t_{\text{em}}}^{t_0}\frac{\mathrm{d}t}{a(t)}.$$

By the chain rule,

$$\mathrm{d}a = \frac{\mathrm{d}a}{\mathrm{d}t}\mathrm{d}t = \dot{a}\mathrm{d}t \quad \to \quad \mathrm{d}t = \frac{\mathrm{d}a}{\dot{a}} = \frac{\mathrm{d}a}{aH},$$

where we have used the definition of the Hubble parameter $H = \frac{\dot{a}}{a}$. Then the r integral is

$$r = \int_{a_{\mathrm{em}}}^{a_0=1} \frac{\mathrm{d}a}{a^2 H}.$$

Finally, changing variables in terms of redshift,

$$a = \frac{1}{1+z}, \quad \mathrm{d}a = -\frac{\mathrm{d}z}{(1+z)^2} = -a^2 \mathrm{d}z,$$

the integral becomes

$$r = -\int_z^0 \frac{\mathrm{d}z}{H} = \int_0^z \frac{\mathrm{d}z'}{H(z')}.$$

The expression for luminosity distance is therefore.

$$D_L = (1+z) \int_0^z \frac{\mathrm{d}z'}{H(z')}$$

Recall that the expression for H is given by Eq. (6.21),

$$H = H_0 \sqrt{\Omega_{\mathrm{m}} a^{-3} + \Omega_\Lambda + \Omega_k a^{-2}} = H_0 \sqrt{\Omega_{\mathrm{m}} (1+z)^3 + \Omega_\Lambda + \Omega_k (1+z)^2}.$$

Therefore the expression for luminosity distance is

$$D_L = (1+z) \int_0^z \frac{\mathrm{d}z'}{H_0 \sqrt{\Omega_{\mathrm{m}} (1+z)^3 + \Omega_\Lambda + \Omega_k (1+z)^2}}. \tag{6.33}$$

In a FLRW cosmology with parameters Ω_{m}, Ω_Λ, and Ω_k, the luminosity distance of a star with redshift z has luminosity distance D_L given above.

For technical reasons, astronomers prefer to use the *distance modulus* defined by

$$\mu = 25 + 5 \log_{10} \frac{D_L}{\mathrm{Mpc}}. \tag{6.34}$$

Since we are assuming the universe is flat, $\Omega_k = 0$. So, for a given set $(\Omega_{\mathrm{m}}, \Omega_\Lambda)$, we get a particular graph of μ vs z. These two can be measured observationally. Note that D_L depends on L/F. The flux F is what a telescope on Earth measures. If the L of an emmitter is known, the observational value of D_L is determined from $D_L^{\mathrm{obs}} = \sqrt{L/4\pi F}$. This can be compared with Eq. (6.33).

The problem now is how do we know L of an emitter? There are so many kinds of stars in space with various complicated structures, various factors will affect its emission rate L. In order to make progress, we need to seek a type of object that always emits at a known luminosity. These are called *standard candles* of cosmology. According to astrophysics, Type Ia supernova serves as standard candles. So astronomers look for Type Ia supernova events in distant galaxies, measure the flux F, then use it to get $D_L^{\mathrm{obs}} = \sqrt{L/4\pi F}$ and subsequently the observed distance modulus μ_{obs}. This is a daunting task, as supernova events are rare. So one has to search the sky very thoroughly. In the late 90s, two independent research collaborations managed to do so. They are the Supernova Cosmology Project [30] and the High-z Supernova Search Team [31]. With this data, one can adjust the parameters $(\Omega_{\mathrm{m}}, \Omega_\Lambda)$ to find a curve that best fits the data.

It turns out the result is

$$(\Omega_{\mathrm{m}}, \Omega_\Lambda) \simeq (0.3, 0.7).$$

The data and fitted curve is shown in Fig. 6.8, using more updated dataset of nearly 580 Type Ia supernovae and a rough value of $H_0 = 70 \times 10^3\ \mathrm{m\,s^{-1}\,Mpc^{-1}}$.[4]

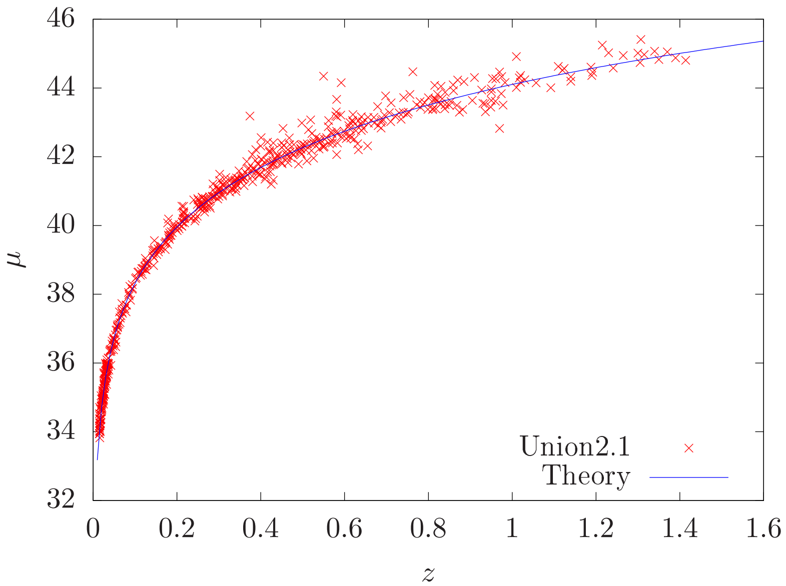


Figure 6.8: The distance modulus vs redshift, using the 2011 Union2.1 dataset from the Supernova Cosmology Project [32]. `https://supernova.lbl.gov/Union/`. The theoretical curve is plotted using Eq. (6.33) with $H_0 = 70 \times 10^3\ \mathrm{m\,s^{-1}\,Mpc^{-1}}$, $\Omega_{\mathrm{m}} = 0.3$, and $\Omega_\Lambda = 0.7$.

What does this mean for the evolution of the universe? Going back to the Einstein equation for $\ddot{a}$, Eq. (6.15) is for pressureless dust $w = 0 \to p = 0$,

$$\frac{\ddot{a}}{a} = \frac{\Lambda}{3} - \frac{8\pi G}{3}\rho = H_0^2 \left(\frac{\Lambda}{3H_0} - \frac{8\pi G\rho}{3H_0^2} \right).$$

[4] Which is $H_0 = 2.43 \times 10^{-4}\ \mathrm{Mpc}^{-1}$ in lightspeed units.

For the present day, $a = a_0 = 1$ and $\rho = \rho_0$ and we have

$$\ddot{a} = H_0^2 \left(\Omega_\Lambda - \Omega_\mathrm{m}\right) \simeq 0.4 H_0^2 > 0.$$

This means our universe is accelerating! Furthermore, this means the cosmological constant Λ has a value of

$$\begin{aligned}
\Lambda = 3\Omega_\Lambda H_0^2 &= 3 \times 0.7 \times \left(2.43 \times 10^{-4}\ \mathrm{Mpc}^{-1}\right)^2 \\
&= 3 \times 0.7 \times \left(\frac{2.43 \times 10^{-4}}{3.0857 \times 10^{16} \times 10^{6}\ \mathrm{m}}\right)^2 \\
&= 8.02 \times 10^{-52}\ \mathrm{m}^{-2}.
\end{aligned}$$

(using SI $H_0 = 70 \times 10^3\ \mathrm{m\,s^{-1}\,Mpc^{-1}}$, which is $H_0 = 2.43 \times 10^{-4}\ \mathrm{Mpc}^{-1}$ in lightspeed units.) Therefore the cosmological constant has a small but non-zero value. The leaders of the two groups (Perlmutter from SCP, Schmidt and Reiss in High-z) won the 2011 Nobel prize for this discovery.

Chapter 7 Weak fields and gravitational waves

7.1 Linearised Einstein equation

We now attempt to simplify Einstein's equation by considering the case where gravity is weak. That is, the spacetime is a small perturbation of Minkowski space,

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}. \tag{7.1}$$

The idea is to formulate the mathematics such that the main spacetime is still Minkowski, and we seek the evolution of $h_{\mu\nu}$ as a tensor field on Minkowski spacetime. When calculating the Ricci tensor for the metric $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$, we will ignore terms of h^2, since we consider $h_{\mu\nu}$ to be small. So throughout all the calculations, the equations are always linear in $h_{\mu\nu}$.

We start with the Christoffel symbol,

$$\Gamma^\kappa_{\mu\nu} = \frac{1}{2}\left(\eta^{\kappa\lambda} + h^{\kappa\lambda}\right)\left[\partial_\mu\left(\eta_{\lambda\nu} + h_{\lambda\nu}\right) + \partial_\nu\left(\eta_{\lambda\mu} + h_{\lambda\mu}\right) - \partial_\lambda\left(\eta_{\mu\nu} + h_{\mu\nu}\right)\right].$$

Keeping only terms linear in $h_{\mu\nu}$, and noting that derivatives of the Minkowski metric vanishes, we have

$$\Gamma^\kappa_{\mu\nu} = \frac{1}{2}\eta^{\kappa\lambda}\left(\partial_\mu h_{\lambda\nu} + \partial_\nu h_{\lambda\mu} - \partial_\lambda h_{\mu\nu}\right).$$

The Riemann tensor is $R^\rho{}_{\sigma\mu\nu} = \partial_\mu\Gamma^\rho_{\nu\sigma} - \partial_\nu\Gamma^\rho_{\mu\sigma} + \Gamma^\rho_{\mu\lambda}\Gamma^\rho_{\nu\sigma} - \Gamma^\rho_{\nu\lambda}\Gamma^\lambda_{\nu\mu}$. The last two terms involve $\Gamma\Gamma$, which will lead to $\partial h \cdot \partial h$ so they can be neglected. Therefore

$$\begin{aligned} R^\rho{}_{\sigma\mu\nu} &= \partial_\mu\left[\frac{1}{2}\eta^{\rho\lambda}\left(\partial_\nu h_{\lambda\sigma} + \partial_\sigma h_{\lambda\nu} - \partial_\lambda h_{\nu\sigma}\right)\right] - \partial_\nu\left[\frac{1}{2}\eta^{\rho\lambda}\left(\partial_\mu h_{\lambda\sigma} + \partial_\sigma h_{\lambda\mu} - \partial_\lambda h_{\mu\sigma}\right)\right] \\ &= \frac{1}{2}\left(\partial_\mu\partial_\sigma h^\rho{}_\nu + \eta^{\rho\lambda}\partial_\nu\partial_\lambda h_{\mu\sigma} - \partial_\nu\partial_\sigma h^\rho{}_\mu - \eta^{\rho\lambda}\partial_\mu\partial_\lambda h_{\nu\sigma}\right). \end{aligned}$$

Taking the trace, we get the Ricci tensor,

$$R^\rho{}_{\sigma\rho\nu} = R_{\sigma\nu} = \frac{1}{2}\left(\partial_\rho\partial_\sigma h^\rho{}_\nu + \eta^{\rho\lambda}\partial_\nu\partial_\lambda h_{\rho\sigma} - \partial_\nu\partial_\sigma h^\rho{}_\rho - \eta^{\rho\lambda}\partial_\rho\partial_\lambda h_{\nu\sigma}\right).$$

It would be nicer if we renamed the dummy index $\sigma \to \mu$. We also define the *d'Alembertian*,

$$\Box = \eta^{\rho\lambda}\partial_\rho\partial_\lambda = -\partial_t^2 + \partial_x^2 + \partial_y^2 + \partial_z^2,$$

so that

$$R_{\mu\nu} = \frac{1}{2}\left(\partial_\mu\partial_\rho h^\rho{}_\nu + \partial_\nu\partial_\lambda h^\lambda{}_\mu - \partial_\mu\partial_\nu h^\rho{}_\rho - \Box h_{\mu\nu}\right). \tag{7.2}$$

The Ricci scalar is then

$$R = \frac{1}{2}\left(2\partial_\lambda\partial_\rho h^{\lambda\rho} - 2\Box h^\lambda{}_\lambda\right). \tag{7.3}$$

With these, we can now get the Einstein tensor

$$\begin{aligned}\mathcal{G}_{\mu\nu} &= R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} \\ &= \frac{1}{2}\left[-\Box h_{\mu\nu} - \partial_\mu\partial_\nu h^\lambda{}_\lambda + \partial_\mu\partial_\rho h^\rho{}_\nu + \partial_\nu\partial_\rho h^\rho{}_\mu - \eta_{\mu\nu}\partial_\lambda\partial_\rho h^{\lambda\rho} + \eta_{\mu\nu}\Box h^\lambda{}_\lambda\right].\end{aligned}$$

This still looks pretty complicated. We can simplify the Einstein tensor a bit by considering the trace-reversed form of the perturbation,

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h^\lambda{}_\lambda.$$

Note that by taking the trace, we see $\bar{h}^\lambda_\lambda = -h^\lambda{}_\lambda$. Therefore we have

$$\begin{aligned}h_{\mu\nu} &= \bar{h}_{\mu\nu} + \frac{1}{2}\eta_{\mu\nu}h^\lambda{}_\lambda \\ &= \bar{h}_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h^\lambda{}_\lambda.\end{aligned} \tag{7.4}$$

Substituting into the Einstein tensor, several terms cancel and we get

$$\mathcal{G}_{\mu\nu} = \frac{1}{2}\left[-\Box\bar{h}_{\mu\nu} + \partial_\mu\partial_\rho\bar{h}^\rho_\nu + \partial_\nu\partial_\rho\bar{h}^\rho_\mu - \eta_{\mu\nu}\partial_\lambda\partial_\rho\bar{h}^{\lambda\rho}\right]. \tag{7.5}$$

We can simplify this further by fixing a gauge. A *gauge transformation* is considering an infinitesimal coordinate transformation

$$x^\mu \to x^\mu + \xi^\mu, \tag{7.6}$$

where ξ^μ is small, and we also only keep this to linear order. Now, the metric also depends on coordinates, so

$$\eta_{\mu\nu} + h_{\mu\nu}(x) \to \eta_{\mu\nu} + h_{\mu\nu}(x+\xi) = \eta_{\mu\nu} + h_{\mu\nu}(x) + \xi^\lambda\partial_\lambda h_{\mu\nu},$$

where we have performed a Taylor expansion of $h_{\mu\nu}(x+\xi)$. Also the differentials transform according to

$$\mathrm{d}x^\mu \to \mathrm{d}\left(x^\mu + \xi^\mu\right) = \mathrm{d}x^\mu + \mathrm{d}\xi^\mu = \mathrm{d}x^\mu + \mathrm{d}x^\lambda \partial_\lambda \xi^\mu.$$

Therefore, keeping only to the lowest order,[1] the line element transforms according to

$$\begin{aligned}
\mathrm{d}s^2 \to & \left(\eta_{\mu\nu} + h_{\mu\nu}\xi^\lambda \partial_\lambda h_{\mu\nu}\right)\left(\mathrm{d}x^\mu + \mathrm{d}x^\lambda \partial_\lambda \xi^\mu\right)\left(\mathrm{d}x^\nu + \mathrm{d}x^\lambda \partial_\lambda \xi^\nu\right) \\
&= \left(\eta_{\mu\nu} + h_{\mu\nu}\right)\mathrm{d}x^\mu \mathrm{d}x^\nu + \eta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\lambda \partial_\lambda \xi^\nu + \eta_{\mu\nu}\mathrm{d}x^\lambda \partial_\lambda \xi^\mu \mathrm{d}x^\nu \\
&= \left(\eta_{\mu\nu} + h_{\mu\nu}\right)\mathrm{d}x^\mu \mathrm{d}x^\nu + \partial_\lambda \xi_\mu \mathrm{d}x^\mu \mathrm{d}x^\lambda + \partial_\lambda \xi_\nu \mathrm{d}x^\lambda \mathrm{d}x^\nu .
\end{aligned}$$

Renaming dummy indices, the transformation is

$$\begin{aligned}
\mathrm{d}s^2 \to & \left(\eta_{\mu\nu} + h_{\mu\nu}\right)\mathrm{d}x^\mu \mathrm{d}x^\nu + \partial_\nu \xi_\mu \mathrm{d}x^\mu \mathrm{d}x^\nu + \partial_\mu \xi_\nu \mathrm{d}x^\mu \mathrm{d}x^\nu \\
&= \underbrace{\left(\eta_{\mu\nu} + h_{\mu\nu} + \partial_\mu \xi_\nu + \partial_\nu \xi_\mu\right)}_{h'_{\mu\nu}}\mathrm{d}x^\mu \mathrm{d}x^\nu .
\end{aligned}$$

In other words, an infinitesimal transformation $x^\mu \to x^\mu + \xi^\mu$ is equivalent to changing the metric by

$$h_{\mu\nu} \to h'_{\mu\nu} = h_{\mu\nu} + \partial_\mu \xi_\nu + \partial_\nu \xi_\mu .$$

Replacing the metric in this way is equivalent to a coordinate transformation, and doesn't affect the spacetime physically.

We will take advantage of this to simplify Eq. (7.5). Suppose we do such a gauge transformation, and the new metric is $h'_{\mu\nu} = h_{\mu\nu} + \partial_\mu \xi_\nu + \partial_\nu \xi_\mu$. The trace-reversed form is

$$\begin{aligned}
\bar{h}'_{\mu\nu} &= h'_{\mu\nu} - \frac{1}{2} h'^\lambda{}_\lambda \eta_{\mu\nu} \\
&= h_{\mu\nu} + \partial_\mu \xi_\nu + \partial_\nu \xi_\mu - \frac{1}{2}\left(h^\lambda{}_\lambda + 2\partial_\lambda \xi^\lambda\right)\eta_{\mu\nu} \\
&= h_{\mu\nu} - \frac{1}{2} h^\lambda{}_\lambda \eta_{\mu\nu} + \partial_\mu \xi_\nu + \partial_\nu \xi_\mu - \eta_{\mu\nu}\partial_\lambda \xi^\lambda .
\end{aligned}$$

Raising one index, we have

$$\bar{h}'^\mu_\nu = \bar{h}^\mu_\nu + \partial^\mu \xi_\nu + \partial_\nu \xi^\mu - \delta^\mu_\nu \partial_\lambda \xi^\lambda .$$

[1] Any product of the form $h \cdot h$, $\xi \cdot \xi$, and $\xi \cdot h$ are negligible.

Taking the divergence,

$$\partial_\mu \bar{h}'^\mu_\nu = \partial_\mu \bar{h}^\mu_\nu + \Box \xi_\nu + \partial_\mu \partial_\nu \xi^\mu - \partial_\mu \partial_\lambda \xi^\lambda.$$

The last two terms cancel each other. Therefore we end up with

$$\partial_\mu \bar{h}'^\mu_\nu = \partial_\mu \bar{h}^\mu_\nu + \Box \xi_\nu. \tag{7.7}$$

Now, the transformation ξ^μ is completely arbitrary, and so we are free to choose it. Suppose we purposely choose ξ^μ to satisfy $\Box \xi_\nu = -\partial_\mu \bar{h}^\mu_\nu$, then we get

$$\partial_\mu \bar{h}'^\mu_\nu = 0.$$

In summary, we can always choose an infinitesimal coordinate transformation $x^\mu \to x^\mu + \xi^\mu$ to make $\partial_\mu h^\mu{}_\nu = 0$. This is called the Lorenz gauge.[2]

Under the Lorenz gauge, the last three terms in Eq. (7.5) are zero and the Einstein tensor simplifies to

$$\mathcal{G}_{\mu\nu} = -\frac{1}{2} \Box \bar{h}_{\mu\nu}. \tag{7.8}$$

Finally, with this Einstein tensor we can write down the *linearised Einstein equation*. In SI units,

$$\Box \bar{h}_{\mu\nu} = -\frac{16\pi G}{c^4} T_{\mu\nu}. \tag{7.9}$$

7.2 Newtonian limit

Let us consider the linearised Einstein equation for the isotropic perfect fluid.

$$\Box \bar{h}_{\mu\nu} = -\frac{16\pi G}{c^4} \left[\left(\rho + \frac{p}{c^2} \right) u_\mu u_\nu + p g_{\mu\nu} \right].$$

Consider a co-moving, pressureless fluid. So $u^\mu = (c, 0, 0, 0)$ and $p = 0$. Then we have

$$T_{00} = \rho c^2, \quad T_{ij} = 0.$$

[2] Named after Ludvig Lorenz, not to be confused with Hendrik Loren**t**z associated with the Lorentz transformations.

We also assume the spacetime is static at the Newtonian limit. So the time derivative vanishes and we have $\Box h_{\mu\nu} = \vec{\nabla}^2 h_{\mu\nu}$. The Einstein equations are now

$$\vec{\nabla}^2 \bar{h}_{00} = -\frac{16\pi G}{c^2}\rho, \quad \vec{\nabla}^2 \bar{h}_{ij} = 0.$$

The second equation is solved by $\bar{h}_{ij} = 0$. Comparing the first equation with the Newtonian Poisson equation $\vec{\nabla}^2 \Phi = 4\pi G\rho$, we conclude that

$$\bar{h}_{00} = -4\frac{\Phi}{c^2}. \tag{7.10}$$

Be careful to remember that this is the trace-reversed form. Not $h_{\mu\nu}$. To get the latter, we take the trace

$$\begin{aligned} \bar{h}^\lambda_\lambda &= \eta^{\mu\nu}\bar{h}_{\mu\nu} = -h_{00} + \delta^{ij} \underbrace{h_{ij}}_{=0} \\ \bar{h}^\lambda_\lambda &= 4\frac{\Phi}{c^2} = -h^\lambda{}_\lambda. \end{aligned}$$

Then, using the definition of the trace reversal,

$$h_{\mu\nu} = \bar{h}_{\mu\nu} - \frac{1}{2}\bar{h}^\lambda_\lambda \eta_{\mu\nu}.$$

The components are

$$h_{00} = \bar{h}_{00} - \frac{1}{2}\left(4\frac{\Phi}{c^2}\right)(-1) = -4\frac{\Phi}{c^2} + \frac{1}{2}\left(4\frac{\Phi}{c^2}\right) = -2\frac{\Phi}{c^2},$$

and

$$\begin{aligned} h_{ij} &= \bar{h}_{ij} - \frac{1}{2}\left(4\frac{\Phi}{c^2}\right)\delta_{ij} \\ &= -2\frac{\Phi}{c^2}. \end{aligned}$$

Therefore, the metric components are

$$g_{00} = \eta_{00} - \frac{2\Phi}{c^2}, \quad g_{ij} = \eta_{ij} - \frac{2\Phi}{c^2}.$$

Reconstructing the line element, we have

$$\mathrm{d}s^2 = -\left(1 + \frac{2\Phi}{c^2}\right)c^2\mathrm{d}t^2 + \left(1 - \frac{2\Phi}{c^2}\right)\left(\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2\right). \tag{7.11}$$

This recovers the Newtonian potential for a point mass

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

which solves the

Chapter A Constants and formulae

Speed of light. $c = 2.99\,792\,458 \times 10^8\ \mathrm{m\,s^{-1}}$,

Gravitational constant. $G = 6.6743 \times 10^{-11}\ \mathrm{kg^{-1}\,m^3\,s^{-2}}$.

Particle motion.

$$\text{Geodesic equation:} \quad u^\mu \nabla_\mu u^\nu = 0 \quad \leftrightarrow \quad \ddot{x}^\kappa + \Gamma^\kappa_{\mu\nu}\dot{x}^\mu \dot{x}^\nu = 0.$$

$$\text{Lagrangian:} \quad \mathcal{L} = \frac{1}{2} k g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu, \quad k = \begin{cases} m, & \text{time-like,} \\ \alpha, & \text{null.} \end{cases}$$

$$\text{Canonical momenta and energy:} \quad p_\mu = \frac{\partial \mathcal{L}}{\partial \dot{x}^\mu}, \quad E = -V^\mu p_\mu.$$

Perfect fluid stress tensor. $T^{\mu\nu} = \left(\rho + \frac{p}{c^2}\right) u^\mu u^\nu + p g^{\mu\nu}$.

Connections and curvature tensors.

$$\text{Christoffel symbol, } \Gamma^\kappa_{\mu\nu} = \frac{1}{2} g^{\kappa\lambda}\left(\partial_\mu g_{\lambda\nu} + \partial_\nu g_{\lambda\mu} - \partial_\lambda g_{\mu\nu}\right),$$

$$\text{Riemann tensor, } R^\rho{}_{\sigma\mu\nu} = \partial_\mu \Gamma^\rho_{\nu\sigma} - \partial_\nu \Gamma^\rho_{\mu\sigma} + \Gamma^\rho_{\mu\lambda}\Gamma^\lambda_{\nu\sigma} - \Gamma^\rho_{\nu\lambda}\Gamma^\lambda_{\mu\sigma},$$

$$\text{Ricci tensor and scalar, } R_{\sigma\nu} = R^\mu{}_{\sigma\mu\nu}, \quad R = g^{\mu\nu} R_{\mu\nu} = R^\mu{}_\mu.$$

Trigonometric and hyperbolic identities.

- $\cos^2\theta + \sin^2\theta = 1, \qquad \sin 2\theta = 2\sin\theta\cos\theta$
- $\cos 2\theta = \cos^2\theta - \sin^2\theta = 2\cos^2\theta - 1 = 1 - 2\sin^2\theta$. Conversely, $\cos^2\theta = \frac{1}{2} + \frac{1}{2}\cos 2\theta$, and $\sin^2\theta = \frac{1}{2} - \frac{1}{2}\cos 2\theta$.
- $\int_0^{2\pi} \sin^2\theta\,\mathrm{d}\theta = \int_0^{2\pi} \cos^2\theta\,\mathrm{d}\theta = \pi$.
- $\sin(A \pm B) = \sin A\cos B \pm \cos A \sin B, \qquad \cos(A \pm B) = \cos A\cos B \mp \sin A\sin B$.
- $\cosh\alpha = \frac{1}{2}\left(\mathrm{e}^\alpha + \mathrm{e}^{-\alpha}\right), \qquad \sinh\alpha = \frac{1}{2}\left(\mathrm{e}^\alpha - \mathrm{e}^{-\alpha}\right)$.
- $\cosh^2\alpha - \sinh^2\alpha = 1$.

Lightspeed units.

Quantity	SI dim.	Lightpeed. dim.	Conversion
Length	L	L	$x = \tilde{x}$
Time	T	L	$t = c\tilde{t}$
Velocity	LT^{-1}	1	$v = \frac{\tilde{v}}{c}$
Acceleration	LT^{-2}	L^{-1}	$a = \frac{\tilde{a}}{c^2}$
Mass	M	M	$m = \tilde{m}$
Mass density	ML^{-3}	ML^{-3}	$\rho = \tilde{\rho}$
Momentum	MLT^{-1}	M	$p_\mu = \frac{1}{c}\tilde{p}_\mu$
Energy	ML^2T^{-2}	M	$E = \frac{1}{c^2}\tilde{E}$
Pressure	$ML^{-1}T^{-2}$	L^{-2}	$p = \frac{1}{c^2}\tilde{p}$
Grav. constant	$M^{-1}L^3T^{-2}$	$M^{-1}L$	$G = \frac{1}{c^2}\tilde{G}$

Geometric units.

Quantity	SI dim.	Geom. dim.	Conversion
Length	L	L	$x = \tilde{x}$
Time	T	L	$t = c\tilde{t}$
Velocity	LT^{-1}	1	$v = \frac{\tilde{v}}{c}$
Acceleration	LT^{-2}	L^{-1}	$a = \frac{\tilde{a}}{c^2}$
Mass	M	L	$m = \frac{G\tilde{m}}{c^2}$
Mass density	ML^{-3}	L^{-2}	$\rho = \frac{G}{c^2}\tilde{\rho}$
Momentum	MLT^{-1}	L^{-1}	$p_\mu = \frac{G}{c^4}\tilde{p}_\mu$
Energy	ML^2T^{-2}	L	$E = \frac{G}{c^4}\tilde{E}$
Pressure	$ML^{-1}T^{-2}$	L^{-2}	$p = \frac{G}{c^4}\tilde{p}$

Bibliography

[1] I. Newton, *Philosophiæ Naturalis Principia Mathematica.* England, (1687).

[2] M. P. Hobson and G. Efstathiou and A. N. Lasenby, *General Relativity: An introduction for physicists.* Cambridge University Press, Cambridge, (2006).

[3] B. Schutz, *A First Course in General Relativity.* Cambridge University Press, Cambridge, (2009).

[4] J. B. Hartle, *Gravity: An Introduction to Einstein's General Relativity.* Addison–Wesley, San Francisco, (2003).

[5] J. A. Foster and J. D. Nightingale, *A short course in general relativity 3ed.* Springer, , (2005).

[6] S. Carroll, *Spacetime and Geometry: An Introduction to General Relativity.* Addison Wesley, San Fransisco, (2004).

[7] R. Wald, *General Relativity.* The University of Chicago Press, Chicago, (1984).

[8] S. Chandrasekhar, *The mathematical theory of black holes.* The international series of monographs on physics. Oxford University Press, Chicago, (1983).

[9] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation.* W. H. Freeman and Company, San Francisco, (1973).

[10] J. B. Griffiths and J. Podolsý, *Exact Space-times in Einstein's General Relativity.* Cambridge University Press, Cambridge, (2009).

[11] E. Poisson, *A Relativist's Toolkit: The mathematics of black-hole mechanics.* Cambridge University Press, Cambridge, (2004).

[12] R. McNees, "Useful formulae." `http://jacobi.luc.edu/Useful.html`, 2019. [Online; accessed 24 March 2023].

[13] M. Blau, "Lecture Notes on General Relativity." `http://www.blau.itp.unibe.ch/GRLecturenotes.html`, 2022. [Online; accessed 24 March 2023].

[14] D. Baumann, "General Relativity." `http://cosmology.amsterdam/wp-content/uploads/2021/10/GR-Oct19.pdf`, 2021. [Online; accessed 31 Aug 2024].

[15] T. Hartmann, "Lectures on Quantum Gravity and Black Holes." http://www.hartmanhep.net/topics2015/gravity-lectures.pdf, 2022. [Online; accessed 24 March 2023].

[16] D. Tong, "Lectures on General Relativity." http://www.damtp.cam.ac.uk/user/tong/gr.html, 2019. [Online; accessed 31 Aug 2024].

[17] Bioware, "Normandy (Mass Effect)." https://shorturl.at/IfIDN, 2023. [Online; accessed 13 Aug 2024].

[18] Cadman97, "Mass Effect - Collector Ship ." https://shorturl.at/jkr7K, 2023. [Online; accessed 13 Aug 2024].

[19] R. V. Pound and G. A. Rebka, Jr., *Apparent Weight of Photons*, Phys. Rev. Lett. **4** (1960) 337–341.

[20] R. V. Pound and J. L. Snider, *Effect of Gravity on Nuclear Resonance*, Phys. Rev. Lett. **13** (1964) 539–540.

[21] K. Schwarzschild, *On the gravitational field of a mass point according to Einstein's theory*, Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.) **1916** (1916) 189–196, [physics/9905030].

[22] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series, and Products.* Elsevier Science, Amsterdam, (2002).

[23] Y.-K. Lim and Z. C. Yeo, *Energies and angular momenta of periodic Schwarzschild geodesics*, Phys. Rev. D **109** (2024), no. 2 024037, [arXiv:2401.13894].

[24] G. W. Gibbons and M. Vyska, *The Application of Weierstrass elliptic functions to Schwarzschild Null Geodesics*, Class. Quant. Grav. **29** (2012) 065016, [arXiv:1110.6508].

[25] G. W. Gibbons and C. M. Warnick, *Aspherical Photon and Anti-Photon Surfaces*, Phys. Lett. B **763** (2016) 169–173, [arXiv:1609.01673].

[26] Y.-K. Lim, *Null geodesics in the C metric with a cosmological constant*, Phys. Rev. D **103** (2021), no. 2 024007, [arXiv:2011.05503].

[27] **Event Horizon Telescope** Collaboration, K. Akiyama et al., *First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole*, Astrophys. J. Lett. **875** (2019) L1, [arXiv:1906.11238].

[28] H. Kragh, *Conceptions of Cosmos, From Myths to the Accelerating Universe: A History of Cosmology.* Oxford University Press, Oxford, (2007).

[29] G. Hinshaw, D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, J. Dunkley, M. R. Nolta, M. Halpern, R. S. Hill, N. Odegard, L. Page, K. M. Smith, J. L. Weiland, B. Gold, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, E. Wollack, and E. L. Wright, *Nine-year wilkinson microwave anisotropy probe (wmap) observations: Cosmological parameter results*, The Astrophysical Journal Supplement Series **208** (Sept., 2013) 19.

[30] **Supernova Cosmology Project** Collaboration, S. Perlmutter et al., *Measurements of* Ω *and* Λ *from 42 High Redshift Supernovae*, Astrophys. J. **517** (1999) 565–586, [astro-ph/9812133].

[31] **Supernova Search Team** Collaboration, P. M. Garnavich et al., *Constraints on cosmological models from Hubble Space Telescope observations of high z supernovae*, Astrophys. J. Lett. **493** (1998) L53–57, [astro-ph/9710123].

[32] **Supernova Cosmology Project** Collaboration, N. Suzuki et al., *The Hubble Space Telescope Cluster Supernova Survey: V. Improving the Dark Energy Constraints Above* $z > 1$ *and Building an Early-Type-Hosted Supernova Sample*, Ap. J **746** (2011) 85, [arXiv:1105.3470].