Part II Relativity

William Royce

June 14, 2024

Part II Physics, The University of Cambridge

Abstract

> The views of space and time which I wish to lay before you have sprung from the soil of experimental physics, and therein lies their strength. They are radical. Henceforth space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality.

General relativity is the theory of space and time and gravity. The essence of the theory is simple: gravity is geometry. The effects that we attribute to the force of gravity are due to the bending and warping of spacetime, from falling cats, to orbiting spinning planets, to the motion of the cosmos on the grandest scale. The purpose of these lectures is to explain this.

Before we jump into a description of curved spacetime, we should first explain why Newton's theory of gravity, a theory which served us well for 250 years, needs replacing. The problems arise when we think about disturbances in the gravitational field. Suppose, for example, that the Sun was to explode. What would we see? Well, for 8 glorious minutes – the time that it takes light to reach us from the Sun – we would continue to bathe in the Sun's light, completely oblivious to the fate that awaits us. But what about the motion of the Earth? If the Sun's mass distribution changed dramatically, one might think that the Earth would start to deviate from its elliptic orbit. But when does this happen? Does it occur immediately, or does the Earth continue in its orbit for 8 minutes before it notices the change?

Of course, the theory of special relativity tells us the answer. Since no signal can propagate faster than the speed of light, the Earth must continue on its orbit for 8 minutes. But how is the information that the Sun has exploded then transmitted? Does the information also travel at the speed of light? What is the medium that carries this information? As we will see throughout these lectures, the answers to these questions forces us to revisit some of our most basic notions about the meaning of space and time and opens the to door to some of the greatest ideas in modern physics such as cosmology and black holes.

Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

1.1 Newtonian Gravity

There is a well trodden path in physics when trying to understand how objects can influence other objects far away. We introduce the concept of a field. This is a physical quantity which exists everywhere in space and time; the most familiar examples are the electric and magnetic fields. When a charge moves, it creates a disturbance in the electromagnetic field, ripples of which propagate through space until they reach other charges. To develop a causal theory of gravity, we must introduce a gravitational field that responds to mass in some way.

It's a simple matter to cast Newtonian gravity in terms of a field theory. A particle of mass m_G experiences a force that can be written as

$$\mathbf{F} = -m_G \boldsymbol{\nabla}\Phi. \tag{1.1}$$

The quantity m_G is the *passive gravitational mass*, and it determines the gravitational force on the particle.The gravitational field $\Phi(\mathbf{r}, t)$ is determined by the surrounding matter distribution which is described by the mass density $\rho(\mathbf{r}, t)$. If the matter density is static, so that $\rho(\mathbf{r})$ is independent of time, then the gravitational field obeys

$$\nabla^2\Phi = 4\pi G\rho, \tag{1.2}$$

with Newton's constant G given by

$$G \approx 6.67 \times 10^{-11}\mathrm{m}^3\mathrm{kg}^{-1}\mathrm{s}^{-2}. \tag{1.3}$$

This equation is simply a rewriting of the usual inverse square law of Newton. For example, if a mass M is concentrated at a single point we have

$$\rho(\mathbf{r}) = M\delta^{(3)}(\mathbf{r}) \implies \Phi = -\frac{GM}{r}, \tag{1.4}$$

which is the familiar gravitational field for a point mass.

The question that we would like to answer is: how should we modify (1.2) when the mass distribution $\rho(\mathbf{r})$ changes with time? Of course, we could simply postulate that (1.2) continues to hold even in this case. A change in ρ would then immediately result in a change of Φ throughout all of space. Such a theory clearly is not consistent with the requirement that no signal can travel faster than light. Our goal is to figure out how to generalise (1.2) in a manner that is compatible with the postulates of special relativity. The end result of this goal will be a theory of gravity that is compatible with special relativity: this is the general theory of relativity.

Fixing this incompatibility will ultimately require a radical modification of how we think about gravity and, indeed, spacetime itself. Sticking with Newtonian gravity for the moment, it is not immediately obvious that the mass density appearing in Poisson's equation should refer to the density of the passive gravitational mass. Rather, let us also introduce the active gravitational mass m_A, so that the relevant mass density for a point particle at position $\mathbf{r}'(t)$ at time t is

$$\rho(\mathbf{r},t) = m_A \delta^{(3)}(\mathbf{r} - \mathbf{r}'(t)). \tag{1.5}$$

For the point particle, the relevant solution of Poisson's equation is

$$\Phi(\mathbf{r},t) = -\frac{G m_A}{|\mathbf{r} - \mathbf{r}'(t)|}. \tag{1.6}$$

It follows that the force on a test particle of passive gravitational mass $m_{G,1}$ at position $\mathbf{r}_1$ at time t due to a particle of active gravitational mass $m_{A,2}$ at position $\mathbf{r}_2$ *at the same time* t is

$$\mathbf{F}_{2\text{ on }1} = -G m_{G,1} m_{A,2} \frac{\mathbf{r}_1 - \mathbf{r}_2}{|\mathbf{r}_1 - \mathbf{r}_2{}^3|}. \tag{1.7}$$

Similarly, the force on the second particle due to the first is

$$\mathbf{F}_{1\text{ on }2} = -G m_{G,2} m_{A,1} \frac{\mathbf{r}_2 - \mathbf{r}_1}{|\mathbf{r}_1 - \mathbf{r}_2{}^3|}. \tag{1.8}$$

If momentum is to be conserved, i.e., $\mathbf{F}_{2\text{ on }1} = \mathbf{F}_{1\text{ on }2}$, we must have

$$m_{G,1} m_{A,2} = m_{G,2} m_{A,1}. \tag{1.9}$$

Since this must hold for arbitrary masses, we must have that the ratio of passive to active gravitational mass is the same for all particles. Thus, we can take these masses to be equal, $m_G = m_A$, for all matter (absorbing their universal ratio in the gravitational constant).

This sort of universality is not unusual in physics – a similar thing happens in electromagnetism, for example, where the passive and active electric charges are equal. However, there is a further equality of masses in Newtonian gravity that is rather more surprising: the equality of gravitational and inertial masses. A particle acted on by a force $\mathbf{F}$ experiences an acceleration such that

$$\mathbf{F} = m_I \frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}, \tag{1.10}$$

where m_I is the *inertial mass*. For the gravitational force, the acceleration is

$$\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} = -\frac{m_G}{m_I} \boldsymbol{\nabla}\Phi. \tag{1.11}$$

It is an experimental fact[1] (known since Galileo's time) that the ratio m_G/m_I is the same for all particles, so we can always take $m_G = m_I$ (further absorbing their universal ratio in the gravitational constant). This means that if two particles of different composition fall freely in a gravitational field, they have the same acceleration. This is often rephrased as the *weak equivalence principle*:

> Freely-falling particles with negligible gravitational self-interaction follow the same path through space and time if they have the same initial position and velocity, independent of their composition.

This property of gravity is in striking contrast to other forces; for example, in electromagnetism the acceleration of a point particle in a given electric field depends on the ratio of the electric charge to inertial mass, which is definitely not universal.

[1] The equality of gravitational and inertial masses is now verified to the level of one part in 10^{13}.

1.2 Implications of the Equivalence Principle

Consider an observer in a free-falling, non-rotating elevator in a uniform gravitational field. Relative to this observer, free-falling particles move on straight lines at constant velocity – the effects of the uniform gravitational field have been removed and the observer perceives that the usual laws of special relativistic kinematics hold. This idea motivates an extension of the weak equivalence principle to what is known as the *strong equivalence principle*:

> In an arbitrary gravitational field, *all* the laws of physics in a free-falling, non-rotating laboratory occupying a sufficiently small region of spacetime look locally like special relativity (with no gravity).

Note how the strong equivalence principle is supposed to apply to all laws of physics, not just the dynamics of free-falling particles. Why the qualification of observations over a sufficiently small region of spacetime?

Consider the same elevator falling freely in the non-uniform gravitational field of the earth. Free particles initially at rest in the elevator will move together over time as they follow radial trajectories towards the centre of the earth. It is these tidal effects that are the physical manifestation of the gravitational field, and that cannot be removed by passing to the free-falling frame. However, for sufficiently local measurements in space and time, these tidal effects are undetectable, and physics relative to the free-falling elevator looks just like special relativistic physics in an inertial frame of reference in the absence of gravity.

The strong equivalence principle implies the local equivalence of a gravitational field and acceleration. In particular, it implies that a constant gravitational field is unobservable – observations in a reference frame at rest in such a field would be indistinguishable from those in a uniformly-accelerating reference frame in the absence of gravity. In special relativity, physics looks simple when referred to an inertial frame, one defined by comoving, unaccelerated observers with synchronised clocks. However, with gravity, the equivalence principle tells us that physics looks equally simple *locally* in a free-falling reference frame, suggesting that we should *define* inertial reference frames locally by free-falling observers. Acceleration should be defined relative to such local inertial frames, so that a particle acted on by no other force (and so free-falling) should be regarded as unaccelerated.

1.2.1 Gravity as Spacetime Curvature

The universality of free fall suggested to Einstein that the trajectories of free-falling particles should be determined by the local structure of spacetime, rather than by the action of a gravitational force with a mysterious universal coupling to matter.

Local inertial reference frames correspond to local systems of coordinates over spacetime so that the geometry over a small region looks like that of the spacetime of special relativity. Gravity manifests itself through our inability to extend such coordinates globally, reflecting the *curvature of spacetime*.

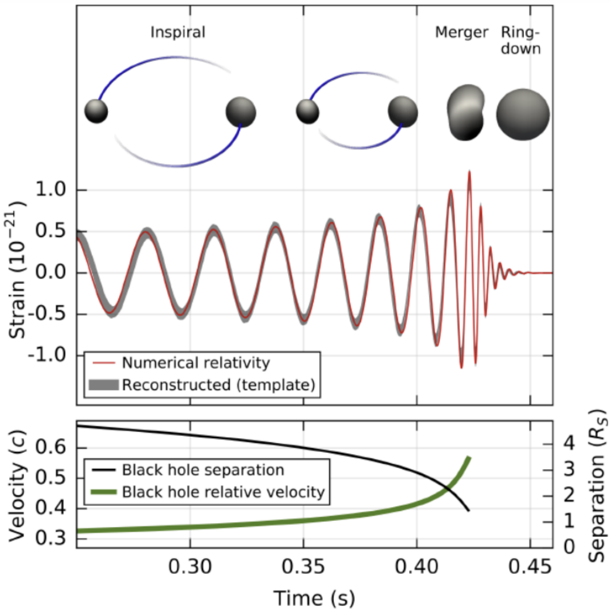

Fig. 1.1: Top: Estimated gravitational wave strain amplitude inferred from the LIGO data for their discovery event. The signal is generated from the inspiral, merger and ring-down of two massive black holes. The properties of the source can be estimated by comparing the measured waveform with detailed calculations in general relativity. Bottom: the relative speed and separation (in units of the Schwarzschild radius, $R_s = 2GM/c^2$) of the blackholes during the event. For reference, the Newtonian potential at R_S away from a mass M is $|\Phi|/c^2 = 1/2$. Figure taken from Abbot et al., Phys. Rev. Lett. 116, 061102 (2016).

General relativity abandons the idea of gravity as a force defined on the fixed spacetime of special relativity, replacing it with a geometric theory in which the geometry of spacetime determines the trajectories of free-falling particles, the geometry itself being curved by the presence of matter.

1.3 Further Motivation: Extreme Gravity

Newtonian gravity is recovered from general relativity in the limit of low relative speeds of particles, $v \ll c$, and weak gravitational fields, typically $|\Phi| \ll c^2$. Note that in situations where speeds are determined by gravity, these two regimes are generally equivalent.

To see this, consider a particle in a circular orbit of radius R around a mass M in Newtonian gravity: the speed is determined by

$$\frac{v^2}{R} = \frac{GM}{R^2}, \tag{1.12}$$

and so

$$\frac{v^2}{c^2} = \frac{GM}{Rc^2} = \frac{|\Phi|}{c^2}. \tag{1.13}$$

However, increasingly we are observing phenomena where Newtonian gravity is a very poor approximation. A striking example is the recent first detection of gravitational waves by the LIGO interferometer; see Fig. 1.1.

Gravitational waves are wavelike disturbances in the geometry of spacetime, which can be detected by looking for their characteristic quadrupole distortion (i.e., a shortening in one direction and stretching in an orthogonal direction) of the two arms of a laser interferometer. Gravitational waves propagate at the speed of light and are a natural prediction of general relativity; they do not arise in Newtonian gravity where the potential responds instantly to distant rearrangements of mass.

The first LIGO signal was generated by a truly extreme astrophysical source: two merging black holes each with a mass around 30 times that of the Sun at a distance from us of around 2 Gly.As the blackholes orbited their common centre of mass, the system radiated gravitational waves causing the blackholes to spiral inwards and increase their speed until they merged to form a single black hole. Such sources probe the strong-field regime of general relativity during the merger phase and involve highly relativistic speeds (see Fig. 1.1). At its peak, the source was losing energy to gravitational waves at a rate of 3.6×10^{49}W, which is equivalent to 200 times the rest mass energy of the Sun per second!

Chapter 2

Recap of Special Relativity

Although Newtonian mechanics gives an excellent description of Nature, it is not universally valid. When we reach extreme conditions — the very small, the very heavy or the very fast — the Newtonian Universe that we're used to needs replacing. You could say that Newtonian mechanics encapsulates our common sense view of the world. One of the major themes of twentieth century physics is that when you look away from our everyday world, common sense is not much use.

One such extreme is when particles travel very fast. The theory that replaces Newtonian mechanics is due to Einstein. It is called *special relativity.* The effects of special relativity become apparent only when the speeds of particles become comparable to the speed of light in the vacuum. The speed of light is

$$c = 299792458 \mathrm{m\,s^{-1}} \tag{2.1}$$

This value of c is exact. It may seem strange that the speed of light is an integer when measured in meters per second. The reason is simply that this is taken to be the definition of what we mean by a meter: it is the distance travelled by light in 1/299792458 seconds. For the purposes of this course, we'll be quite happy with the approximation $c \approx 3 \times 10^8 \mathrm{m\,s^{-1}}$.

The first thing to say is that the speed of light is fast. Really fast. The speed of sound is around $300\mathrm{m\,s^{-1}}$; escape velocity from the Earth is around $104\mathrm{m\,s^{-1}}$; the orbital speed of our solar system in the Milky Way galaxy is around $105\mathrm{m\,s^{-1}}$. As we shall soon see, nothing travels faster than c.

The theory of special relativity rests on two experimental facts. (We will look at the evidence for these shortly). In fact, the first of these is simply the Galilean principle of relativity as in classical Newtonian mechanics. The second postulate is more surprising:

- The principle of relativity: the laws of physics are the same in all inertial frames.
- The speed of light in vacuum is the same in all inertial frames.

On the face of it, the second postulate looks nonsensical. How can the speed of light look the same in all inertial frames? If light travels towards me at speed c and I run away from the light at speed v, surely I measure the speed of light as $c - v$. Right? Well, no.

2.1 Newtonian Geometry of Space and Time

Newtonian theory assumes an absolute time – the same for every observer. This common sense view is encapsulated in the Galilean transformations.. Mathematically, we derive this "obvious" result as follows: two inertial frames, S and S', in *standard configuration*: axes aligned, the same spacetime origin, which move relative to each with velocity $\mathbf{v} = (v, 0, 0)$, have Cartesian coordinates related by

$$x' = x - vt, \qquad y' = y, \qquad z' = z, \qquad t' = t \tag{2.2}$$

If a ray of light travels in the x direction in frame S with speed c, then it traces out the trajectory $x/t = c$. The transformations above then tell us that in frame S0 the trajectory if the light ray is $x'/t' = c - nv$. This is the result we claimed above: the speed of light should clearly be $c - v$. If this is wrong (and it is) something must be wrong with the Galilean transformations (2.2). But what?

Our immediate goal is to find a transformation law that obeys both postulates above. As we will see, the only way to achieve this goal is to allow for a radical departure in our understanding of time. In particular, we will be forced to abandon the assumption of absolute time, enshrined in the equation $t' = t$ above. We will see that time ticks at different rates for observers sitting in different inertial frames.

For two events A and B, the Gallilieean transformation implies that

- the time difference $\Delta t = t_B - t_A$ is invariant; and
- $\Delta r^2 = \Delta x^2 + \Delta y^2 + \Delta z^2$ is invariant for simultaneous events (since Δx, Δy, and Δz are).

Space and time are separate entities in Newtonian theory.

2.2 Lorentz Transformations

We stick with the idea of two inertial frames, S and S', moving with relative speed v. For simplicity, we'll start by ignoring the directions y and z which are perpendicular to the direction of motion. Both inertial frames come with Cartesian coordinates: (x, t) for S and (x', t') for S' . We want to know how these are related. The most general possible relationship takes the form

$$x' = f(x, t), \qquad t' = g(x, t), \tag{2.3}$$

for some function f and g. However, there are a couple of facts that we can use to immediately restrict the form of these functions. The first is that the law of inertia holds; left alone in an inertial frame, a particle will travel at constant velocity. Drawn in the (x, t) plane, the trajectory of such a particle is a straight line. Since both S and S' are inertial frames, the map $(x, t) \mapsto (x', t'))$ must map straight lines to straight lines; such maps are, by definition, linear. The functions f and g must therefore be of the form

$$x' = \alpha_1 x + \alpha_2 t, \qquad t' = \alpha_3 x + \alpha_f t, \tag{2.4}$$

where $\alpha_i = 1, 2, 3, 4$ can each be a function of v.

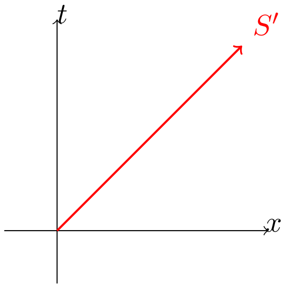

Fig. 2.1: Space-time diagram, representing the motion of a particle at the origin $x' = 0$ in S', which moves along the trajectory $x = vt$ in S.

Secondly, we use the fact that S' is travelling at speed v relative to S. This means that an observer sitting at the origin, $x' = 0$, of S' moves along the trajectory $x = vt$ in S shown in Fig. (2.1). Or, in other words, the points $x = vt$ must map to $x' = 0$. (There is actually one further assumption implicit in this statement: that the origin $x' = 0$ coincides with $x = 0$ when $t = 0$). Together with the requirement that the transformation is linear, this restricts the coefficients α_1 and α_2 above to be of the form,

$$x' = \gamma(x - vt), \tag{2.5}$$

for some coefficient γ. Once again, the overall coefficient can be a function of the velocity: $\gamma = \gamma_v$. (We've used subscript notation v rather than the more standard (v) to denote that depends on v. This avoids confusion with the factors of $(x - vt)$ which aren't arguments of but will frequently appear after like in the equation (2.5)).

There is actually a small, but important, restriction on the form of γ_v: it must be an even function, so that $\gamma_v = \gamma_{-v}$. There are a couple of ways to see this. The first is by using rotational invariance, which states that can depend only on the direction of the relative velocity $\mathbf{v}$, but only on the magnitude $v^2 = \mathbf{v} \cdot \mathbf{v}$. Alternatively, if this is a little slick, we can reach the same conclusion by considering inertial frames $\tilde{S}$ and $\tilde{S}'$ which are identical to S and S' except that we measure the x-coordinate in the opposite direction, meaning $\tilde{x} = -x$ and $\tilde{x}' = -x'$. While S is moving with velocity $+v$ relative to S', $\tilde{S}$ is moving with velocity $-v$ with respect to $\tilde{S}'$ simply because we measure things in the opposite direction. That means that

$$\tilde{x}' = \gamma_{-v}(\tilde{x} + v\tilde{t}). \tag{2.6}$$

Comparing this to (2.5), we see that we must have $\gamma_v = \gamma_{-v}$ as claimed.

We can also look at things from the perspective of S', relative to which the frame S moves backwards with velocity v. The same argument that led us to (2.5) now tells us that

$$x = \gamma(x' + vt'). \tag{2.7}$$

Now the function $\gamma_v = \gamma_{-v}$. But by the argument above, we know that v = v. In other words, the coefficient appearing in (2.7) is the same as that appearing in (2.5).

At this point, things don't look too different from what we've seen before. Indeed, if we now insisted on absolute time, so $t = t'$, we're forced to have $\gamma = 1$ and we get back to the Galilean transformations (2.2). However, as we've seen, this is not compatible with the second postulate of special relativity. So let's push forward and insist instead that the speed of light is equal to c in both S and S'. In S, a light ray has trajectory

$$x = ct. \tag{2.8}$$

While, in S', we demand that the same light ray has trajectory

$$x' = ct'. \tag{2.9}$$

Substituting these trajectories into (2.5) and (2.7), we have two equations relating t and t',

$$ct' = \gamma(c - v)t, \qquad \text{and,} \qquad ct = \gamma(c + v)t'. \tag{2.10}$$

A little algebra shows that these two equations are compatible only if γ is given by

$$\boxed{\gamma = \sqrt{\frac{1}{1 - v^2/c^2}}.} \tag{2.11}$$

We'll be seeing a lot of this coefficient γ in what follows. Notice that for $v \ll c$, we have $\gamma \approx 1$ and the transformation law (2.5) is approximately the same as the Galilean transformation (2.2). However, as $v \to c$ we have $\gamma \to \infty$. Furthermore, becomes imaginary for $v > c$ which means that we're unable to make sense of inertial frames with relative speed $v > c$.

Equations (2.5) and (2.11) give us the transformation law for the spatial coordinate. But what about for time? In fact, the temporal transformation law is already lurking in our analysis above. Substituting the expression for x' in (2.5) into (2.7) and rearranging, we get

$$t' = \gamma\left(t - \frac{v}{c^2}x\right). \tag{2.12}$$

We shall soon see that this equation has dramatic consequences. For now, however, we merely note that when $v \ll c$, we recover the trivial Galilean transformation law $t' \approx t$. Equations (2.5) and (2.12) are the *Lorentz transformations.*

2.2.1 Lorentz Transformations in Three Spatial Dimensions

In the above derivation, we ignored the transformation of the coordinates y and z perpendicular to the relative motion. In fact, these transformations are trivial. Using the above arguments for linearity and the fact that the origins coincide at $t = 0$, the most general form of the transformation is

$$y' = \kappa y, \tag{2.13}$$

But, by symmetry, we must also have $y' = \kappa y$. Clearly, we require $\kappa = 1$. (The other possibility $\kappa = -1$ does not give the identity transformation when $v = 0$. Instead, it is a reflection).

With this we can write down the final form of the Lorentz transformations. Note that they look more symmetric between x and t if we write them using the combination ct,

$$\begin{aligned} x' &= \gamma\Big(x - \frac{v}{c}ct\Big), \\ y' &= y, \\ z' &= z, \\ ct' &= \gamma\Big(ct - \frac{v}{c}x\Big), \end{aligned} \tag{2.14}$$

where γ is given by (2.11). These are also known as Lorentz boosts. Notice that for $v/c \ll 1$, the Lorentz boosts reduce to the more intuitive Galilean boosts. (We sometimes say, rather sloppily, that the Lorentz transformations reduce to the Galilean transformations in the limit $c \to \infty$).

It's also worth stressing again the special properties of these transformations. To be compatible with the first postulate, the transformations must take the same form if we invert them to express x and t in terms of x' and t', except with v replaced by $-v$. And, after a little bit of algebraic magic, they do.

Secondly, we want the speed of light to be the same in all inertial frames. For light travelling in the x direction, we already imposed this in our derivation of the Lorentz transformations. But it's simple to check again: in frame S, the trajectory of an object travelling at the speed of light obeys $x = ct$. In S', the same object will follow the trajectory $x' = \gamma(x - vt) = \gamma(ct - vx/c) = ct'$

What about an object travelling in the y direction at the speed of light? Its trajectory in S is $y = ct$. From (2.14), its trajectory in S' is $y' = ct'/\gamma$ and $x' = vt'$. Its speed in S' is therefore $v'^2 = v_x^2 x + v_y^2$, or

$$v'^2 = \left(\frac{x'}{t'}\right)^2 + \left(\frac{x'}{t'}\right)^2 = v^2 + \frac{c^2}{\gamma^2} = c^2. \tag{2.15}$$

Note how time and space are mixed by the Lorentz transformation. However, for two events, the (squared) interval

$$\boxed{\Delta s^2 = c^2 \Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2.} \tag{2.16}$$

is *invariant* under any Lorentz transformation. In special relativity, space and time are united into a four-dimensional continuum called spacetime with invariant geometry characterised by Δs^2. The spacetime of special relativity is topologically $\mathbf{R}^4$. When endowed with the measure of distance (2.16), this spacetime is referred to as Minkowski space. Although topologically equivalent to Euclidean space, distances are measured differently. To stress the difference between the time and spatial directions, Minkowski space is sometimes said to have dimension $d = 1 + 3$. (For once, it's important that you don't do this sum!). In later courses – in particular General Relativity – you will see the invariant interval written as the distance between two infinitesimally close points. In practice that just means we replace all the Δ(something)s with d(something)s.

$$\mathrm{d}s^2 = c^2\,\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2\,. \tag{2.17}$$

In this infinitesimal form, $\mathrm{d}s^2$ is called the *line element*.

2.2.2 Lorentz Transformations as 4D 'Rotations'

Different Cartesian inertial frames S and S' simply relabel events in Minkowski spacetime, i.e., perform a coordinate transformation $(ct, x, y, z) \to (ct', x', y', z')$. It is often convenient to define the rapidity parameter ψ (which runs from $-\infty$ to ∞) by $v/c = \tanh\psi$, so that

$$\gamma = \cosh\psi, \qquad \text{and,} \qquad \gamma v/c = \tanh\psi \tag{2.18}$$

For S and S' in standard configuration, we have

$$\begin{aligned} ct' &= ct\cosh\psi - x\sinh\psi \\ x' &= -ct\sinh\psi + x\cosh\psi \\ y' &= y \\ z' &= z \end{aligned} \tag{2.19}$$

These are like a rotation in the $ct-x$ plane, but with hyperbolic rather than trigonometric functions. The hyperbolic functions are necessary to ensure the invariance of Δs^2 given the minus signs in its definition.

2.2.3 More Complicated Lorentz Transformations

More generally, the relation between two Cartesian inertial frames S and S' can differ from that for the standard configuration since[1]:

- the $4D$ origins may not coincide, i.e., the event at $ct = x = y = z = 0$ may not be at $ct' = x' = y' = z' = 0$;
- the relative velocity of the two frames may be in an arbitrary direction in S, rather than along the x-axis; and
- the spatial axes in S and S' may not be aligned, e.g., the components of the relative velocity in S' may not be minus those in S.

We can always deal with the origins not coinciding (known as inhomogeneous Lorentz transformations or Poincaré transformations) by appropriate temporal and spatial displacements. We can find the form of the remaining Lorentz transformation in the general case by decomposing as follows.

1. Apply a purely spatial rotation in the frame S to align the new x-axis with the relative velocity of the two frames.
2. Apply a standard Lorentz transformation as in Eq. (2.12) and (2.5).
3. Apply a spatial rotation in the transformed coordinates to align the axes with those of S'.

[1]Lorentz transformations can be considered more formally as linear transformations that preserve the interval Δs^2. In this case, the definition admits transformations that are not continuously connected to the identity; i.e., parity transformations and/or time reversal. We shall not consider such transformations further.

Given a reference frame S, the *Lorentz boost* of this frame for a general relative velocity $\mathbf{v}$ is obtained by rotating the spatial axes of S so that the relative velocity is along the new x-axis, applying the standard Lorentz transformation, and applying the inverse spatial rotation in the transformed frame.If the relative velocity is along the original x-axis, this reduces to the standard Lorentz transformation.

More generally, reference frames connected by a Lorentz boost have their spatial axes as aligned as possible given the relative velocity of the frames, i.e., they are generated by hyperbolic "rotations" in the plane defined by the ct-axis and the relative velocity.

2.2.4 The Interval

As we have seen, the interval is invariant under Lorentz transformations. This is particularly transparent using the hyperbolic form of the standard transformation:

$$\begin{aligned}\Delta s^2 &= c^2(\Delta t')^2 - (\Delta x')^2 - (\Delta y')^2 - (\Delta z')^2 \qquad (2.20)\\ &= [(c\Delta t)\cosh\psi - (\Delta x)\sinh\psi]^2 - [-(c\Delta t)\sinh\psi + (\Delta x)\cosh\psi]^2 - \Delta y^2 - \Delta z^2\\ &= c^2\Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2\end{aligned}$$

The interval is also invariant under more general Lorentz transformations since a shift in origin does not alter the differences $(c\Delta t, \Delta x, \Delta y, \Delta z)$, and rotations preserve the spatial interval $\Delta x^2 + \Delta y^2 + \Delta z^2$.

The invariant interval provides an observer-independent characterisation of the distance between any two events. However, it has a strange property: it is not positive definite. Two events whose separation is $\Delta s^2 > 0$ are said to be *timelike* separated. They are closer together in space than they are in time. Pictorially, such events sit within each others light cone.

In contrast, events with $\Delta s^2 < 0$ are said to be *spacelike* separated. They sit outside each others light cone. Two observers can disagree about the temporal ordering of spacelike separated events. However, they agree on the ordering of timelike separated events. Note that since $\Delta s^2 < 0$ for spacelike separated events, if you insist on talking about Δs itself then it must be purely imaginary. However, usually it will be perfectly fine if we just talk about Δs^2.

Finally, two events with $\Delta s^2 = 0$ are said to be *lightlike* separated. Notice that this is an important difference between the invariant interval and most measures of distance that you're used to. Usually, if two points are separated by zero distance, then they are the same point. This is not true in Minkowski spacetime: if two points are separated by zero distance, it means that they can be connected by a light ray.

2.2.5 Space-Time Diagrams

We'll find it very useful to introduce a simple spacetime diagram to illustrate the physics of relativity. In a fixed inertial frame, S, we draw one direction of space – say x – along

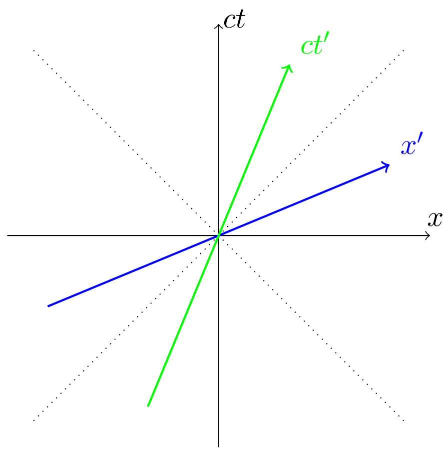

Fig. 2.2: Space-time diagram with axes corresponding to an inertial frame S' moving with a relative velocity. They can be the thought of as the x and ct axes, rotated by an equal amount towards the diagonal light ray. The fact the axes are symmetric about the light ray reflects the fact that the speed of light is equal to c in both frames.

the horizontal axis and time on the vertical axis. But things look much nicer if we rescale time and plot ct on the vertical instead. In the context of special relativity, space and time is called *Minkowski space*.

This is a spacetime diagram. Each point, P, represents an event. In the following, we'll label points on the spacetime diagram as coordinates (ct, x) i.e. giving the coordinate along the vertical axis first. This is backwards from the usual way coordinates but is chosen so that it is consistent with a later, standard, convention.

A particle moving in spacetime traces out a curve called a worldline as shown in the figure. Because we've rescaled the time axis, a light ray moving in the x direction moves at 45°. We'll later see that no object can move faster than the speed of light which means that the worldlines of particles must always move upwards at an angle steeper than 45°.

The horizontal and vertical axis in the spacetime diagram are the coordinates of the inertial frame S. But we could also draw the axes corresponding to an inertial frame S' moving with relative velocity $\mathbf{v} = (v, 0, 0)$. The t' axis sits at $x' = 0$ and is given by $x = vt$. Meanwhile, the x' axis is determined by $t' = 0$ which, from the Lorentz transformation (2.14), is given by the equation $ct = \frac{v}{c}x$.

These two axes are drawn on the Fig. (2.2). They can be the thought of as the x and ct axes, rotated by an equal amount towards the diagonal light ray. The fact the axes are symmetric about the light ray reflects the fact that the speed of light is equal to c in both frames.

2.2.6 Causality and the Lightcone

We start with a simple question: how can we be sure that things happen at the same time? In Newtonian physics, this is a simple question to answer. In that case, we have an absolute time t and two events, P_1 and P_2, happen at the same time if $t_1 = t_2$. However,

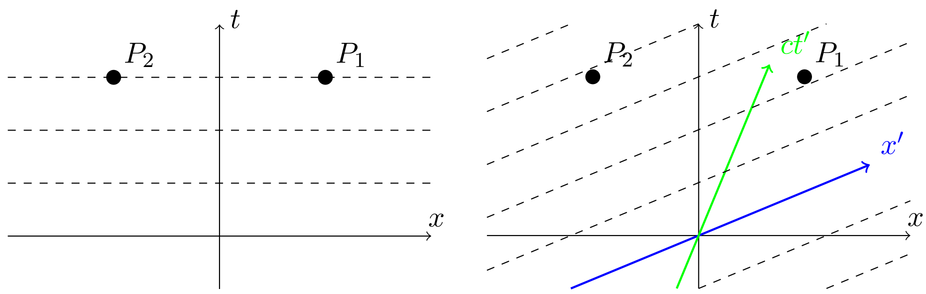

Fig. 2.3: Simultaneity is relative.

in the relativistic world, things are not so easy.

We start with an observer in inertial frame S, with time coordinate t. This observer sensibly decides that two events, P_1 and P_2, occur simultaneously if $t_1 = t_2$. In the spacetime diagram on the left of Fig. 2.3 we have drawn lines of simultaneity for this observer.

But for an observer in the inertial frame S', simultaneity of events occurs for equal t'. Using the Lorentz transformation, lines of constant t' become lines described by the equation $t - vx/c^2 =$ constant. These lines are drawn on the spacetime diagram on the right of Fig. 2.3.

The upshot of this is that two events simultaneous in one inertial frame are not simultaneous in another. An observer in S thinks that events P_1 and P_2 happen at the same time. All other observers disagree.

We've seen that different observers disagree on the temporal ordering of two events. But where does that leave the idea of causality? Surely it's important that we can say that one event definitely occurred before another. Thankfully, all is not lost: there are only some events which observers can disagree about.

To see this, note that because Lorentz boosts are only possible for $v < c$, the lines of simultaneity cannot be steeper than 45°. Take a point A and draw the 45° light rays that emerge from A. This is called the *light cone*. In more than a single spatial dimension, the light cone is really two cones, touching at the point A. They are known as the future light cone and past light cone.

For events inside the light cone of A, there is no difficulty deciding on the temporal ordering of events. All observers will agree that B occurred after A. However, for events outside the light cone, the matter is up for grabs: some observers will see D as happening after A; some before.

This tells us that the events which all observers agree can be causally influenced by A are those inside the future light cone. Similarly, the events which can plausibly influence A are those inside the past light cone. This means that we can sleep comfortably at night, happy in the knowledge that causality is preserved, only if nothing can propagate outside the light cone. But that's the same thing as travelling faster than the speed of light.

The converse to this is that if we do ever see particles that travel faster than the speed of light, we're in trouble. We could use them to transmit information faster than light. But another observer would view this as transmitting information backwards in time. All our ideas of cause and effect will be turned on their head. We will show later why it is impossible to accelerate particles past the light speed barrier.

There is a corollary to the statement that events outside the lightcone cannot influence each other: there are no perfectly rigid objects. Suppose that you push on one end of a rod. The other end cannot move immediately since that would allow us to communicate faster than the speed of light. Of course, for real rods, the other end does not move instantaneously. Instead, pushing on one end of the rod initiates a sound wave which propagates through the rod, telling the other parts to move. The statement that there is no rigid object is simply the statement that this sound wave must travel slower than the speed of light.

Finally, let me mention that when we're talking about waves, as opposed to point particles, there is a slight subtlety in exactly what must travel slower than light. There are at least two velocities associated to a wave: the group velocity is (usually) the speed at which information can be communicated. This is less than c. In contrast, the phase velocity is the speed at which the peaks of the wave travel. This can be greater than c, but transmits no information.

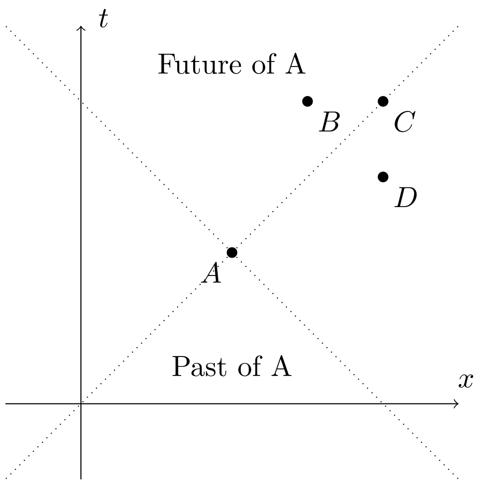

Fig. 2.4: Lightcone structure around the event A. Events B and A are separated by a timelike interval, and B lies in the forward lightcone of A. The events could be causally connected. Events C and A are separated by a null (or lightlike) interval and could be connected by a light signal. Events D and A are separated by a spacelike interval and cannot be causally connected.

2.2.6.1 A Potential Confusion: What the Observer Observes

We'll pause briefly to press home a point that may lead to confusion. You might think that the question of simultaneity has something to do with the finite speed of propagation. You don't see something until the light has travelled to you, just as you don't hear something until the sound has travelled to you. This is not what's going on here! A look at the spacetime diagram in Figure 48 shows that we've already taken this into account when deciding whether two events occur simultaneously. The lack of simultaneity between moving observers is a much deeper issue, not due to the finiteness of the speed of light but rather due to the constancy of the speed of light.

The confusion about the time of flight of the signal is sometimes compounded by the common use of the word observer to mean "inertial frame". This brings to mind some guy sitting at the origin, surveying all around him. Instead, you should think of the observer more as a Big Brother figure: a sea of clocks and rulers throughout the inertial frame which can faithfully record and store the position and time of any event, to be studied at some time in the future.

2.3 Length Contraction and Time Dilation

2.3.1 Time Dilation

We'll now turn to one of the more dramatic results of special relativity. Consider a clock sitting stationary in the frame S' which ticks at intervals of T'. This means that the tick events in frame S' occur at $(ct'_1, 0)$ then $(ct'_1 + cT', 0)$ and so on. What are the intervals between ticks in frame S?

We can answer immediately from the Lorentz transformations (2.14). Inverting this gives

$$t = \gamma\left(t' + \frac{vx'}{c^2}\right). \tag{2.21}$$

The clock sits at $x' = 0$, so we immediately learn that in frame S, the interval between ticks is

$$T = \gamma T' \tag{2.22}$$

This means that the gap between ticks is longer in the stationary frame. A moving clock runs more slowly. But the same argument holds for any process, be it clocks, elementary particles or human hearts. The correct interpretation is that time itself runs more slowly in moving frames. This is *time dilation.*[2]

Note that, throughout this course, we shall consider only *ideal clocks* – clocks that are unaffected by acceleration – for example, the half-life of a decaying particle.

2.3.2 Length Contraction

We've seen that moving clocks run slow. We will now show that moving rods are shortened. Consider a rod of length L' sitting stationary in the frame S'. What is its length in frame S?

[2] I'm not sure that this is a helpful description: what exactly is dilated?? It is better, as always in Special Relativity, to fix on a precise space-time description of the situation: what events we are considering and in which frame.

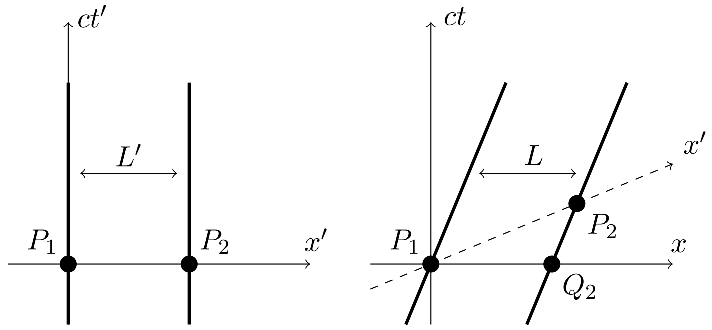

Fig. 2.5: The length L measured in frame S is $L = L'/\gamma$. It is shorter than the length of the rod in its rest frame by a factor of γ. This phenomenon is known as Lorentz contraction.

To begin, we should state more carefully something which seems obvious: when we say that a rod has length L', it means that the distance between the two end points at equal times is L'. So, drawing the axes for the frame S', the situation looks like the left diagram in Fig. 2.5. The two, simultaneous, end points in S' are P' and P'. Their coordinates in S' are $(ct', x') = (0, 0)$ and $(0, L')$ respectively.

Now let's look at this in frame S, illustrated in Fig. 2.5. Clearly P_1 sits at $(ct, x) = (0, 0)$. Meanwhile, the Lorentz transformation gives us the coordinate for P_2

$$x = \gamma L', \qquad \text{and,} \qquad t = \frac{\gamma v L'}{c^2}. \tag{2.23}$$

But to measure the rod in frame S, we want both ends to be at the same time. And the points P_1 and P_2 are not simultaneous in S. We can follow the point P_2 backwards along the trajectory of the end point to Q_2, which sits at

$$x = \gamma L' - vt. \tag{2.24}$$

We want Q_2 to be simultaneous with P_1 in frame S. This means we must move back a time $t = \gamma v L'/c^2$, giving

$$x = \gamma L' - \frac{\gamma v^2 L'}{c^2} = \frac{L'}{\gamma}. \tag{2.25}$$

This is telling us that the length L measured in frame S is

$$L = \frac{L'}{\gamma} \tag{2.26}$$

It is shorter than the length of the rod in its rest frame by a factor of γ. This phenomenon is known as *Lorentz contraction*. The rod suffers no contraction in the y- and z-directions (i.e., perpendicular to its velocity)

But hold on! If we look at the lengths of the rod, marked in the two diagrams in Fig. 2.5, it seems that we have got it the wrong way round: the length in frame S' definitely looks longer than in frame S. This is a trap: lengths in space-time diagrams are not like lengths in the more familiar x-y plane and we must rely on our calculations.[3]

[3]Lengths are shorter the closer the inclination to the 45° of the null cone. This is because instead of the Euclidean norm (Pythagoras), one must use the norm $|(ct, x)| = (c^2t^2 - x^2)^{1/2}$.

It follows that the volume V' of a moving object is related to proper volume V by $V = V'/\gamma$. Since the total number of objects in a system is Lorentz invariant, number densities thus transform from the rest frame as $n = n'/\gamma$.

2.3.3 The Ladder-and-Barn Non-Paradox

Take a ladder of length $2L$ and try to put it in a barn of length L. If you run fast enough, can you squeeze it? Here are two arguments, each giving the opposite conclusion

- From the perspective of the barn, the ladder contracts to a length $2L/\gamma$. This shows that it can happily fit inside as long as you run fast enough, with $\gamma \geq 2$.
- From the perspective of the ladder, the barn has contracted to length L/γ. This means there's no way you're going to get the ladder inside the barn. Running faster will only make things worse.

What's going on? The answer stems, as is often the case with apparent paradoxes in relativity, from loose use of language. As usual, to reconcile these two points of view we need to think more carefully about the question we're asking. What does it mean to 'fit a ladder inside a barn'?

In this case, it is the use of the word 'fit'; what does it mean to say the ladder 'fits' exactly into the barn? Clearly, we mean that the two events:

- front end of ladder hits back of barn;
- back end of ladder goes through the door.

are simultaneous. Any observer will agree that we've achieved this if the back end gets in the door before the front end hits the far wall. But we know that simultaneity of events is not fixed, as observers in different frames do not agree on simultaneity, so 'fit into' is a frame-dependent concept: we should not expect observers in different frames to agree so there is no paradox to account for. The two statements are true and compatible and that is really the end of the story. However, we can investigate further.

The spacetime diagram (see Fig. 2.6) in the frame of the barn is drawn in the figure with $\gamma > 2$. We see that, from the barn's perspective, both back and front ends of the ladder are happily inside the barn at the same time. We've also drawn the line of simultaneity for the ladder's frame. This shows that when the front of the ladder hits the far wall, the back end of the ladder has not yet got in the door. Is the ladder in the barn? Well, it all depends who you ask.

2.3.4 The Twins Non-Paradox

Twins Alice and Bob synchronise watches in an inertial frame and then Bob sets off at speed $\sqrt{3}c/2$, which corresponds to $\gamma = 2$. When Bob has been travelling for a time T

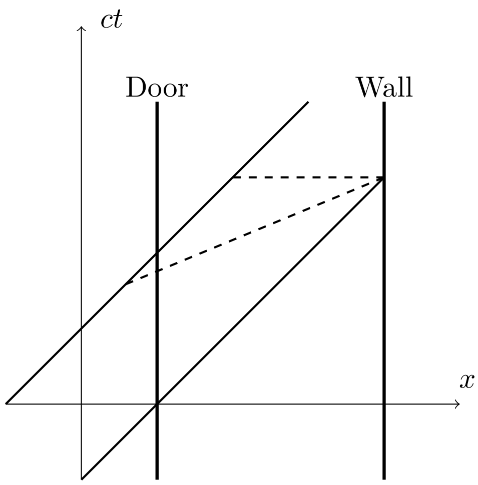

Fig. 2.6: The Ladder-and-Barn Non-Paradox: Regarded from the point of view of a space-time diagram, the paradox dissolves. One consequence of time not being invariant under Lorentz transformations is that the ladder 'fits in' the barn in one frame but does not 'fit in' in another.

according to Alice, he reaches Proxima Centauri[4] and turns round by means of accelerations that are very large in his frame and goes back to Alice at the same speed. Since Bob is in a moving frame, relative to Alice, his time runs slower by a factor of γ than Alice's, so he will only have aged by $2T \times \frac{1}{2}$ on the two legs of the journey. Thus when they meet up again, Alice has aged by $2T$ but Bob has aged only by T. *This is not the paradox: it is just a fact of life.*[5]

The difficulty some people have with Alice and Bob is the apparent symmetry: surely exactly the same argument could be made, from Bob's point of view, to show that Alice would be the younger when they met again? But the same argument *cannot* be made for Bob because the situation is not symmetric: Alice's frame is inertial, whereas Bob has to accelerate to turn round: while he is accelerating, his frame is not inertial.

BUT, some people might say, suppose we just consider the event of Bob's arrival at Proxima Centauri, so as not to worry about acceleration. Now the situation is symmetric. Surely from Alice's point of view, when Bob arrives he will have aged half as much as Alice, and from Bob's point of view, when he arrives, Alice will have aged half as much as Bob? The answer to this is a simple 'yes'. Surely, they would then say, this doesn't make sense? But it does, as long as you are careful about the word 'when'.

In the diagram in Fig. 2.7, Alice's world line is the ct (containing points A, B and C) axis and Bob's world line is the line containing A and P. P represents the event 'Bob arrives at Proxima Centauri'.

[4] The closest star to the Sun: about 4.2 light years away

[5] In 1971, Hafele and Keating packed four atomic (caesium) clocks into suitcases and went round the Earth, in different directions, on commercial flights. When they returned, they found that the clocks were slightly behind a clock remaining at the first airport. The result was somewhat inconclusive. The calculations are complicated by the fact that the rate of the clocks is also affected by the gravitational field: clocks run slower in stronger fields, and in fact the two affects balance at $3R/2$ (where R is the radius of the Earth). Thus the heights of the aircraft had to be taken into account as well as their speeds, and it turns out that the two effects are of comparable magnitude, namely of the order of 100 nanoseconds.

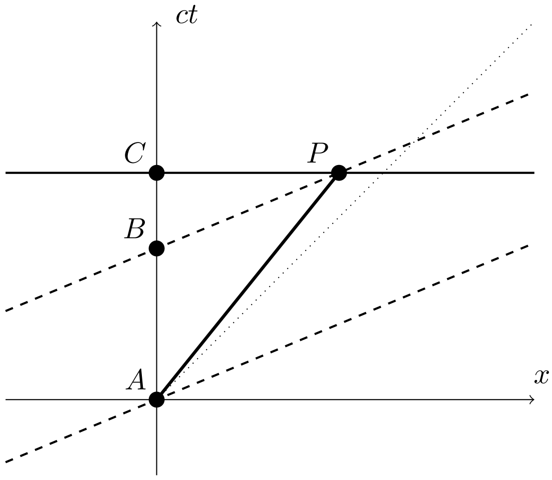

Fig. 2.7: The Twins Non-Paradox: Alice's world line is the ct (containing points A, B and C) axis and Bob's world line is the line containing A and P. P represents the event 'Bob arrives at Proxima Centauri'.

The line CP is a line of simultaneity in Alice's frame and C is the event 'Alice is at this point in space-time *when* – according to Alice – Bob arrives at Proxima Centauri'; the first use of the word 'when'.

The line BP is a line of simultaneity in Bob's frame and B is the event 'Alice is at this point in space-time when – according to Bob – he arrives Proxima Centauri'; the second use of the word 'when'. The two 'whens' don't mean the same thing, since one is a 'when' in Alice's frame the other is a 'when' in Bob's frame.

We can do the calculation. Let us assume for simplicity that Bob sets off the moment he is born. The event C has coordinates $(cT, 0)$ in Alice's frame, and the event P has coordinates (cT, vT). In Bob's frame, the elapsed time T' is given by the Lorentz transformation:

$$T' = \gamma(T - v^2T/c^2) = T/\gamma = \tfrac{1}{2}T. \tag{2.27}$$

This is just the usual time dilation calculation. Thus Bob and Alice agree that Bob's age at Proxima Centauri is $\frac{1}{2}T$. In Alice's frame, Bob has aged half as much as Alice.

We now work out the coordinates of the event B, sticking with Alice's frame. The line of simultaneity, BP has equation $t' = \frac{1}{2}T$, i.e. (using a Lorentz transformation)

$$\gamma(t + vx/c^2) = \tfrac{1}{2}T, \tag{2.28}$$

so the point B, for which $x = 0$, has coordinates $(\frac{1}{2}cT/\gamma, 0)$, i.e. $(\frac{1}{4}cT, 0)$. Alice's age when, according to Bob, he arrives at Proxima Centauri is therefore $\frac{1}{4}T$, which is indeed half of Bob's age. So no paradox there either.

BUT, some other people might say, suppose Bob does not turn round but just synchronises his watch at Proxima Centauri with that of another astronaut, Bob$'$, who is going at speed v in the opposite direction (like two trains passing at a station). Each leg of the journey is then symmetric, so why should Alice age faster or slower Bob and Bob$'$ during their legs of the journey? There's no mystery here, either: the situation is indeed symmetric and Alice does indeed age by the same amount as Bob+Bob$'$. But at the synchronisation event, Bob and Bob$'$ do not agree on Alice's age, because in their

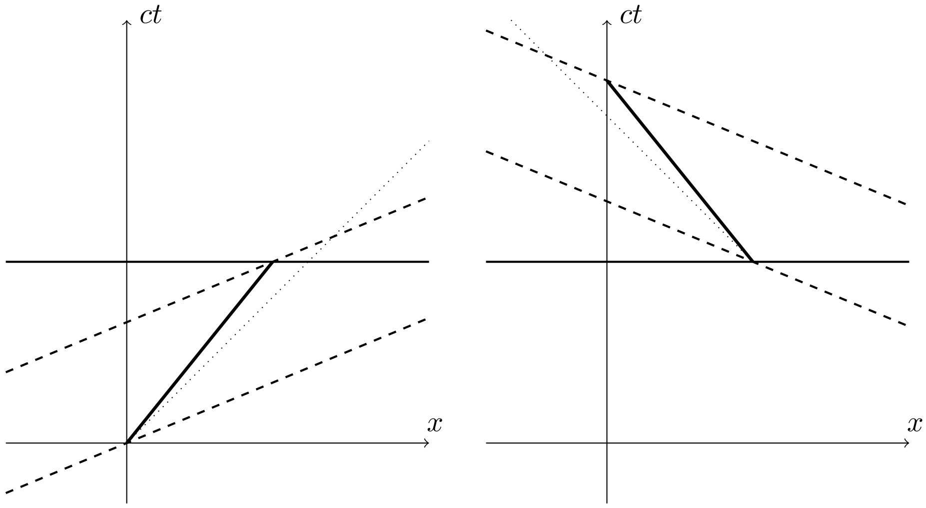

Fig. 2.8: Left: The outward journey. The heavy line is Bob's world line. The dotted line through the origin is the light cone. The dashed lines are the lines of simultaneity in Bob's frame. Right: The return journey. The heavy line is the world line of Bob′. The dotted line through the turn-round event is the light cone. The dashed lines are the lines of simultaneity in the frame of Bob′.

different frames the synchronisation event is simultaneous with different times in Alice's life.

Let us see how this looks in a space-time diagram with figures 2.7–2.8.

As before, Bob ages by $\frac{1}{2}T$ on the outward journey to Proxima Centauri. By symmetry Bob' ages by $\frac{1}{2}T$ on the inward journey from Proxima Centauri.

However, according to Bob's idea of time, the clock synchronisation occurs when Alice is at B, and according to Bob′'s it occurs when Alice is at D. Thus Bob′'s clock will read time T when he meets Alice and Alice's clock will read $2T$. But the time Alice spends between B and D is accounted for by Bob in his journey *after* Proxima Centauri and by Bob′ in his journey *before* reaching Proxima Centauri, so the two Bobs would say that, while they were travelling between Earth and Proxima Centauri, Alice travelled from A to B and then from D to E, taking on her clock a total time T – the same as the journey time of the two Bobs.

Finally, we see that if, instead of meeting Bob′, Bob turns round at Proxima Centauri, Alice ages rapidly (according to Bob) from B to D while he is changing direction.

2.4 Paths in spacetime

2.4.1 Minkowski Spacetime Line Element

As we first introduced in Eq. (2.16), the invariant interval $\Delta s^2 = c^2\Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2$ corresponds to the "distance" in spacetime between two events A and B measured along the straight line connecting them.

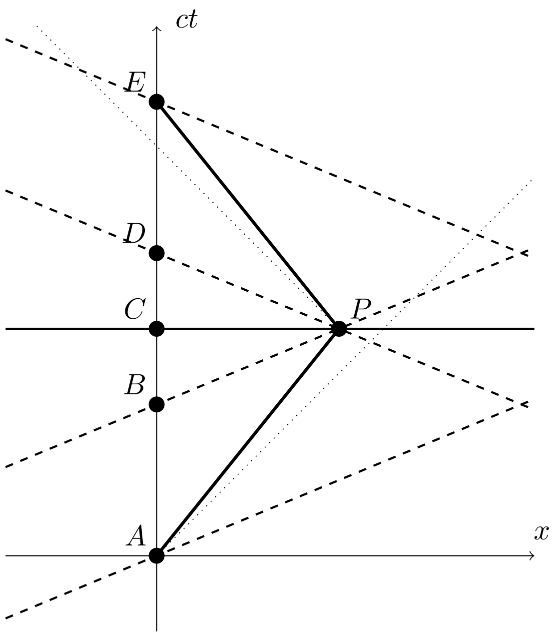

Fig. 2.9: The superposition of the previous two space-time diagrams in Fig. 2.8, representing together both the outward journey of Bob and the return journey of Bob′.

For a general, arbitrary path through spacetime, we must express the intrinsic geometry of Minkowski spacetime in infinitesimal form using the invariant Minkowski line element for infinitesimally-separated events:

$$\boxed{\mathrm{d}s^2 = c^2\,\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2\,.} \tag{2.29}$$

For a path connecting events A and B, the invariant "distance" along the path is given by the line integral

$$\Delta s = \int_A^B \mathrm{d}s\,. \tag{2.30}$$

The advantage of the invariant interval is that it is something all observers agree upon.

2.4.2 Particle Worldlines and Proper Time

A particle describes a *worldline* in spacetime. For a massive particle passing through an event A, the particle's worldline must be inside the lightcone through A and each infinitesimal step must lie within the lightcone at each point. For a photon or other massless particle, the worldline will be tangent to the lightcone.

The fact that the concept of time is frame dependent can be rather unsettling. It would be good to have some quantity that corresponds to time but does not vary at the whim of the observer. Such a quantity exists and is called *proper time*.

We can write the spacetime path as $x(t)$, $y(t)$, $z(t)$ or, parametrically, as $t(\lambda)$, $x(\lambda)$, $y(\lambda)$, $z(\lambda)$ for some parameter λ. The most natural parameter for a massive particle is *proper time* – the time measured by an ideal clock carried by the observer comoving with the particle (i.e. particle is at rest in the observer's frame).

The increment in proper time, $\mathrm{d}\tau$, is just the increment in time in the *instantaneous rest frame* of the particle, where $\mathrm{d}x' = \mathrm{d}y' = \mathrm{d}z' = 0$. It follows that $c^2\,\mathrm{d}\tau^2 = \mathrm{d}s^2$ and so, for two infinitesimally close events on the particle's worldline separated by $\mathrm{d}t$, $\mathrm{d}x$, $\mathrm{d}y$, $\mathrm{d}z$ in some inertial frame,

$$c^2\,\mathrm{d}\tau^2 = c^2\,\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2 \tag{2.31}$$

$$\Longrightarrow\ \mathrm{d}\tau = \mathrm{d}t\,/\gamma, \tag{2.32}$$

i.e.

$$\frac{\mathrm{d}t}{\mathrm{d}\tau} = \gamma. \tag{2.33}$$

Note that here in (2.33), γ is not linked to the velocity between two frames explicitly, though of course it is implicitly related to the velocity between the rest frame of the observer and the lab frame.

The total time that elapses on the world-line of an observer moving with (not necessarily constant) velocity in a frame S is given by

$$\Delta\tau = \int \mathrm{d}\tau = \int \gamma^{-1}\,\mathrm{d}t\,; \tag{2.34}$$

this is the observer's actual time (clock or biological).

We can use proper time to derive the velocity addition formula (2.46) for an observer moving with non-constant velocity. We parameterise the observer's world line by τ:

$$x = x(\tau), \quad t = t(\tau), \qquad \text{in } S \tag{2.35}$$

$$x' = x'(\tau), \quad t' = t'(\tau), \qquad \text{in } S' \tag{2.36}$$

and

$$u = \frac{\mathrm{d}x}{\mathrm{d}\tau}\bigg/\frac{\mathrm{d}t}{\mathrm{d}\tau}, \qquad u' = \frac{\mathrm{d}x'}{\mathrm{d}\tau}\bigg/\frac{\mathrm{d}t'}{\mathrm{d}\tau}. \tag{2.37}$$

We can differentiate the Lorentz transformation (2.12) and (2.5) to obtain

$$\frac{\mathrm{d}x'}{\mathrm{d}\tau} = \gamma\left(\frac{\mathrm{d}x}{\mathrm{d}\tau} - v\frac{\mathrm{d}t}{\mathrm{d}\tau}\right) = \gamma(u-v)\frac{\mathrm{d}t}{\mathrm{d}\tau} \tag{2.38}$$

$$\frac{\mathrm{d}t'}{\mathrm{d}\tau} = \gamma\left(\frac{\mathrm{d}t}{\mathrm{d}\tau} - \frac{v}{c^2}\frac{\mathrm{d}x}{\mathrm{d}\tau}\right) = \gamma(1 - \frac{uv}{c^2})\frac{\mathrm{d}t}{\mathrm{d}\tau} \tag{2.39}$$

and dividing these expressions gives

$$u' = \frac{u - v}{1 - uv/c^2}. \tag{2.40}$$

2.4.3 Doppler Effect

Consider an observer $\mathcal{E}$ who moves at speed v along the x-axis of an intertial frame S in which an observer $\mathcal{O}$ is at rest at position x_o (see Fig. 2.10). Let successive wavecrests be emitted by $\mathcal{E}$ at events A and B, which are separated by proper time $\Delta\tau_{AB}$; this is the *proper period* of the source.

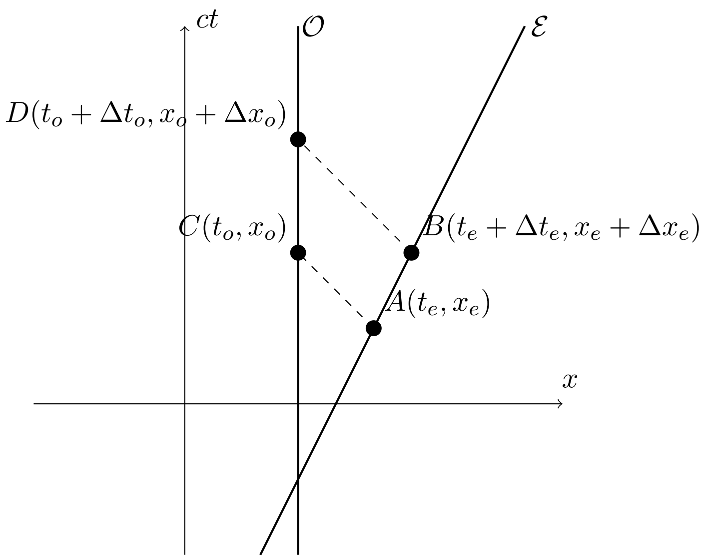

Fig. 2.10: Spacetime diagram of the Doppler effect. An observer $\mathcal{E}$ moves at speed v along the x-axis of an inertial frame S in which an observer $\mathcal{O}$ is at rest at position x_o. A wavecrest is emitted by $\mathcal{E}$ at the event A with coordinates (t_e, x_e) in S and is received by $\mathcal{O}$ at the event C with coordinates (t_o, x_o). A second crest is emitted by $\mathcal{E}$ at the event B, which occurs at a time Δt_e later than A in S, and is received by $\mathcal{O}$ at the event D a time Δt_o later than C.

The relation between $\Delta\tau_{AB}$ and the time Δt_e between the emission events in S is

$$\Delta\tau_{AB} = \Delta t_e/\gamma. \tag{2.41}$$

The wavecrests are received by $\mathcal{O}$ at the events C and D, which are separated by time Δt_o in S; since $\mathcal{O}$ is at rest in S, the proper time between C and D is $\Delta\tau_{CD} = \Delta\tau_o$. In time Δt_e, the source $\mathcal{E}$ moves a distance $\Delta x_e = v\Delta t_e$ along the x-axis in S, and the second wavecrest has to travel Δx_e further than the first to be received by $\mathcal{O}$ at x_o. It follows that

$$\Delta t_o = \left(1 + \frac{v}{c}\right)\Delta t_e, \tag{2.42}$$

so that the ratio $\Delta\tau_{AB}/\Delta\tau_{CD}$ of proper times is

$$\frac{\Delta\tau_{AB}}{\Delta\tau_{CD}} = \frac{\sqrt{1 - \frac{v^2}{c^2}}\Delta t_e}{(1 + \frac{v}{c})\Delta t_e} = \sqrt{\frac{1 - \frac{v}{c}}{1 + \frac{v}{c}}}. \tag{2.43}$$

This ratio is also the ratio of the received frequency, as measured by $\mathcal{O}$, to the proper frequency (i.e., the frequency in the rest-frame of the source $\mathcal{E}$).

2.4.4 Addition of Velocities

A particle moves with constant velocity u' in frame S' which, in turn, moves with constant velocity v with respect to frame S. What is the velocity u of the particle as seen in S?

The Newtonian answer is just $u = u' + v$. But we know that this can't be correct because it doesn't give the right answer when $u' = c$. So what is the right answer?

The worldline of the particle in S' is

$$x' = ut'. \tag{2.44}$$

So the velocity of the particle in frame S is given by

$$u = \frac{x}{t} = \frac{\gamma(x' + vt')}{\gamma(t' + vx'/c^2)}, \tag{2.45}$$

which follows from the Lorentz transformations (2.14). (Actually, we've used the inverse Lorentz transformations since we want S coordinates in terms of S' coordinates, but these differ only changing v to $-v$). Substituting (2.44) into the expression above, and performing a little algebra, gives us the result we want:

$$u = \frac{u' + v}{1 + u'v/c^2}. \tag{2.46}$$

Note that when $u' = c$, this gives us $u = c$ as expected. We can also show that if $|u'| < c$ and $|v| < c$ then we necessarily have $-c < u < c$. The proof is simple algebra, if a little fiddly

$$c - u = c - \frac{u' + v}{1 + u'v/c^2} = \frac{c(c - u')(c - v)}{c^2 + u'v} > 0, \tag{2.47}$$

where the last equality follows because, by our initial assumptions, each factor in the final expression is positive. An identical calculation will show you that $-c < u$ as well. We learn that if a particle is travelling slower than the speed of light in one inertial frame, it will also be travelling slower than light in all others.

It follows that the velocity components in S are given by

$$u_x = \frac{\mathrm{d}x'}{\mathrm{d}t'} = \frac{u'_x + v}{1 + u'_x v/c^2}, \tag{2.48}$$

$$u_y = \frac{\mathrm{d}y'}{\mathrm{d}t'} = \frac{u'_y}{\gamma(1 + u'_x v/c^2)}, \tag{2.49}$$

$$u_z = \frac{\mathrm{d}z'}{\mathrm{d}t'} = \frac{u'_z}{\gamma(1 + u'_x v/c^2)}. \tag{2.50}$$

The appropriate velocity transformations from frame S to frame S' are obtained by replacing v with $-v$ (and switching u'_i and u_i).

These results replace the "common-sense" addition of velocities in Newtonian mechanics; they reduce to the Newtonian results in the limit $v/c \to 0$ (or equivalently as $c \to \infty$).

Now consider three inertial frames S, S' and S'', where S' and S are related by a standard boost along the x-direction with speed v, and S'' and S' are related by a standard boost along the x'-direction with speed u'.

If we write the velocities u, u', and v in terms of rapidities:

$$\frac{u}{c} = \tanh\beta, \qquad \frac{u'}{c} = \tanh\beta', \qquad \frac{v}{c} = \tanh\alpha. \tag{2.51}$$

We can then find the composition of the two Lorentz transforms in terms of the rapidities by substitution into (2.40) to give

$$\tanh \beta' = \frac{\tanh \beta - \tanh \alpha}{1 - \tanh \alpha \tanh \beta} = \tanh(\beta - \alpha). \tag{2.52}$$

so an alternative form of the transformation law is

$$\beta' = \beta - \alpha \tag{2.53}$$

i.e.

$$\tanh^{-1}\left(\frac{u'}{c}\right) = \tanh^{-1}\left(\frac{u}{c}\right) + \tanh^{-1}\left(\frac{v}{c}\right). \tag{2.54}$$

We see that the composition of two colinear boosts is another boost along the same direction and the rapidities add (like adding angles for rotations about a common axis).

2.5 Acceleration in Special Relativity

It is often said, erroneously, that Special Relativity cannot deal with acceleration because it deals only with inertial frames, and that therefore acceleration must be the preserve of General Relativity. We must, of course, only allow transformations between inertial frames; the frames must not accelerate, but the observers in the frame can move as the please. Special Relativity can deal with anything kinematic but General Relativity is required when gravitational forces are present.

As an example of non-uniform motion, we consider an observer who is moving with constant acceleration.

The first step is to define what we mean by 'constant acceleration' which is certainly a frame-dependent concept. The most common situation is that of an observer in a rocket experiencing a constant 'G-force' due to the rocket thrust. This corresponds to the acceleration measured in the instantaneous (inertial) rest frame of the rocket being constant (acceleration having the usual definition of $\mathrm{d}v \,/\, \mathrm{d}t$), so we take this to be our definition.

For reasons that will later become clear (see section **??**), we need to determine the way that acceleration transforms under Lorentz transformations. We can do this in a number of ways. We will here start with the velocity transformation law (2.46) for an observer with world line given in S by $(ct(\tau), x(\tau))$ and in S' by $(ct'(\tau), x'(\tau))$. Forgetting the acceleration problem for the moment, we assume that these frames have a constant relative velocity v.

The velocities u and u' in the two frames are related by

$$u' = \frac{u - v}{1 - uv/c^2} \equiv \frac{(c^2/v)(1 - v^2/c^2)}{1 - uv/c^2} - \frac{c^2}{v}. \tag{2.55}$$

(the equivalent form is just a bit of algebra to obtain a useful expression). Differentiating this with respect to τ gives

$$\frac{\mathrm{d}u'}{\mathrm{d}\tau} = \frac{1 - v^2/c^2}{(1 - uv/c^2)^2} \frac{\mathrm{d}u}{\mathrm{d}\tau}. \tag{2.56}$$

The acceleration, a, in S is by definition $\mathrm{d}u\,/\,\mathrm{d}t$ and similarly for S' so

$$
\begin{aligned}
a' &= \frac{\mathrm{d}u'}{\mathrm{d}t'} \\
&= \frac{\mathrm{d}u'}{\mathrm{d}\tau}\bigg/\frac{\mathrm{d}t'}{\mathrm{d}\tau} \\
&= \frac{1-v^2/c^2}{(1-uv/c^2)^2}\frac{\mathrm{d}u}{\mathrm{d}\tau}\bigg/\frac{\mathrm{d}t'}{\mathrm{d}\tau} && \text{(using (2.56))} \\
&= \frac{1-v^2/c^2}{(1-uv/c^2)^2}\frac{\mathrm{d}u}{\mathrm{d}\tau}\bigg/\gamma(1-uv/c^2)\frac{\mathrm{d}t}{\mathrm{d}\tau} && \text{(using (2.38))} \\
&= \frac{(1-v^2/c^2)^{3/2}}{(1-uv/c^2)^3}a \\
&= \frac{1}{\gamma^3(1-uv/c^2)^3}a. && (2.57)
\end{aligned}
$$

As mentioned above there are other ways of obtaining this result; for example, more elegantly using four-vectors (see section **??**).

In the situation we have in mind, S' is the instantaneous rest frame of the accelerating observer, so that $u' = 0$ and $u = v$, and the acceleration a' in this frame is constant (i.e. independent of v). Thus (2.57) becomes

$$a = a'/\gamma^3. \tag{2.58}$$

Now

$$a = \frac{\mathrm{d}u}{\mathrm{d}\tau}\bigg/\frac{\mathrm{d}t}{\mathrm{d}\tau}, \tag{2.59}$$

and using (2.33), so we can find the parameterised equation of the world line by integrating

$$\frac{\mathrm{d}u}{\mathrm{d}\tau} = a\frac{\mathrm{d}t}{\mathrm{d}\tau} = a'/\gamma^2. \tag{2.60}$$

This gives

$$u = c\tanh\left(a'\tau/c\right), \tag{2.61}$$

choosing the origin of τ so that $u = 0$ when $\tau = 0$, and hence

$$\gamma = \cosh\left(a'\tau/c\right). \tag{2.62}$$

Then from $\mathrm{d}t\,/\,\mathrm{d}\tau = \gamma$, we find that

$$t = c/a'\sinh\left(a'\tau/c\right), \tag{2.63}$$

choosing the origin of t such that $t = 0$ when $\tau = 0$. Finally,

$$
\begin{aligned}
\frac{\mathrm{d}x}{\mathrm{d}\tau} &= \frac{\mathrm{d}x}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{d}\tau} && (2.64) \\
&= u\gamma && (2.65) \\
&= c\sinh\left(a'\tau/c\right), && (2.66)
\end{aligned}
$$

so, choosing the origin of x such that $x = c^2/a'$ when $t = 0$,

$$x = c^2/a'\cosh\left(a'\tau/c\right). \tag{2.67}$$

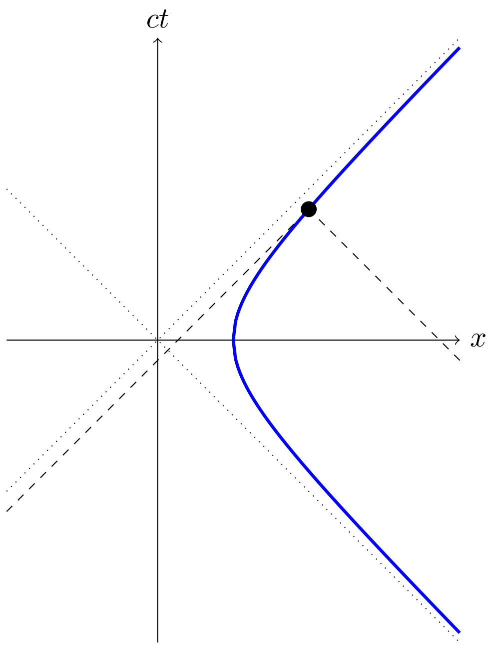

Fig. 2.11: The space-time diagram for an accelerated observer. The thick hyperbola is the observer's world line. An observer 'below' the dashed lines could in principle send a message to the observer marked as a heavy dot; other observers could not.

Uniformly accelerated particles therefore move on rectangular hyperbolas of the form

$$x^2 - (ct)^2 = (c^2/a')^2. \tag{2.68}$$

The diagram in Fig 2.11 shows the trajectory. The dotted lines are the light cones. An event taking place within the dashed lines can influence an accelerated observer at the position shown, but events taking place outside the dashed lines would have to move faster than the speed of light to do so. As $\tau \to \infty$, the whole of the space-time to the left of the dotted line $x = ct$ would be inaccessible to the observer. This line is called the *Rindler event horizon* for the accelerated observer. In some ways, it performs the same function as the event horizon of a black hole. In particular, the observer has to accelerate to avoid falling through it and anything happening on the other side would be hidden to the observer. Of course, the accelerating observer could just stop accelerating whereas the observer in a black hole space-time can do nothing to affect the event horizon.

Moreover, the accelerated observer sees the emitted light Doppler shifted to longer and longer wavelengths as the object approaches the event horizon and is observed as $\tau \to \infty$.

Chapter 3

Introducing Differential Geometry

Gravity is geometry. To fully understand this statement, we will need more sophisticated tools and language to describe curved space and, ultimately, curved spacetime. This is the mathematical subject of differential geometry and will be introduced in this section and the next.

Our discussion of differential geometry is not particularly rigorous. We will not prove many big theorems. Furthermore, a number of the statements that we make can be checked straightforwardly but we will often omit this. We will, however, be careful about building up the mathematical structure of curved spaces in the right logical order. As we proceed, we will come across a number of mathematical objects that can live on curved spaces. Many of these are familiar – like vectors, or differential operators – but we'll see them appear in somewhat unfamiliar guises. The main purpose of this section is to understand what kind of objects can live on curved spaces, and the relationships between them. This will prove useful for both general relativity and other areas of physics.

Moreover, there is a wonderful rigidity to the language of differential geometry. It sometimes feels that any equation that you're allowed to write down within this rigid structure is more likely than not to be true! This rigidity is going to be of enormous help when we return to discuss theories of gravity.

3.1 Concept of a Manifold

The stage on which our story will play out is a mathematical object called a *manifold*. We will give a precise definition below, but for now you should think of a manifold as a curved, n-dimensional space. If you zoom in to any patch, the manifold looks like $\mathbf{R}^n$. But, viewed more globally, the manifold may have interesting curvature or topology.

To begin with, our manifold will have very little structure. For example, initially there will be no way to measure distances between points. But as we proceed, we will describe the various kinds of mathematical objects that can be associated to a manifold, and each one will allow us to do more and more things. It will be a surprisingly long time before we can measure distances between points!

You have met many manifolds in your education to date, even if you didn't call them by name. Some simple examples in mathematics include Euclidean space $\mathbf{R}^n$, the sphere $\mathbf{S}^n$, and the torus $\mathbf{T}^n = \mathbf{S}^1 \times \cdots \times \mathbf{S}^1$. Some simple examples in physics include the configuration space and phase space that we use in classical mechanics and the state space of thermodynamics. As we progress, we will see how familiar ideas in these subjects can be expressed in a more formal language. Ultimately our goal is to explain how spacetime is a manifold and to understand the structures that live on it.

We now come to our main character: an n-dimensional manifold is a space which, locally, looks like $\mathbf{R}^n$. Globally, the manifold may be more interesting than $\mathbf{R}^n$, but the idea is that we can patch together these local descriptions to get an understanding for the entire space.

Informally, an N-dimensional manifold is a set of objects that locally resembles ND Euclidean space $\mathbf{R}^n$. In relativity, the objects are events and the set of events is spacetime. What "locally resembles" means is that there exists a map ϕ from the ND manifold $\mathcal{M}$ to an *open subset* of $\mathbf{R}^n$ that is one-to-one and onto.[1]

Under the map ϕ, a point $P \in \mathcal{M}$ maps to a point in the open subset U of $\mathbf{R}^n$ with *coordinates* x_a , $a = 1, \cdots, N$. Generally, we cannot cover the entire manifold with a single map ϕ (or, equivalently, set of coordinates), but it is sufficient if we can subdivide $\mathcal{M}$ and map each piece separately onto open subsets of $\mathbf{R}^n$.

The manifold is *differentiable* if these subdivisions join up smoothly so that we can define scalar fields on the manifold that are differentiable everywhere.

We can generally think of manifolds as surfaces embedded in some higher-dimensional Euclidean space, and we shall often do so, but it is important to appreciate that a given manifold exists independent of any embedding. A non-trivial example of a manifold is the set of rotations in 3D; these can be parameterised by three Euler angles, which form a coordinate system for the 3D manifold.

3.2 Coordinates

As we have just seen, points in an ND manifold can be labelled by N real-valued coordinates $(x_1, x_2, \cdots, x_N)$. We shall denote these collectively by x_a with $a = 1, \cdots, N$. The coordinates are not unique: think of them as labels of points in the manifold that can change under a coordinate transformation (i.e., a change of map ϕ) while the point itself does not.

We have also noted that, generally, it will not be possible to cover a manifold with a single *non-degenerate* coordinate system, i.e., one where the correspondence between points and coordinate labels is one-to-one. In such cases, multiple coordinate systems are required to cover the whole manifold.

Here are a few simple examples of differentiable manifolds:

- Coordinates (ρ, ϕ) in the plane $\mathbf{R}^2$: The Euclidean plane $\mathbf{R}^2$ is a 2D manifold that can be covered globally with the usual Cartesian coordinates. However, we could instead use plane-polar coordinates, (ρ, ϕ) with $0 \leq \rho \leq \infty$ and $0 \leq \phi < 2\pi$. Plane-polar coordinates are degenerate at $\rho = 0$ since ϕ is indeterminate there.

[1] An open subset U of $\mathbf{R}^n$ is such that for any point one can construct a sphere centred on the point whose interior lies entirely inside U. A map from $\mathcal{M}$ to U is one-to-one and onto if every element of U is mapped to by exactly one element of $\mathcal{M}$.

- Coordinates (θ, ϕ) on the 2-sphere $\mathbf{S}^2$: The 2-sphere is the set of points in $\mathbf{R}^3$ with $x^2+y^2+z^2=1$. It is an example of a 2D manifold. The spherical polar coordinates (θ, ϕ), with $0 \leq \theta \leq \pi$ and $0 \leq \phi < 2\pi$, are degenerate at the poles $\theta = 0$ and $\theta = \pi$, where ϕ is indeterminate. For $\mathbf{S}^2$, there is no single coordinate system that covers the whole manifold without degeneracy: at least two coordinate patches are required.

3.2.1 Curves and Surfaces

Subsets of points in a manifold define *curves* and *surfaces*. These are usually defined parametrically for some coordinate system, e.g., for a curve with parameter u:

$$x^a = x^a(u) \qquad (a = 1, 2, \cdots, N). \tag{3.1}$$

For a *submanifold* (or surface) of M $(M < N)$ dimensions, we need M parameters:

$$x^a = x^a(u^1, u^2, \cdots, u^M) \qquad (a = 1, 2, \cdots, N). \tag{3.2}$$

The special case $M = N - 1$ is called a hypersurface. In this case, we can eliminate the $N - 1$ parameters from the N equations (3.2) to give

$$f(x^1, x^2, \cdots, x^N) = 0, \tag{3.3}$$

for some function f.

Similarly, points in an M-dimensional surface can be specified by $N-M$ (independent) constraints

$$f_1(x^1, x^2, \cdots, x^N) = 0, \cdots, f_{N-M}(x^1, x^2, \cdots, x^M) = 0, \tag{3.4}$$

i.e., by the intersection of $N - M$ hypersurfaces, as an alternative to the parametric representation of Eq. (3.2).

3.2.2 Coordinate Transformations

Coordinates are used to label points in a manifold, but the labelling is arbitrary. Later, we shall learn how to construct geometric objects that are independent of the way we assign coordinates, and that express the true physical content of the theory (think vectors in $\mathbf{R}^n$).

We can relabel points by performing a coordinate transformation given by N equations

$$x'^a = x'^a(x^1, x^2, \cdots, x^N) \qquad (a = 1, 2, \cdots, N). \tag{3.5}$$

We shall view coordinate transformations as *passive*, i.e., assigning new coordinates x'^a to a given point in terms of the original coordinates x^a. We shall further assume that the functions $x'^a(x^1, x^2, \cdots, x^N)$ are single-valued, continuous and differentiable.

Consider two neighbouring points P and Q with coordinates x^a and $x^a + \mathrm{d}x^a$. In the new (primed) coordinates,

$$\mathrm{d}x'^a = \sum_{b=1}^{N} \frac{\partial x'^a}{\partial x^b}\,\mathrm{d}x^b\,, \tag{3.6}$$

where the partial derivatives are evaluated at the point P. This defines an $N \times N$ transformation matrix at the point P with elements

$$J^a_b = \frac{\partial x'^a}{\partial x^b} = \begin{pmatrix} \frac{\partial x'^1}{\partial x^1} & \cdots & \frac{\partial x'^1}{\partial x^N} \\ \vdots & & \vdots \\ \frac{\partial x'^N}{\partial x^1} & \cdots & \frac{\partial x'^N}{\partial x^N} \end{pmatrix}, \tag{3.7}$$

where the numerator (index a) labels the rows and the denominator (index b) the columns. The determinant of $J \equiv \det(J^a_b)$ is the *Jacobian* of the transformation. If $J \neq 0$ for some range of the coordinates, the coordinate transformation can be inverted locally to give x^a as a function of the x'^a.

The transformation matrix for the inverse

$$x^a = x^a(x'^1, x'^2, \cdots, x'^N) \tag{3.8}$$

is the inverse of J^a_b; this follows from the chain rule for partial derivatives,

$$\sum_{b=1}^{N} \frac{\partial x'^a}{\partial x^b}\frac{\partial x^b}{\partial x'^c} = \frac{\partial x'^a}{\partial x'^c} = \delta_{ac}. \tag{3.9}$$

It also follows that the determinant of the inverse transformation is $1/J$.

3.2.3 Einstein Summation Convention

It will rapidly get cumbersome to include the summation over indices explicitly, as in Eq. (3.6). We therefore introduce the Einstein summation convention: Whenever an index occurs twice in an expression, once as a subscript and once as a superscript, summation over the index from 1 to N is implied.

For example, for an infinitesimal displacement

$$\mathrm{d}x'^a = \frac{\partial x'^a}{\partial x^b}\,\mathrm{d}x^b\,. \tag{3.10}$$

Here, the index a is a free index and may take any value from 1 to N, while the index b is summed over 1 to N. Note the following points about the summation convention.

- A superscript in the denominator of a partial derivative is considered a subscript, which is why the index b in Eq. (3.10) is summed over.
- Indices that are summed over are called dummy indices because can be replaced by any other index not already in use, e.g.,

$$\frac{\partial x'^a}{\partial x^b}\,\mathrm{d}x^b = \frac{\partial x'^a}{\partial x^c}\,\mathrm{d}x^c\,. \tag{3.11}$$

- In any term, an index should not occur more than twice, and any repeated index must occur once as a subscript and once as a superscript (and is summed over).

3.3 Local Geometry of Riemannian Manifolds

The general definition of a differentiable manifold does not define its *geometry*. To do so requires introducing additional structure to the manifold. Consider two neighbouring points P and Q in a manifold, i.e., points with coordinates x^a and $x^a + \mathrm{d}x^a$, in some coordinate system, which differ infinitesimally. The *local geometry* near P is specified by giving the invariant "distance" or "interval" between the points. In a *Riemannian manifold*, the interval takes the form (summation convention!)

$$\boxed{\mathrm{d}s^2 = g_{ab}(x)\,\mathrm{d}x^a\,\mathrm{d}x^b\,,} \tag{3.12}$$

i.e., the interval is quadratic in the coordinate differentials. The coefficients $g_{ab}(x)$ contain information about the local geometry but also depend on the particular coordinate system. Strictly, the geometry is Riemannian if $\mathrm{d}s^2 > 0$ and pseudo-Riemannian otherwise (the latter being the relevant case for spacetime). It is also possible to consider more general intervals, but these are not relevant for general relativity because of the equivalence principle.

3.3.1 The Metric

The metric functions relate infinitesimal changes in the coordinates to invariantly-defined "distances" in the manifold. In general relativity, these will be proper distances and times. The metric functions $g_{ab}(x)$ can always be chosen symmetric, $g_{ab}(x) = g_{ba}(x)$. To see this, note that we can write a general gab as the sum of a symmetric and antisymmetric part:

$$g_{ab}(x) = \tfrac{1}{2}[g_{ab}(x) + g_{ba}(x)] + \tfrac{1}{2}[g_{ab}(x) - g_{ba}(x)]. \tag{3.13}$$

The contribution of the antisymmetric part to $\mathrm{d}s^2$ vanishes since

$$\begin{aligned}(g_{ab} - g_{ba})\,\mathrm{d}x^a\,\mathrm{d}x^b &= g_{ab}\,\mathrm{d}x^a\,\mathrm{d}x^b - g_{ba}\,\mathrm{d}x^b\,\mathrm{d}x^a \\ &= (g_{ab} - g_{ab})\,\mathrm{d}x^a\,\mathrm{d}x^b \\ &= 0\end{aligned} \tag{3.14}$$

where we have relabelled the dummy indices $a \leftrightarrow b$ in the first line on the right.

It follows that in an N-dimensional Riemannian manifold there are $N(N+1)/2$ independent metric functions at each point. Given two neighbouring points, the interval between them is independent of the coordinate system used. Since the coordinate differentials change under a change of coordinates, so must the metric functions, i.e.,

$$\begin{aligned}\mathrm{d}s^2 &= g_{ab}(x)\,\mathrm{d}x^a\,\mathrm{d}x^b \\ &= \frac{\partial x^a}{\partial x'^c}\frac{\partial x^b}{\partial x'^d}\,\mathrm{d}x'^c\,\mathrm{d}x'^d \\ &= g'_{cd}(x')\,\mathrm{d}x'^c\,\mathrm{d}x'^d\,,\end{aligned} \tag{3.15}$$

where the metric functions in the new coordinates at the same physical point are $g'_{cd}(x')$.

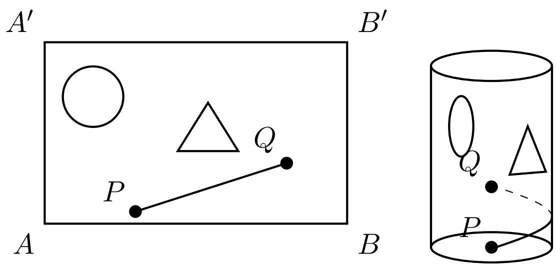

Fig. 3.1: The Euclidean plane $\mathbf{R}^2$ can be rolled up into a cylindrical surface without distortion. The intrinsic geometry of the cylindrical surface is therefore the same as the plane. In particular, a bug confined to the surface would measure the sum of the angles of a triangle to be 180° and the circumference of a circle to be 2π times its radius.

We can read off from Eq. (3.15) that the metric functions must transform as

$$\boxed{g'_{cd}(x') = g_{ab}(x(x'))\frac{\partial x^a}{\partial x'^c}\frac{\partial x^b}{\partial x'^d}.} \tag{3.16}$$

Since there are N arbitrary coordinate transformations that we can make, there are really only $N(N-1)/2$ independent functional degrees of freedom associated with $g_{ab}(x)$.

3.3.2 Intrinsic and Extrinsic Geometry

The interval (or line element) $\mathrm{d}s^2$ characterises the local geometry (or curvature), which is an *intrinsic* property of the manifold independent of any possible embedding in some higher-dimensional space.

Intrinsic properties are those that can be determined by a "bug" confined to the manifold – the bug can set up a coordinate system, measure physical distances and hence determine the metric functions.

As an example of the distinction between intrinsic and extrinsic geometry, consider the surface of a cylinder of radius a embedded in $\mathbf{R}^3$ (see Fig. **??**).

3.4 Lengths and volumes

3.4.1 Lengths along curves

3.4.2 Volumes of regions

3.4.3 Invariance of the volume element

3.5 Local Cartesian coordinates

3.5.1 Proof of existence of local Cartesian coordinates

3.6 Pseudo-Riemannian manifolds

3.7 Topology of manifolds

Chapter 4

Vector Tensor Algebra

4.1 Scalar and vector fields on manifolds

4.1.1 Scalar fields

4.1.2 Vector fields and tangent spaces

4.1.3 Vectors as differential operators

4.1.4 Dual vector fields

4.2 Tensor fields

4.2.1 Tensor equations

4.2.2 Elementary operations with tensors

4.2.3 Quotient theorem

4.3 Metric tensor

4.3.1 Inverse metric

4.4 Scalar products of vectors revisited

CHAPTER 5

Vector and Tensor Calculus on Manifolds

5.1 Covariant derivatives

5.1.1 Derivatives of scalar fields

5.1.2 Covariant derivatives of tensor fields

5.1.3 The connection

5.1.3.1 Extension to other tensor fields

5.1.4 The metric connection

5.1.4.1 Other useful properties of the metric connection

5.1.5 5 Relation to local Cartesian coordinat

5.1.6 Divergence, curl and the Laplacian

5.2 Intrinsic derivative of vectors along a curv

5.3 Parallel transpor

5.3.1 Properties of parallel transport

5.4 Geodesic curves

5.4.1 Tangent vectors

5.4.2 Stationary property of non-null geodesics

5.4.3 Relation to parallel transport

5.4.4 Alternative “Lagrangian” procedure

5.4.5 Conserved quantities along geodesics

Chapter 6

Chapter

Chapter 7

Chapter

Chapter 8

Chapter

Chapter 9

Chapter

Chapter **10**

Chapter

Chapter 11

Chapter

Chapter **12**

Chapter

Chapter 13

Chapter

Appendix A

Code