Chapter 2 Tools for mapping space and time

The science of cosmology involves attempting to map the Universe in space and to consider its evolution in time. You saw in the previous chapter that distance and time measurements in astronomy are closely linked: the finite speed of light means that distant parts of the Universe can only be observed as they appeared at times in the long-ago past. Our maps of the night sky are therefore maps of time as well as space, recording the history of the Universe.

In this chapter you will learn about a further way in which space and time are entwined. The sections that follow summarise some of the key concepts of special relativity and some of their non-intuitive consequences. The discussion of special relativity in this chapter forms the starting point for building the conceptual framework that underpins cosmological theory.

Online resources: introductory special relativity

If you have not previously studied Einstein's theory of special relativity, or would like a refresher, you may find the online resources for this chapter helpful.

Objectives

Working through this chapter will enable you to:

- explain the importance of reference frames for making measurements related to space and time intervals
- manipulate the Lorentz transformations to compare measurements in different frames and explore some non-intuitive consequences of special relativity
- define the concept of a metric, and apply it to examples of two- and three-dimensional geometries and four-dimensional spacetime
- explain how the curvature of geometric spaces can be measured, and describe the key differences between flat and curved spaces
- solve numerical problems relating to the curvature of two-dimensional surfaces.

2.1 Understanding space and time

2.1.1 Reference frames and relativity

If you have ever sat in a moving car watching roadside trees appear to zoom away from you, or in a train carriage looking at another train through the window and been unsure which train is stationary and which is moving, then you have everyday experience of the concept of **reference frames**. Any time we measure a speed it is in relation to a particular reference frame. You are unlikely to be reading this book from anywhere other than one particular rapidly rotating planet in a solar system moving at around $200\,\mathrm{km\,s^{-1}}$ around the centre of our Galaxy. However, the speed that you are moving relative to other parts of the Universe is usually irrelevant for considerations of how objects move in your local environment.

A basic concept in relativity is the **inertial frame**. An inertial frame is a reference frame that is not accelerating, in which the laws of motion (e.g. Newton's laws) apply.

Figure 2.1 shows two inertial frames, which are moving relative to each other at a constant speed, V. Each frame is defined by a set of coordinate axes: x, y and z in frame S, and x', y' and z' in frame S$'$. It is usual in special relativity problems to define the x-axis as the direction of relative motion between the frames. Likewise, in the standard configuration for special relativity problems the frame origins are assumed to coincide at a particular time, $t = t' = 0$.

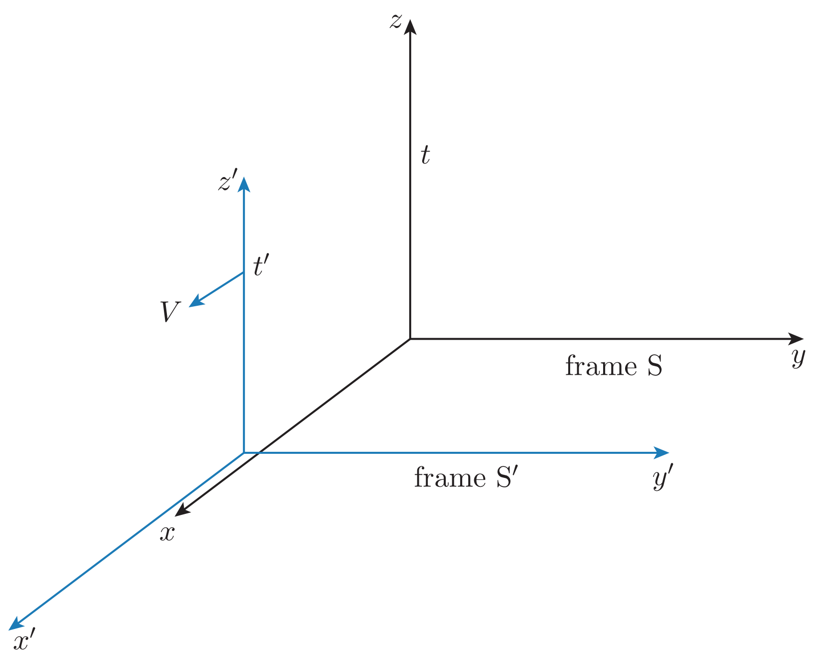

Figure 2.1 Two inertial references frames with coordinate axes labelled.

The special theory of relativity was born from the need to reconcile two well-evidenced statements ('postulates'):

The postulates of special relativity

- The laws of physics operate in the same way in all inertial frames.
- The speed of light in a vacuum has the same constant value, $c = 3 \times 10^8\,\mathrm{m\,s^{-1}}$, in all inertial frames.

Both of these statements can be experimentally verified. The constant speed of light was first established via experiments in the late nineteenth century. A famous example is the Michelson–Morley experiment, which ruled out the presence of an 'ether': an all-pervading material, thought at the time to be a medium through which light waves travelled.

But the constancy of the speed of light is hard to reconcile with our intuitive understanding of how inertial reference frames behave. To understand why this is a problem, we need to introduce (or perhaps remind you of) some key definitions.

Special relativity terminology

- An **event** is an occurrence that takes place instantaneously (i.e. it does not extend for a significant length of time) at a fixed location in space.
- A **coordinate system** is a set of axes that define an inertial reference frame, e.g. x-, y- and z-axes, together with a time axis, t.
- An **observer** is someone who can make measurements of events, noting down the coordinates at which they occur within their own inertial frame.

In the following example you will consider the problem of how a moving object is measured in two frames, S and S$'$, where those frames are defined in the same way as in Figure 2.1.

Example 2.1

Figure 2.2 depicts a scenario at a train station, in which two observers experience an event from different reference frames.

- Observer B is on a train that is slowly passing through the station at a speed of $V = 20\,\mathrm{km\,h^{-1}}$.
- Observer A is standing still on the station platform.

Observer B throws a ball down the train carriage at a speed of $10\,\mathrm{km\,h^{-1}}$ in the direction opposite to the train's direction of motion, i.e. in the $-x$ direction.

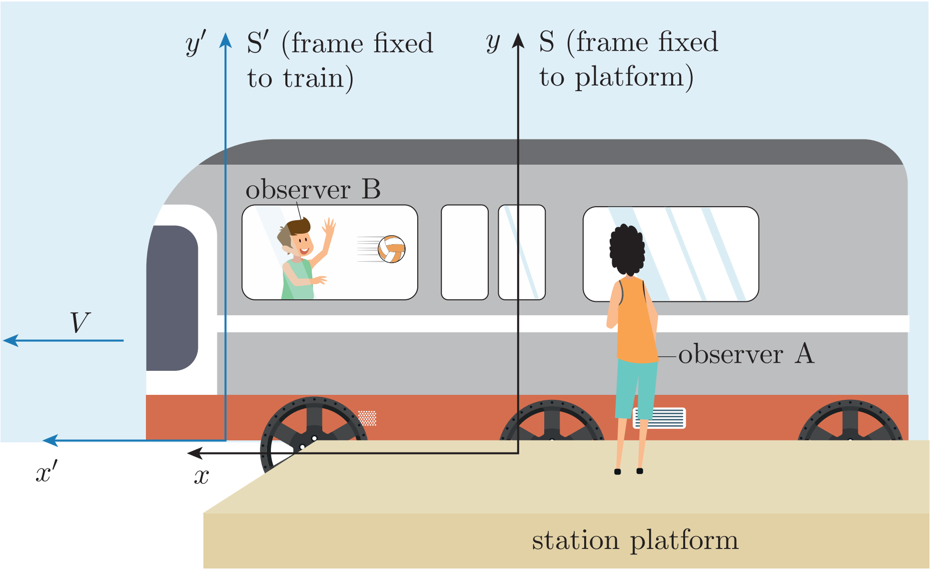

Figure 2.2 The reference frames of observers A (on a station platform) and B (on a train leaving the station at speed V in the x direction).

Taking S as the reference frame of observer A (on the platform), and S′ as the frame of observer B (on the train), determine the speed of the ball as measured by each of the two observers. In other words, calculate and compare $\mathrm{d}x/\mathrm{d}t$ and $\mathrm{d}x'/\mathrm{d}t$ by considering how the coordinate position in each frame must change with time.

Solution

The ball's speed in B's frame, S′, is $\mathrm{d}x'/\mathrm{d}t$, where the prime ($'$) superscript indicates that we are referring to coordinates in frame S′. For A (whose frame is S), the ball's speed is given by its change of position relative to their coordinate x, so $\mathrm{d}x/\mathrm{d}t$.

The question tells us that in frame S′ the ball's velocity $\mathrm{d}x'/\mathrm{d}t$ is $-10\,\mathrm{km\,h^{-1}}$. To work out what A sees in frame S, we need to think about how the coordinates x and x' are related. If Δt is the time elapsed since $x = x'$ (i.e. the point at which the two sets of coordinates coincided), then in that time interval observer A will see a given location on the train changing position by $V\Delta t$. The x-coordinate location of the end of the train will be increasing, whereas its x'-coordinate will not, because the S′-coordinates are fixed to the train. Therefore, the x- and x'-coordinates are related by:

$$x = x' + V\Delta t$$

where $V = 20\,\mathrm{km\,h^{-1}}$ is the relative speed of the two frames in the x direction.

Because we want to know the rates of change of x and x', we can take the derivative of the relation above to get

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\mathrm{d}x'}{\mathrm{d}t} + V$$

Therefore, we find that the velocity of the ball as observed by A is $-10\,\mathrm{km\,h^{-1}} + 20\,\mathrm{km\,h^{-1}} = 10\,\mathrm{km\,h^{-1}}$. In other words, A will see the ball moving away from them, in the same direction that the train is moving, but it will be moving away more slowly than the train itself. So for A, the ball still appears to be moving down the carriage away from observer B, as expected on the basis of everyday experience.

The example showed that rates of change in the position of objects will be different when measured from different frames. But what if, instead of considering an object like a ball, we consider how a beam of torchlight produced by B travels down the train?

If we apply the same logic as we did in the example of the ball we should conclude that A observes the light beam to be travelling at a speed of $-c + 20\,\mathrm{km\,h^{-1}}$, but this isn't consistent with the second postulate of special relativity. Both A and B *must* observe the light to travel at the same speed, so the logic we have applied in the example must have a gap in it. It is this gap that shows the need for a more sophisticated theory to describe the relationship between how different observers measure physical behaviour.

- In comparing measurements in two reference frames, what assumption was made in Example 2.1 about how the observers measure time intervals?

- We assumed that observers A and B measure time progressing (e.g. their clocks 'ticking') at the same rates.

In the next section you will see that the assumption that all observers measure time in the same way has to be sacrificed to enable the two postulates of special relativity to be reconciled in a logical way. The consequences of this are at odds with our everyday experience, but are well verified by a variety of experiments.

2.1.2 Transformations between reference frames

The example in the previous section used everyday intuition to consider how distance intervals and speeds are measured differently from two reference frames. This intuition can be summarised by writing a set of equations to transform coordinates between two inertial frames whose relative velocity in the direction of the x-coordinate axis is V (e.g. as shown in Figure 2.1):

$$
\begin{aligned}
x' &= x - Vt \\
y' &= y \\
z' &= z \\
t' &= t
\end{aligned}
$$

These equations are the **Galilean transformations**. But, as you saw in the previous section, they run into trouble when we try to consider the behaviour of light. Instead, Einstein showed in 1905 that there is a set of transforms between reference frames that avoid this logical contradiction. These are known as the **Lorentz transformations** – after Hendrik Lorentz, who first determined them (though with a somewhat different interpretation to our modern understanding). They are summarised as follows:

The Lorentz transformations

$$x' = \gamma(x - Vt) \tag{2.1}$$

$$y' = y \tag{2.2}$$

$$z' = z \tag{2.3}$$

$$t' = \gamma(t - Vx/c^2) \tag{2.4}$$

where the standard assumption is made that the x-coordinate axis is aligned with the direction of relative motion. The quantity γ, known as the **Lorentz factor**, depends on the relative speed of motion V, and is defined as

$$\gamma = \frac{1}{\sqrt{1 - V^2/c^2}} \tag{2.5}$$

- What are the two key differences between the Galilean and Lorentz transformations?

- Although the relation between x and x' has a similar form for the two transformations, in the case of the Lorentz transformations it includes the additional factor of γ, which depends on speed. The other difference is that the time coordinate undergoes a transformation in the Lorentz case and not the Galilean one.

The following exercise will help you to explore further how these two sets of transformations relate to the world around us.

Exercise 2.1

Calculate the Lorentz factor when the relative speed is: (a) an everyday speed of $20\,\mathrm{km\,h^{-1}}$; (b) 90% of the speed of light.

Optional Python extension:* you may find it interesting to write a short piece of Python code to plot how the Lorentz factor depends on speed, for values ranging from $0.1c$ to c.

*We are using the term 'Python' throughout this module to refer to the Python® programming language, as developed by the Python Software Foundation.

The exercise above shows that our intuitive Galilean view of relativity works reasonably well for speeds we experience in everyday life, because the Lorentz transformations reduce to the Galilean forms at low speeds. However, the large value of the Lorentz factor for objects travelling close to the speed of light leads to effects that disrupt our everyday expectations. In particular, you have now seen that measurements of time and spatial coordinates in different reference frames differ dramatically when the relative speed of the different frames becomes large.

2.1.3 Consequences of special relativity

By breaking our assumption that all observers must agree about the time intervals between events, the Lorentz transformations resolve the contradiction between the behaviour of ordinary objects (e.g. the train and the ball in Example 2.1) and that of light.

To better understand the consequences of the Lorentz transformations, it is necessary to consider how they apply to **intervals** of space and time. Consider two events, labelled 1 and 2, which occur in frame S at times t_1 and t_2, and at two separate locations, x_1 and x_2. The events are observed from two frames, S and S$'$, the relative speed of the latter being V in the x direction. In frame S, the time interval between the two events is $\Delta t = t_2 - t_1$, and the space interval, i.e. the difference between the locations of the two events, is $\Delta x = x_2 - x_1$.

We can use the Lorentz transformation equations to obtain expressions for $\Delta t'$ and $\Delta x'$, namely the intervals between the same two events as measured from frame S$'$. We can first write separate expressions for the coordinates of events 1 and 2 in S$'$ (noting that there is no change in the y- and z-coordinates, so they can be omitted):

$$x'_1 = \gamma(x_1 - Vt_1)$$
$$x'_2 = \gamma(x_2 - Vt_2)$$
$$t'_1 = \gamma(t_1 - Vx_1/c^2)$$
$$t'_2 = \gamma(t_2 - Vx_2/c^2)$$

Defining $\Delta t' = t'_2 - t'_1$ and $\Delta x' = x'_2 - x'_1$, we can write expressions for these two intervals in frame S$'$ by taking, in turn, the difference between the two x expressions and then the two t expressions:

$$\Delta x' = x'_2 - x'_1 = \gamma(x_2 - Vt_2) - \gamma(x_1 - Vt_1)$$
$$\Delta t' = \gamma(t_2 - Vx_2/c^2) - \gamma(t_1 - Vx_1/c^2)$$

These expressions can be simplified to give two equations for the transformation of intervals, in a form similar to the original transforms:

$$\Delta x' = \gamma(\Delta x - V\Delta t) \tag{2.6}$$
$$\Delta t' = \gamma(\Delta t - V\Delta x/c^2) \tag{2.7}$$

Exercise 2.2

Show that Equations 2.6 and 2.7 can be rearranged to find similar expressions for Δx and Δt, each only in terms of coordinates in the S' frame. Comment on how the resulting expressions compare to those for $\Delta x'$ and $\Delta t'$.

Now that we have expressions for how intervals of time and space transform between reference frames, we can investigate one of the most interesting and well-known consequences of special relativity – the idea that time is relative, and 'moving clocks run slow'. Example 2.2 explores this idea further.

Example 2.2

A short-lived particle is created in a particle physics experiment, and travels at very high speed relative to an observer in the lab. Its time and position of creation are t_1 and x_1, respectively. The particle is then seen by the same observer to decay at a later time, t_2, when it has moved (only in the x-direction) to a new location, x_2. The particle's lifetime, defined for a particle at rest, is known from theory to be $2.2 \times 10^{-6}\,\mathrm{s}$.

Find an expression for the time interval between particle creation and decay, as measured by the lab observer, assuming the particle is travelling at a constant speed V in the x-direction.

Solution

It is always important in special relativity to first think carefully about defining reference frames. The measured lifetime corresponds to a time interval in a reference frame in which the particle is at *rest*, and so we define reference frame S' as a frame that moves with the particle. Therefore, this reference frame moves at a constant speed of V with respect to the laboratory (which we define as the reference frame S of the observer).

The next step is to determine what information we have, and what information we need to work out. Because we only have two events to consider (particle creation and particle decay) there are only two intervals to evaluate in each reference frame, as summarised in Table 2.1.

Table 2.1 Information relating to particle observations in a lab.

Interval	Description	Expression	Value
Δx	distance travelled in S	$x_2 - x_1$	unknown
$\Delta x'$	distance travelled in S'	$x_2' - x_1'$	0
Δt	particle lifetime in S	$t_2 - t_1$	requested quantity
$\Delta t'$	particle lifetime in S'	$t_2' - t_1'$	2.2 μs

We need to find an expression for the particle lifetime observed in S, namely Δt. A crucial piece of information is that $\Delta x' = 0$: because the particle is not moving in its own frame, S′, the locations of the particle's creation and decay must be *the same*. This means that we can determine Δt if we have an expression that involves any of the intervals apart from the unknown Δx.

We can use the interval transforms from the solution to Exercise 2.2 to show that

$$\Delta t = \gamma(\Delta t' + V\Delta x'/c^2)$$

and because $\Delta x' = 0$ in this scenario, we find that

$$\Delta t = \gamma \Delta t' \qquad (2.8)$$

Because $\gamma > 1$ for all values of V (see Equation 2.5), the lab-based observer will always measure a longer lifetime for the particle than the lifetime of the particle at rest (i.e. that period measured in a reference frame in which the particle is stationary).

Example 2.2 provides an illustration of **time dilation**: for two events observed to occur at the same location in a reference frame S′, an observer in reference frame S will measure a longer time interval than elapses in frame S′.

- If there are observers in both frames S and S′, and each sees the *other* frame in relative motion, doesn't this lead to a logical contradiction in which each observer measures a longer time interval than the other?

- No – the situation is not symmetric: in one frame the two events occur at the same x-coordinate (e.g. in a frame moving with a relativistic particle, the particle's location does not change with time), while in the other frame the x-coordinate changes, meaning that setting up the reverse problem and applying the Lorentz transformations will give a different result.

It is important to recognise this asymmetry in how time dilation works: the dilation is measured by the observer who *does not* see the two events as being co-located.

Special relativistic effects such as time dilation (and a similar effect known as length contraction) seem odd and counter-intuitive. It is important to emphasise that these effects only become important for relative speeds that are a significant fraction of the speed of light. But there are many contexts – both in everyday life and in physics research – where they matter, and they have been repeatedly tested and confirmed. These contexts include particle physics experiments, global positioning system (GPS) technology, and a variety of well-studied astrophysical situations including jets from black holes and the explosions of massive stars.

Online resources: special relativity applications

We are only able to touch on the implications and applications of special relativity in this chapter. For more information and opportunities to put these ideas into practice, see the online resources for this chapter, which include further content on special relativity from the OU Stage 2 physics curriculum.

2.2 Spacetime and metrics

You have seen in the previous sections that space and time cannot be fully separated in physics – there is no *absolute* time that all observers can agree on, because it depends on your frame of reference. This idea led to the fundamental concept of spacetime, a four-dimensional union of space and time, introduced by Hermann Minkowski (a teacher of Einstein). The mathematical framework developed to describe spacetime makes it possible to set out physical laws that *can* be agreed on by all observers. In this section you will explore the geometry of spacetime and some of its implications for cosmology.

2.2.1 Spacetime diagrams

Some of the biggest challenges in cosmology come from trying to understand **causality** in the history of the Universe. How did the vast and varied Universe we observe today evolve from the very different conditions that observations show existed in the early Universe? As we noted in Chapter 1, a particularly interesting question arises from our observations of the CMB – how can it be so uniform when it originates from regions now vastly separated in space?

Minkowski introduced a helpful tool for visualising the possible connection (or lack of connection) between particular events. Figure 2.3 shows an example of a **spacetime diagram**, plotting one spatial dimension (x) against time. The vertical axis corresponds to time measured in an inertial reference frame S, and is plotted as ct. This approach gives the vertical axis the same units as distance x, plotted on the horizontal axis. A second inertial frame, S$'$ is represented by the blue, diagonal axes labelled as ct' and x'.

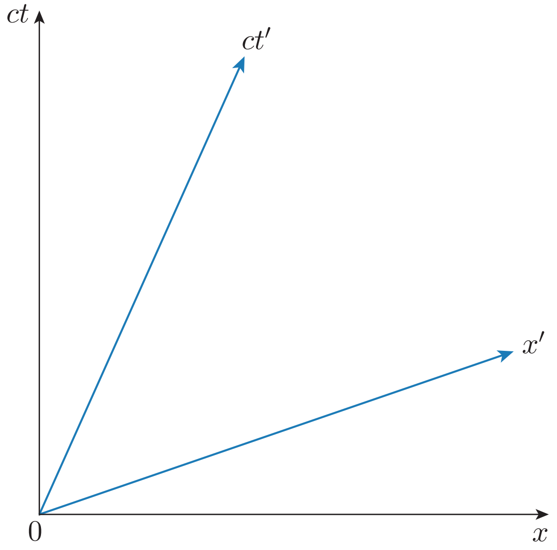

Figure 2.3 Example of a Minkowski spacetime diagram for an inertial reference frame S, with additional (blue) axes representing a second reference frame, S′.

Any fixed point on the diagram with coordinates (x,ct) corresponds to an event. But these coordinates refer only to that event in frame S. The next example explains the meaning of the diagonal axes shown in Figure 2.3.

Example 2.3

Consider an inertial frame S′ with velocity V relative to frame S. The x'-axis of a spacetime diagram consists of all the events for which $ct' = 0$, and conversely the ct'-axis will be a line describing where $x' = 0$. (There is a matching relationship between the x- and ct-axes in relation to frame S.) Use the Lorentz transformations to find mathematical expressions for the straight lines corresponding to the x'- and ct'-axes on a plot of x vs ct.

Solution

Because the spacetime diagram is drawn with x, ct as its primary axes, we are first looking for the relationship between x and ct that holds true for all events on the x'-axis (and then similarly for the events on the ct'-axis).

Starting from the Lorentz transformations (multiplying the t' expression by the constant c) we can deduce the following:

$$ct' = \gamma(ct - Vx/c)$$
$$x' = \gamma(x - Vt)$$

The question noted that the x'-axis is where $ct' = 0$. We can therefore set the left-hand side of the ct' transformation to zero to get an expression that must hold true for all events on the x'-axis:

$$0 = \gamma(ct - Vx/c)$$

Helpfully, this describes a relation between the S-coordinates of the events (x, ct), which is what we are looking for. Rearranging to the familiar form for a straight line gives:

$$ct = (V/c)x$$

In other words, the x'-axis on the diagram is a straight line that goes through the origin and has slope V/c.

The ct'-axis can be found similarly, by setting $x' = 0$ in the second Lorentz transformation equation above (for x') and multiplying by c, to derive $ct = (c/V)x$.

The time and position axes of the S′ frame therefore appear as diagonal lines with a slope that depends on the relative speed between the two frames, V.

- How does the appearance of the spacetime diagram alter if the relative speed of the frames, V, is increased?

- A higher value of V increases the slope of the x'-axis and decreases that of the ct'-axis, so that they move closer together. The equations for both of the S′-axes are shown in Figure 2.4.

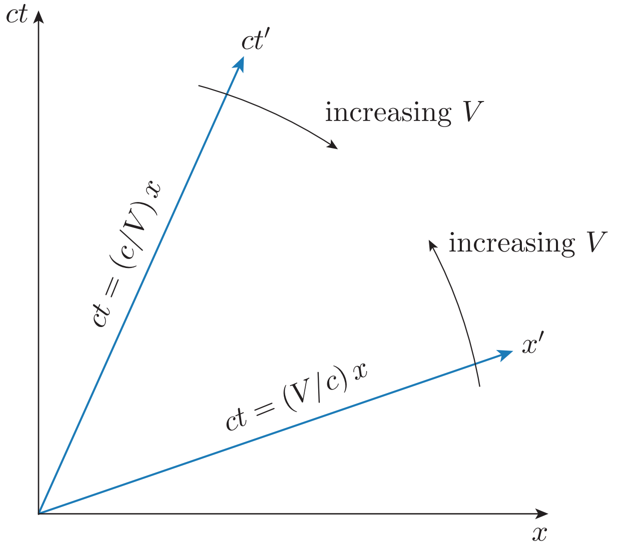

Figure 2.4 The effect on the S′-axes in a spacetime diagram of changing V.

Figure 2.5 shows how to read off the coordinates of an event from the diagram in both the S and S′ frames. Three events are shown, two of which lie on the S′ axes. Let's consider how to read off the coordinates of event 1 in each frame.

In frame S, event 1 has coordinates (x_1, ct_1), but its coordinates in frame S′ must be different. To read them off the diagram it is necessary to draw lines that intersect at event 1 and are parallel to each of the S′-axes; then the coordinates (x'_1, ct'_1) can be read from where these parallel lines intersect with the other primed axes, as shown in the diagram. The result is that (coincidentally) event 1 has the same x'-coordinate as event 3, and the same ct'-coordinate as event 2.

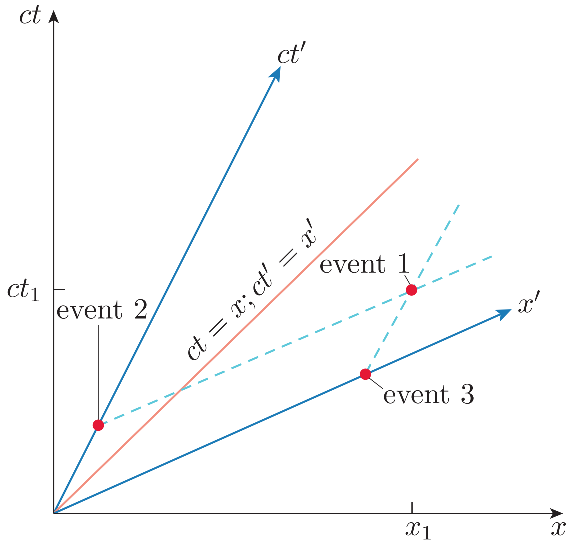

Figure 2.5 A spacetime diagram recording the coordinates of three events in the reference frames S and S′. A red line is also shown that plots $ct = x$ (which is equivalent to $ct' = x'$).

Another useful concept when using spacetime diagrams is the **world line**. A world line is the path that a particle (or another object; for example, a spacecraft) travels through spacetime. This too can be represented as a line on a spacetime diagram.

- What would the world line look like on a spacetime diagram for a particle measured in frame S′ to be travelling at $0.9999c$?
- A particle travelling very close to the speed of light will travel along a line of slope ~ 1 on the diagram – in other words, its world line is very close to the red line $ct = x$ (and $ct' = x'$) in Figure 2.5.

Figure 2.6 shows the world lines for two particles travelling between the same two locations in spacetime. One particle has travelled at a constant speed (as represented by path A) while the other has taken a less direct route with varying speeds (path B). You will see in later sections of this chapter that, while world lines for objects can take any route through spacetime that does not involve travelling faster than the speed of light, the shortest route between two locations plays an important role in defining the geometry of spacetime.

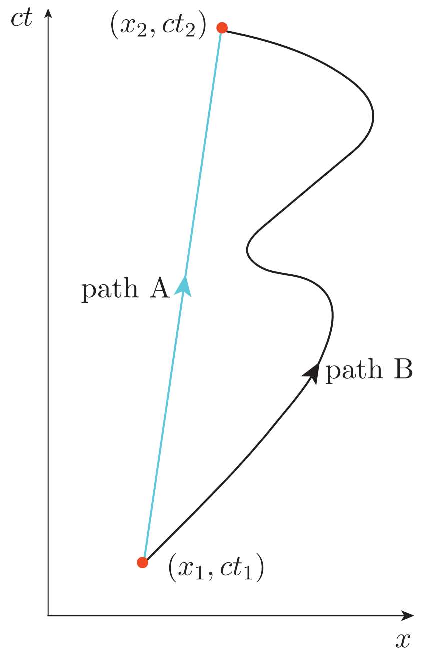

Figure 2.6 The world lines for two particles travelling between the same two points.

2.2.2 Causality and simultaneity

By visualising the relationship between events and how they are seen from different reference frames, spacetime diagrams make clear the significance of the finite speed of light. That the S′-axes converge around the line of $ct = x$ as relative speeds increase reflects the fact that the behaviour of light does not alter between the two reference frames. This has implications for causality and the connection between different events.

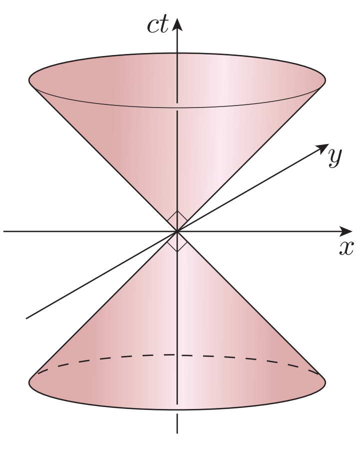

Figure 2.7 A light cone illustrating the region of a spacetime diagram that encompasses event locations that could be causally connected to the origin.

Figure 2.7 illustrates the concept of a **light cone**, via a spacetime diagram to which an extra spatial axis (y) has been added to make a three-dimensional representation. If we take the coordinate origin to be the location of a particular event, then the time axis extends both into the past (lower half of the plot, below the x-axis) and future (upper half) relative to that event. The shaded cone area encompasses all events that could be **causally connected** to the event at the origin.

- Consider two events, A $(x_\mathrm{A}, ct_\mathrm{A})$, and B $(x_\mathrm{B}, ct_\mathrm{B})$; for example, A could be an explosion, and B could be a window shattering. What condition needs to be met for A to have caused B?

- Event A could only be the cause of B if some type of 'information', such as a transfer of force, can travel between the two locations x_A and x_B within the time interval $\Delta t = t_\mathrm{B} - t_\mathrm{A}$. The required speed of information transfer, $\Delta x/\Delta t$, must be less than c, because information cannot travel faster than the speed of light.

The edges of a light cone have a slope of $c\Delta t/\Delta x = 1$, because world lines with slopes < 1 would correspond to motion at speeds faster than c. This means that there is no path by which information could travel between the origin of the spacetime diagram and a location outside the light cone unless it was travelling faster than c. Hence, if an event is within the light cone of event A, it may be causally connected to A; if not, then the two events cannot be connected.

- Is it possible for observers in different reference frames to disagree about whether events are causally connected?

- No. It would be impossible to develop logically consistent laws of physics if one event could cause another according to some observers, but not to others. This logical contradiction is avoided because all observers measure the speed of light as having the same value: the $ct = x$ line is the same in all frames, and therefore the light cone is too.

One reason that it is important to consider the question of causality in special relativity is because observers *can* disagree about the *order* of events in some situations, as the following example demonstrates.

Example 2.4

Figure 2.8 shows five events, A–E, as observed from two reference frames.

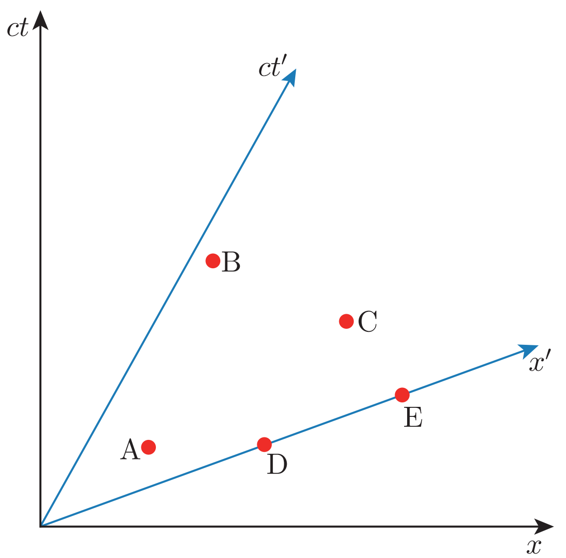

Figure 2.8 Five events plotted on a spacetime diagram representing two reference frames.

State the order in which the events occur according to: (a) an observer in frame S; (b) an observer in frame S′.

Solution

(a) In frame S, the events are observed to occur in a sequence defined by increasing values of ct: A = D (simultaneous), E, C, B.

(b) In frame S′, where the time coordinate can be read off in the direction perpendicular to the x'-axis, the order is: D = E (simultaneous), A, C, B.

In other words, the observers will disagree about which events are simultaneous, and the ordering of some events. This situation is referred to as the **relativity of simultaneity**.

How can Example 2.4 be reconciled with the idea that observers must agree about whether events are causally connected? In the example above, it may at first glance look as though an observer in frame S could think that event A causes event E, whereas an observer in S′ would think this cause and effect to be the other way round.

In fact, there is a straightforward solution. In Figure 2.8, a line connecting events A and E would have a very shallow slope, so any information travelling from A to E would need to cover a large x interval in a very small time interval – it would need to travel faster than the speed of light, so events A and E do not fall inside each other's respective light cones and so are not causally connected.

More generally, it turns out that situations in which observers disagree about the order of two events only *ever* arise where the two events are not contained within each other's light cone, so all observers will agree that one event cannot be the cause of the other (i.e. they will agree that it would require faster-than-light travel for information from one event to arrive at the location of the other before it happens). Therefore, provided information cannot travel faster than the speed of light, special relativity does not require us to abandon our sense of the reality of cause and effect.

2.2.3 Metrics in space and spacetime

In the chapter so far we have referred to 'distances' between events in a somewhat imprecise way. To extend our understanding of spacetime to encompass the behaviour of gravity and the expansion of the Universe, we need to define formally the spacetime separation between events. Our starting point is the concept of a **line element**: a small separation between coordinate locations. It is simplest to first consider the geometry of line elements in two and three dimensions, before coming back to four-dimensional spacetime.

Differential notation

In the previous sections we used notation such as Δx and Δt to indicate intervals in space and time. The mathematics of line elements (a form of differential geometry) is based on considering *infinitesimally* small coordinate intervals, and so from this point onwards the module materials will use the differential notation of 'd' to indicate such intervals (so $\mathrm{d}x$ and $\mathrm{d}t$ in the example of space and time).

The next example provides some practice in working with line elements.

Example 2.5

In two-dimensional space, the line element length $\mathrm{d}l$ shown in Figure 2.9 can be expressed in terms of the infinitesimal separations $\mathrm{d}x$ and $\mathrm{d}y$, such that

$$\mathrm{d}l^2 = \mathrm{d}x^2 + \mathrm{d}y^2 \tag{2.9}$$

By writing expressions for a coordinate location (x,y) in terms of the plane polar coordinates r and θ and then differentiating them, find an expression for $\mathrm{d}l^2$ in terms of $\mathrm{d}r$ and $\mathrm{d}\theta$.

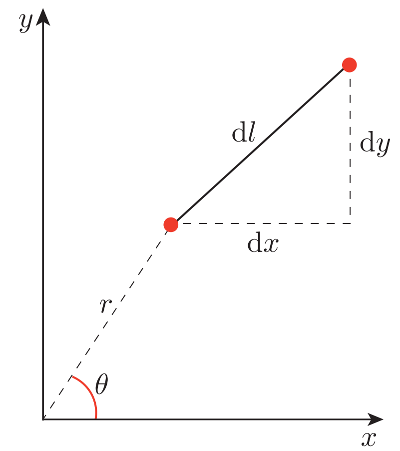

Figure 2.9 A line element $\mathrm{d}l$ in two-dimensional space, with x- and y-coordinate axes. The polar coordinates (r, θ), corresponding to one end of the line element, are also shown.

Solution

First, write x and y in polar coordinates as

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

Using the total derivative rule, we can find expressions for $\mathrm{d}x$ and $\mathrm{d}y$:

$$\mathrm{d}x = \mathrm{d}r\cos\theta - r\sin\theta\,\mathrm{d}\theta$$
$$\mathrm{d}y = \mathrm{d}r\sin\theta + r\cos\theta\,\mathrm{d}\theta$$

The total derivative rule says that for any function $f(r, \theta)$

$$\mathrm{d}f = \frac{\partial f}{\partial r}\,\mathrm{d}r + \frac{\partial f}{\partial \theta}\,\mathrm{d}\theta,$$

where r and θ could be any independent variables.

Squaring each expression and substituting into the original equation for $\mathrm{d}l^2$ (Equation 2.9) gives:

$$\mathrm{d}l^2 = \mathrm{d}r^2\cos^2\theta - 2r\cos\theta\sin\theta\,\mathrm{d}r\,\mathrm{d}\theta + r^2\sin^2\theta\,\mathrm{d}\theta^2$$
$$+ \mathrm{d}r^2\sin^2\theta + 2r\cos\theta\sin\theta\,\mathrm{d}r\,\mathrm{d}\theta + r^2\cos^2\theta\,\mathrm{d}\theta^2$$

Recalling that $\sin^2\theta + \cos^2\theta = 1$, and noting that the two terms with $\mathrm{d}r\,\mathrm{d}\theta$ in them cancel out, this simplifies to:

$$\mathrm{d}l^2 = \mathrm{d}r^2 + r^2\,\mathrm{d}\theta^2 \tag{2.10}$$

The idea of a line element can be extended to three and four dimensions, and is related to a definition of the **spacetime separation** between two events, Δs, as follows:

$$\Delta s^2 = (c\Delta t)^2 - \Delta x^2 - \Delta y^2 - \Delta z^2 \tag{2.11}$$

This definition looks similar to the equation for a simple three-dimensional (3D) distance, but with an additional term relating to how time intervals change, and with the spatial distances having negative sign. This negative sign ensures that Δs is invariant under Lorentz transformations.† All observers will agree about the spacetime separation between two events, which makes it a useful universal description of the geometry of spacetime.

†Mathematically, it doesn't matter whether the negative sign is applied to the time component or the spatial one; you may find examples outside the module material in which the signs of all of the terms in Equation 2.11 are reversed – this can be a source of confusion, so it is important to check which convention is being used.

When infinitesimal intervals in spacetime are considered, the spacetime separation becomes:

$$ds^2 = (c\,dt)^2 - dx^2 - dy^2 - dz^2 \tag{2.12}$$

Exercise 2.3

In the case where $dy = 0$ and $dz = 0$, show that the spacetime separation ds *is* invariant under Lorentz transformations (i.e. show that $ds = ds'$).

The expression for spacetime separation in Equation 2.12 is an example of a **metric** – a mathematical formulation of the relationship between coordinate intervals in a particular geometry. Equation 2.12 describes the **Minkowski metric**, which is the first of several metrics you will meet in this module. Metrics can also be written in the form of a summation:

$$ds^2 = \sum_{\mu,\nu=i}^{n} g_{\mu\nu}\,dx^\mu\,dx^\nu \tag{2.13}$$

where for 4D spacetime $i = 0$ and $n = 3$, and dx^0, dx^1, dx^2, dx^3 are the four-dimensional components of a vector $dx^\mu = (cdt, dx, dy, dz)$. This expression makes use of a tensor: a mathematical object that provides a concise way of manipulating multidimensional geometric relationships. The term $g_{\mu\nu}$ is known as the **metric tensor**,[‡] encapsulating the **metric coefficients** that apply to each coordinate. The metric tensor can be represented as a matrix with dimensions μ and ν. In the case of the Minkowski metric (Equation 2.12) this is written as:

$$g_{\mu\nu} = \begin{pmatrix} g_{00} & g_{01} & g_{02} & g_{03} \\ g_{10} & g_{11} & g_{12} & g_{13} \\ g_{20} & g_{21} & g_{22} & g_{23} \\ g_{30} & g_{31} & g_{32} & g_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{2.14}$$

The non-zero diagonal elements of the metric tensor (i.e. g_{00}, $g_{11}, \ldots$) are the coefficients of the coordinates for the ds^2 interval – compare them with the factors by which the coordinates $(c\,dt)^2, dx^2, dy^2$ and dz^2 are multiplied in Equation 2.12. The non-diagonal elements are all zero, because the metric doesn't include any terms involving (for example) $dx\,dy$.

The Minkowski metric, set out in two different forms by Equations 2.12 and 2.14, is a description of the geometry of space and time (in locally inertial frames). It enables calculations and measurements to be made about the properties and relations of lines, surfaces and volumes. Many other possible metrics exist. Note that it is conventional to use the index '0' for the time coordinate only, and so the summation in Equation 2.13 conventionally runs from 1 to n for an n-dimensional spatial metric.

[‡]Unfortunately, the tensor notation used in Equation 2.13 is easily confused with exponents when applied to individual elements – e.g. here dx^2 does not mean the square of dx, but the x-coordinate element with label '2'. A full introduction to tensors is beyond the scope of this module.

2.3 Curved space and spacetime

The ability to map the behaviour of space, or spacetime, using metric descriptions like those you were introduced to in the previous section is essential for astrophysics and cosmology. A metric helps to define the shortest distance between any two locations, which is important for understanding how light travels across the Universe.

Although the Minkowski metric applies to many useful situations, it is not the only metric we need to consider if we want to be able to tackle the subject of cosmology. The key insight of Einstein's theory of general relativity was a connection between gravity and geometry. To describe the Universe and its expansion (including the behaviour of light under the influence of gravity), it is necessary to extend our descriptions of geometry and metrics to encompass **curvature** of space and spacetime.

2.3.1 Flat and curved geometries

To introduce curved (i.e. **non-Euclidean**) geometries we start by taking a step back from the discussion of spacetime and consider further the geometry of two- and three-dimensional spaces. In the mathematical field of differential geometry, the type of smoothly varying spaces we will explore are referred to as **manifolds** – we will use both 'manifold' and 'geometric space' interchangeably.

In three dimensions, we have an intuitive idea of the difference between a flat and a curved surface. A piece of paper on a desk has a flat geometry. The surface of a balloon or a globe cannot be flattened in such a way as to lie flat on a desk and still retain the original distances between points and angles between lines.

● Does the surface of a cylinder have a flat or curved geometry?

○ The surface of a cylinder has a flat geometry. A cylindrical tube can be unrolled to form a flat surface without distorting the relationships between points on its surface. An intrinsically curved surface such as the surface of a sphere cannot be flattened in this way (e.g. consider how maps of the world become distorted in flat projections).

Whether a particular surface (manifold) is flat or curved can be determined mathematically. To explore this, we will consider the metric that describes the surface of a sphere. Spherical coordinates are commonly used in physics because they simplify the mathematics of problems that are spherically symmetric.

Figure 2.10 shows how the three-dimensional spherical coordinates r, θ and ϕ are defined. They are related to x, y and z (Cartesian coordinates) as follows: $x = r\sin\theta\cos\phi$; $y = r\sin\theta\sin\phi$; and $z = r\cos\theta$.

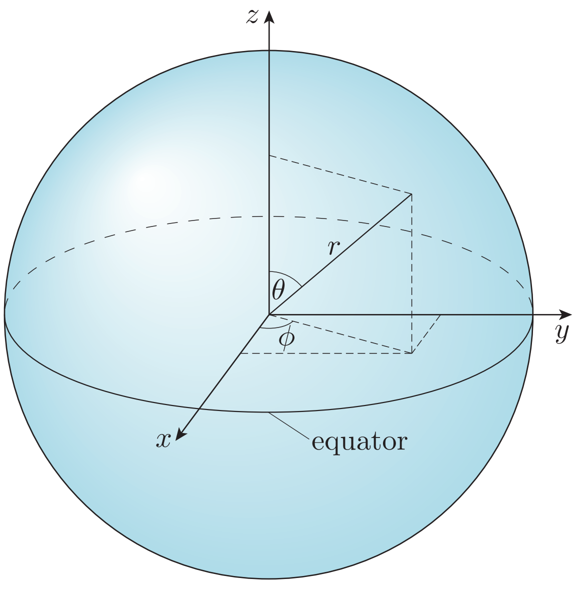

Figure 2.10 The 3D relation between spherical and Cartesian coordinates.

The length of a line element in three-dimensional space is given by $\mathrm{d}l^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2$. Using the relations above it is possible to show that this is equivalent to:

$$\mathrm{d}l^2 = \mathrm{d}r^2 + r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\,\mathrm{d}\phi^2 \tag{2.15}$$

The surface of a sphere corresponds to the situation $r = R$, where R is a constant. In other words, we consider locations on a surface to be at a fixed distance R from the origin, as shown in Figure 2.11.

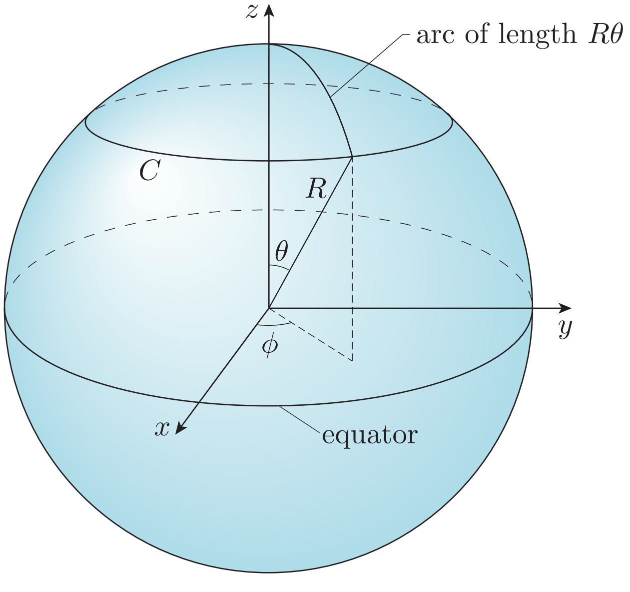

Figure 2.11 A circle C of constant θ coordinate drawn on the surface of a sphere.

In this geometry, $\mathrm{d}r = 0$ (because $r = R =$ constant), and so the metric for this geometry can be described by the line element:

$$\mathrm{d}l^2 = R^2\mathrm{d}\theta^2 + R^2\sin^2\theta\,\mathrm{d}\phi^2 \tag{2.16}$$

- How many dimensions does the space described by the metric in Equation 2.16 have?

- It describes a 2D space – the two variables θ and ϕ are the only dimensions. The third spatial dimension (the freedom to move in the r direction) does not exist in this geometry.

In the following example you will consider how the geometry described by Equation 2.16 affects the properties of circles.

Example 2.6

Consider the circle, C, drawn on the upper half of the surface of the sphere in Figure 2.11, which has its centre at the 'north pole'. Every point on the circle has the same value of θ. Use the concept of a line element and the metric of the spherical surface to derive an expression for the circumference of the circle, C.

Solution

Because θ is constant, $\mathrm{d}\theta = 0$, and so the line element that represents an infinitesimal distance on the sphere's surface reduces to:

$$\mathrm{d}l^2 = R^2\sin^2\theta\,\mathrm{d}\phi^2$$

The circumference of C is the sum of all of the infinitesimal distances around the circle, so is given by the integral of the line element summed over all values of ϕ (the angle around the circle), which (in radians) ranges from 0 to 2π. This gives:

$$C = \int_0^{2\pi} \mathrm{d}l = \int_0^{2\pi} R\sin\theta\,\mathrm{d}\phi = R\sin\theta[\phi]_0^{2\pi} = 2\pi R\sin\theta$$

- Considering the same circle C, defined on the surface of the sphere shown in Figure 2.11, what is its radius?

- Remembering that the circle is defined *on* the curved surface, with its centre at the 'north pole', the radius is the length of the arc shown in the figure, i.e. $R\theta$.

If the surface had the same flat geometry as a plane, we would expect the circumference to be $2\pi \times$ radius, i.e. $2\pi R\theta$, whereas the example found a circumference of $2\pi R\sin\theta$. So the circumference on the curved surface is *smaller* than that of a circle in flat space. This makes intuitive sense if we consider Figure 2.11 again, focusing now on the circle around the equator. Its circumference is simply $2\pi R$ (because $\theta = \pi/2$, and so $\sin\theta = 1$). But the circle's radius, as drawn on the surface, is the arc from the 'north pole' to the equator, which is clearly much longer than R.

Hopefully, you are now persuaded that geometrical relationships differ in curved geometries. The reason for this is that the geometry determines the shortest distance between two points (known as a **geodesic**). In flat space the geodesic is always a straight line, but when movement is confined to a curved surface, the shortest route between two points on that surface always has to account for its curvature.

In the case of the surface of a sphere, shortest paths are always part of a **great circle**: a circle defined by the intersection of the surface with a plane that passes through the sphere's centre; for example, the Earth's Equator. Another example you may have come across is great-circle aeroplane flight paths, which trace the shortest route between two points on the Earth's surface, but can seem surprising when shown on a map that is a flat projection of the globe. (If you have studied observational astronomy previously, you will have encountered similar ideas in the context of astronomical coordinate systems defined on the celestial sphere.) The next exercise explores an example of distances in curved geometries.

Exercise 2.4

The cities of Calgary and Brussels are both at a latitude of 51° north, and are separated in longitude by 118°. The shortest (great-circle) distance between the two cities is $\sim 7300\,\mathrm{km}$, which involves a route that passes over Greenland, Iceland and the very north of Scotland. Calculate the distance between the two cities if, instead, a route that follows the 51° latitude line is taken (which would pass through eastern Canada and Southern England), and comment on how the two distances compare.

Figure 2.12 illustrates three ways in which curved 2D spaces – such as those of a spherical or a saddle-shaped surface – behave differently to flat ones. These differences concern: the relation between the circumference and the radius of circles (labelled C and b respectively in the diagram); the geometry of triangles; the behaviour of initially parallel lines. The figure also shows that it is possible to identify two different directions in which curved geometry can deviate from that of flat space – non-planar surfaces can have **positive curvature** or **negative curvature**, typically parameterised by a **curvature parameter**, k:

- In positively curved 2D spaces ($k > 0$), the interior angles of triangles sum to *greater than* 180°, circumferences of circles are smaller, and lines of shortest distance (geodesics) that are initially parallel may *converge*.
- In negatively curved 2D spaces ($k < 0$), the interior angles of triangles sum to *less than* 180°, circumferences of circles are larger, and initially parallel geodesics may *diverge*.

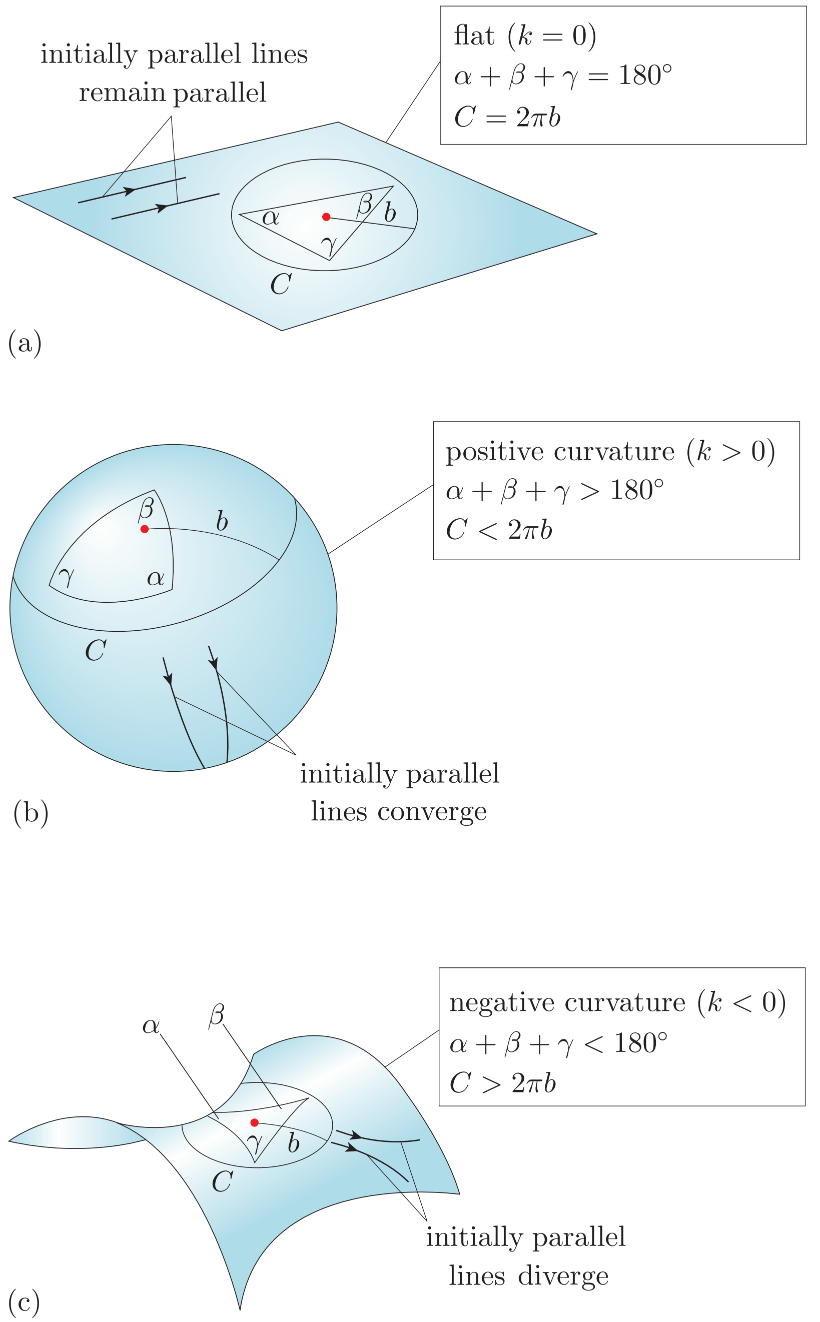

Figure 2.12 Differences in geometry of 2D spaces: (a) a flat space; (b) a space with positive curvature; (c) a space with negative curvature.

Similar deviations from Euclidean geometry occur for 3D and 4D curved spaces, and have important implications for the behaviour of light, and so for our understanding of spacetime in cosmology.

2.3.2 Defining and measuring curvature

How can curvature be measured? First consider a curved line, as shown in Figure 2.13.

- Use the geometric information in Figure 2.13 to determine whether the line is most curved at point A, B or C. Briefly explain your reasoning.
- The line is more curved at A than at B or C. At point A, the line deviates the most over a short distance from a straight tangent line, whereas at C the curve remains closest over a large distance to a straight tangent line.

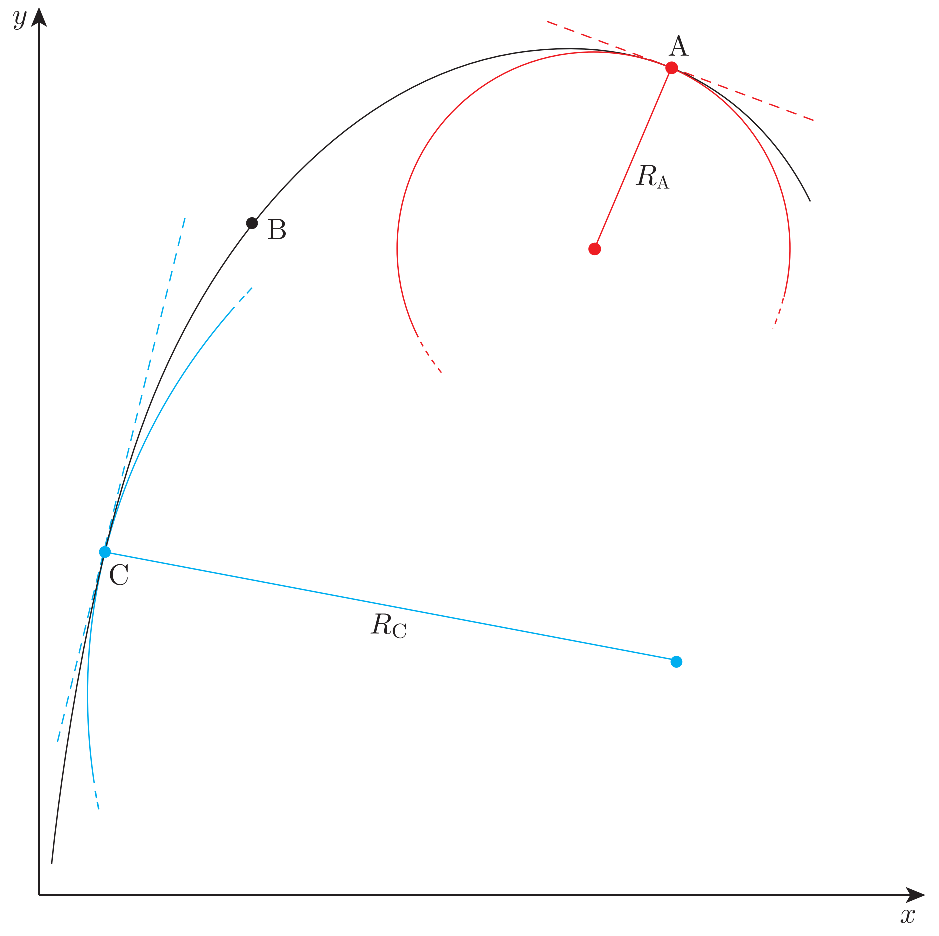

Figure 2.13 A line with varying curvature, marked with points A, B and C.

Figure 2.13 also shows that a line segment that is *more* curved can be approximated at a particular location by a segment of a circle of *smaller* radius, whereas a less curved line requires a circle of much larger radius to approximate its shape (compare the lengths of lines R_A and R_C). This leads to the curvature k_x, at a point x, being defined as the inverse of the radius of a circle that best matches the curve at that location:

$$k_x = \frac{1}{R_x} \tag{2.17}$$

Exercise 2.5

(a) What shape is described by a line of constant curvature $k_x = 0.2\,\mathrm{cm}^{-1}$?

(b) What is the curvature k_x of a straight line (measured at any point, x, along it)?

For cosmology, the curvatures of both three-dimensional spaces and four-dimensional spacetime are relevant. These properties influence the paths taken by light and the distances between objects such as galaxies as the Universe expands. We therefore need to define curvature in a way that can be applied to a variety of geometries in multiple dimensions.

Figure 2.13 demonstrates that curvature is connected in some way to the behaviour of tangent lines. Where the curvature is high, the slopes of the tangent lines change rapidly as small steps are taken along the curve, whereas for low curvature the slopes do not change much as you move along the curve. In a 2D geometry, the slope is a first derivative of the coordinates, e.g. $\mathrm{d}y/\mathrm{d}x$, and so the rate at which the slope changes is a second derivative, e.g. $\mathrm{d}^2y/\mathrm{d}x^2$.

Derivatives are quantities that connect how different coordinates change in relation to each other, and they can be used to construct a definition of curvature that is *intrinsic* to the geometry being measured. This means that the curvature can be determined without making measurements in a higher-dimensional space (in contrast to what is done in Figure 2.13, where the 1D curve is investigated with 2D (x,y) coordinates).

- What is the mathematical entity that describes the intrinsic geometry of a particular space (manifold)?
- The intrinsic geometry of a manifold is described by its metric.

Therefore, metric coefficients and their derivatives provide us with a way to determine intrinsic curvature. In a two-dimensional space, we can write the following (non-intuitive!) expression for intrinsic curvature, $\mathcal{K}$, in terms of metric coefficients and their derivatives:§

$$\begin{aligned}\mathcal{K} = {} & \frac{1}{4g_{11}g_{22}}\left[\frac{1}{g_{11}}\frac{\mathrm{d}g_{11}}{\mathrm{d}x^1}\frac{\mathrm{d}g_{22}}{\mathrm{d}x^1} + \frac{1}{g_{22}}\frac{\mathrm{d}g_{11}}{\mathrm{d}x^2}\frac{\mathrm{d}g_{22}}{\mathrm{d}x^2} + \frac{1}{g_{11}}\left(\frac{\mathrm{d}g_{11}}{\mathrm{d}x^2}\right)^2 + \frac{1}{g_{22}}\left(\frac{\mathrm{d}g_{22}}{\mathrm{d}x^1}\right)^2\right] \\ & - \frac{1}{2g_{11}g_{22}}\left[\frac{\mathrm{d}^2g_{22}}{(\mathrm{d}x^1)^2} + \frac{\mathrm{d}^2g_{11}}{(\mathrm{d}x^2)^2}\right]\end{aligned} \tag{2.18}$$

where $g_{\mu\nu}$ refers to the metric elements as defined in Equation 2.13 and its '1' and '2' labels refer to coordinates (e.g. $x^1 = x$ and $x^2 = y$ for a 2D Cartesian geometry).

§Note again the potential source of confusion that in this context '$\mathrm{d}x^1$' and '$\mathrm{d}x^2$' indicate derivatives with respect to co-ordinates x^1 and x^2, but the superscript '2's in the final two terms indicate second-order derivatives.

For a subset of 2D metrics in which g_{11} is constant, and g_{22} depends only on x^1, Equation 2.18 reduces to a simpler form:

$$\mathcal{K} = \frac{1}{4g_{11}(g_{22})^2}\left(\frac{\mathrm{d}g_{22}}{\mathrm{d}x^1}\right)^2 - \frac{1}{2g_{11}g_{22}}\left[\frac{\mathrm{d}^2 g_{22}}{(\mathrm{d}x^1)^2}\right] \tag{2.19}$$

- What are g_{11} and g_{22} for a 2D plane polar geometry?

- Referring back to Example 2.5, $g_{11} = 1$ and is the metric coefficient for r (it is the multiplier of the $\mathrm{d}r^2$ term in the line element form of the metric, Equation 2.10), and $g_{22} = r^2$ and is the metric coefficient for θ (it is the multiplier of the $\mathrm{d}\theta^2$ term in the same form of the metric).

Example 2.7

Find an expression for $\mathcal{K}(\theta, \phi)$ for the surface of a sphere, and show that the curvature does not depend on the coordinate location (θ, ϕ) – i.e. that the curvature of a spherical surface is the same at every location.

Solution

The metric for the surface of a sphere is given by Equation 2.16:

$$\mathrm{d}l^2 = R^2\mathrm{d}\theta^2 + R^2\sin^2\theta\,\mathrm{d}\phi^2$$

To calculate $\mathcal{K}$ (using Equation 2.19) we need the two metric coefficients, g_{11} and g_{22} (where the '1' and '2' labels refer to θ and ϕ, respectively). From Equation 2.16, $g_{11} = R^2$ and $g_{22} = R^2\sin^2\theta$.

To evaluate the equation, we need to calculate the first and second derivatives of g_{22} with respect to x^1, i.e. θ:

$$\frac{\mathrm{d}g_{22}}{\mathrm{d}x^1} = \frac{\mathrm{d}g_{22}}{\mathrm{d}\theta} = 2R^2\sin\theta\cos\theta$$

and so

$$\frac{\mathrm{d}^2 g_{22}}{(\mathrm{d}x^1)^2} = 2R^2\left(-\sin^2\theta + \cos^2\theta\right) = 2R^2\left(\cos^2\theta - \sin^2\theta\right)$$

Because Equation 2.19 is algebraically messy, we will first work out each of its two terms individually. Substituting the first derivative of g_{22} into the first term in the equation gives:

$$\frac{1}{4g_{11}\left(g_{22}\right)^2}\left(\frac{\mathrm{d}g_{22}}{\mathrm{d}x^1}\right)^2 = \frac{1}{4R^2\left(R^2\sin^2\theta\right)^2}\left(2R^2\sin\theta\cos\theta\right)^2$$

Doing the same for the second term, dependent on the second derivative of g_{22}, gives:

$$\frac{1}{2g_{11}g_{22}}\left[\frac{\mathrm{d}^2 g_{22}}{(\mathrm{d}x^1)^2}\right] = \frac{1}{2R^2R^2\sin^2\theta}\left[2R^2\left(\cos^2\theta - \sin^2\theta\right)\right]$$

We can now use the two expressions above to find an expression for $\mathcal{K}$:

$$\mathcal{K}(\theta,\phi) = \frac{1}{4R^2\left(R^2\sin^2\theta\right)^2}\left(2R^2\sin\theta\cos\theta\right)^2 - \frac{1}{2R^2R^2\sin^2\theta}\left[2R^2\left(\cos^2\theta-\sin^2\theta\right)\right]$$

which simplifies to

$$\mathcal{K}(\theta,\phi) = \frac{\cos^2\theta}{R^2\sin^2\theta} - \left(\frac{\cos^2\theta-\sin^2\theta}{R^2\sin^2\theta}\right) = \frac{\cos^2\theta}{R^2\sin^2\theta} - \frac{\cos^2\theta}{R^2\sin^2\theta} + \frac{\sin^2\theta}{R^2\sin^2\theta}$$

The first two terms cancel out, and the final one simplifies further to give

$$\mathcal{K} = \frac{1}{R^2}$$

You can see that – in the end, after much algebra – the curvature of the spherical surface does not depend on the location (θ, ϕ position). For the surface of a sphere of particular radius R, the curvature is the same everywhere, and depends only on the sphere's radius.

Example 2.7 was intended to give you a flavour of the mathematical and physical meaning of spatial curvature. Curvature becomes even more complex to calculate for higher-dimensional spaces. $\mathcal{K}$ is a simplified form of the **Riemann curvature tensor**, $R^{\lambda}_{\mu\rho\nu}$, sometimes just referred to as the Riemann tensor. The Riemann tensor is a set of equations analogous to Equation 2.18 that encapsulates the curvature at any location in a multidimensional space.

Online resources: Riemannian geometry

The mathematical toolkit that underpins the theory of general relativity is Riemannian geometry. Our primary aim in this module is to develop and apply cosmological models, and so – although it is a fascinating topic – introducing the wealth of definitions and terminology required for a full discussion of Riemannian geometry is beyond our scope. The online resources for this chapter provide some additional information about the subject for interest only.

The crucial point to take away, and the reason we have shown you the somewhat complicated form for the two-dimensional case in Equation 2.18, is that, like $\mathcal{K}$, the curvature tensor is constructed entirely from *derivatives of the metric coefficients*. You will not be asked to calculate or manipulate it directly.

The essential concepts to remember from this section are summarised in the following box.

Curvature of geometric spaces

- Curvature affects the shortest distance between two points in space.
- Geometrical relationships, such as the properties of circles and triangles, differ in flat and curved spaces.
- Curvature is an *intrinsic* property of a geometric space, which can be determined at a given location in space if the metric is known.
- Intrinsic curvature is determined from a combination of derivatives of the metric coefficients, given for 2D geometries by Equation 2.18.
- The Riemann curvature tensor is a more complex combination of derivatives of metric coefficients. It encapsulates the curvature at any location of a manifold of arbitrary dimensions (and so can be used to quantify curvature in the 3D and 4D situations relevant for general relativity and cosmology).

So far, we have considered only spatial dimensions, but you have seen that special relativity requires us to work with four-dimensional spacetime, in which spatial and time dimensions can become intertwined. In the next section we consider how the concepts of curvature explored here can be applied to 4D spacetime.

2.3.3 Curved spacetime and geodesics

Curvature is an important concept in four-dimensional spacetime as well as in the spatial geometries discussed so far. Although, formally, spacetime metrics are termed 'pseudo-Riemannian' (a distinction not important for this module), the Riemann tensor and the arguments about curved geometries in the previous section can be applied to spacetime metrics as well as to purely spatial geometries.

● Curvature depends on combinations of the derivatives of metric components with respect to different coordinates. Considering the Minkowski metric (Equation 2.12), what basic statement must be true about any derivatives of its components?

○ All of the components of the Minkowski metric are constants (either 1 or -1), and so their derivatives with respect to another coordinate must be zero.

Because all of the partial derivatives of the components of its metric must be zero, we can conclude that Minkowski spacetime has zero curvature: it is a flat geometry. However, it is possible to construct spacetime geometries that are curved, and you will see in the next chapter that such geometries are a consequence of the theory of general relativity.

Another concept from this chapter that is important to general relativity is the geodesic: the path of shortest distance between two points for a particular metric (where distance is the spacetime invariant distance $\mathrm{d}s$).

Geodesics have the form of straight lines in flat (Minkowski) spacetime, but take different forms in curved geometries, such as the great circles around the Earth's surface. Geodesics can also be divided into three categories, depending on the nature of the spacetime separation (or metric interval) between the two points considered.

- **Null geodesics** are curves for which $\mathrm{d}s^2 = 0$ for every interval along the curve. Only particles travelling at the speed of light (i.e. photons) can travel along null geodesics, which are also known as 'light-like' geodesics.
- **Time-like geodesics** are curves where $\mathrm{d}s^2 > 0$ for every interval along the curve. The world lines of particles (moving at $v < c$) must follow time-like paths.
- **Space-like geodesics** are curves where $\mathrm{d}s^2 < 0$ for every interval along the curve. These correspond to paths between events that are outside one another's light cones – particles (including photons) cannot travel along a space-like path.

● Can observers logically disagree about the order of two events that are connected by a geodesic that is: (i) null, (ii) time-like, and (iii) space-like?

○ Observers must agree about the order of events in situations where one event could be the physical cause of another. Therefore, two events that can be connected by a null or a time-like geodesic must have a fixed order for all observers. Observers may disagree, however, about the order of events for which the separation is space-like because there can be no causal relationship between them if neither event is inside the other's light cone.

The world lines along which particles travel are always time-like. The earlier discussion of special relativity implies that – whether in flat or curved spacetime – observers in different reference frames may disagree about how quickly time elapses along those paths, i.e. how changes in the t-coordinate along a path relate to changes in the other coordinates. In the next section you will meet a definition of time that all observers in a region described by the same metric can agree on – a concept that will be essential for building geometric models of the Universe.

2.3.4 Proper time

Metrics are defined such that all observers can agree on the invariant spacetime interval between two events, $\mathrm{d}s$. An invariant time interval, called the **proper time**, $\mathrm{d}\tau$, can be defined in relation to the spacetime interval, $\mathrm{d}s$, in a very simple way:

$$\mathrm{d}\tau = \mathrm{d}s/c \tag{2.20}$$

Because $\mathrm{d}s$ and c are both the same for all observers, $\mathrm{d}\tau$ is a measure of time that everyone can agree on.

It turns out that proper time has another useful property, shown in the following example.

Example 2.8

Consider a particle travelling along a time-like geodesic in Minkowski (i.e. flat) space. An observer who is not travelling with the particle sees it move a very short distance between two locations that have coordinates of $(x_{\rm A}, y_{\rm A}, z_{\rm A})$ and $(x_{\rm B}, y_{\rm B}, z_{\rm B})$ in the observer's reference frame.

Write down expressions for the proper time interval that elapses when the particle travels between these two locations, both in the reference frame of the observer and in the reference frame of the particle. Comment on how proper time relates to how time is perceived in the reference frame of a moving object.

Solution

We start by writing down expressions for the interval $\mathrm{d}s$ in the two reference frames. Firstly, for that of the observer:

$$\mathrm{d}s^2 = c^2\,\mathrm{d}t^2 - (x_{\rm B} - x_{\rm A})^2 - (y_{\rm B} - y_{\rm A})^2 - (z_{\rm B} - z_{\rm A})^2$$

where we have used Equation 2.12 (the Minkowski metric), and substituted the individual coordinate separations between events into the spatial separation terms.

For the reference frame of the particle we will use primed notation to denote the different reference frame:

$$\mathrm{d}s'^2 = c^2\,\mathrm{d}t'^2 - \mathrm{d}x'^2 - \mathrm{d}y'^2 - \mathrm{d}z'^2$$

We can immediately simplify this expression hugely because, in the reference frame attached to the moving particle, its spatial coordinates do not change at all: it is at rest. So

$$\mathrm{d}s'^2 = c^2\,\mathrm{d}t'^2$$

If we now apply Equation 2.20 to the two expressions, we find that

$$\mathrm{d}\tau^2 = \mathrm{d}t^2 - [(x_{\rm B} - x_{\rm A})^2 - (y_{\rm B} - y_{\rm A})^2 - (z_{\rm B} - z_{\rm A})^2]/c^2$$

and

$$\mathrm{d}\tau'^2 = \mathrm{d}t'^2$$

The spacetime separation is the same for both frames (i.e. $\mathrm{d}s = \mathrm{d}s'$), and so from the definition of proper time we can conclude that $\mathrm{d}\tau = \mathrm{d}\tau' = \mathrm{d}t'$.

In other words, for a time-like world line, the proper time interval between two events on that world line is equivalent to the coordinate time interval $\mathrm{d}t$ measured in the frame at rest in relation to those events.

The example provides a very useful result, which can be generalised to apply to curved as well as flat spacetimes.

Proper time and coordinate time

For events on the world line of a moving observer – meaning that they all occur at the same, unchanging spatial coordinate location in the observer's frame – the proper time corresponds to the coordinate time that would be measured by a clock travelling with the observer.

For an observer in a different reference frame, the proper time between events does not correspond to the coordinate time, but can be determined from the metric based on the spatial coordinates of those events.

The concept of proper time, and its relation to coordinate time, will be important for building and interpreting metric descriptions of curved and flat spacetime in different regions of the Universe. This is the topic of the next chapter.

2.4 Summary of Chapter 2

- Measurements of intervals in space and time, and related quantities such as velocity, depend on the **reference frame** of the observer who is measuring them.
- The theory of special relativity enables a consistent framework for applying the laws of physics in all **inertial frames**, in which the **Lorentz transformations** can be used to relate measurements of distance and time **intervals** in different frames:

$$\Delta x' = \gamma(\Delta x - V\Delta t)$$
$$\Delta y' = \Delta y$$
$$\Delta z' = \Delta z$$
$$\Delta t' = \gamma(\Delta t - V\Delta x/c^2)$$

  where the **Lorentz factor**

$$\gamma = \frac{1}{\sqrt{1 - V^2/c^2}} \qquad \text{(Eqn 2.5)}$$

- Where two reference frames are moving at different speeds from one another, the time interval between two **events** observed to occur at the same location in one frame will be measured as different in the other frame, and this **time dilation** is an example of the consequences of special relativity.
- The relationship between events and the **world lines** representing the paths of moving objects can be investigated using **spacetime diagrams**.
- **Light cones** provide a way of conceptualising which events can be **causally connected** to each other. If two events A and B take place, then event A cannot have been caused by event B if B occurs at a location outside A's light cone.

- **Metrics** define the geometric relationship between distance and time intervals, with the **Minkowski metric** defining a **spacetime separation** on which all observers can agree:

$$\mathrm{d}s^2 = (c\,\mathrm{d}t)^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2 \qquad \text{(Eqn 2.12)}$$

- More generally, the **metric coefficients** of a geometric space with n dimensions can be defined by

$$\mathrm{d}s^2 = \sum_{\mu,\nu=i}^{n} g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu \qquad \text{(Eqn 2.13)}$$

where $g_{\mu\nu}$ is the **metric tensor** and x^μ are the coordinates for each dimension, which are conventionally labelled from 1 to n for an n-dimensional space, and from 0 to 3 for 4-dimensional spacetime.

- **Curvature** of space is an intrinsic property of a particular metric tensor (or just 'metric'). It affects the paths of shortest distance between points, as well as geometric relationships such as the properties of circles and triangles.
- The curvature of a 2D surface where g_{11} is constant and g_{22} depends only on x^1 can be calculated from the metric coefficients according to:

$$\mathcal{K} = \frac{1}{4g_{11}(g_{22})^2}\left(\frac{\mathrm{d}g_{22}}{\mathrm{d}x^1}\right)^2 - \frac{1}{2g_{11}g_{22}}\left[\frac{\mathrm{d}^2 g_{22}}{(\mathrm{d}x^1)^2}\right] \qquad \text{(Eqn 2.19)}$$

- The **Riemann curvature tensor** is a more complex combination of derivatives of the metric. It provides a rigorous definition of curvature that can be applied to any multidimensional **manifold**.
- A **geodesic** is the path of shortest distance between two points for a particular metric. Geodesics can be classed as **null** ($\mathrm{d}s^2 = 0$), **time-like** ($\mathrm{d}s^2 > 0$) or **space-like** ($\mathrm{d}s^2 < 0$).
- **Proper time** is an invariant measure of time intervals between events. It is defined as

$$\mathrm{d}\tau = \mathrm{d}s/c \qquad \text{(Eqn 2.20)}$$

and corresponds to the coordinate time interval measured by an observer who is co-located with both events.