The Geometry of Spacetime

Author: J Oliver Linton

Address: Pentlands, Keasdale Road, Milnthorpe, LA7 7LH

Email: jolinton@btinternet.com

Abstract: Gravity is explained these day in terms of the 'warping of spacetime' but it not easy to visualise exactly what this means. A simple argument involving geodesics in a metric space is used to explain how an inertial particle will move in a spacetime whose metric varies from place to place. This leads to a simple derivation of the formula for the time dilation factor at the surface of a star.

There can be few people these days who are unaware of the fact that Newton's view of gravity as a force acting instantaneously across space has been replaced by a completely different explanation which involves the 'warping' of something called 'spacetime'. If you have ever watched television programmes or visited a science museum you will probably be familiar with the following image (*Fig*. 1) which purports to show how the distortion of space can cause a moon to orbit a planet.

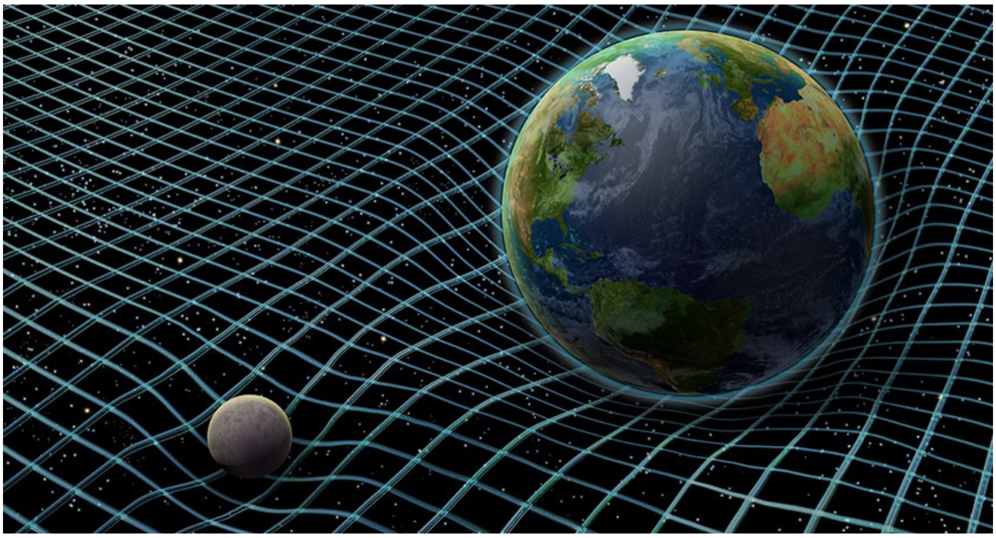

Fig. 1: The 'warping' of 'space-time' around a massive object

Space is imagined to be a rubber sheet which is stretched and pulled out of shape by the massive object representing the planet placed on it and this distortion causes a ball bearing projected on to its surface to follow a pleasingly elliptical orbit round the planet, just like a moon.

The problem with this image is not that the rubber sheet 'space' is two-dimensional – it is not too difficult to imagine a block of rubber which is somehow distorted in three dimensions by an internal tension – nor do we have to worry too much about the fact that it is gravity (the very thing we are trying to explain) which causes the rubber sheet to distort in the first place. The real problem lies with the fact that massive objects do not distort *space* at all. They distort *spacetime*. The difference is crucial, but a proper understanding of the difference between the four dimensional spacetime of Einstein and Minkowsky and Newton's concept of a three dimensional space moving in time is not easy to acquire. It is the purpose of this article to help sort out this mystery.

Let us start with a more easily visualizable problem. What is the distance between London and San Francisco?

Geodesics

Google maps gives the answer as 5360 miles and the shortest route passes over Greenland. The reason for this becomes immediately clear if you stretch a piece of string over a globe. The string automatically finds the shortest route and it is easy to see that this route forms part of a circle whose centre is at the centre of the Earth. Such a circle is called a Great Circle and the route is known as a geodesic. (*Fig*. 2)

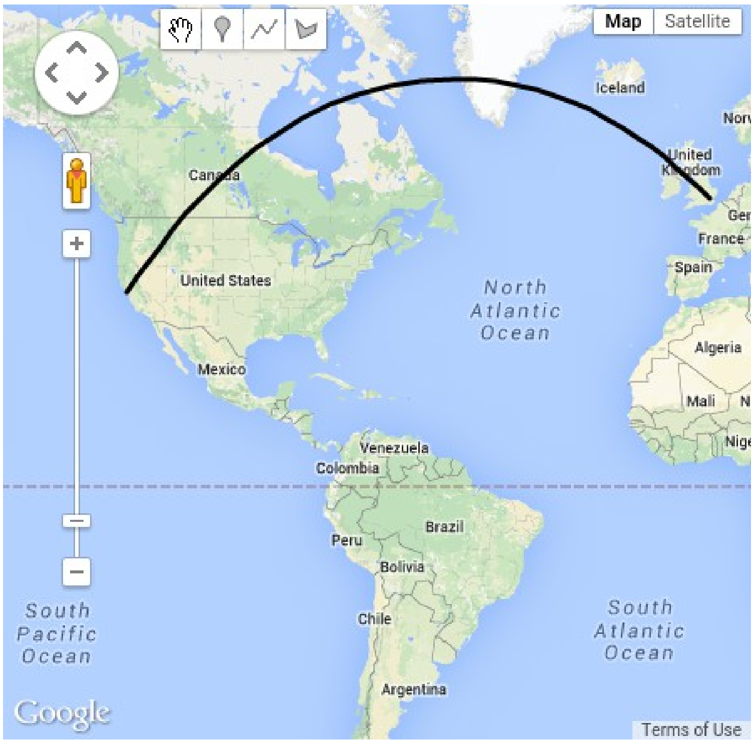

Fig. 2: The Great Circle route from London to San Francisco

In spite of its etymology, geodesics are not confined to spherical objects like the Earth but can be generalised to refer to the shortest route through any kind of terrain. Suppose, for example, you want to walk the shortest possible route from A to B in the mountainous country shown in *Fig*. 3. Pond Mountain stands in your way and it is obvious (quite apart from the extra effort involved) that the direct (red) route over the top of the mountain is going to be longer than one or other of the two blue routes, both of which can be regarded as geodesics (the important point being that any *small deviations* from either route will always result in a slightly *longer* distance travelled).

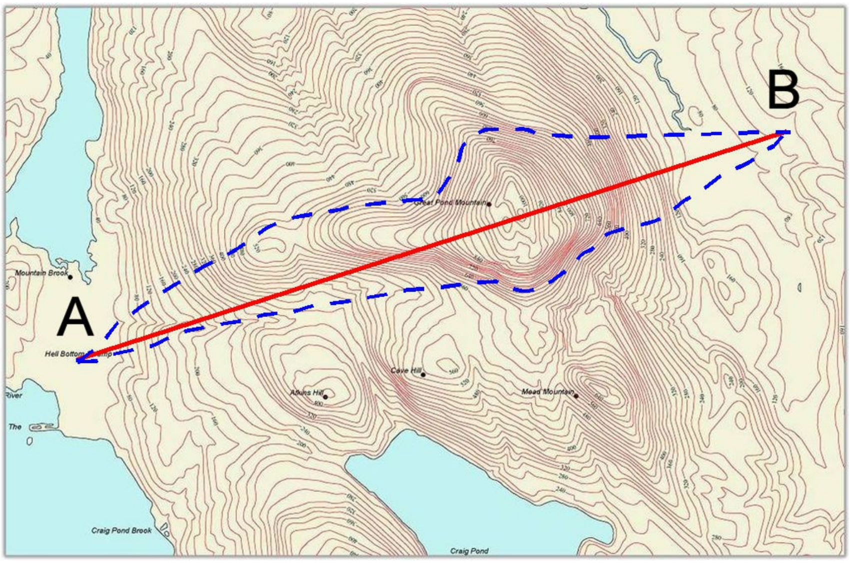

Fig. 3: Two possible geodesic routes across mountainous country

It is possible to extend the idea to flat terrain with different characteristics. There is a famous problem involving the best path for a life saver to take across a beach to a swimmer in trouble off shore. If the lifeguard can run at 12 mph but only swim at 3 mph, his best route looks something like the route depicted in *Fig*. 4:

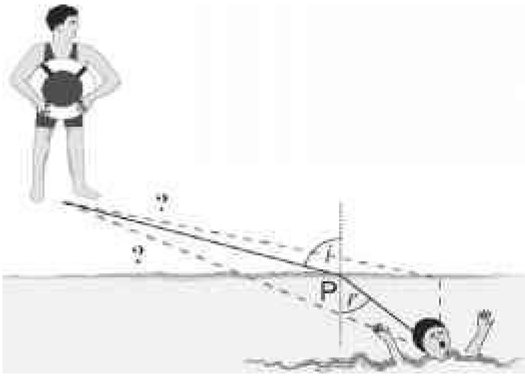

Fig. 4: The lifeguard problem

You might even recall a similar problem in optics to do with the way a ray of light enters glass. Amazingly the problem turns out to be identical and it is easily shown that rays of light always follow lines which take the shortest time. How they 'know' which way to go is a bit of a mystery but they do. Newton's First Law of Motion – the one that states that objects will continue at rest or travel in a straight line unless acted on by an external force – can also be stated in the language of geodesics. Left to themselves, projectiles follow a geodesic path.

The geometry of (flat) two dimensional space

If you like hiking you will know that any point on an OS map can be specified (to a precision of 100 m) by means of a six-figure grid reference. For example, Scafell Pike has a grid reference of NY215072. What this actually means is that the summit is 21.5 km East of the origin of the map NY and 7.2 km North. (The origin of the NY map is actually a point a couple of miles off Seascale in the Irish sea) The coordinates of Scafell Pike can therefore be written as (21.0, 7.2).

The grid reference of the Old Man of Coniston is SD272978 but since the origin of the SD map is 10 km south of the NY map, we must subtract 10 km from the northing. This means that the coordinates of the Old Man are (with reference to the NY origin) (27.2, -2.2).

Now let us calculate the (straight line) distance between Scafell Pike and the Old Man of Coiniston. To do this we use Pythagoras' Theorem as follows:

$$\begin{aligned} s &= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \\ s &= \sqrt{(21.5 - 27.2)^2 + (7.2 - -2.2)^2} = 11.0 \end{aligned} \qquad (1)$$

which comes to almost exactly 11 km.

The point I am making here is simply this. Euclidean space obeys Pythagoras' theorem – and conversely, any space which obeys Pythagoras' theorem is Euclidean.

The geometry of spherical two dimensional space

In order to specify any point on a sphere (for example the Earth) you need to state two *angles* – the longitude θ (Easting) and latitude φ (Northing) but you cannot simply use Pythagoras' theorem on the two sets of coordinates like you can with flat space. There are two problems to solve. The first is that you cannot simply subtract the longitude and latitude coordinates of two points to get the 'longitude' and 'latitude' of their separation because the physical distance between lines of longitude depends on latitude. Secondly, you have to convert angular distances to 'straight line' distances, then use Pythagoras' theorem before converting back to to angular separation. This involves knowing the radius of the sphere. Nevertheless, there exists a well defined, if complicated, procedure for calculating the angular distance (along a geodesic) between any two points on a sphere.

Any space which embodies such a procedure for calculating distances is called a *metric space*. Metric spaces include Euclidean space, the surface of a sphere, the topography of the Lake district and many, many other mathematical creations.

Minkowski Spacetime

Einstein published his seminal paper on Special Relativity in 1905. Three years later, Hermann Minkowski (who had already anticipated much of Einstein's work) showed how all of Einstein's results could be interpreted in terms of a four dimensional metric space involving three spatial dimensions and one temporal dimension. He began his lecture to the German Society of Scientists and Physicians on 21st September 1908 with the following famous words:

"*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*"

This quote has often been misinterpreted as meaning that the temporal dimension is qualitatively no different from the other three spatial ones and that, by changing your system of coordinates appropriately, you can turn space into time and vice versa, just as by rotating a map by 90 degrees, you can change an Easting into a Northing. This is only partially true. In special relativity, by moving at a sizeable fraction of the speed of light, two observers in relative motion will indeed use different coordinate systems to record events which they see in the universe around them and while two spatially separated events may appear simultaneous to one observer, they will not so appear to the other. However, it is *never* the case that two events will appear simultaneous but spatially separated to one observer – and – in the same place but temporally separated to the other. The reason for this lies in the metric of spacetime.

The metric of spacetime

The distance s between the points (x_1, y_1, z_1) and (x_2, y_2, z_2) in Euclidean space is given by the formula:

$$s = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2} \qquad (2)$$

and it is this formula which defines the *metric* of Euclidean space.

The *metric* of Minkowski space is defined by the following formula:

$$s = \sqrt{\Delta(ct)^2 - \Delta x^2 - \Delta y^2 - \Delta z^2} \qquad (3)$$

where s is the separation (technically called the *interval*) between the two events (t_1, x_1, y_1, z_1) and (t_2, x_2, y_2, z_2)[1]

We can now see exactly how and in what way the temporal dimension differs from the spatial ones – they differ in sign.

Now consider two events which (for Ann) occur at the same place but four years apart. The coordinates of the first event A (the birth of a daughter) are (0, 0, 0, 0) and the coordinates of the second B (the birth of a son) are (4, 0, 0, 0). The interval between these events is $\sqrt{4^2 - 0 - 0 - 0} = 4$ light years. (We shall use the year as our unit of time and the light year as our unit of distance. The speed of light c is of course 1 light year per year.)

1 Many authors define the *interval* between two events as s^2 rather than s. I prefer to use the latter so that the unit of interval is the same as the unit of length and successive intervals can be added together.
I also define s as $\sqrt{\Delta(ct)^2 - \Delta x^2 - \Delta y^2 - \Delta z^2}$ not $\sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2 - \Delta(ct)^2}$ so that the interval between events occurring in the absolute past and the absolute future (i.e. time-like events) is real rather than imaginary.

Now consider the situation from the point of view of her brother Bob who speeds past Earth (in the X direction) at 80% of the speed of light at the very instant that Ann gives birth to her first child. As he flies past he sets his clocks and rulers to zero so according to him the coordinates of A are also (0, 0, 0, 0). Due to the effects of time dilation, he considers that his sisters clocks are running slow by a factor of $\sqrt{1^2 - 0.8^2} = 0.6$ So by his reckoning, the birth of Ann's second child occurs at $t = 6.67$ years. In this time he has travelled $6.67 \times 0.8 = 5.33$ light years so according to Bob, the second event occurs at (6.67, -5.33, 0, 0).

Now let us calculate the *interval* between the two events according to Bob. This works out as follows:

$$s = \sqrt{(6.67)^2 - (-5.33)^2 - (0)^2 - (0)^2} = 4 \text{ light years} \tag{4}$$

As you can see, it all works out very nicely. The *interval* between the two events is always the same, whoever does the measuring.

Now I said earlier that two events which occur in the same place but which are temporally separated (like the birth of Ann's children) could never appear simultaneous but spatially separated to someone else. The reason for this is that the interval (as I have defined it) would be real for the first observer but imaginary for the second.

Visualizing spacetime

We now come to the really difficult bit. How can we picture spacetime so that the constancy of the interval makes sense?

Well, the first thing to do is to forget about the y and z spatial dimensions and just concentrate on t and x. In effect, we shall consider a two-dimensional spacetime. We can therefore plot everything which happens on an ordinary piece of paper with the spatial axis horizontal and the temporal one vertical. Ann's view of the two most important events in her life is shown in *Fig.* 5:

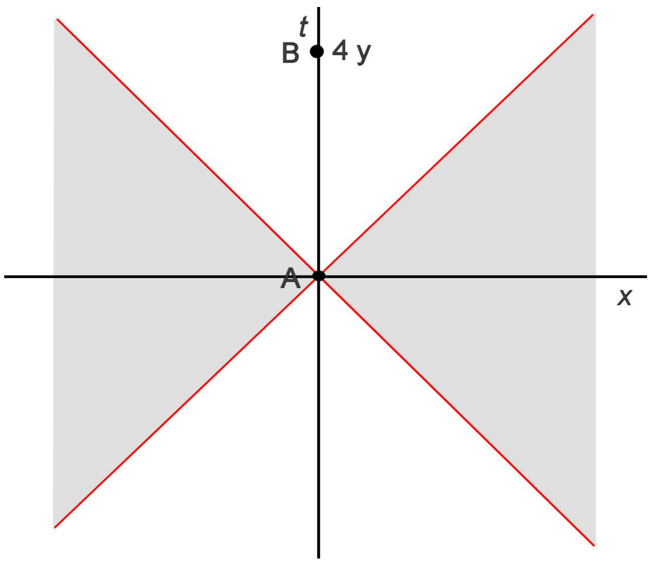

Fig. 5: Ann's view of the two events

(The shaded areas represent all those events which are causally disconnected from Ann's origin because light cannot reach them in time. The red diagonal lines represent the future and past light cones with the future at the top.)

Now what about Bob's view? We know that Bob sees exactly the same events as Ann but he plots them differently on his x and t axes. We know where Ann would plot Bob's t axis because she knows that he is moving away from her at 80% of the speed of light. In her view, Bob's origin is moving upwards and to the right. It is not so easy to see where she would plot his spatial axis but it turns out that (provided we use appropriate units) Bob's axes act like a pair of scissors, making the same angle with the 45° line: (see *Fig*. 6)

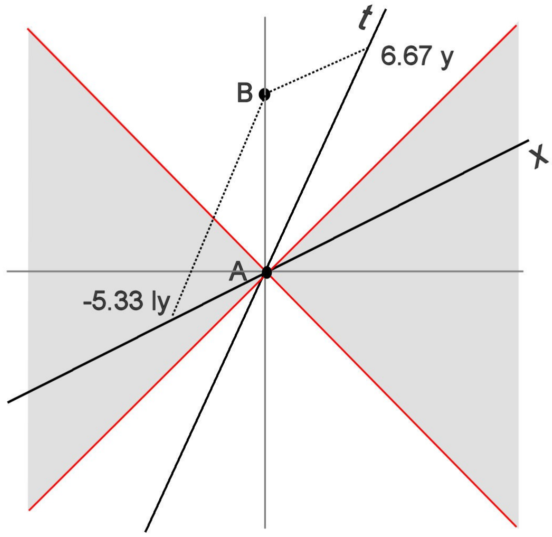

Fig. 6: Ann's view of Bob's axes

We can now see immediately why Bob sees the birth of Ann's second child as happening 6.67 rather than 4 years later and why the event happens 5.33 light years behind him. We can also see why, as Bob's speed approaches that of light, the spatial and temporal axes come closer and closer together but it is equally clear that they can never swap places. The temporal axis always lies in the absolute future and the spatial axis always lies in the potential present.

Spacetime and geodesics

Suppose that after travelling for 3.33 years, Bob decides to about turn and go home. He will, of course arrive back just in time to witness the birth of Ann's second child. *Fig*. 7 shows his journey as plotted by Ann:

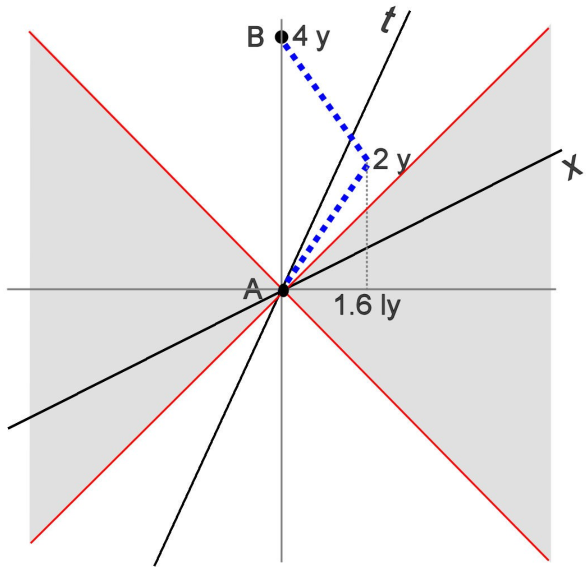

Fig. 7: Bob's return trip

According to Ann, the coordinates of the three points in Bob's journey are (0, 0), (2, 1.6) and (0, 4) and the *interval* between the first two waypoints is $\sqrt{2^2 - 1.6^2} = 1.2$ light years. Similarly the interval between the second pair is also 1.2 light years. The *total interval* for the whole journey is therefore 2.4 light years. Now this is rather surprising. We know that the interval between the first and last event is 4 light years. How can the interval as measured along Bob's journey be *less* than this?

The answer is simple. It just is. We must not expect spacetime to act just like space. In space there is a shortest route between A and B and all other routes are longer; in *spacetime* there is a *longest* interval between two events and all other ways of getting from one event to the other involve a *smaller* interval. It is just a consequence of that negative sign in the metric formula. Live with it!

Of course, this is just the famous twins paradox. Since at 80% of the speed of light, Bob's clocks appear to Ann to going at 60% of the correct speed, she will not be surprised to find that Bob is only 2.4 years older when he returns. What we are saying here is that an *inertial* observer (i.e. an observer who travels from event A to event B at constant speed) will travel along a geodesic and he or she will age in years equal to the interval between the events in light years.

The warping of spacetime

We are now in a position to get a glimpse of what is meant by the phrase – the warping of spacetime.

The difference between the surface of the Earth and a flat sheet of paper is completely described by the difference in the way in which distances are calculated – i.e. by the metric of the space. Likewise, the shape of surface of the Lake District hills is completely described by the list of geodesic distances between every pair of points. In a flat Minkowski spacetime, we have a simple formula which enables us to calculate intervals but what would happen in a Minkowski spacetime which wasn't 'flat'? Clearly the answer must be that the formula for calculating the interval must vary from event to event. Now in a 2 dimensional Minkowski space the metric is $s = \sqrt{(ct)^2 - x^2}$ and in the units which Ann is using, $c = 1$. But suppose that, 2 light years away from Ann there was a region of space where $c = 2$. (I know what you are thinking but bear with me.) In this region the metric is $s = \sqrt{(2t)^2 - x^2}$.

Now suppose that, as before, Bob sets out at 80% of the speed of light. He will reach this new region of space (shown shaded in Fig. 8) in 2.5 light years. What happens then?

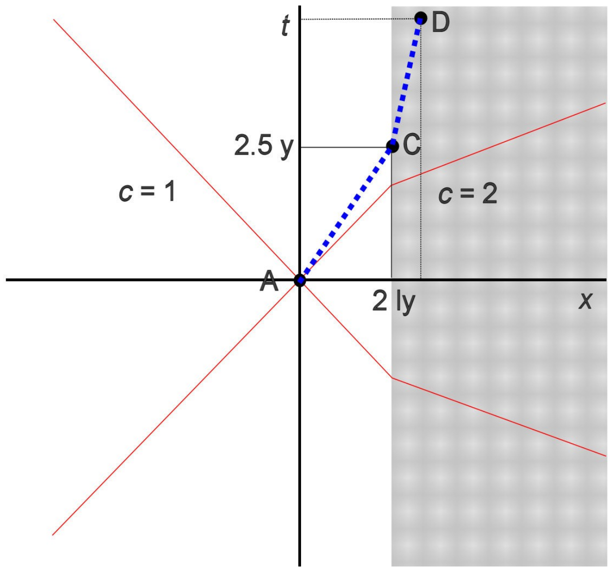

Fig. 8: Bob's trip in a non flat spacetime

To resolve this question we only have to recall that Bob will follow a geodesic line through spacetime. If, after travelling a while longer he reaches a point D, it must be because the total interval along his path ACD (shown dotted) is *longer* than any other nearby path. The situation is very similar to the problem of working out the best route for the lifeguard to take to reach the drowning swimmer. Without doing any calculations, it is easy to see that in order to *maximise* the total interval, it is better for Bob to spends *less* time in the region where $c = 1$ and *more* time in the region where $c = 2$. This explains the kink in his geodesic and why (according to Ann) his speed suddenly appears to decrease when he enters the region. Naturally (according to Ann) the speed of light will appear to increase which is why I have put a kink in the red lines as well.[2]

(Of course, as far as Bob is concerned, he feels nothing when he enters the shaded region and, naturally, any measurements he makes of the speed of light with his rulers and clocks will still give the result he expects. But if he were to stop at his destination D and radio back to Ann, she would notice something rather odd. Bob's radio waves would be *blue* shifted by a factor of two and his speech would be very *rapid*. It is perhaps better to think of time being *speeded up* (from Ann's point of view remember) in the shaded region rather than the speed of light being doubled. Similarly, Bob will consider that Ann's clocks are running slow and when he (metaphorically) watches a light beam bouncing back and forwards between a pair of Ann's mirrors, he will conclude that where Ann lives, the speed of light is only half what it should be. In other words, the speed of light does not really change in the shaded region – it only appears to change to an observer outside the region. In fact it would be better to stop referring to c as the speed of light altogether. Let us call it the *metric constant* from now on. I hope this assuages the doubts you may have had earlier.)

The spacetime metric around a gravitating body

We are now in a position to tackle the question of how a change in metric can explain gravity.

If you were to release a small particle from a great height above gravitating body (assumed to have negligible dimensions) it would, according to Newton's law, execute a perpetual oscillatory motion. On a spacetime diagram, this motion would look something like *Fig*. 9: (Note that the X axis has been much expanded. If we were to adopt units in which $c = 1$, the wiggles would, in practice, be very tiny.)

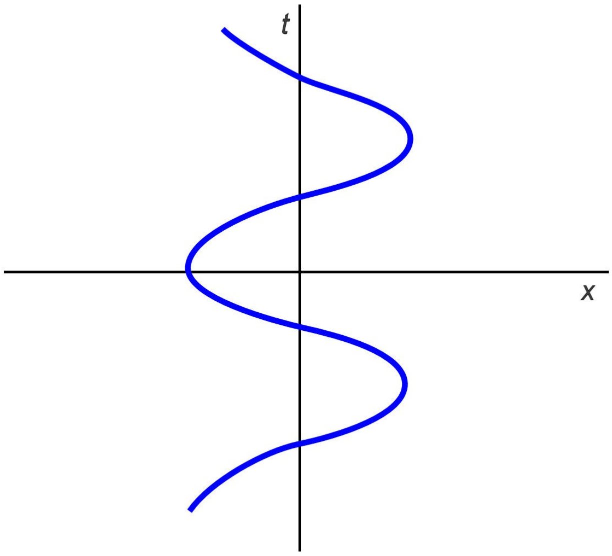

Fig. 9: Spacetime diagram of an oscillating particle

The question now arises – what sort of metric could give rise to a curve like this? We can get a clue by considering how a beam of light is constrained inside an optical fibre. These are made with a core of glass with a high refractive index surrounded by a cladding with a low refractive index. Alternatively, in what is called a graded-index fibre, the refractive index varies smoothly from high in the centre to low at the edge. When the light beam deviates off the centre line it is bent back into

2 It may appear odd that while light appears (to Ann) to increase in speed, Bob appears to slow down but his is just the way 'metric refraction' works. If Bob travels at very near the speed of light his trajectory will, of course, be very close to that of light. There is a slightly slower speed at which his geodesic trajectory is 'straight'. At slower speeds, the trajectory bends in the direction illustrated.

the fibre by simple refraction as shown in Fig 10.

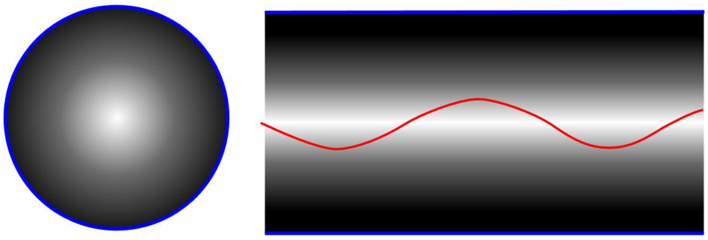

Fig 10: Graded-index fibre

Looking at the way Bob's geodesic was bent when he entered the shaded region it is clear that to get the desired effect, the speed of light – sorry, the *metric constant* – must *increase* with increasing distance from the planet. Alternatively, we could say that time must *slow down* the closer you get to the planet.

What we need can therefore be depicted as follows where the density of the shading gives an indication of the value of *c* which must be used to calculate the interval – the darker the shade the *larger* the value of *c*. (see Fig 10) It is not difficult to imagine this extended to two spatial dimensions with the geodesic path of the satellite spiralling up the temporal axis.

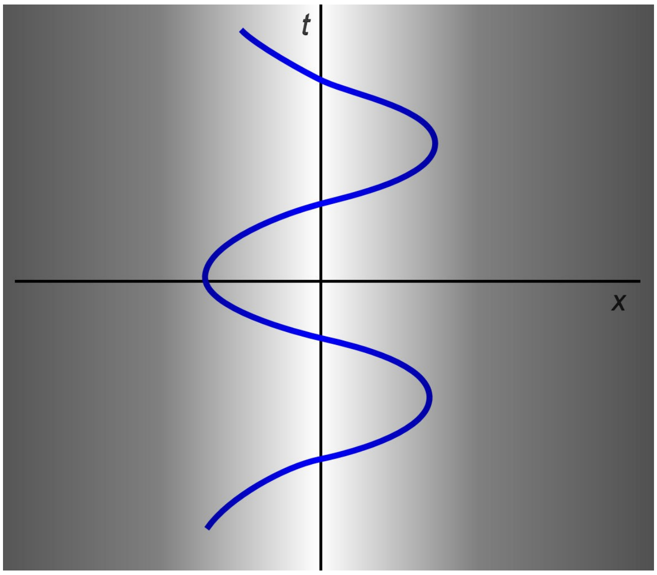

Fig. 10: Spacetime with a graded metric

All that remains to be done is to deduce the exact relation between the metric constant *c* and the radial distance *r* from the planet which is necessary to reduce to Newton's Law in the classical limit and, basically, you have the equation which tells you everything you need to know about how a massive object affects spacetime in the vicinity. Here goes:

First, the curvature of the geodesic in spacetime is equal to the acceleration due to gravity of the particle in space and can be written as

$$g = c\frac{dc}{dr} = \frac{GM}{r^2} \qquad (5)$$

where *c* is the metric constant, *G* is the gravitational constant and *M* is the mass of the planet.

Separating the variables and integrating leads us to

$$\frac{1}{2}c^2 = A - \frac{GM}{r} \qquad (6)$$

Let us require that a long way from the planet $c = c_0$. Then:

$$\frac{1}{2}c^2 = \frac{1}{2}c_0^2 - \frac{GM}{r} \tag{7}$$

and

$$c = c_0\sqrt{1 - \frac{2GM}{r\,c_0^2}} \tag{8}$$

In order to put this formula to some practical use, it is better to reformulate it in terms of the time dilation factor which you would expect to observe as your friend approaches closer to the planet. Since as c increases, clocks slow down the correct formula is

$$t' = t\,\frac{1}{\sqrt{1 - 2GM/r\,c_0^2}} \tag{9}$$

Putting in some figures for the Sun we find that the time dilation factor at the surface of the Sun is

$$\frac{1}{\sqrt{1 - 2GM/r\,c_0^2}} = \frac{1}{\sqrt{1 - 2\times 6.7\times 10^{-11}\times 2\times 10^{30}/7\times 10^{8}\times (3\times 10^{8})^2}}$$

$$= \frac{1}{\sqrt{1 - 4.25\times 10^{-6}}} = 1.0000021$$

What this means is that a clock on the surface of the sun will lose time by 1 second every 5½ days.

There is just one last thing I would like to point out. If M is sufficiently large and r is sufficiently small, the expression $2GM/r\,c_0^2$ can equal unity and the time dilation factor becomes infinite. I do not need to tell you what sort of object this represents.

Carr Bank, July 2018