General Relativity: The Equivalence Principle and Spacetime Curvature

The origins of general relativity can be traced to two considerations that were well-known to Galileo 300 years earlier, but for 250 or so of them do not seem to have been regarded as worthy of special attention:

(1) (In the absence of air resistance, etc.): All bodies, irrespective of their weight, composition, shape etc., fall at the same rate in a gravitational field (inertial mass = (passive) gravitational mass: compare the Eötvös and "5th-force" experiments).

(2) There exists (in both Galilean mechanics and special relativity) a "special" class of reference frames, namely the inertial frames.

Before special relativity, in the latter half of the nineteenth century, Ernst Mach took the second point seriously and asked: What makes the "special" class of inertial frames special? Weinberg's "pirouette" experiment and Newton's "bucket" experiment both demonstrate that there does appear to be something "special" about the frame in which the so-called "fixed" stars appear to be at rest. However, Mach objected that no one really knew what would happen, if in the latter, the thickness of the walls of the bucket were increased to several kilometers: "no one is competent to say…". He postulated that inertial frames are determined by the rest frame of the (local) matter of the Universe ("Mach's principle"). (Isotropy of black-body radiation.)

Connected with this is the so-called "principle of gravitational induction". Circulating electric currents produce a magnetic field, which "favors" rotation in the sense that the magnetic force can be removed by going to a rotating frame. Can we do the same for gravity? The gravitational force between two static bodies (the analog of the Coulomb force in electromagnetism) has the form $F_{\text{stat}} = G\,\frac{m_1 m_2}{r_{12}^2}$: suppose there also exists the analog of the Ampere force between currents, i.e., $F_{\text{dyn}} = G\,\frac{m_1 m_2}{r_{12}^2}\,\frac{v_1 v_2}{c^2}$.

Now, we know that in a rotating system of coordinates (e.g., on a merry-go-round) there exists, in addition to the "centrifugal" force, a "Coriolis" force proportional to velocity:

$$F_{Cor} = 2m\omega v$$

where ω is the angular velocity of rotation; this Coriolis force is in a direction perpendicular both to the axis of rotation and to the velocity. Is it possible that this is just F_{dyn} in disguise? Well, suppose that in the formula for $F_{dyn}$the subscript 1 refers to "us" (i.e., a massive object on Earth) and 2 refers to the fixed stars, and we assume for the moment that we are at their "center"; then, if we go into a frame of reference rotating with angular velocity ω, the fixed stars will appear to move with respect to us with velocity $v_2 = \omega r_{12}$ where r_{12} is some kind of average radius. Hence the "gravito-magnetic" force F_{dyn} would be given (at least in order of magnitude) by

$$F_{dyn} \sim \frac{G m_1 m_2 v_1 \omega r_{12}}{r_{12}^2 c^2} \quad = \quad \frac{G m_1 m_2 v_1 \omega}{r_{12} c^2}$$

In the same notation the Coriolis force is (leaving out the 2)

$$F_{Cor} \sim m_1 v_1 \omega$$

So if the two are to be equal we must have

$$\frac{Gm_2}{r_{12}c^2} \quad \sim \quad 1.$$

But to the extent that we can define the "radius" $R \sim r_{12}$ of the Universe, its mass m_2 is $\sim \rho R^3$, where ρ is the average mass density; hence we would predict that

$$G\rho R^2/c^2 \quad \sim \quad 1$$

Remarkably, the cosmological evidence does seem to suggest that to the extent that R can be defined (cf. lecture 26) this is about right (to an order of magnitude!). The direction of F_{dyn} is also that of F_{Cor} (though this may not be immediately obvious).

Einstein postulated: No *mechanical* experiment in a freely falling elevator can reveal the fact that it is falling rather than floating freely in space in the absence of gravity. Alternatively, one can consider the "box-with-rope" experiment (Einstein p. 66): an inhabitant of the box will experience the acceleration as *equivalent to a gravitational field.* (Modern implementation: aeroplane during takeoff.)

Generalization: *All* physical phenomena in a freely falling frame are identical to what they would be in free space (”equivalence principle”). Conversely, all (local) effects of acceleration can be attributed to a gravitational field. Thus, the role of *inertial* frames in special relativity is replaced by *freely falling* frames in general relativity. (Note: it is not trivial, but nevertheless true, that any "smooth" local frame can be transformed into a freely falling one).

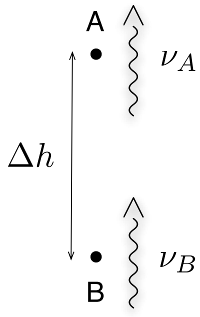
So the prescription is: to work out how things happen *locally*, go over to a freely falling frame, do the calculations as in Minkowski space, then transform back to the original frame!

Immediate applications:

- Gravitational redshift.

  Imagine a source B, emitting some kind of periodic signal (one usually thinks of a light wave, but it need not be), which is picked up by A who is sitting in a gravitational potential which is higher by an amount $g\Delta h$. Suppose A is using a standard "clock" (e.g., a cesium atom) and B compares the frequency of the signal as reported to him by A with that of his own exactly similar standard clock. Obviously, if no gravitational field were involved they should see the same.

  To find the correct result we must visualize the situation as viewed by a freely falling observer. He sees A and B accelerating *upwards* with acceleration g. Suppose we



arrange (as we certainly can) that he is stationary with respect to B at the moment B emits the signal. Then if the signal is received (according to him) by A at a time Δt later, by that time A will be moving upwards with velocity $v = g\Delta t$, and hence will see the signal Doppler-shifted downwards in frequency by an amount which (for $v \ll c$) is $-v/c = -g\Delta t/c$. Now if v is indeed $\ll c$, then Δt is approximately $\Delta h/c$, and thus:

$$\frac{\Delta\nu}{\nu_0} \simeq -g\Delta h/c^2 = -\Delta\phi/c^2$$

which is just the difference in gravitational potential divided by c^2. So, A will see the signal emitted by B "redshifted".* This prediction has been verified, e.g., in the Harvard Tower experiment (and also in astrophysics).

The general formula for the redshift, under conditions when the difference in gravitational potential $\phi_A - \phi_B$ is not necessarily small, turns out to be

$$\nu_A/\nu_B = \sqrt{\frac{1 + 2\phi_B/c^2}{1 + 2\phi_A/c^2}}$$

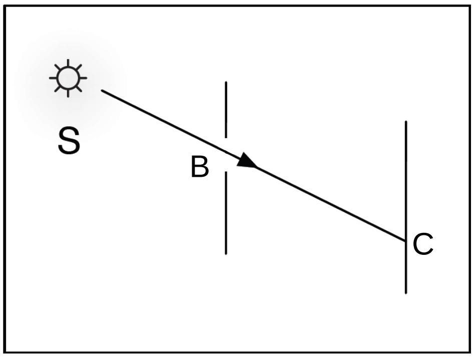
I quote this result† without proof because it will be useful later. See below for a discussion of the general-relativistic "twins paradox".

- Bending of light (qualitative).

  Imagine a source of light, a screen with a slit, and a fixed screen on which the light is received, all inside an elevator. According to the principle of equivalence, if (and only if) the elevator is in free fall, the light will appear to travel in a straight line, i.e., it will traverse the spatial points S, B, C. But for an observer fixed on Earth, these points at the times the light traverses them will *not* appear to lie in a straight line: in fact, the light ray will appear to accelerate downwards, with acceleration g! Thus, light should be bent in the field of massive bodies like the Sun. (This

*Caution: This has nothing to do directly with the cosmological redshift.

†The alert reader will notice that if the difference $\phi_A - \phi_B$ is small but ϕ_B is not itself small, this result does not reduce to the one stated above. I hope to have time to discuss this; it is associated with the question of the choice of an "absolute" zero for the gravitational potential.



effect was already predicted, on the basis of a "particle" picture of light, by Sollner in 1802: but as we shall see there is a catch!)

Digression:

Why are these effects not seen in everyday life? Answer: generally they are of order ϕ/c^2 which is small for any reasonable "everyday-life" situation (for the Harvard Tower experiment [$g \sim 10$ m/s^2, $\Delta h \sim 70$ ft, $c^2 \sim 10^{17}$ m/s], it is of order of 10^{-15}!). The predicted order of magnitude of the bending of light by the Sun is ~ 1 second of arc. One might reasonably ask what happens when ϕ/c^2 becomes of order 1: see below.

Geodesics

A statement which is true for both massive bodies and light, in both Newtonian physics and special relativity, is the following (Newton's first law, plus):

As viewed from any *inertial* frame, a body on which no forces act travels in a straight line with constant velocity; so does light. Special relativity adds the statement that the velocity of light is always isotropic and equal to c.

A way of reformulating this statement, which at first sight (only!) looks rather artificial, is the following: define the (Minkowskian) *spacetime interval* between any two events in the standard way (for 3 space dimensions):[‡]

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

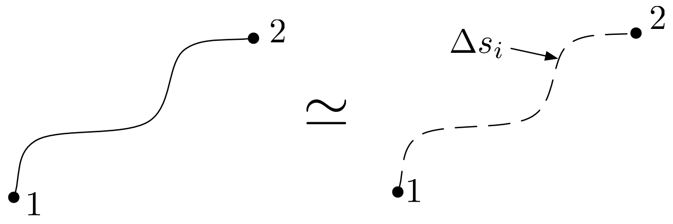
It turns out to be convenient to define also the "proper time" interval $\Delta\tau$ between a pair of events by

$$\Delta\tau = \Delta s/c$$

i.e., $\Delta\tau$ is (apart from the factor of c) the interval as defined in lecture 12. Note that the "proper" time interval is that measured by a clock travelling uniformly between the events in question (or equivalently, that measured in the rest frame of that clock). We

[‡]It is conventional in GR to take the sign of Δs^2 (which is a matter of convention) to be positive for *spacelike* intervals (contrary to the convention used in lecture 12).

know that this (the numerical value of Δs^2) is agreed upon by all inertial observers. Consider now a body that starts at point (x_1, y_1, z_1) at time t_1 and ends at (x_2, y_2, z_2) at time t_2, but can follow any path between these two spacetime points. Suppose it does follow some convoluted path (I draw only the "space" part!) and break it up into small



parts of length of order ϵ: define a "total spacetime path length" as

$$\Delta s = \lim_{\epsilon \to 0} \sum_i \Delta s_i, \quad \Delta s \equiv \sqrt{\Delta x_i^2 + \Delta y_i^2 + \Delta z_i^2 - c^2 \Delta t_i^2}$$

Then the above statement of Newton's first law is simply: The actual path followed between spacetime points 1 and 2 is such that the total path length Δs is a minimum (such a path is called a "geodesic"). In addition, for light Δs is zero.

Let's verify that this does give back Newton's first law. First, assume that we "forget about" the time variable: then we simply have to minimize the actual path length in space between the two points. But the shortest path between two points in (Euclidean) space is precisely a straight line! Hence part (1) of Newton's first law follows, and we have (e.g., in two dimensions) simply

$$\Delta y_i = \text{const} \cdot \Delta x_i$$

When we know x_1, y_1 and x_2, y_2, the constant is of course uniquely determined as $(y_2 - y_1)/(x_2 - x_1)$.

To investigate the second part, let's assume we choose our coordinate system so that motion is only in the x direction. Then, we have

$$\Delta s_i = \sqrt{\Delta x_i^2 - c^2 \Delta t_i^2}$$

This expression differs from the "Pythagorean" one only by the factor of c^2 and the $-$ sign: both can be formally removed by defining a "pseudo-Euclidean" coordinate $\tau \equiv ict$, so $\Delta s_i = \sqrt{\Delta x_i^2 + \Delta \tau_i^2}$. It is then clear from the analogy to ordinary Euclidean space that the optimum path is the "straight line" $\Delta x = \text{const}\ \Delta\tau$ or $\Delta x =$ another const. Δt, i.e.:

$$(x_2 - x_1) = \text{const}\ (t_2 - t_1) \equiv v(t_2 - t_1)$$

which is just the statement that the body moves at constant velocity. In the special case of light (in special relativity) we know that this velocity has to be c, and we therefore recover the result $\Delta s = 0$.

So far, so good: we have really learned nothing we didn't know. But we note a great advantage of the "geodesic" formulation: it is totally independent of what coordinate system we choose to employ! (we can not only use different inertial frames, but having chosen a frame can use, e.g., spherical polar coordinates instead of Cartesian ones). Thus, we can for example choose to use the frame of reference appropriate to an observer who is stationary on the Earth's surface. This is *not* a freely falling frame, and hence we expect that in general the trajectories of "free" (i.e, not acted on by any nongravitational forces) bodies would not be straight lines and/ or would not be traversed at constant velocity: this is, of course, what we observe. However, the Minkowskian expression for the space-time interval, $\Delta s^2 = \Delta x^2 - c^2 \Delta t^2$, automatically leads to linear motion with constant velocity: thus, *the expression for the interval in this noninertial frame must be non-Minkowskian.*

Before embarking on the ensuing argument, let us mention one important point: in writing down a proposed metric, we *must use a system of coordinates that refer to a single reference frame*, even though we are trying to describe the dynamics throughout all of space. Thus, the Δx's, etc., that appear in the metric must be those measured by a single observer, irrespective of whether he/ she is freely falling. If we want to look at dynamics near Earth's surface, at first sight the most "natural" observer to choose would be one stationary on Earth, and we could indeed do this; however, for reasons we will see below, it is actually more convenient to choose an observer (Alice below) stationary at infinity. (For the correction necessary with the former choice, see below.)

If the general expression for the metric (spacetime interval) is to scale linearly with the space and time dimensions, it must be bilinear in Δx, Δy, etc., i.e., Δs^2 should involve only terms like Δx^2, Δy^2, $\Delta x \times \Delta y \ldots$ (each possibly multiplied by some function of space and time). Since the horizontal motion is not affected by a gravitational field, it seems natural to assume that the corresponding part of Δs^2 is unchanged from its simple Minkowskian form $\Delta x^2 + \Delta y^2$. This leaves the possibility of terms proportional to Δz^2, Δt^2, and $\Delta z \times \Delta t$; however, the last can be excluded by considerations of time-reversal invariance. Thus, since the situation is independent of the x and y coordinates and of time, the most general expression we are allowed to have is

$$\Delta s^2 = \Delta x^2 + \Delta y^2 + f(z)\Delta z^2 - g(z)c^2 \Delta t^2$$

where $f(z)$ and $g(z)$ are for the moment arbitrary functions of z. Now, the redshift experiment (or the associated "GR twins paradox") will allow us to infer the form of $g(z)$; however,considerable caution is necessary, and it is easy to reach a qualitatively wrong result: suppose that the gravitational field actually falls off with z (as is in fact the case for a finite body like the Earth), so that "at infinity" we can take it (and the gravitational potential) to be zero. Imagine a twins-paradox thought experiment in which Alice and Barbara both start "at infinity" and Alice stays there throughout, while Barbara travels to a point low gravitational potential $\phi(z) < 0$, spends a substantial time there, and eventually returns to join Alice. Inverting the frequency ratio obtained above, we see that the time intervals between events in Barbara's region (such as the flashes emitted by her flashlight), as estimated by Alice (A) and Barbara (B), have the relation:

$$\Delta t_A = \frac{\Delta t_B}{(1 + 2\phi(z)/c^2)c^2\Delta t^2}$$

Note that had we decided to use the reference frame of an observer stationary at the surface of Earth, let us say at height z_0, the effect would have been to change the last term to:

$$\frac{(1 + 2\phi(z)/c^2)}{(1 + 2\phi(z_0)/c^2)c^2\Delta t^2}$$

and Δt^2 would now mean the time interval measured by Barbara. The metric is still non-Minkowskian!

Now comes the tricky point: which of the two (if either) is an inertial observer and therefore entitled to treat the time interval she observes as the true space-time interval (note both agree that Δz is zero)? It is at first sight tempting to argue that since Alice is stationary in an inertial frame (as no gravitational field exists in her region), while Barbara is not "freely falling" in the nonzero gravitational field that exists in her region, Alice is an inertial observer and Barbara is not. *But this is wrong*; we are discussing what happens in Barbara's region, so that it is the gravitational acceleration *there* that is relevant, and Alice is certainly *not* falling freely with that! On the other hand, somewhat surprisingly, *for the purposes of calculating time intervals* (only), Barbara *can* be treated as an inertial observer! To see this, imagine a freely "falling" (i.e., accelerating) observer whose initial upward velocity makes him exactly coincident with both flashes. The true space-time interval is just the time interval as measured by him – but, since he sees Barbara moving, first downward then upward, by symmetry, this is just the time interval measured by her![§] Thus, it is Barbara's time that is the "proper" time – i.e., apart from trivial factors, the true spacetime interval (and so the lack of aging by Barbara, whose biological clocks presumably go by proper time, is a real effect, just as in the SR twins paradox). Further, since the above formula for the spacetime interval is that which has to be used *by Alice*, the Δt in it is Δt_A. Putting these considerations together, we find:

$$g(z) = 1 + \frac{2\phi(z)}{c^2}$$

i.e.,

$$\Delta s^2 = \Delta x^2 + \Delta y^2 + f(z)\Delta z^2 - (c^2 + 2\phi(z))\Delta t^2$$

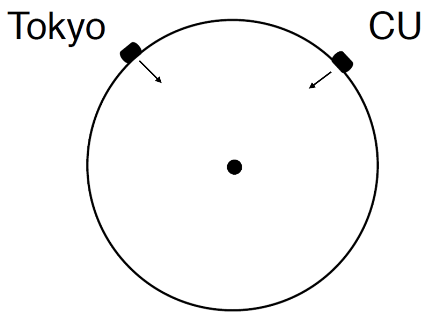
We do not yet know what $f(z)$ is: let us for the moment tentatively set it equal to 1. A standard variational calculation[¶] then shows that provided $|\phi(z)| \ll 1$, an observer

[§]At first sight this argument might seem inconsistent with that used in the discussion of the *special-relativistic* twins paradox, where clearly it matters that one twin is accelerating relative to an inertial frame. However, in the present case, the needed velocity of the freely-falling observer is proportional to Δt, so that any correction to Δt is of order (at most) Δt^2, and can be made negligibly small by taking Δt small enough.

[¶]Which, however, requires more math than assumed in the course.

using this expression and requiring that Δs be a minimum will indeed see the trajectories predicted by "Newtonian gravity".

Now, it might be objected that all of the last two pages is a sledgehammer to crack a nut. After all, in this case it is perfectly straightforward to transform to a freely falling frame (which is the same everywhere in space), simply posit that in this frame free bodies move in straight lines, and then transform back to the frame of the stationary observer. To see the advantage of the "metric" (or "geodesic") formulation we have to consider a somewhat more complicated problem, e.g., the gravitational field of the spherical earth. Since the direction of the gravitational acceleration is *different* in Tokyo from that in Champaign-Urbana, it is clear that it is impossible to transform *globally* to a single freely falling (inertial) frame, and thus we would expect that there exists no frame in which motion is globally in a straight line. We *must* then use the more general expression, with $\phi(z) \to \phi(r) = -GM/r$. This begins to suggest that space-time might be "intrinsically curved".



Digression on curved space (–time)

We know that the three-dimensional space around us is (at least to a very good approximation) Euclidean. But imagine for a moment that we were ants crawling on the two-dimensional surface of a spherical football (soccer ball), and had no means of knowing about the direction perpendicular to the surface. We would then measure all distances as "distances traversed along the surface", and we could find a number of "non-Euclidean" features:

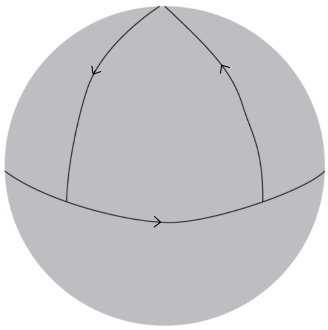
1. The angles of a triangle need not add up to 180° (consider, e.g., a triangle based on the North Pole and the Equator).

2. The circumference of a circle is in general less than $2\pi \times$ its radius.

3. More than one "straight line" can be drawn through the same two points.

4. The space is in an obvious intuitive sense "closed" (in fact its area is $4\pi R^2$, though the ants would not know that R is a "radius").

5. "Parallel transport is nontrivial": if, e.g., I start from the North Pole, carrying a stick that always points in the some direction as viewed in the local coordinate system (e.g., always points south), walk as far as the Equator, turn east and go along the Equator for a few thousand miles, then turn back north and return to the North Pole – is the direction of the stick the same as it started?



Now of course *we* know, though the ants don't, that the two-dimensional surface is "embedded" in a three-dimensional Euclidean continuum. But, in fact, it is possible to describe the "intrinsic" properties of the surface *independently* of this knowledge; the ants could infer everything about it by making sufficiently careful measurements, and indeed there are some two-dimensional "surfaces" that *cannot* be embedded in Euclidean space. So the questions are: is our "four-dimensional" space-time in some sense "curved", and if so, what is it that determines the degree and nature of the curvature?

Einstein's answers were as follows: (1) yes; (2) the presence of mass. In fact, in a frame of reference where all mass is stationary, one can say crudely that

$$\boxed{\text{curvature is proportional to mass density}}$$

It should be emphasized here that the "curvature" that occurs in this statement is the appropriately defined curvature of *four-dimensional spacetime.* Thus, this is non-zero inside (e.g.) the Earth (mass density $\neq 0$), but zero in the free space outside. It then turns out (lots of math!) that this determines the function $f(r)$ to be the inverse of $g(r)$; that is, the correct expression for the metric outside a spherical massive body as a function of r (independent of t!) is‖

$$\Delta s^2 = \Delta s_\perp^2 + \left(1 - \frac{2GM}{rc^2}\right) c^2 \Delta t^2 + \frac{\Delta r^2}{1 - \frac{2GM}{rc^2}}$$

– the famous *Schwarzschild metric.* Note that the (appropriately defined) four dimensional curvature of this metric is zero. However, let us consider the three dimensional space obtained by "slicing" at a definite time (for the observer in question) (i.e, considering events with $\Delta t = 0$). It turns out that this three dimensional space *does* have a

‖Here $\Delta s_\perp^2$ the "distance on the surface of the sphere" (technically, in spherical polar coordinates we have $\Delta s_\perp^2 = r^2(\Delta\theta^2 + \sin^2\theta\Delta\phi^2)$).

finite curvature! This is no different from the fact that in "everyday life", while ordinary three dimensional space is (very nearly) Euclidean (zero curvature), the two dimensional subspace corresponding to a "slice" at $r = \text{const}$ (e.g. the surface of a football) has finite curvature.

Some consequences of the Schwarzschild metric:

1. Bending of light in the field of the Sun (the result turns out to be a factor of 2 times the result of the "naïve" (Sollner-type) argument; Lenard). This was verified in the famous Royal Society expedition to West Africa to observe the 1919 eclipse.

2. The precession of perihelion of Mercury (space curvature contributes 1/2 of the effect of the redshift).

3. Effects in radar echoes between planets.

One obvious question: What happens if we have so large a mass, in so small a radius, that $2GM/rc^2 = 1$ (so that the redshift becomes "imaginary")? Answer: a black hole!

Application to cosmology: Suppose the Universe is filled with some kind of matter whose average density, in its rest frame, is ρ. Then at "radius" R we have $2GM/R = (8\pi/3)\,G\rho R^2 \equiv \Omega$. It looks, therefore, as if it ought to "matter" whether the actual "radius" of the Universe[**] is such that this parameter is greater or less than 1. In fact, this is true: in the simplest model of the Universe (the so-called Friedmann-Robertson-Walker model, see lecture 26), we have the following situation.

$\Omega < 1$: the Universe is a space of constant negative curvature, and is infinite ("open" Universe).

$\Omega = 1$: the Universe is a space of flat space-time, infinite ("flat" Universe).

$\Omega > 1$: the Universe is a space of constant positive curvature and is hence the three dimensional analog of a spherical surface ("closed" Universe).

What is the experimental value of Ω? See lecture 26.

Finally, a note: If the mass is stationary, the freely falling bodies have no "rotation". However, if the mass is *rotating*, in general they do: there is a tendency to "drag" projectiles into rotation though it is generally not complete. (More in the next lecture.)

[**] Or more generally the "scale factor" (see lecture 26).