General relativity and its tests

Reference webpages:

http://en.wikipedia.org/wiki/General_relativity
and Chapter 2 in Thorne

Questions to keep in mind are:

1. How has GR, or quantum mechanics, changed our overall philosophy? Should it?
2. As we will see later, evidence for GR in its most extreme manifestations is very meager. How much should we believe it? Should aesthetic attraction play a role?
3. What is the most important aspect of a physical theory: the predictions it makes or the framework it suggests?
4. In the year 2500, assuming we have a quantum theory of gravity, will general relativity still be useful?

Observational situation before general relativity

In this class thus far we have emphasized the important role played by observations and experiment. Indeed, we would say that this is a hallmark of the scientific method. However, in general relativity we have a subject that was introduced by Einstein not because of observation, but because of philosophy. This does not make general relativity nonscientific now, because various tests have been performed that corroborate the predictions of GR. However, at the time there were precious few tests of the theory, and this therefore brings up questions about the role that aesthetics play.

To set the stage, we will review the status of Newton's law of gravity in the early 20th century. You recall that in the 19th century Newton's laws had passed two very important tests: Neptune was successfully predicted and discovered, and so was the dim companion of Sirius. As telescopes improved and observations became more precise, astronomers continued to evaluate the agreement between Newton's theories and their data. The agreement was outstanding, even as more moons of planets were discovered and even as the number of known asteroids grew. There were, however, two minor discrepancies that existed in the late 19th and early 20th century.

The first had to do with the motion of Mercury. Mercury is the innermost, smallest, and most eccentric of our major planets (now that Pluto has been demoted). Because it has no moons and has such a short orbital period (just 88 days), it could be monitored with great precision. If Mercury orbited only around the Sun, and if the Sun were a perfect sphere, then

Mercury would trace over its elliptical orbit exactly; no drift. The Sun really is close enough to a perfect sphere that we can discount deviations, but the effect of the other planets is to make Mercury drift compared to a perfect, self-overlapping ellipse. By the end of the 19th century the masses and distances of the other planets were known well enough to make a very precise prediction of this drift (also called precession), and yet Mercury's orbit was not quite in line with the prediction. Instead, it drifted by 43 arcseconds per century relative to that prediction. Now, this is a really tiny number! It means that it would take about three million years to precess by 360 degrees relative to the prediction. This is so small that people didn't worry overmuch about it; besides, maybe there was another small planet inside the orbit of Mercury that would explain the discrepancy.

The second related to the orbit of the Moon around the Earth. As mentioned in passing in Thorne's book, it appeared at the beginning of the 20th century that the Moon was moving progressively faster in its orbit, an effect that could not be explained in Newtonian theory.

If you had been presented with these two problems, how would you have reacted? In the light of hindsight we know that the precession of Mercury is a real effect, and indeed it took a radical revision of our knowledge of space, time, and gravity to explain it. In contrast, the apparent speedup of the Moon occurs relative to the Earth's rotation because tides from the Moon are slowing the rotation down; in reality, the Moon is moving farther away from us and thus is orbital motion is slowing down. However, in the early 20th century neither discrepancy was resolved, and both were small enough that they played little role in Einstein's thinking.

Einstein's philosophical motivations

Instead, Einstein was motivated almost entirely by philosophy. His theory of special relativity, which gained gradual but not decisive support from physicists within the first few years after its proposal, was "special" because it dealt only with comparisons of reference frames that moved uniformly past each other. It had nothing to say about accelerated reference frames. In addition, Newton's law of gravity was subtlely inconsistent with special relativity. Why? Because the force law

$$F = \frac{GMm}{r^2} \tag{1}$$

(in magnitude) makes no mention of time. Therefore, if we experience gravity from the Sun of a certain magnitude and the Sun moves, we know about it instantly. Special relativity says that nothing can move faster than light, so it would predict that it would take at least the light travel time from the Sun (a little over eight minutes) for us to experience the change. Something new was needed, and that something was general relativity. We will now discuss the principles of general relativity and some of the relevant mathematics, and then discuss the tests that existed upon its proposal, the conceptual changes that were required, and whether the observations warranted such an extreme shift in perspective.

The mathematics and principles of general relativity

Einstein's first attempt to modify the gravitational law came in 1907, and he kept at it until he presented his final version in November 1915. Later he said that when working on the 1907 paper he had the key insight, which he described as the happiest thought of his life: "The gravitational field has only a relative existence... Because for an observer freely falling from the roof of a house - at least in his immediate surroundings - there exists no gravitational field." Whazzat? Gravity has only a relative existence? What the heck does that mean??

This is called the *equivalence principle*, and it is the most fundamental principle of general relativity. Let's take this by steps, starting with the elevator analogy. Suppose that you are in an elevator that is in free space, far from any gravity. You therefore float freely. The door closes, and you feel pressed against the floor. There are two effects that you could imagine causing this sensation. One is that someone has slipped a source of gravity below the elevator, so gravity is pulling you down. The other is that the elevator is accelerating upwards, perhaps because a spaceship is pulling it along at an ever-faster rate.

Einstein argued that for a sufficiently small elevator there is no experiment you can perform inside the elevator to distinguish between those two possibilities. You can drop balls, roll them, do experiments with light, anything, and the outcomes will be the same. Now, if you have a large enough elevator that won't quite be the case. For example, consider two objects that you drop from the same radius on Earth. Both objects will fall towards the center of the Earth, and therefore their relative separation will decrease and their acceleration will increase with time. In contrast, two balls dropped in a uniformly accelerating elevator will maintain the same separation and acceleration.

This change in perspective can be related to everyday experiences. Suppose you jump off a diving board. Do you feel the force of gravity when you are in midair? No. This may not be trivial to imagine, because we are born in a gravitational field and therefore are used to its effects, but you are not being pushed by anything when you are in the air (if you neglect air resistance). To put it another way: it's not the fall that kills you, it's the landing! Einstein supposedly had this thought when he asked a worker who had fallen off a roof (onto a bale of hay, so he was safe) whether he had felt the force of gravity as he was falling.

Even with this insight, however, Einstein had a long way to go before he was able to turn this into a quantitative theory. More specifically, he had to find the right mathematics with which to describe his theory. Then, he had to make predictions using that mathematics and compare them with observations.

It was a long and tortuous path for Einstein. Ultimately, he realized that gravity would warp spacetime itself. This is the realization that eventually leads to black holes, so let's explore some of the basics of the mathematics. We will return to the invariant interval we discussed in the last lecture, but first, to set the stage, we will take an apparent diversion into geometry.

Flat Geometry

Remember high school geometry? Pretty simple and elegant, as compiled by Euclid more than two thousand years ago. In any plane in a three-dimensional flat space, any triangle has interior angles that sum to 180 degrees, independent of size. The ratio of the circumference to radius of any circle is 2π. If you take a vector and move it around so that at all times you keep its orientation fixed (this is called "parallel transport"), then when you return it to its original location it will always be in the same orientation as before. In addition, in a two-dimensional flat space if there are two points that are separated by dx in the x direction and by dy in the y direction (perpendicular to x), then the distance ds between those points is given by the Pythagorean theorem: $ds^2 = dx^2 + dy^2$. These properties follow from the infamous fifth postulate of Euclid: "If a straight line crossing two straight lines makes the interior angles on the same side less than two right angles, the two straight lines, if extended indefinitely, meet on that side on which are the angles less than the two right angles."

Urk! Stated more simply, this says that if two lines are parallel at one point, they never meet, otherwise they do. This is infamous because it's a lot more complicated than Euclid's other postulates and as a result mathematicians through the ages spent a lot of time trying to prove it from the first four. Instead, to the consternation of the community, Riemann showed that one could develop perfectly fine geometries in which the fifth postulate does not hold.

I've always been surprised that it took this long, because in two dimensions there is an everyday counterexample that would easily have been familiar to the Greeks: the surface of a sphere. That's what we'll treat next.

Spherical Geometry

Technically what we really mean is "sphere-like" in the sense that the surface is of "positive curvature". There are other surfaces with this property, but a sphere is a good thing to keep in mind so let's examine it a bit.

Suppose we confine ourselves to the surface of a sphere. This is therefore a two-dimensional surface. Suppose we draw a really big triangle, which has one apex at the North Pole and the others at the equator, 90 degrees of longitude apart. It is easy to see that this triangle has three right angles, for a total of 270 degrees! How about a circle? Suppose we take the equator. Its circumference is $2\pi R$, where R is the radius of the sphere. You might be tempted to say that its radius is R, but remember that we are confined to the surface, so the actual radius of the circle *in the surface* is $(\pi/2)R$, so that the ratio of circumference to radius is 4! One can also show that parallel transport can lead to rotations of a vector.

Obviously, we have something different going on. Note, though (and this is really important for general relativity), that if you confine yourself to a very small portion of the sphere, the geometry is *almost* that of a flat plane. This is one reason that it actually took some pretty decent observation and arguments from the Greeks to realize that the Earth is closer to a sphere than to a plane. It isn't obvious from local observations (for example, I presume that you navigate the campus using a map designed for Euclidean flatness!).

However, you cannot map a plane to the surface of a sphere globally. That's why, e.g., Mercator projections make Greenland look so huge. There is, in fact, no coordinate transformation that makes the surface of a sphere look exactly like a plane everywhere. These are genuinely different geometries.

The universe, of course, has three spatial dimensions instead of the two on the surface of a sphere. We can, however, make the extrapolation to another dimension. If a hypersphere has four dimensions, its boundary will have three. To us, locally, such a boundary will look like a standard Euclidean space. Globally, however, it will be very different. Just as it is difficult for us to prove the Earth's sphericity, it is tough to probe the geometry of the universe from home. Nonetheless, our ability to observe very distant things has allowed us to place serious constraints on the geometry.

Hyperbolic Geometry

These are also called spaces of negative curvature. A commonly used example is a saddle, but something that is more representative is a hyperboloid, which is a hyperbola rotated around its minor axis. In two dimensions, the interior angles of any triangle add to less than 180 degrees and there are an infinite number of lines through some point P not on line L that never intersect L (as opposed to there being no such lines in a spherical geometry, and exactly one in flat geometry). As before, however, in small regions on a hyperboloidal surface, the geometry is close to flat.

Now let's return to the invariant interval.

The Invariant Interval in General Relativity

When we discussed special relativity we introduced the concept of an *invariant interval.*

We imagined that two *events* happen close to each other in space and time. For example, if we use a Cartesian coordinate system, we can suppose that the x coordinates of the events are separated by a small distance dx, the y coordinates by dy, and so on. Special relativity says that the invariant interval between these events is given by

$$ds^2 = -c^2dt^2 + dx^2 + dy^2 + dz^2 \ . \tag{2}$$

If we wanted to express this in spherical coordinates (useful if we are interested in the gravity around a spherically symmetric object), this becomes

$$ds^2 = -c^2dt^2 + dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2) \tag{3}$$

but the physical content is unchanged. The "invariant" part of "invariant interval" means that another observer moving relative to the first will measure the same ds^2, even if the relative motion or a different orientation of coordinates means that dt, dx, dy, and dz are all different. As we discussed, this is not fundamentally any more mysterious than the statement that the distance between two points on a paper remains the same no matter how you rotate the paper.

But gravity introduces something different. In Einstein's view, gravity warps spacetime, giving it a curvature that it does not possess in the absence of gravity. Objects then naturally follow this warp. To use the phrase of John Wheeler: "Matter tells spacetime how to curve, and spacetime tells matter how to move." From the mathematical point of view, one consequence of this is that the invariant interval takes a different form in the presence of objects with mass-energy. For example, a distance r from the center of a spherically symmetric object of mass M we have instead

$$ds^2 = -c^2[1 - 2GM/(rc^2)]dt^2 + dr^2/[1 - 2GM/(rc^2)] + r^2(d\theta^2 + \sin^2\theta d\phi^2) \ . \tag{4}$$

Newton's constant G is small, and the speed of light c is huge, so unless r is really small this is actually quite close to the special relativistic form. For example, consider the Earth's orbit around the Sun. Putting in the Sun's mass for M and the Earth's orbital radius for r, $2GM/(rc^2) = 2 \times 10^{-8}$. No wonder it is so difficult to discern the effects of general relativity.

On the other hand, there are some troubling aspects about this. Look at the invariant interval when $r = 2GM/c^2$. Then the coefficient of the dt^2 term goes to zero, and the coefficient of the dr^2 term goes to infinity! That can't be good. What does this mean? Is the theory wrong? As always, we would like to reassure ourselves that at least the theory makes verified predictions, so that we can accept the conceptual weirdness. How does general relativity do?

Tests of general relativity; not many!

The answer is that, even now, general relativity is the least-tested of our fundamental theories. This was even more true when Einstein proposed it. In November 1915 there was a grand total of one prediction made with the theory that could be tested: the precession of the orbit of Mercury. It did get the right answer, which is nice, but given the overwhelmingly new aspects of the theory that isn't much, and it was pre-existing data and thus didn't truly count as a prediction. A real prediction was verified in 1919, when Sir Arthur Stanley Eddington led an expedition to observe the deflection of light caused by gravitational fields. This is what made Einstein into an international popular superstar, so let's discuss it in some more detail.

Newton considered light to be made of particles that happened to move at the speed of light. With this in mind, his theory would have predicted that a ray of light that passed just by the limb of the Sun as we saw it would be deflected by about 0.9 arcseconds (the apparent diameter of a penny at a distance of 4.4 kilometers [2.7 miles]!). Einstein's theory predicted double this value. Great, but how are you supposed to observe the direction to a star when its light passes that near the Sun? Wait for a total eclipse, of course! Of course, the measurements had to wait for the cessation of World War I, but Eddington and others went to different sites for the May 29, 1919 eclipse. The measurements were very tricky, and the equipment and weather were not perfect, but Eddington and crew got results that agreed with Einstein's prediction. This result, in which an English astronomer confirmed a German physicist's overthrow of an English physicist, must have seemed to many a salve on the wounds left from WWI (in which England and Germany were on opposite sides).

As with many stories, however, this one isn't so simple. The data were far from perfect, and in fact it has been alleged that Eddington cherry-picked the data that supported Einstein's prediction (although the most recent analysis suggests that his work was legitimate). It was actually during the 1922 eclipse that definitive confirmation was achieved, but Einstein was already a star by then.

A third classical test of general relativity is the gravitational redshift of light. This essentially comes from the conservation of energy. To update Einstein's example a bit, consider the following. Suppose that you have a particle of matter, and a particle of antimatter. You drop them in a gravitational field, so they acquire speed and thus kinetic energy. At the bottom of your apparatus you let them annihilate with each other, producing two gamma rays. Using mirrors or something you make the gamma rays go back up again, then you convert them back to a particle of matter and a particle of antimatter.

Got the picture? Using this thought experiment, we can argue that when light (gamma rays in this case) moves upwards in a gravitational field it must lose energy. Otherwise, you could use the arrangement we just described to run a perpetual motion machine. Suppose no energy was lost. Since the matter and antimatter gain kinetic energy when they are dropped, if the gamma rays don't lose energy when they move upwards then the total energy of the system increases; for example, you could bleed off the extra energy into a machine, create matter and antimatter with the remainder, and do it again. This tells us quantitatively how much energy the light must lose going up, or gain going down.

When light has longer wavelengths it has less energy, and in the visible part of the spectrum red light has the longest wavelength and blue light has among the shortest. Therefore, a "redshift" indicates a process that decreases the energy of light, e.g., light from a source that is moving away from us or light that climbs out of a gravitational field. Similarly a "blueshift" indicates a process that increases the energy of light, e.g., light from a source moving toward us or light that falls into a gravitational field. It is important to note that the actual energy of the light does not change, because energy is conserved. However, the energy measured by a local observer does change.

Lovely, except that this is a very difficult effect to measure. An astronomer claimed to measure gravitational redshift from white dwarfs in 1925, but his data were actually faulty. The first conclusive measurement of the effect occurred in a laboratory in 1959, four years after Einstein's death!

Therefore, within a few years after Einstein's proposal of GR there were only two tests that it had passed. As we will see in future classes there have been some spectacular predictions confirmed since, but let's put ourselves in the position of physicists of the time. Why should we have believed general relativity? Was it enough that it explained the bending of light and the anomalous precession of Mercury?

Certainly not. If that were all that it did, its high level of complexity would not be worth the effort. There is, however, an extra point that must be kept in mind: **in the everyday weak-gravity limit, general relativity makes exactly the same predictions as Newtonian gravity**. Therefore, *all* of the successful predictions of Newtonian gravity are also made by general relativity. This is a point often missed by people who consider the history of science. In most cases it isn't that you simply trash previous ideas (although this does happen). Instead, you generalize and incorporate them. We will now discuss this in the context of the philosophical changes that are required by general relativity, and then ask whether aesthetic considerations (which motivated Einstein and which attracted many physicists to GR) are appropriate when judging between theories.

Philosophical changes due to general relativity

Philosophers of science sometimes take a point of view that I consider unproductive. The idea is that as each new theory comes in (Newton's theory; general relativity; quantum gravity if and when it is developed) the old theory is discarded. For example, in Newton's mechanics time and space are absolute; in Einstein's theory time and space are relative, with different observers measuring different distances and intervals in time between the same two events. Therefore, Newton's ideas are simply wrong.

This is *not* the way it works. Instead, in most cases a new theory generalizes and incorporates previous ideas. In *all* cases, a new theory that hopes to supplant a previous one has to be able to make the predictions of the previous theory, and ideally need to make new ones that can be tested.

As a way of understanding this more concretely, I will borrow from Isaac Asimov's outstanding essay "The Relativity of Wrong". What is the shape of the Earth? The Mesopotamians saw hills and valleys, but suggested that on average the Earth is flat. That's not a bad approximation over small scales; for example, when you navigate around the campus of the University of Maryland, you can use a map that assumes that this small part of the Earth is flat. Over larger scales this doesn't work as well. The ancient Greeks knew this, and had classical arguments that indicated that the Earth is a sphere. But this sphere doesn't deviate much from flatness over small distances; if you were to travel a mile over such a sphere, it would deviate by about 8 inches from a perfect flat plane. So the apparently overwhelming change (flat to sphere) makes little difference over the domain where people first made this hypothesis, and makes just as little difference over small regions today.

Is the Earth a sphere? No. One of the early successes of Newton's theory was to predict the next-order deviation from this model: he predicted that because the Earth spins, it should be slightly farther around at the equator than it is pole to pole (whereas these would be exactly equal for a sphere). This is in fact the case: this shape is called an oblate spheroid. But the deviation from a perfect sphere is small. If you smooth out the hills and valleys, the curvature per mile on the surface of the Earth ranges from 7.973 inches per mile to 8.027 inches per mile. Note that the correction here is much smaller than the correction from a plane to a sphere.

So is the Earth a perfect oblate spheroid? No! The southern hemisphere bulges a little more than the northern hemisphere, so you can think of this as pear-shaped. Again smoothing out the hills, you find that the deviation from an oblate spheroid is millionths of of an inch per mile. Thus there is a correction, but this correction is tiny compared to our previous corrections.

The point is that although a naive philosopher will say that we have made major qualitative steps (flat to sphere to oblate spheroid to ...) and therefore that previous models of the Earth are simply wrong, a more reasonable and productive point of view is that our predictions are being refined, and that out mental picture (which might change substantially, as it did for general relativity versus Newton's theories) is simply a guide to allow us to reason our way through different situations. Thus when quantum mechanics tells us that an apparently solid chair is actually composed of fuzzy electron orbitals, or general relativity tells us that space and time are intertwined and warped by the presence of masses, this does not change the fact that I can sit on a chair without falling through, or the paths of baseballs in the air.

Basically, in our everyday life we encounter a limited range of speeds (much less than the speed of light), sizes (much smaller than the universe, and much larger than atoms), and gravitational fields (one Earth gravity). In these realms, Newton's laws work very well; we need highly precise experiments to find discrepancies. However, when objects move close to the speed of light, or we look at very small or large things or objects with very strong gravity, it happens that Newton's laws do not work well and we need to seek generalizations. One reason that astronomers and physicists are interested in extreme realms (near black holes, close to the beginning of the universe, etc.) is that historically these are the places where we may find deviations from our current predictions and thus be led to greater generalizations.

How much of a role should aesthetic attraction play?

With all that in mind, in general relativity we have a theory that was poorly tested early on, and which still has few tests of its most extreme predictions. Now we have large amounts of effort expended in trying to find theories that unify all basic forces, e.g., superstring theory. Currently these theories make no predictions that can be tested, but like GR and Newton, they reduce to our current laws in realms we have probed. The only current way to evaluate such theories is to appeal to aesthetics: is the theory beautiful.

How much of a role do you think this should play?