

Chapter 8

Introduction to General Relativity

8.1 The Problem

After 1905 and the success of the Special Theory of Relativity, Einstein turned his attention to the problem of making the other known fundamental force of his time, gravitation, consistent with Special Theory of Relativity. Remember that the electromagnetic theory of Maxwell was consistent with the Special Theory from the start. The other force systems that we now know about such as the strong or nuclear force and weak force had not yet been identified. At this time, gravitation was still described by the action at a distance formulation identified with Newton, see Section B.1. This theory was intrinsically inconsistent with speed of light restrictions on the propagation of energy and momentum. In Newtonian Gravity, the acceleration of the moon due to the presence of the earth as $a_{moon} = -G_N \frac{M_{earth}}{r} \vec{r}$, where \vec{r} is the separation vector between the moon and the earth. If, for some reason, the mass of the earth would change, the acceleration of the moon is instantly changed to accommodate the new mass. The moon instantly changes its orbit to a new one to accommodate the change. In essence, there is momentum and energy transferred to the moon. This implies that the information about the earth's mass in the form of energy and momentum is propagated to the moon faster than the speed of light. This violates the basic premises of the Special Theory.

The theory that he developed was rather long in gestation. It was not until 1916–1917 that he was finally able to articulate the basic principles of what is now called the General Theory of Relativity. This name is both a

misnomer and yet an insightful appellation. It was a modern theory of the effects of gravitation and thus should be called by that name – The Modern Theory of Gravity. But it was only after he took the fullest advantage of the underlying concepts of relativity that he was able to find the correct formulation of the theory and, in fact, it was through a generalization of the principles of relativity that he was able to develop the theory. We will follow this thread of development. The problem is that it is rather abstract and there is some tendency to lose track of the fact that it is a theory of gravity. On the other hand, it has the advantage of making it clear that a modern theory of gravitation is, in fact, a theory of the structure of space-time.

8.2 Free Fall Observers and the Equivalence Principle

In Section 1.2, we discussed the physical implications of Galilean invariance. One of the ways of describing the meaning of this invariance was that you were always at rest in your own rest frame. In other words, there was an infinite set of related observers all of whom thought that they were at rest. Their world was isotropic. An object held out and released would remain there. If the object was given an initial relative velocity, it maintained that velocity. Yet these observers were moving relative to each other. On the other hand, the accelerated observer finds that a released piece of chalk will drift in some direction. The space is no longer isotropic. There are any number of experiments that the inertial, uniformly moving observer, and accelerated observer can perform to note their difference. It is in this sense that we say that, although you cannot measure velocity, you can measure acceleration. There are no speedometers on the starship Enterprise but it can have an accelerometer. It can even integrate the accelerations over time to find a velocity relative to some initial velocity but it cannot know its velocity in any absolute sense. Beside noticing the important point of the unmeasurability of absolute velocity, it is important to appreciate the fact that being inertial is a knowable fact. If you hold out a piece of chalk and release it. It will stay fixed in position. If it suddenly begins to move, you can know that you accelerated. Even more fundamentally, you feel a jolt. We should be a little more careful here. How do you know that it was you and not the chalk that was suddenly accelerated away from you? Putting us again, back in the box of knowing only relative effects, in this case acceleration. It is the jolt that is relevant here. Not only does the chalk start accelerating but a mass and spring held by you changes its configuration –

8.2. FREE FALL OBSERVERS AND THE EQUIVALENCE PRINCIPLE 133

a jolt. In other words, you can build an Inertiality Maintenance Detector. For instance, using identical masses and springs build a three axis stretch meter, see Figure 8.1. To an unaccelerated observer, these six mass-spring systems are all identical. If there is a difference between them, there is an acceleration. This is what is meant by a “jolt.” Thus we can tell if it is us or the chalk that is accelerating. Thus inertiality is an experimentally determined state.

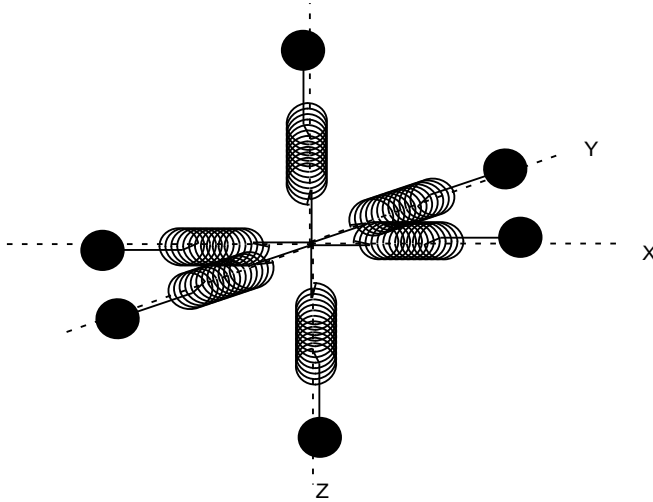


Figure 8.1: **Inertiality Maintenance Detector** Using three pair of springs and masses each pair arranged along each of three axis, we can construct an instrument to detect whether or not we are accelerating or, better said, whether or not we are inertial. Differences in the configuration of the springs will indicate the magnitude and direction of an acceleration.

There is another situation in which there is a detection of inertiality but there is acceleration. If an observer with an Inertiality Maintenance Detector, IMD, is in a gravitational field but is also in “free fall.” This is, for instance, the case when near the surface of the earth an observer is falling or an astronaut is in near earth orbit. For all these cases, an IMD would not show any preferred direction. This statement is actually not quite true and we will have to clarify it later, see Section 8.5. Note, that in any of these free fall situations, there are actually an infinite family of “free fall” observers.

In fact, all observers connected by a Lorentz transformation are equally free fall. This is the difference between the astronaut and the observer that is just falling, a Lorentz boost. These free fall observers are interestingly like the inertial observers that we dealt with in Special Relativity. In the absence of gravity, these free fall observers are the same as our inertial observers of special relativity. We can now make the statement of the first principle of General Relativity. Free fall observers have the same laws of physics as the inertial observers of Special Relativity. This principle is called the Equivalence Principle.

8.3 The Equivalence Principle

The Equivalence Principle states that locally the effects of gravity are indistinguishable from those of an acceleration. This is the same as the observation in the previous section that an observer with an IMD that registers inertiality has the same laws of physics as an inertial observer in Special Relativity.

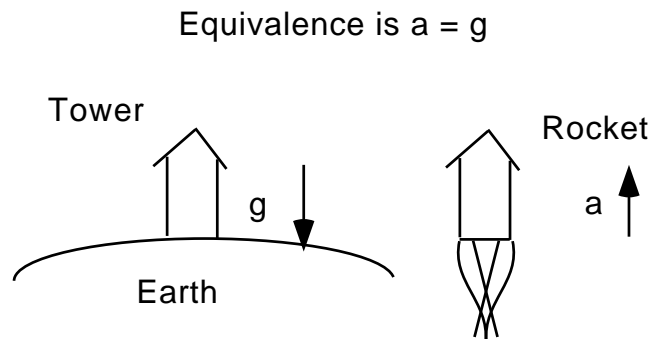


Figure 8.2: **Equivalence Principle** The Equivalence Principle states that locally there is no experiment that can differentiate between the effects of gravity and the a rocket ship with $a = g$.

The Equivalence Principle allows us to now identify some of the important effects of gravity. Using our knowledge of the the physics of accelerated motion in Relativity, see Chapter 6 and particularly Section 6.5. These will be examined in more specific contexts later, see Chapter 10, but for now we will review the simplest implications. Before we get into these cases, let's look at a very popular lecture/demonstration, The Monkey and the Hunter.

8.3.1 The Monkey and the Hunter

There is a very popular demonstration that is performed in most high school and college introductory physics classes. There is a gun of some type that launches a projectile and a target object, usually a toy monkey, that can fall some distance. The gun and the monkey are rigged so that at the instant the gun fires the monkey is released to start to fall.

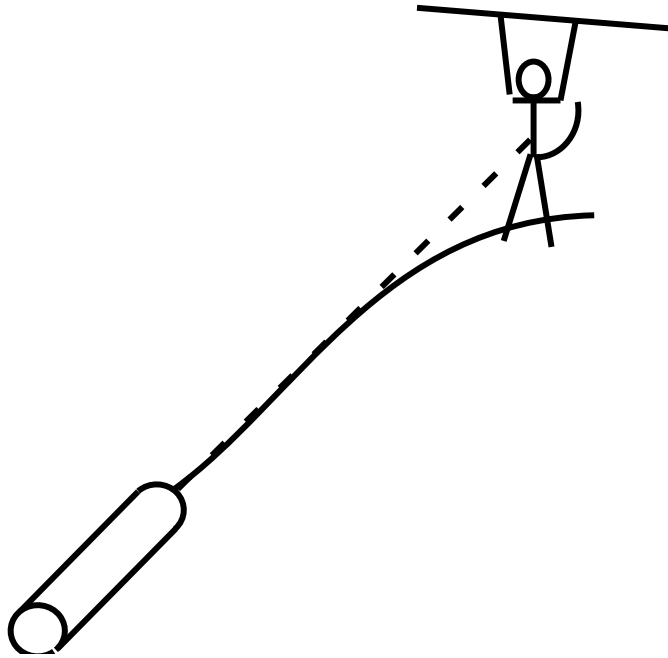


Figure 8.3: **Monkey and the Hunter** A popular lecture demonstration is to fire a projectile at a hanging toy monkey. The monkey is released at the instant that the gun is fired.

The class is usually asked where does the hunter aim. Since the monkey is falling, there is an argument that the hunter should aim below the initial position of the monkey to compensate for the finite time of flight of the projectile. On the other hand, the projectile has an arced trajectory and thus the aim should be above the current position. The correct answer is that the hunter should aim at the present position of the monkey. This is because once the gun is fired both the projectile and monkey are falling with an acceleration of g . In the frame accelerating down at a rate g , the effects of gravity are cancelled and thus neither the monkey nor the projectile have accelerated motion. In that frame, the projectile travels in a straight line

and the monkey never moves. Note that if the aim is correct, no matter how small the projectile muzzle velocity, it will ultimately hit the monkey. This is an interesting pre-relativity example of the equivalence principle. **Also it is important to note that in the free fall frame, the one in which the effects of gravity are removed, the projectile and the monkey have trajectories that are straight lines in space-time.**

8.4 Direct Effects from the Equivalence Principle

The statement of the Equivalence Principle above that the effects of gravity are indistinguishable from those of an acceleration is valid only locally. Measurements over extended regions of space and time can and as we will see show a difference between an acceleration and gravity but the Equivalence Principle provides a basis for some of the more direct effects of gravity. In real situations of mass distributions leading to gravitational effects there are two things that make the following discussion approximate. First, gravity is a field and thus takes on values at all points in space and time. It is just a fact that the dynamics of the gravitational field, called Einstein's Equations, do not admit solutions that are uniform in space and time. There is a similar circumstance in the case of the electromagnetic field. Maxwell's Equations do not admit solutions that are uniform in space and time. For applications of the Equivalence Principle since there is only one acceleration that the frame can have, it can only match a gravitational field at some point. Nearby points will have different values of g and thus will not be eliminated. We will see a case of this in our discussion of the the Gravity Detector, see Section 8.5. Regardless, there will be many cases when the gravitational system of interest can be well approximated by a uniform field and we will do so in the following. Second, any measurement apparatus will have some extension and thus the effects will have to take into account the extended effects of gravity. Again, in many circumstances, the measuring apparatus is small in extent compared to the region of interest and the measurement can be considered local. Clearly, a legitimate approximation. With these provisos, we proceed to look at some of the simple direct effects of gravity.

8.4.1 Universality and Eötvös–Dicke

One of the most striking features of Newton's Theory of Gravity is its universality. The great idea that behavior of apples falling from trees and the moon in orbit were two aspects of the same law was one of its first significant philosophical and phenomenological successes of the theory. Not

only does it effect all things, it effects them in the same way. Again, an interesting lecture demonstration is almost always performed in high school physics classes. A penny and a feather are enclosed in an evacuated glass tube. Inside the tube where the only significant forces on the penny and the feather are gravity, they fall together. The Equivalence Principle gives an immediate explanation to these two simple aspects of the universality of the Newton's Theory. All objects move the same in gravity because it is the observer that is accelerated.

Interestingly, Newton achieves universality in an indirect way and in several steps. First, gravity sees only mass and no other attribute of the object. Then it identifies two distinct roles for mass and then arbitrarily equates them. This issue of the relationship between gravitational and inertial mass was discussed earlier, see Section ???. Let's be more specific. Newton first ascribes the force of gravity to an action at a distance force law, see Section B.1, that is based on the identification of mass as the source of the strength of the force. The gravitational force between two bodies labeled 1 and 2 is

$$\vec{F}_{Grav12} = -G \frac{m_{gr1} m_{gr2}}{r_{12}} \vec{r}_{12}, \quad (8.1)$$

where \vec{r}_{12} is the displacement vector from body two to body one. The two masses in this equation are called gravitational masses, indicated by the subscript gr , and are the source fo the gravitational force. These masses are measured for instance in a balance scale. Body one reacts to the force by having an acceleration according to Newton's Laws as

$$m_{i1} \vec{a}_1 = \vec{F}_{Grav12} = -G \frac{m_{gr1} m_{gr2}}{r_{12}} \vec{r}_{12}. \quad (8.2)$$

where the m_{i1} in the first part of Equation 8.2 is the inertial mass of body one. One way in which this mass could be measured is by collision with a standard mass. The next step is to invoke the magic idea that these two very different concepts of mass are identical, $m_{i1} = m_{gr1}$, and cancel them from the two sides of the equation so that the acceleration no longer depends on the mass of body one. As is emphasized in Chapter ??, the standards and protocol for measuring something is its definition. Here we have two very different definitions of mass that would have two very different protocols for measurement. This equality of the two masses is even more striking in light of the mass energy relationship, Equation ??, and the constituent nature of matter. Yet these two different things are the same, strikingly the same.

The equality of the gravitational and inertial masses was tested in a classic experiment in 1889 by Roland von Eötvös and recently improved by

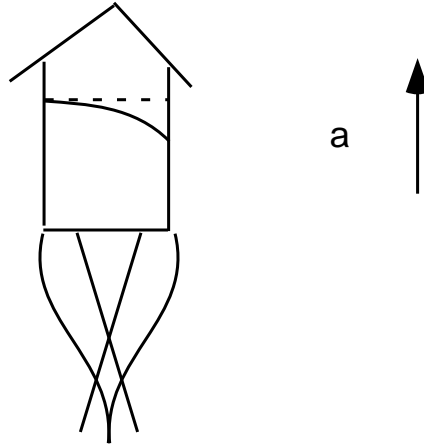
R. H. Dicke, [Eötvös 1922, Dicke 1967]. The idea is that the measurement of the gravitational force in a rotating system is influenced by the non-inertiality of the laboratory and the effects of each are proportional to the gravitational to inertial masses and thus the corrections are proportional to the ratio of the inertial mass to the gravitational mass. Newtonian universality required that this ratio be one. Using different materials in each leg of a torsion bar, the experiment could detect differences between the ratio for these materials. Eötvös found that the difference between wood and platinum was less than 10^{-9} and Dicke improved this limit for aluminum and gold to 10^{-11} . These small differences are very impressive especially in light of our new understanding about mass and energy as discussed in Section 7.4. Gold or aluminum atoms have very different atomic and nuclear structures and thus different energies of binding. These differences are well within the measured precision of this experiment. Thus even if protons, neutrons, and electrons have identical inertial and gravitational masses, in these systems, the binding would manifest a detectable difference.

The Equivalence Principle directly requires the Universality of Newton's gravity and includes the result of the Eötvös–Dicke experiment without further assumptions.

8.4.2 Bending of Light Rays

Consider a rocket accelerating in a region of space that has no nearby masses and thus is free of gravitational effects. In Figure 8.4, it is clear that a light beam entering one side of the rocket perpendicular to one wall will have a bent trajectory as measured in the rocket.

Using the Equivalence Principle, light in the neighborhood of a massive body must also bend. We can even be more quantitative. The time of passage of the beam across a rocket of width L is $\frac{L}{c}$. If the acceleration is g , the deflection on the far side of the rocket is $\frac{g}{2} \frac{L^2}{c^2}$. This calculation can be carried out with more care using the information that we have on accelerated observers, see Section 6.5, but this result is certainly the correct order of magnitude. For the earth, the deflection in a one kilometer size room, a big room, is 5×10^{-11} meters, too small to be measured. This effect though has been measured for the case of the bending of star light by the sun. A classic experiment using a total eclipse of the sun was among the first verifications of the General Theory of Relativity of Einstein. It is important to note that the bending of star light predicted by the equivalence principle alone does not produce the full bending but, to get the correct value, will require that we use the full metric theory that is developed later, see Section 9.9,



Light Rays in an accelerating rocket

Figure 8.4: **Bending of Light** A light ray entering one side of an accelerating rocket will be seen in the rocket as bending down. The Equivalence Principle then requires that a beam of light bend in the presence of a massive body.

[Will 1986], and [Weinberg 1972].

8.4.3 Clocks and Accelerations in Towers

In Section 6.4.2, we study the behavior of clocks in an accelerated rocket. There we find that clocks at the top and bottom run at different rates and that the relationship between them is given in Equation 6.9 as

$$\tau_{top} = \left(1 + \frac{ha_{bottom}}{c^2} \right) \tau_{bottom} \quad (8.3)$$

where h is the length of the rocket and a_{bottom} is the acceleration. It is important to realize that the top of the rocket is a fixed proper distance from the bottom. This keeps the rocket a fixed length as measured on the rocket. Because of this requirement is also important to note that the acceleration of the top of the rocket is not the same as that measured at the bottom of the rocket. Using the relationship between the acceleration of the uniformly accelerated observer and the distance to the magic point, Equation 6.2, these accelerations are related by

$$a_{top} = \frac{c^2}{d_{top}} \quad (8.4)$$

but d_{top} is $d_{bottom} + h$ so

$$\begin{aligned}
 a_{top} &= \frac{c^2}{d_{bottom} + h} \\
 &= \frac{\frac{c^2}{d_{bottom}}}{1 + \frac{h}{d_{bottom}}} \\
 &= \frac{a_{bottom}}{1 + \frac{h a_{bottom}}{c^2}}
 \end{aligned} \tag{8.5}$$

These phenomena associated with an accelerated rocket are nicely summarized in the examples on accelerated rockets and Bell's Problem, see Section 6.4.2 and Section 6.4.3.

Thus, again invoking the Equivalence Principle, we have that, in a tower near or on a massive body, clocks at the top and bottom of the tower run at different rates. Setting $a_{bottom} = g$ the local gravitational field, these rates follow directly from Equation 8.3 where h is the height of the tower and g is the gravitational field at the location of the tower. A simple insertion of values into this equation would seem to indicate that it is not testable in an earth based laboratory. For a tower of height h in meters, the fractional change in rate between the bottom and top clock is $\frac{\Delta t}{t} = \frac{10 \frac{\text{m}}{\text{sec}^2}}{9 \times 10^{16} \frac{\text{m}^2}{\text{sec}^2}} h \approx 10^{-16} \text{ m}^{-1} h$. This would appear to be a forbidding shift to measure but since precision time sources are available and, with a clever trick to identify the signal from the background, Pound and Rebka have measured this shift, [Pound & Rebka 1959].

Of course, if we could have towers several kilometers tall, there would be no problem in conducting these experiments. The trouble is that our formula is valid only in the cases in which the gravitational field strength is a constant. The effect though is universal. In the presence of gravity, clocks run slower at the bottom than clocks at the top. If we use the full power of the Einstein Equations, Section 9.11, cases of a varying field strength can be treated and this shift to lower frequencies called a red shift is observed in radar ranging experiments to the moon. In addition, with the advent of earth satellites in low earth orbit, these effects will also be realized. In fact, the GPS positioning system has to be corrected for these effects. An application of General Relativity in everyday life.

8.5 Intrinsic Effects of Gravity

Consider an observer near a massive body, a free fall observer above the surface of the earth for instance. Using an IMD, Inertial Maintenance Detector, see Section 8.2 and Figure 8.1, the observer concludes that he/she is in free fall. There is no distortion of the masses in the IMD that would indicate an unbalanced force. A piece of chalk released by the observer at his/her location hovers where it is released. This is the essence of the Equivalence Principle. The acceleration has removed the effects of gravity. Despite this, the IMD is not uniform in all six axis. If the IMD is oriented so that one of the axis is along the line to the massive body, the the two masses along that axis are slightly further apart then the mass pairs in the two other axis directions that are in the plane parallel to the surface of the massive body. The elongation is twice the compression. There is no state of motion that the observer can carry out that eliminates this distortion. Even if he/she accelerates, there will be a distortion identified with the unbalanced gravitational force but there will also be this unusual distortion. Thus we conclude that the Equivalence Principle cannot remove all the effects of gravity. There always remains a distortion which stretches along the axis directed at the mass and compresses half that amount in the plane perpendicular to that axis. The magnitude of the distortion is proportional to the mass of the gravitating body, inversely proportional to the cube of the distance from the gravitating body, and the size of the IMD. A distortion of a elastic system in which the system is stretched in one direction and compressed in the other two is called a tidal distortion.

8.5.1 Distortion of Elastic Bodies

In an elastic mechanical system, there can be distortions of the system in which is little net motion but only relative motion between parts. The system is deformable. An elastic rubber band stretches, see Figure 8.5.

Another simple distortion is shear. A common and simple way to produce this distortion is by placing a large phone book, not really elastic since the phone book will retain the distortion, on a table face up and on the top of the phone slide your hand across the flat top surface. The cross section of the phone book will change from a rectangle to a rhombus. This is shear, see Figure 8.6.

A general property of a shear distortion is that although there is relative displacement of the parts the enclosed volume is retained as the distortion takes place. In deformable bodies, shear is a very common phenomena. An

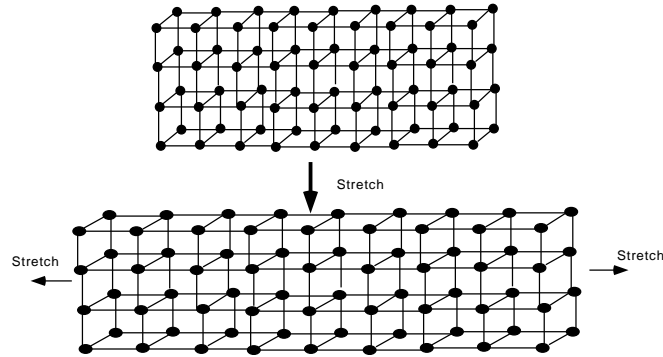


Figure 8.5: **Stretch of an Elastic Solid** The stretch distortion of an elastic solid. A pulled rubber band is an example of a stretched elastic solid

interesting example is that a Pascal or perfect fluid in hydrostatics can be defined as one which will not sustain a shear forming distortion. This is the reason that you cannot pile water. This is the direct manifestation that pressure is a scalar quantity,

$$\vec{F} = P\vec{A} \quad (8.6)$$

and that the hydrostatic force is directed along the normal to the area.

The stress that leads to shear deformations is called a shear stress. Generally, these are couples, a pair of equal and opposite pair of forces acting at a slight separation, not a single force.

The distortion that is manifest in our IMD is a tidal distortion, see Figure 8.7. In this case, the body extends along one axis and compresses on the other two orthogonal axis. As shown below, Section 8.5.2, to first order in the stretch, this distortion also has the property that it is volume preserving. It stretches twice the contraction but there are two contraction directions so that the total effect does not change the volume.

Again, this distortion is reasonably common and the most well known manifestation is the oceanic tides of the earth and thus the name. Although most of the explanations of the origin of the tides is a complex analysis of the the gravitational attraction of the moon and the center of mass motion of the earth due to the earth moon orbit, it is really that the ocean is an IMD, a bunch of water—an incompressible perfect fluid, for the earth and that the earth is in free fall in the gravitational field of the moon. Thus the direct acceleration effects of the moons gravity are eliminated but the tidal distortion of the gravitational field remains. You can find the shape of the tides on the earth by combining the gravitation from the earth to the tidal

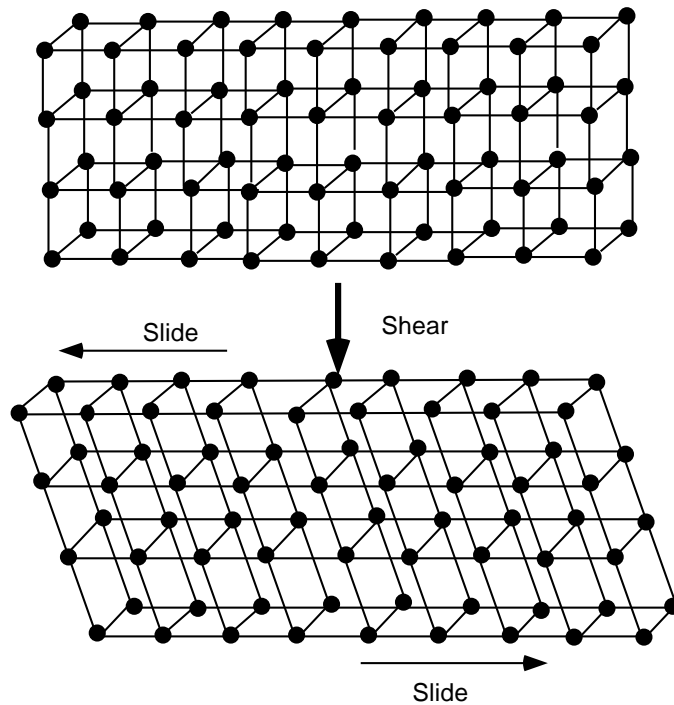


Figure 8.6: **Shear of an Elastic Solid** The shear distortion of an elastic solid. In shear two parallel planes are displaced relative to each other. Volumes are preserved.

force. The shape of the surface of the liquid is the one that everywhere has its normal along the net gravitational force and has the correct volume.

8.5.2 Gravitation and Tidal Forces

Returning to our main theme, we now understand that the Equivalence principle provides a means for the elimination of gravitational forces but does not eliminate the tidal stress that is the intrinsic signature of the presence of gravity. No motion based or any other type of coordinate relabeling can eliminate this aspect of gravity. More on this later, see Section 9.9. In order to understand its implications better, let's look at a system like our Inertiality Maintenance Detector but actually a little simpler, basically no springs, and do this a bit above the surface of the earth. A free fall observer places several independent masses in a sphere that surrounds him/her and one at his/her location, see Figure 8.8. The released masses are independent

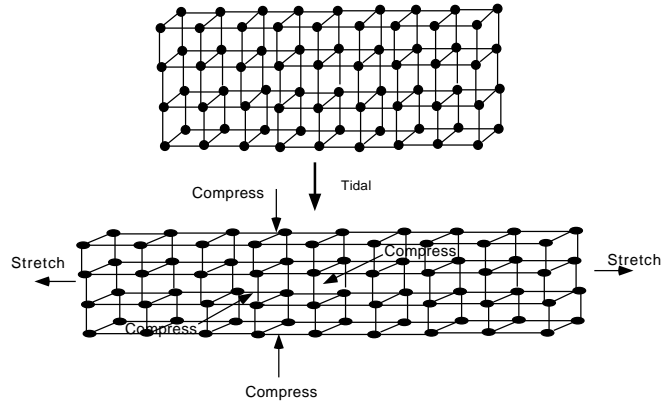


Figure 8.7: **Tidal Distortion of an Elastic Solid** The tidal distortion of an elastic solid is an extension along one axis and compressions in the other two. Volumes are preserved.

in the sense that, once placed, they are in free fall. There are no external forces acting on them except gravity but they are released so that they are not moving relative to the original observer. They are commoving at $t = 0$.

As time develops, there is a tidal distortion of the sphere with the expanding axis along the line to the center of the earth. It is easy to understand the basis for this distortion in the conflict between nature of the gravitational field and the direct application of the Equivalence Principle. First consider the three masses along the axis to the center of the earth, called the in/out axis. For definiteness, let's use δ as the initial radius of the sphere. Since the gravitational field is different at the three locations, the corresponding free fall accelerations are different and there is a relative acceleration between them. The three gravitational field strengths are $g_{top} = g \left(\frac{R_e}{R_e+h+\delta} \right)^2$, $g_{center} = g \left(\frac{R_e}{R_e+h} \right)^2$, and $g_{bottom} = g \left(\frac{R_e}{R_e+h-\delta} \right)^2$ where R_e is the radius of the earth and h is the height above the earth of our free fall observer, the center of the sphere. Since all three are in free fall these must be their accelerations. Thus the two relative accelerations between the center and top and bottom are

$$a_{rel\ top} = 2g_{center} \frac{\delta}{R_e + h} \tag{8.7}$$

and

$$a_{rel\ bottom} = -2g_{center} \frac{\delta}{R_e + h} \tag{8.8}$$

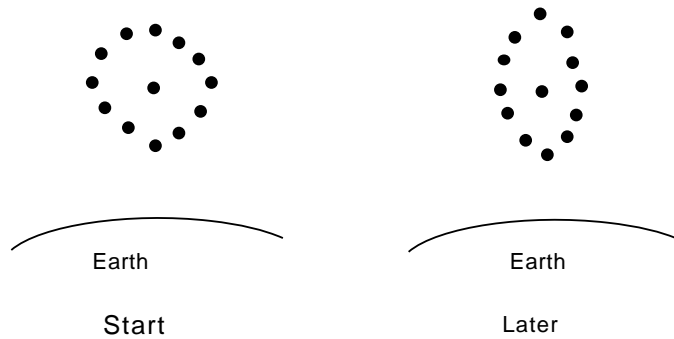


Figure 8.8: **Tidal Distortion of Free Fall Masses** A free fall observer arranges a sphere of identical free fall masses that are commoving. After a time, the masses are no longer spherical but the masses along the line to the center of the earth begin to separate and the masses in the plane tangent to the line to the center begin to come closer together. As the motion develops, the volume of the sphere is preserved.

to first order in $\frac{\delta}{R_{e+h}}$. In this case the masses move apart.

Thus a plot of these trajectories as observed by the central free faller is shown in Figure 8.9.

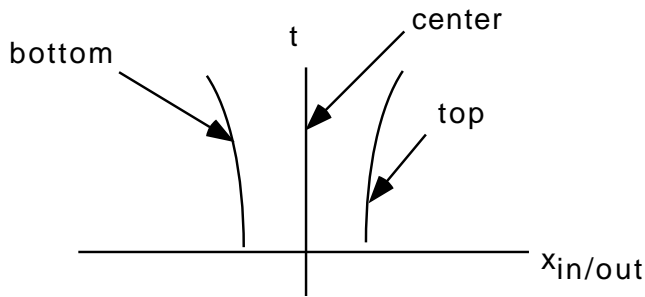


Figure 8.9: **Trajectories of Top and Bottom Free Fall Masses** Over time, the three free fall masses along the axis to the center of the earth, the $x_{in/out}$ axis, move apart. They are initially commoving.

The important point to note is that all three of these trajectories are for free fall objects, objects that have simple physics and thus these trajectories are of objects that are “inertial” and are straight line trajectories. These are three straight lines that start out parallel and as time develops drift apart. This is clearly not the geometry of Euclid. In Section 9, we will discuss the

meaning of this non-Euclidean geometry.

Picking an axis in the plane tangent to the earth and calling it the sideways axis, the three masses have their free fall accelerations directed differently. All three point to the center of the earth, see Figure 8.10.

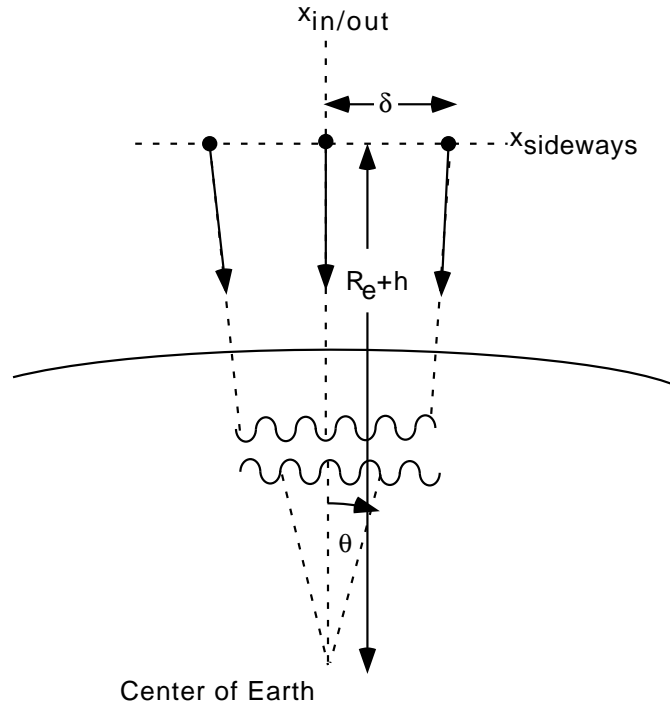


Figure 8.10: **Free Fall Accelerations in the Sideways Direction** For masses released along an axis in the plane tangent to the earth's surface, the free fall accelerations which are directed to the center of the earth are in different directions. Thus there is a relative acceleration among the masses.

Thus although the magnitudes of the free fall accelerations are the same, there is a relative acceleration between the masses. Projecting along the in/out and sideways direction and using the fact that θ is small and that $\sin(\theta) \approx \tan(\theta) = \frac{\delta}{R_e+h}$, the relative acceleration to first order in $\frac{\delta}{R_e+h}$ of either sideways mass relative to the center mass is

$$a_{rel \pm sideways} = \mp g_{center} \frac{\delta}{R_e + h}, \quad (8.9)$$

where, again, δ is the radius of the initial sphere of free fall masses. Thus, the

sideways masses move toward the center, see Figure 8.11. Again, we have

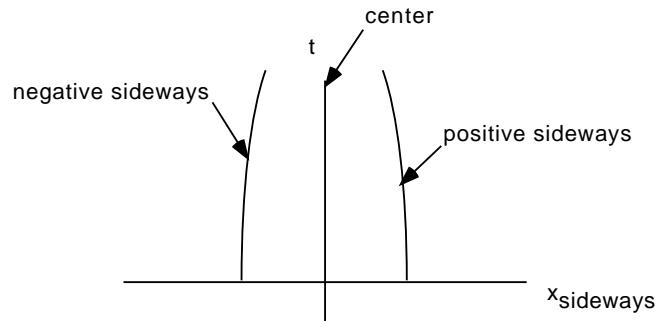


Figure 8.11: **Trajectories of Sideways Free Fall Masses** Over time, the three free fall masses along the axis in a plane tangent to the surface of the earth, the x_{sideways} axis, move apart. They are initially commoving.

a situation in which three initially commoving free fall objects, straight line trajectories, are moving toward each other. Again, a direct violation of Euclid's axioms and thus the geometry that is non-Euclidean.

Thus, we see that in all three space-time two planes the geometry is non-Euclidean. To make progress, we will have to understand a little bit of geometry.

