3.5: General Relativity

The Equivalence Principle

One thing that comes up over and over in special relativity is the role of inertial frames. We talked about accelerated frames as well, but those followed different rules. The most notable of these is the effect that the acceleration of a frame has on the measurement of the proper time. If an observer carries a clock in their non-inertial frame and measures the time between two events that occur at the same place in that frame with it, the time measured is less than when an observer in an inertial frame carries a clock and does the same with those two events.

This phenomenon is reflected in the shape of the world line of the origin of the non-inertial frame in a spacetime diagram in the rest frame of an inertial observer. By contrast, the shape of the world line of the origin of an inertial frame is always straight when drawn in an inertial observer's space time diagram. The intervals defined by all the straight world lines come out the same.

Einstein must have worried if perhaps it is possible for the world line of an inertial frame to be curved in the spacetime diagram of another inertial observer (it is unlikely he worried about it in exactly these terms, but we'll run with this). This would cause problems, because the Lorentz transformation that links inertial frames only rotates world lines, it doesn't cause them to curve. It turns out that Einstein came up with an example that causes this exact problem!

To see how this problem can arise, let's first return to something we studied in Physics 9HA. When we talked about free-fall, we noted that because every object in an enclosed space experiences the same acceleration when that space is in free-fall, an observer in this frame will assume that nothing in the space is experiencing a force. Einstein elevated this to a fundamental principle that sounds very much like the relativity principle, and called this one the *equivalence principle*:

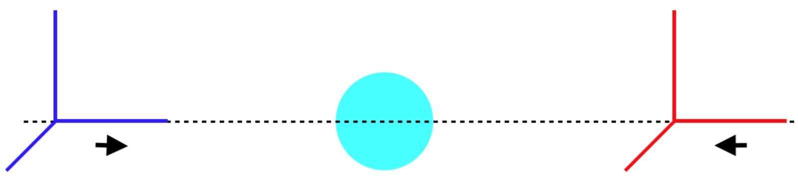
No experiment can be performed that will distinguish a frame in gravitational free-fall from an inertial frame.

The Problem

If this principle is correct, then the following thought experiment raises a problem...

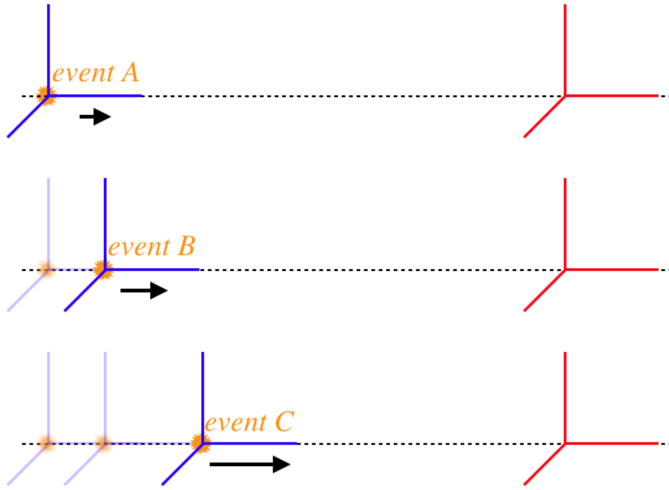
Suppose we have two free-falling frames on opposite sides of a planet. They are moving toward each other in a symmetric fashion, with their common x-axes passing through the center of the planet:

Figure 3.5.1 – Two Free-Falling Frames



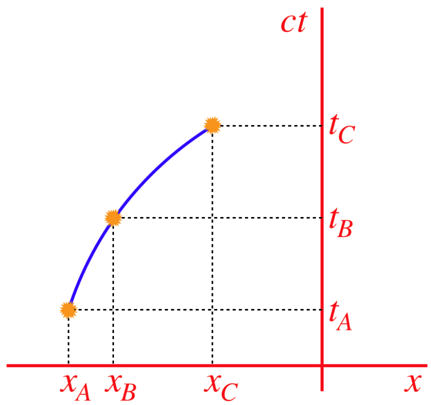
According to the equivalence principle, *both* of these frames are inertial. Now suppose the observer in one of these frames witnesses three events that occur at the origin of the other frame separated by equal coordinate time intervals. The other frame is speeding up relative to the observer's frame, which means that the spatial separation of the first and second events is smaller than that of the second and third events. We'll choose the red frame as the observer frame, and obscure the planet for clarity:

Figure 3.5.2 – Three Events in Free-Falling Frames



To see why this causes us a problem, we now have a look at the spacetime diagram of these events in the observer frame:

Figure 3.5.3 – Spacetime Diagram of Free-Falling Frame



The first thing we notice is that the slope of the world line of the origin of the other frame must be changing, because the three spacetime events are not aligned in the diagram. If we fill-in lots of other events as well, we see that the observed frame has a world line that is a continuous curve that is *not straight*. We have seen this before: Whenever the world line of an object not in an inertial frame is drawn in the spacetime diagram of an inertial frame, we get a curved path through spacetime.

The problem here is that the observed frame *is inertial*, according to the equivalence principle! In fact, we can reverse the roles of these frames – the blue frame can claim to be stationary, and then the spacetime diagram of the world line of the red frame's origin will be curved. Essentially what we have here is two frames that can both claim to be inertial, but when one measures the motion of the other, the second derivative of position with respect to time does not vanish. By all we have studied so far, this constitutes acceleration, which is synonymous with non-inertial.

Einstein's Solution

Clearly gravity is the culprit here – somehow it is redefining how inertial frames relate to one another when they are separated in a region where gravity exists. Our use of Lorentz transformations to related frames needs to be generalized to include cases that involve gravity. This is why what we have studied so far is called *special* relativity. Incorporating gravitation into the same theoretical relativity framework is called general relativity.

In the thought experiment above, the blue frame is inertial, and the events all occur at the same position, which means it measures the spacetime interval (the longest proper time interval) between events. For this to be an invariant, it means that the interval measured in the red frame following the path shown in the diagram above must be the same value. But this means that we cannot equate the intervals in the way we have done previously:

$$\Delta s^2 = c^2\Delta\tau^2 \neq c^2\Delta t^2 - \Delta x^2 \tag{3.5.1}$$

The reason this is no longer equal is that this relation only holds if the world line is straight (see the discussion of Equation 2.1.7 for a refresher on this). Recall that this same relation applied to the invariance of the lengths of position 4-vectors, and by extension all other 4-vectors as well. We certainly don't want the whole theory to come crashing down, so what is the solution?

Taken in small enough steps, the world line does "look straight," but the machinery we have created around the need for straight world lines requires that it be straight everywhere. So let's instead make the Minkowski metric (which we use in scalar products, including those that define magnitudes of 4-vectors) able to change from place-to-place. That is, instead of the dot product incorporating the simple Minkowski matrix shown in Equation 3.2.3, let's introduce one that can change from one point to the next in spacetime:

$$X\cdot X=\left(ct\;\;x\;\;y\;\;z\right)\begin{pmatrix} g_t\left(ct,x\right) & 0 & 0 & 0 \\ 0 & -g_x\left(ct,x\right) & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}\begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix}=g_tc^2t^2-g_xx^2-y^2-z^2=\Delta s^2 \tag{3.5.2}$$

This new matrix that defines dot products represents what is called the *metric tensor*. What we have done here is a greatly over-simplified version (so much that we won't be doing any specific mathematics in this area), but the basic idea is correct. In essence, the presence of gravity changes the way intervals are measured, making it different from the usual Minkowski case. The simplest way to describe this change is to say that the *spacetime itself is curved* by the presence of gravity. When there is no gravity present, then the metric coefficients g_t and g_x converge back to 1, and we say that the space is "flat," giving us once again special relativity.

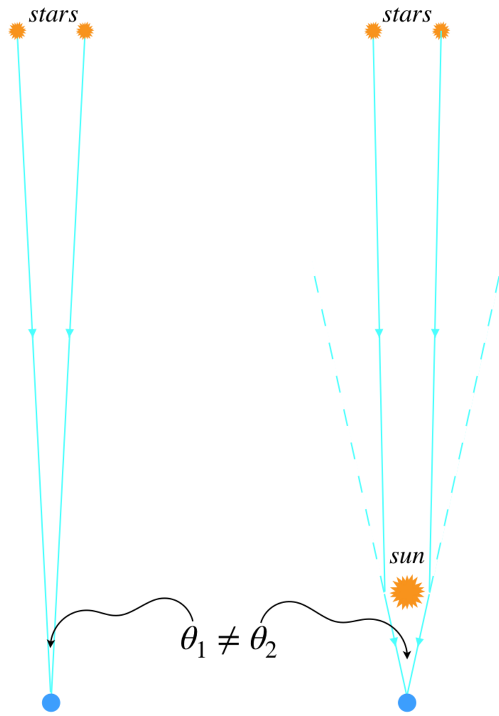
Of course the "hard part" of general relativity is determining how the metric coefficients arise from what we already know is the source of gravity – mass-energy. This is something Einstein spent so many years on (it was 10 years between his publications of special and general relativity), largely because he had to learn a field of mathematics called differential geometry, which essentially deals with curved surfaces. Unfortunately, this lies well beyond the scope of this course.

Effects of Gravity

Now that we know gravity curves spacetime itself, we can discuss a few of its measurable effects. Perhaps the most direct evidence came from the result of the experiment that is generally considered to be the premier confirmation of the theory, and it came more than three years after Einstein's publication of the theory. Light is massless, so from Newton's theory, it should not be effected by gravity. According to general relativity, light should follow a specific world line through spacetime, and gravity bends world lines, so we should be able to measure this bend.

The experiment that made this measurement looked at two stars in the night sky in two ways: The first way was directly – in the night sky. The angular separation of the positions of the stars was noted. The second viewing of these stars was during a full solar eclipse, when they could be seen on opposite sides of the sun. The effect of the sun's gravity on the light coming from these stars would be to bend the spacetime such that the light rays deflect slightly compared to when the sun is not present. This would be manifested as an apparent shift in the angular separation of the stars. Obviously the figure below is not drawn to scale – the dimensions are exaggerated to clearly show the effect.

Figure 3.5.4 – Bending of Starlight



One might try to argue that we can give light an "effective mass" using mass-energy equivalence, then use Newton's theory to predict the effect that gravity has on it, but in fact this approach makes the wrong prediction for the deflection of the light!

Of course gravity also has effects on measured spatial lengths and intervals of time, the latter of which has been measured by atomic clocks placed at different elevations on the Earth (thereby experiencing different gravitational fields), it creates a doppler-like effect on light escaping a gravitating body (called gravitational red-shift), and it makes new predictions for planetary orbits that are different from Newton's (which have also been confirmed), but the relation of these phenomena to the curving of spacetime is not as dramatic as the case of the deflection of starlight.
