Massachusetts Institute of Technology
Department of Physics
8.033 Fall 2024

Lecture 17
Goodbye global Lorentz frames, hello principle of equivalence
Initial considerations on relativistic gravity

17.1 Farewell to global Lorentz frames

What is it that puts the "special" in special relativity? The key concept that we come back to again and again is the notion of a *Lorentz frame*: A frame of reference in which things move at constant velocity if no forces act on them. Such a frame is an inertial frame; we move between different Lorentz frames using Lorentz transformations.

What is particularly special about special relativity is that it assumes that we can "cover" all of spacetime — all events, all time and all space — using a single Lorentz frame. In other words, special relativity tells us that it makes sense for there to be *global* Lorentz frames.

Gravity breaks this. Once we begin including gravity in our model of physics, we cannot have a global Lorentz frame that covers all events. This is actually fairly easy for us to see based on things that we have already learned about the nature of Lorentz frames, and the influence of gravity on light.

Imagine a pulse of light that propagates from the surface of the Earth to a height H. Let us imagine the trajectory that one crest of a light wave in this pulse follows through spacetime. We do not yet know exactly how gravity will affect the pulse's path through spacetime, but we can imagine that the trajectory is "bent" essentially, perhaps moving a little slower near the surface than it moves after propagating to greater heights:

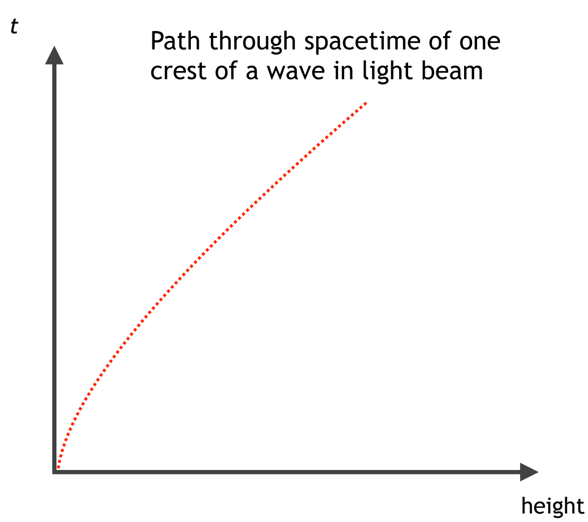

Figure 1: A plausible path for the crest of a light wave in a pulse propagating vertically from the Earth's surface.

Given this behavior of the first crest, what is the behavior of the second crest? Well, *if* we require spacetime to be Lorentz everywhere, then there is nothing special about any particular time or place. The path of the second crest must be identical to the path of the first one, simply shifted later in time.

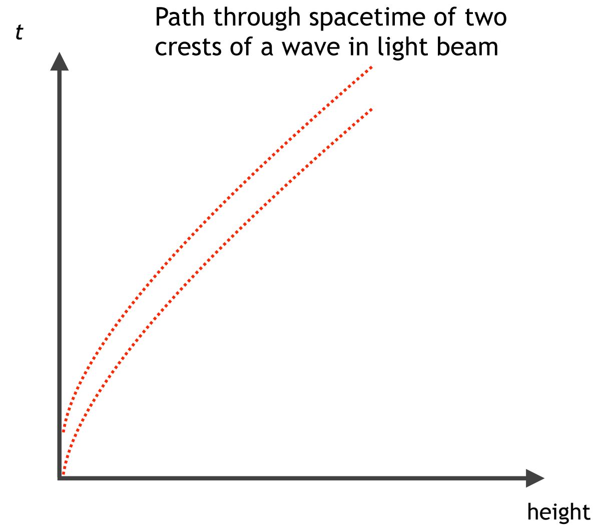

Figure 2: If the trajectory through spacetime of crest 1 looks like Figure 1 *and* we assume spacetime has Lorentzian behavior everywhere, then the trajectories of crests 1 and 2 together will look this. Figure made by duplicating the first crest and sliding it slightly in the t direction.

If this is true, then it must be the case that the wave period at the bottom (height 0) must be identical to the wave period at the top (height H). The two trajectories are *congruent* with each other, simply shifted a bit along the time direction. But if the periods T_H and T_0 are identical, then the frequencies at the top and the bottom are identical: $\nu_H = \nu_0$. **This contradicts the gravitational redshift that we argued must exist** (and that, indeed, experiments have demonstrated does in fact exist), which tells us that $\nu_H = \nu_0(1 - gH/c^2)$. Our starting assumption must be incorrect: *In the presence of gravity, we cannot have global Lorentz reference frames.*

Perhaps we could "rescue" special relativity with the Rindler coordinate system. Rindler coordinates express how things look in special relativity according to a uniformly accelerated observer; we saw that an analysis of light measured by such an observer looks very similar to the expressions we derived for the impact of gravity on light. However, the Rindler coordinate system describes uniform acceleration along a particular direction in space. With a little thought, we can convince ourselves that a Rindler coordinate system cannot describe *all* the measurements that we can make on the Earth's surface.

Consider an observer on the equator who measures the gravitational redshift. They can interpret their measurements as consistent with a Rindler coordinate system that is accelerating "up," i.e., outwards from the equator. Consider a second observer at the North Pole who measures exactly the same gravitational redshift. They likewise may want to interpret the redshift as due a Rindler coordinate system that is accelerating "up." However, their "up" is 90° different from the "up" of the equatorial observer! Consider a third observer at the South Pole. They also want a Rindler observer accelerating "up," but their "up" is 180° different from the North Pole's "up." None of these observers are in fact moving with respect to one another: they are widely separated, but their separations are not changing. This is starkly different from accelerations in three different directions which the Rindler hypothesis requires.

We need a new idea in order to incorporate gravity in the framework of relativity.

17.2 The principle of equivalence

Let's go back to the foundation of what an inertial frame has meant: *In the absence of external forces, all objects maintain their relative velocities.* Is there any way in which the essence of this idea can be captured when we include the action of gravity?

One of Einstein's core insights was that we observe exactly the same thing when we do our analysis in a *Freely Falling Frame*, or FFF. All objects feel the same acceleration due to gravity, thanks to the fact that $\mathbf{F} = m\mathbf{a} = m\mathbf{g}$. The equivalence of "gravitational mass" and "inertial mass" means that gravity effectively cancels out as long as we can work entirely in the FFF. The notion of a Lorentz frame is now upgraded to a Freely Falling Frame, and the rule that we will use is: *In the absence of non-gravitational forces, objects maintain their relative velocities in a Freely Falling Frame.*

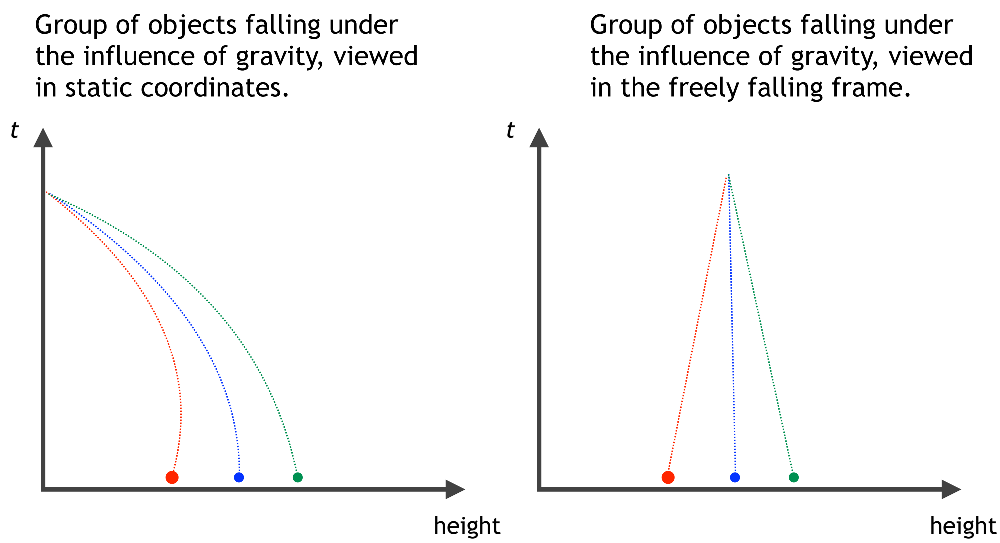

Figure 3: Three objects falling under the influence of gravity. In "static" coordinates (e.g., coordinates at rest with respect to the Earth's surface, shown on the left), the three bodies follow parabolic trajectories before meeting later at height 0. In the freely falling frame, the observer follows the same trajectory as the blue object. All three objects move along straight lines in this frame. Motion in the freely falling frame duplicates the essential features of unaccelerated motion in an inertial frame in the absence of gravity.

The key intuition for this is that, as summarized by Einstein, we cannot distinguish between gravity and uniform acceleration. It is important to bear in mind, however, that in realistic situations gravity is never perfectly uniform. As we move away from the Earth's center, the gravitational force gets weaker. This variation in gravity from the Earth, or from any realistic finite-sized source, is responsible for *tides*.

Tides are responsible for a key aspect of how we describe gravity in relativistic language. In special relativity, if two objects started out moving parallel to one another and no force acted on them, their trajectories would *always* remain parallel. This is a statement that trajectories in spacetime obey what is known as "Euclid's parallelism postulate," an aspect of Euclidean geometry which confused mathematicians and geometers for centuries. Unlike

Euclid's other postulates, the parallelism postulate was not considered to be self evident, and could not be proved under the assumption of Euclid's other postulates. Work by the Russian mathematician Lobachevsky first showed that one could set up a logically consistent framework for geometry without assuming this postulate; in such a geometry, lines which start out parallel later cross or diverge from one another. The German mathematician Riemann later worked out rules describing such geometries.

In modern language, we now say that if Euclid's parallelism postulate holds then it means that one is working in a geometry that is *flat*. In two and three spatial dimensions, a flat geometry in which the Pythagorean theorem holds; in space and time, it is a geometry with the metric $\eta_{\alpha\beta}$ that we have been working with for most of this semester.

On the other hand, if Euclid's parallelism postulate does not hold, then one is working in a geometry that is *curved*. An example is the surface of a sphere. Consider two observers standing on the Earth's equator. Both begin walking due north — perfectly parallel to one another. They walk in a perfectly straight line on the surface, never bending their path from one moment to the next. Despite beginning on parallel trajectories, and despite moving along perfectly straight lines, their trajectories cross when they reach the North Pole.

Tides cause trajectories which are initially parallel in spacetime to either focus or diverge from one another. This tells us that *when we have gravity with tides, spacetime must be curved*. We cannot use the metric $\eta_{\alpha\beta}$ anymore; we need something new.

17.3 How to describe relativistic gravity I: Initial considerations

Let's think about Newtonian gravity for a moment. Begin by considering the potential outside of a spherical mass M,

$$\Phi = -\frac{GM}{r} \; . \tag{17.1}$$

This gravitational potential has the same mathematical form as the electrostatic potential that arises from a spherical charge Q:

$$\Phi^E = \frac{1}{4\pi\epsilon_0}\frac{Q}{r} \; ; \tag{17.2}$$

we just need to replace $Q \to M$ and $1/(4\pi\epsilon_0)$ with $-G$.

In an in-depth study of electrostatics, we learn that for a *general* distribution of charge, the electrostatic potential Φ^E is the function that solves *Poisson's equation*:

$$\nabla^2\Phi^E = -\frac{\rho_Q}{\epsilon_0} \; , \tag{17.3}$$

where ρ_Q is charge density. This can be proven by combining $\mathbf{E} = -\nabla\Phi^E$ with $\nabla\cdot\mathbf{E} = \rho_Q/\epsilon_0$. In the same way, one can show that in Newtonian gravity, the gravitational potential Φ^G is the function that solves a slightly different version of Poisson's equation:

$$\nabla^2\Phi^G = 4\pi G\rho_M \; , \tag{17.4}$$

where ρ_M is mass density.

Let's begin here as we start thinking about how to bring gravity into a relativistic framework. We start by cataloguing the ways in which Eq. (17.4) falls short as a relativistic equation, and imagine ways in which we could perhaps "upgrade" it to something better for our purposes.

- The left-hand side of Eq. (17.4) involves spatial derivatives in a particular reference frame. This is not a Lorentz-covariant derivative operator. One idea for upgrading this: replace ∇^2 with the relativistic wave operator $\Box$. Past studies of gravity were dominated by sources that were static or very slowly varying; perhaps the most important aspects of gravity have been determined from sources for which $\partial(\text{gravity})/\partial t \approx 0$ in the frames in which we did these studies.

- The right-hand side of Eq. (17.4) involves the mass density ρ_M. We argued (and much later, experiments verified) that gravity must also act upon massless energy. However, past studies of gravity were dominated by sources for which the rest energy was the largest part of the source's energy budget. Perhaps we can replace ρ_M with ρ/c^2, where ρ is the source's energy density.

This suggests that perhaps our relativistic gravity equation should look something like

$$\Box\Phi^G \stackrel{?}{=} \frac{4\pi G\rho}{c^2} \ . \tag{17.5}$$

This perhaps looks plausible, but on reflection hopefully you'll notice that it has some issues. Chief among them is that, as we discussed several lectures ago, the energy density ρ is one component in a specified reference frame of the stress-energy tensor. Any theory of physics that picks out a particular component of a tensor as playing a special role is, for lack of a better term, a "sick" theory. If we want gravity to be describable from the viewpoint of different reference frames, then the right-hand side of Eq. (17.5) won't do it.

The left-hand side of (17.5) has problems as well. The derivative operator is a scalar, but what is Φ^G? Is it a scalar (as it appeared to be in Newtonian physics)? Is it one component of a tensor, as the right-hand side seems to suggest? If so, what is the rest of the tensor?

This is roughly where Einstein was in the early 20th century, trying to imagine how to fold gravity into the framework of relativity that so successfully merged Maxwell's electrodynamics with mechanics. Getting from there to the general theory of relativity took Einstein about 10 years, much of which was spent learning what was for him an entirely new field of mathematics (Riemannian geometry), and figuring out how to connect this to the core physical concepts that describe gravity. There were multiple wrong turns along the way; in the meantime, others proposed different ways of making relativistic gravity which in the end did not agree with experimental tests.

In 8.033, we don't have the time to explore all of the wrong turns and hypotheses that were proposed but fell short (although we briefly discuss some highlights of interesting "wrong turns" in a short section of supplementary material). Instead, we will elide many details and compress all of the history and thought processes into a few bullet points:

- In special relativity, an unaccelerated trajectory is one that moves on a straight line. If a pair of unaccelerated trajectories start out parallel, then they will remain forever parallel. This is consistent with the idea that the metric $\eta_{\alpha\beta}$ describes a "flat" spacetime geometry.

- When gravity is included, we can introduce principles that allow to recover much of that core idea. We *define* an unaccelerated trajectory in the freely falling frame as the one that feels no non-gravitational forces.

- Because gravity is never perfectly uniform — it exhibits *tidal* variations — we expect a pair of unaccelerated trajectories that start out parallel to not remain parallel; in almost all cases, they will eventually diverge from one another, or perhaps cross. This suggests that gravity can be modeled by thinking about spacetimes that are not flat, but that have *curvature*.

17.4 How to describe relativistic gravity II: Putting the pieces together

Now let's synthesize these ingredients and bullet points to see how, after 10 years of effort, Einstein managed to develop the relativistic theory of gravity that (so far, at least) has passed all experimental tests. Begin by going back to the Newtonian field equation:

$$\nabla^2\Phi^G = 4\pi G\rho_M \; . \tag{17.6}$$

We've already argued that the right-hand side should be something that involves ρ/c^2 rather than ρ_M, where ρ is energy density. But that ρ is itself one component of the stress-energy tensor $T^{\mu\nu}$. A covariant relativistic formulation cannot pick out one component of a tensor as "the" quantity of interest. Whatever goes on the left-hand side of the relativistic "gravity equation," the right-hand side should be something that is proportional to $T^{\mu\nu}$.

To get some idea of how to handle the left-hand side, note that $\nabla^2\Phi^G$ can be regarded as $-\nabla\cdot\mathbf{g}$, where $\mathbf{g} = -\nabla\Phi^G$ is the gravitational field that arises from the potential Φ^G. The left-hand side is thus something like the divergence of the gravitational field. Derivatives of the gravitational field tell us about how this field varies in space — which tells us about the behavior of gravitational tides. So the physical content of Eq. (17.6) can be regarded, schematically, as

$$(\text{“Quantity related to gravitational tides”}) = (\text{“numerical factor times } G\text{”})(\text{“source”}) \; . \tag{17.7}$$

For the source on the right-hand side of our equation, we've already decided to use the stress-energy tensor $T^{\mu\nu}$. Figuring out how to do the left-hand side is a little more complicated. We begin with the idea that a body which moves under the influence of no forces but gravity follows a trajectory of maximal aging through spacetime. Such a trajectory is called a *geodesic*. We will examine geodesics for specific spacetimes soon enough; in the general case (which we will *not* consider in detail in 8.033), a body's geodesic motion in some coordinate system turns out to be governed by a three-index tensor-like object. The differential equation governing the body's 4-velocity takes the form

$$\frac{du^\alpha}{d\tau} + \Gamma^\alpha{}_{\mu\nu}u^\mu u^\nu = 0 \; . \tag{17.8}$$

The quantity $\Gamma^\alpha{}_{\mu\nu}$ (called a connection coefficient or Christoffel symbol) is built from derivatives of the spacetime metric $g_{\mu\nu}$.

If the spacetime describes gravity with tides, then two nearby geodesics that start parallel to one another will eventually become non-parallel. Suppose that two geodesics each have 4-velocity u^μ, and are separated in our coordinates by δx^α. Then, the action of tides will cause their separation to evolve. This evolution is governed by an equation that takes the form

$$\frac{D^2(\delta x^\alpha)}{d\tau^2} = R^\alpha{}_{\mu\nu\beta}u^\mu u^\nu(\delta x^\beta) \; . \tag{17.9}$$

The operator $D/d\tau$ is a special kind of derivative that takes into account the fact that, in a spacetime with curvature, the unit vectors themselves vary from position to position. The 4-index tensor $R^\alpha{}_{\mu\nu\beta}$ (called the Riemann curvature tensor) describes how nearby geodesics deviate from one another due to variations in spacetime — i.e., how tidal variations in gravity make initially parallel trajectories become non parallel. This curvature tensor is built from derivatives of the Christoffel symbol; we can think of it as expressing (in a rather complicated way) two derivatives of the spacetime metric $g_{\mu\nu}$.

Einstein's hypothesis was that the "right" way to upgrade Eq. (17.6) into a relativistic form was to replace the left-hand side with a curvature tensor which is closely related to $R^\alpha{}_{\mu\nu\beta}$, and to replace the right-hand side with the stress-energy tensor:

$$G^{\mu\nu} = \kappa T^{\alpha\beta} . \tag{17.10}$$

The tensor $G^{\mu\nu}$ is known as the Einstein curvature tensor[1]. It is found by combining the Riemann tensor with the metric in a such a way that the result is a 2nd-rank tensor with zero divergence (the stress-energy tensor on the right-hand side has zero divergence, so whatever we put on the left-hand side must have zero divergence as well). You can think of it as a very complicated set of second derivatives which act on the metric.

The constant κ is determined by demanding that, in the correct limit, this equation's predictions agree with the predictions of Newtonian gravity. Doing so, we at last obtain the *Einstein field equation*:

$$G^{\mu\nu} = \frac{8\pi G}{c^4} T^{\mu\nu} . \tag{17.11}$$

This equation can be regarded as a set of partial differential equations for the spacetime metric, given a stress-energy tensor which describes the flow of energy and momentum in that spacetime. Notice that there is a sense in which (17.11) is similar in physical structure to Eq. (17.6): both have "two derivatives of potential" on the left-hand side (provided we now think about the metric of spacetime as playing the role of the potential), and a source that describes energy density on the right-hand side (picking out the dominant component in the Newtonian version, but using the full tensor-valued mathematical object in Einstein's).

Developing the Einstein field equation takes roughly half of the term in 8.962. The other half is spent figuring out how to solve it, and to examining the nature of its solutions. In 8.033, we will jump straight to looking at some of the solutions (though the story behind how some of those solutions were found is pretty interesting, and we'll at least discuss some anecdotes around them). We will then study these solutions in order to tell us about the nature of gravity with relativity included. A very nice feature of what we have done so far is that, with the principle of equivalence and the calculus of variations in our toolkit, it's a relatively simple step for us to examine motion in some spacetime that is provided to us.

[1]Sadly, the notation overlaps with the dual Faraday tensor we discussed in the E&M section of this course. Context generally makes it clear which is which.