Lecture 1: Equivalence Principle and Tensor Analysis

Andy: *So if I am in an elevator and I feel really heavy, I can't know whether the elevator is accelerating or you mischievously put me on Jupiter?*

Lenny: *That's right, you can't.*

Andy: *But, at least on Jupiter, if I keep still, light rays won't bend.*

Lenny: *Oh yes they will.*

Andy: *Hmm, I see.*

Lenny: *And if you are falling into a black hole, beware, things will get really strange. But, don't worry, I'll shed some light on this.*

Andy: *Er, bent or straight?*

Introduction
Equivalence principle
Accelerated reference frames
Curvilinear coordinate transformations
Effect of gravity on light
Tidal forces
Non-Euclidean geometry
Riemannian geometry
Metric tensor
Mathematical interlude: Dummy variables
Mathematical interlude: Einstein summation convention
First tensor rule: Contravariant components of vectors
Mathematical interlude: Vectors and tensors
Second tensor rule: Covariant components of vectors
Covariant and contravariant components of vectors and tensors

Introduction

General Relativity is the fourth volume in The Theoretical Minimum (TTM) series. The first three were devoted respectively to classical mechanics, quantum mechanics, and special relativity and classical field theory. The first volume laid out the Lagrangian and Hamiltonian description of physical phenomena and the principle of least action, which is one of the fundamental principles underlying all of physics (see volume 3, lecture 7 on fundamental principles and gauge invariance). They were used in the first three volumes and will continue in this and subsequent ones.

Physics extensively uses mathematics as its toolbox to construct formal, quantifiable, workable theories of natural phenomena. The main tools we used so far are trigonometry, vector spaces, and calculus, that is, differentiation and integration. They have been explained in volume 1 as well as in brief refresher sections in the other volumes. We assume that the reader is familiar with these mathematical tools and with the physical ideas presented in volumes 1 and 3. The present volume 4, like volumes 1 and 3 (but unlike volume 2), deals with classical physics in the sense that no quantum uncertainty is involved.

We also began to make light use of tensors in volume 3 on special relativity and classical field theory. Now with general relativity we are going to use them extensively. We shall study them in detail. As the reader remembers, tensors generalize vectors. Just as vectors have different representations, with different sets of numbers (components of the vector) depending on the basis used to chart the vector space they form, this is true of tensors as well. The same tensor will have different components in different coordinate systems. The rules to go from one set of components to another will play a fundamental role. Moreover, we will work mostly with *tensor fields*, which are sets of tensors, a different tensor attached to each point of a space. Tensors were invented by Ricci-Curbastro and Levi-Civita[1] to develop work of Gauss[2] on curvature of sur-

[1] Gregorio Ricci-Curbastro (1853–1925) and his student Tullio Levi-Civita (1873–1941) were Italian mathematicians. Their most important joint paper is "Méthodes de calcul différentiel absolu et leurs applications," in *Mathematische Annalen* 54 (1900), pp. 125–201. They did not use the word *tensor*, which was introduced later by other people.

[2] Carl Friedrich Gauss (1777–1855), German mathematician.

faces and Riemann[3] on non-Euclidean geometry. Einstein[4] made extensive use of tensors to build his theory of general relativity. He also made important contributions to their usage: the standard notation for indices and the Einstein summation convention.

In *Savants et écrivains* (1910), Poincaré[5] writes that "in mathematical sciences, a good notation has the same philosophical importance as a good classification in natural sciences." In this book we will take care to always use the clearest and lightest notation possible.

Equivalence Principle

Einstein's revolutionary papers of 1905 on special relativity deeply clarified and extended ideas that several other physicists and mathematicians – Lorentz,[6] Poincaré, and others – had been working on for a few years. Einstein investigated the consequences of the fact that the laws of physics, in particular the behavior of light, are the same in different inertial reference frames. He deduced from that a new explanation of the Lorentz transformations, of the relativity of time, of the equivalence of mass and energy, etc.

After 1905, Einstein began to think about extending the principle of relativity to any kind of reference frames, frames that may be accelerating with respect to one another, not just inertial frames. An *inertial frame* is one where Newton's laws, relating forces and motions, have simple expressions. Or, if you prefer a more vivid image, and you know how to juggle, it is a frame of reference in which you can juggle with no problem – for instance in a railway car moving uniformly, without jerks or accelerations of any sort. After ten years of efforts to build a theory extending the principle of relativity to frames with acceleration and taking into account gravitation in a novel way, Einstein published his work in November 1915. Unlike special relativity, which topped off the work of many, general relativity is essentially the work of one man.

---

[3] Bernhard Riemann (1826–1866), German mathematician.

[4] Albert Einstein (1879–1955), German, Swiss, German again, and finally American physicist.

[5] Henri Poincaré (1854–1912), French mathematician.

[6] Hendrik Antoon Lorentz (1853–1928), Dutch physicist.

We shall start our study of general relativity pretty much where Einstein started. It was a pattern in Einstein's thinking to start with a really simple elementary fact, which almost a child could understand, and deduce these incredibly far-reaching consequences. We think that it is also the best way to teach it, to start with the simplest things and deduce the consequences.

So we shall begin with the *equivalence principle*. What is the equivalence principle? It is the principle that says that *gravity is in some sense the same thing as acceleration*. We shall explain precisely what is meant by that, and give examples of how Einstein used it. From there, we shall ask ourselves: what kind of mathematical structure must a theory have for the equivalence principle to be true? What kind of mathematics must we use to describe it?

Most readers have probably heard that general relativity is a theory not only about gravity but also about geometry. So it is interesting to start at the beginning and ask what is it that led Einstein to say that gravity has something to do with geometry. What does it mean to say that "gravity equals acceleration"? You all know that if you are in an accelerated frame of reference, say, an elevator accelerating upward or downward, you feel an effective gravitational field. Children know this because they feel it.

What follows may be overkill, but making some mathematics out of the motion of an elevator is useful to see in a very simple example how physicists transform a natural phenomenon into mathematics, and then to see how the mathematics is used to make predictions about the phenomenon.

Before proceeding, let's stress that the following study on an elevator, and the laws of physics as perceived inside it, is simple. Yet it is a first presentation of very important concepts. It is fundamental to understand it very well. Indeed, we will often refer to it. In lectures 4 to 9, it will strongly help us understand acceleration, gravitation, and how gravitation "warps" space-time.

So let's imagine the Einstein thought experiment where somebody is in an elevator; see figure 1. In later textbooks, it got promoted to a rocket ship. But I have never been in a rocket ship, whereas

I have been in an elevator. So I know what it feels like when it accelerates or decelerates. Let's say that the elevator is moving upward with a velocity v.

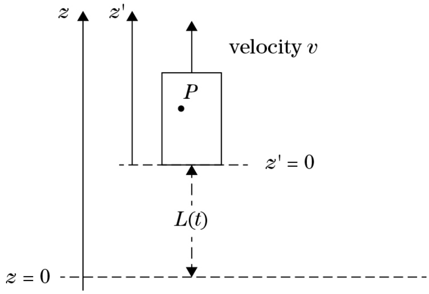

Figure 1: Elevator and two reference frames.

So far the problem is one-dimensional. We are only interested in the vertical direction. There are two reference frames: one is fixed with respect to Earth. It uses the coordinate z. The other is fixed with respect to the elevator. It uses the coordinate z'. A point P anywhere along the vertical axis has two coordinates: coordinate z in the stationary frame, and coordinate z' in the elevator frame. For instance, the floor of the elevator has coordinate $z' = 0$. Its z-coordinate is the distance L, which is obviously a function of time. So we can write for any point P

$$z' = z - L(t) \tag{1}$$

We are going to be interested in the following question: if we know the laws of physics in the frame z, what are they in the frame z'?

One warning about this lecture: at least at the start, we are going to ignore special relativity. This is tantamount to saying that we are pretending that the speed of light is infinite, or that we are talking about motions so slow that the speed of light can be regarded as infinitely fast. You might wonder: if general relativity is the generalization of special relativity, how did Einstein manage to start thinking about general relativity without including special relativity?

The answer is that special relativity has to do with very high velocities, while gravity has to do with heavy masses. There is a range of situations where gravity is important but high velocities are not. So Einstein started out thinking about gravity for slow velocities, and only later combined it with special relativity to think about the combination of fast velocities and gravity. And that became the general theory.

Let's see what we know for slow velocities. Suppose that z' and z are both inertial reference frames. That means, among other things, that they are related by uniform velocity:

$$L(t) = vt \tag{2}$$

We have chosen the coordinates such that when $t = 0$, they line up. At $t = 0$, for any point, z and z' are equal. For instance, at $t = 0$ the elevator's floor has coordinate 0 in both frames. Then the floor starts rising, its height z equaling vt. So for any point we can write equation (1). In view of equation (2), it becomes

$$z' = z - vt \tag{3}$$

Notice that this is a *coordinate transformation* involving space and time. For readers who are familiar with volume 3 of TTM on special relativity, this naturally raises the question: what about time in the reference frame of the elevator? If we are going to forget special relativity, then we can just say that t' and t are the same thing. We don't have to think about Lorentz transformations and their consequences. So the other half of the coordinate transformation would be $t' = t$.

We could also add to the stationary frame a coordinate x going horizontally and a coordinate y jutting out of the page. Correspondingly, coordinates x' and y' could be attached to the elevator; see figure 2. The x-coordinate will play a role in a moment with a light beam. As long as the elevator is not sliding horizontally, x' and x can be taken to be equal. Same for y' and y.

For the sake of clarity of the drawing in figure 2, we offset a bit the elevator to the right of the z-axis. But think of the two vertical axes as actually sliding on each other, and at $t = 0$ the two origins O and O' coincide. Once again, the elevator moves only vertically.

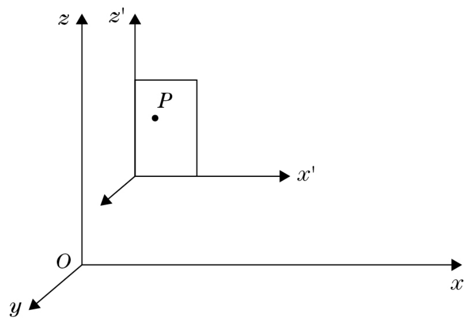

Figure 2: Elevator and two reference frames, three axes in each case.

Finally our complete coordinate transformation is

$$\begin{aligned} z' &= z - vt \\ t' &= t \\ x' &= x \\ y' &= y \end{aligned} \tag{4}$$

It is a coordinate transformation of space-time coordinates. For any point P in space-time, it expresses its coordinates in the moving reference frame of the elevator as functions of its coordinates in the stationary frame. It is rather trivial. Only one coordinate, namely z, is involved in an interesting way.

Let us look at a law of physics expressed in the stationary frame. Take Newton's law of motion $F = ma$ applied to an object or a particle. The acceleration a is $\ddot{z}$, where z is the vertical coordinate of the particle. So we can write

$$F = m\ddot{z} \tag{5}$$

As we know, $\ddot{z}$ is the second time derivative of z with respect to time – it is called the vertical acceleration – and F of course is the vertical component of force. The other components we will take to be zero. Whatever force is exerted, it is exerted vertically. What could this force be due to? It could be related to the elevator or not. There could be some charge in the elevator pushing on the particle. Or it could just be a force due to a rope attached to the ceiling and to the particle that pulls on it. There could be a field force along the vertical axis. Any kind of force could be acting on the particle. Whatever the causes, we know from Newton's law that the equation of motion of the particle, expressed in the original frame of reference, is given by equation (5).

What is the equation of motion expressed in the primed frame? This is very easy. All we have to do is figure out what the original acceleration is in terms of the primed acceleration. What is the primed acceleration? It is the second derivative with respect to time of z'. Using the first equation in equations (4)

$$z' = z - vt$$

one differentiation gives

$$\dot{z'} = \dot{z} - v$$

and a second one gives

$$\ddot{z'} = \ddot{z}$$

The accelerations in the two frames of reference are the same.

All this should be familiar. But I want to formalize it to bring out some points. In particular, I want to stress that *we are doing a coordinate transformation*. We are asking how the laws of physics change in going from one frame to another. What can we now say about Newton's law in the primed frame of reference? We substitute $\ddot{z'}$ for $\ddot{z}$ in equation (5). As they are equal, we get

$$F = m\ddot{z'} \tag{6}$$

We found that Newton's law in the primed frame is exactly the same as Newton's law in the unprimed frame. That is not surprising. The two frames of reference are moving with uniform velocity relative to each other. If one of them is an inertial frame, the other is an inertial frame. Newton taught us that the laws of physics are the same in all inertial frames. It is sometimes called the *Galilean principle of relativity*. We just formalized it.

Let's turn to an accelerated reference frame.

Accelerated Reference Frames

Suppose that $L(t)$ from figure 1 is increasing in an accelerated way. The height of the elevator's floor is now given by

$$L(t) = \frac{1}{2}gt^2 \tag{7}$$

We use the letter g for the acceleration because we will discover that the acceleration mimics a gravitational field – as we feel when we take an elevator and it accelerates. We know from volume 1 of TTM on classical mechanics or from high school, that this is a uniform acceleration. Indeed, if we differentiate $L(t)$ with respect to time, after one differentiation we get

$$\dot{L} = gt$$

which means that the velocity of the elevator increases linearly with time. After a second differentiation with respect to time, we get

$$\ddot{L} = g$$

This means that the acceleration of the elevator is constant. The elevator is uniformly accelerated upward. The equations connecting the primed and unprimed coordinates are different from equations (4). The transformation for the vertical coordinates is now

$$z' = z - \frac{1}{2}gt^2 \tag{8}$$

The other equations in equations (4) don't change:

$$\begin{aligned} t' &= t \\ x' &= x \\ y' &= y \end{aligned}$$

These four equations are our new coordinate transformation to represent the relationship between coordinates that are accelerated relative to each other.

We will continue to assume that in the z, or unprimed, coordinate system, the laws of physics are exactly what Newton taught us. In other words, the stationary reference frame is inertial, and we have $F = m\ddot{z}$. But the primed frame is no longer inertial. It is in uniform acceleration relative to the unprimed frame. Let's ask what the laws of physics are now in the primed frame of reference. We have to do the operation of differentiating twice over again on equation (8). We know the answer:

$$\ddot{z'} = \ddot{z} - g \tag{9}$$

Ah ha! Now the primed acceleration and the unprimed acceleration differ by an amount g. To write Newton's equations in the primed frame of reference, we multiply both sides of equation (9) by m, the particle mass, and we replace $m\ddot{z}$ by F. We get

$$m\ddot{z'} = F - mg \tag{10}$$

We have arrived at what we wanted. Equation (10) looks like a Newton equation, that is, mass times acceleration is equal to some term. That term, $F - mg$, we call the force in the primed frame of reference. You notice, as expected, that the force in the primed frame of reference has an extra term: the mass of the particle times the acceleration of the elevator, with a minus sign.

What is interesting about the "fictitious force" $-mg$, in equation (10), is that it looks exactly like the force exerted on the particle by gravity on the surface of the Earth or the surface of any kind of large massive body. That is why we called the acceleration g. The letter g stood for gravity. It looks like a uniform gravitational field. Let me spell out in what sense it looks like gravity. The special feature of gravity is that gravitational forces are proportional to mass – the same mass that appears in Newton's equation of motion. We sometimes say that *the gravitational mass is the same as the inertial mass*. That has deep implications. If the equation of motion is

$$F = ma \tag{11}$$

and the force itself is proportional to mass, then the mass cancels in equation (11). That is a characteristic of gravitational forces: for a small object moving in a gravitational force field, its motion doesn't depend on its mass. An example is the motion of the Earth about the Sun. It is independent of the mass of the Earth. If you know where the Earth is at time t, and you know its velocity at that time, then you can predict its trajectory. You don't need to know what the Earth's mass is.

Equation (10) is an example of *fictitious force* – if you want to call it that – mimicking the effect of gravity. Most people before Einstein considered this largely an accident. They certainly knew that the effect of acceleration mimics the effect of gravity, but they didn't pay much attention to it. It was Einstein who said: look, this is a deep principle of nature that gravitational forces cannot be distinguished from the effect of an accelerated reference frame.

If you are in an elevator without windows and you feel that your body has some weight, you cannot say whether the elevator, with you inside, is resting on the surface of a planet or, far away from any massive body in the universe, some impish devil is accelerating your elevator. That is the *equivalence principle.* It extends the relativity principle, which said you can juggle in the same way at rest or in a railway car in uniform motion. With a simple example, we have equated accelerated motion and gravity. We have begun to explain what is meant by the sentence: "gravity is in some sense the same thing as acceleration."

We have to discuss this result a bit, though. Do we really believe it totally or does it have to be qualified? Before we do that, let's draw some pictures of what these various coordinate transformations look like.

Curvilinear Coordinate Transformations

Let's first consider the case where $L(t)$ is proportional to t. That is when we have

$$z' = z - vt$$

In figure 3, every point – also called *event* – in space-time has a pair of coordinates z and t in the stationary frame and also a pair of coordinates z' and t' in the elevator frame. Of course, $t' = t$ and we left out the two other spatial coordinates x and y, which don't change between the stationary frame and the elevator. We represented the time trajectories of fixed z with dotted lines and of fixed z' with solid lines.

A fundamental idea to grasp is that events in space-time exist irrespective of their coordinates, just as points in space don't depend on the map we use. Coordinates are just some sort of convenient *tags*. We can use whichever we like. We'll stress it again after we have looked at figures 3 and 4.

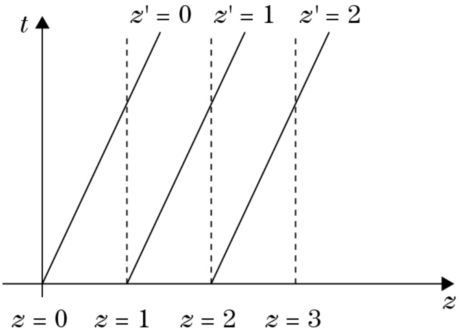

Figure 3: Linear coordinate transformation. The coordinates $(z',\ t')$ are represented in the basic coordinates $(z,\ t)$. An event is a point on the page. It has one set of coordinates in the $(z,\ t)$ frame and another set in the $(z',\ t')$ frame. Here the transformation is simple and linear.

That is called a *linear coordinate transformation* between the two frames of reference. Straight lines go to straight lines, not surprisingly since Newton tells us that free particles move in straight lines in an inertial frame of reference. What is a straight line in one frame had therefore better be a straight line in the other frame. Not only do free particles move in straight lines in space, when we add x and y, but their trajectories are straight lines in space-time – straight in space and with uniform velocity.

Let's do the same thing for the accelerated coordinate system. The transformation equation is now equation (8) linking z' and z. The other coordinates don't change. Again, in figure 4, every point in space-time has two pairs of coordinates $(z,\ t)$ and $(z',\ t')$. The time trajectories of fixed z, represented with dotted lines, don't change. But now the time trajectories of fixed z' are parabolas lying on their side. We can even represent negative times in the past. Think of the elevator that was initially moving downward with a negative velocity but a positive acceleration g (in other words, slowing down). Then the elevator bounces back upward with the same acceleration g. Each parabola is just shifted relative to the previous one by one unit to the right.

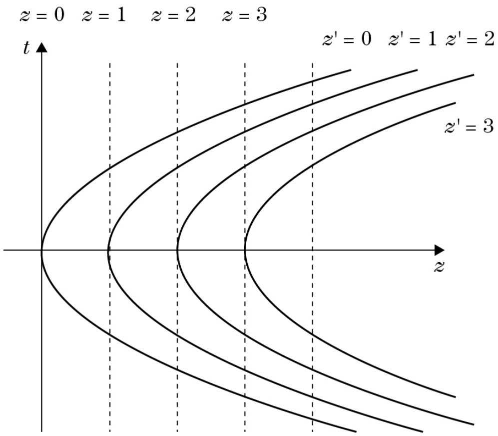

Figure 4: Curvilinear coordinate transformation.

What figure 4 illustrates is, not surprisingly, that straight lines in one frame are not straight lines in the other frame. They become curved lines. As regards the lines of fixed t or fixed t', they are of course the same horizontal straight lines in both frames. We haven't represented them.

We should view figure 4 as just two sets of coordinates to locate each point in space-time. One set of coordinates has straight axes, while the second – represented in the first frame – is curvilinear. Its lines $z' =$ constant are actually curves, while its lines $t' =$ constant are horizontal straight lines. So it is a *curvilinear coordinate transformation.*

Let's insist on the way to interpret and use figure 4 because it is fundamental to understand it very well if we want to understand the theory of relativity – special relativity and even more importantly general relativity. The page represents space-time – here, one spatial dimension and one temporal dimension.

Points (= events) in space-time are points on the page. An event *does not have two positions* on the page, i.e., in space-time. *It has* *only one position* on the page. But this position can be located, mapped, "charted" one also says, using several different systems of reference. A system of reference, also called a frame of reference, is nothing more than a complete set of "labels," if you will, attaching one label (consisting of two numbers, because our space-time here is two-dimensional) to each point, i.e., to each event.

In a two-dimensional space, the system of reference can be geometrically simple, like orthogonal Cartesian axes in the plane. However this is not a necessity. For one thing, on Earth, which is not a plane, the axes are not straight lines. The usual axes used by cartographers and mariners are meridians and parallels. But on a 2D surface, be it a plane or not, we can imagine quite fancy or intricate curvilinear lines to serve as a frame of reference – so long as it attaches unequivocally two numbers to each (by definition, fixed) point. This is what figure 4 does in the space-time made of one temporal and one spatial dimension represented on the page. We will see many more in lecture 2.

Something Einstein understood very early is this:

There is a connection between gravity and curvilinear coordinate transformations of space-time.

Special relativity was only about linear transformations – transformations that take uniform velocity to uniform velocity. Lorentz transformations are of that nature. They take straight lines in space-time to straight lines in space-time. However, if we want to mock up gravitational fields with the effect of acceleration, we are really talking about transformations of coordinates of space-time that are curvilinear. That sounds extremely trivial. When Einstein said it, probably every physicist knew it and thought: "Oh yeah, no big deal." But Einstein was very clever and very persistent. He realized that if he followed very far the consequences of this, he could then answer questions that nobody knew how to answer.

Let's look at a simple example of a question that Einstein answered using the curved coordinates of space-time representing acceleration, and consequently, if the two are the same, gravity. The question is: what is the influence of gravity on light?

Effect of Gravity on Light

When Einstein first asked himself the question "what is the influence of gravity on light"? around 1907, most physicists would have answered: "There is no effect of gravity on light. Light is light. Gravity is gravity. A light wave moving near a massive object moves in a straight line. It is a law of light that it moves in straight lines. And there is no reason to think that gravity has any effect on it."

But Einstein said: "No, if this equivalence principle between acceleration and gravity is true, then gravity must affect light. Why? Because acceleration affects light." It was again one of these arguments that you could explain to a clever child.

Let's imagine that, at $t = 0$, a flashlight (today we might use a laser pointer) emits a pulse of light in a horizontal direction from the left side of the elevator; see figure 5. The light then travels across to the right side with the usual speed of light c. Since the stationary frame is assumed to be an inertial frame, the light moves in a straight line in the stationary frame.

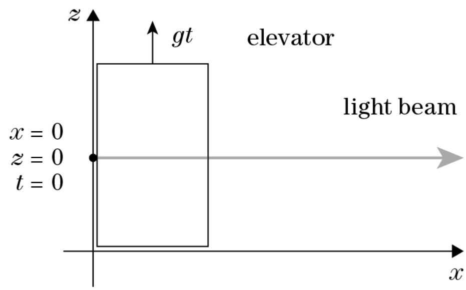

Figure 5: Trajectory of a light beam in the *stationary* reference frame.

The equations for the light ray are

$$\begin{aligned} x &= ct \\ z &= 0 \end{aligned} \tag{12}$$

The first of these equations just says that the light moves across the elevator with the speed of light – no surprise here.

The second says that in the stationary frame the trajectory of the light beam is horizontal.

Let's express the same equations in terms of the primed coordinates. The first equation becomes

$$x' = ct$$

And the second takes the more interesting form

$$z' = -\frac{g}{2}t^2$$

It says that as the light ray moves across the elevator, at the same time the light ray accelerates downward – toward the floor – just as if gravity were pulling it.

We can even eliminate t from the two equations and get an equation for the curved trajectory of the light ray:

$$z' = -\frac{g}{2c^2}x'^2 \tag{13}$$

Thus, the trajectory, in the primed frame of reference, is a parabola, not a straight line.

But, said Einstein, if the effect of acceleration is to bend the trajectory of a light ray, then so must be the effect of gravity.

Andy: *Gee Lenny, that's really simple. Is that all there is to it?*

Lenny: *Yup Andy, that's all there is to it. And you can bet that a lot of physicists were kicking themselves for not thinking of it.*

To summarize, in the stationary frame, the photon trajectory (figure 5) is a straight line, while in the elevator reference frame, it is a parabola (figure 6).

Let's imagine three people arguing. I am in the elevator, and I say: "Gravity is pulling the light beam down." You are in the stationary frame, and you say: "No, it's just that the elevator is accelerating upward; that makes it look like the light beam moves on a curved trajectory." And Einstein says: "They are the same thing!"

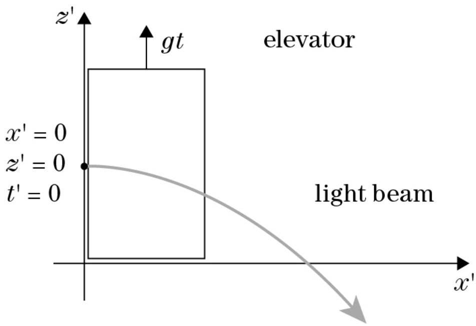

Figure 6: Trajectory of a light beam in the *elevator* reference frame.

This proved to him that a gravitational field must bend a light ray. As far as I know, no other physicist understood this at the time.

In conclusion, we have learned that it is useful to think about curvilinear coordinate transformations in space-time.

When we do think about curvilinear coordinates transformations, *the form of Newton's laws changes.* One of the things that happen is that apparent gravitational fields materialize, which are physically indistinguishable from ordinary gravitational fields.

Well, are they really physically indistinguishable? For some purposes yes, but not for all. So let's turn now to real gravitational fields, namely gravitational fields of gravitating objects like the Sun or the Earth.

Tidal Forces

Figure 7 represents the Earth, or the Sun, or any massive body. The gravitational acceleration doesn't point vertically on the page. It points toward the center of the body.

It is pretty obvious that there is no way that you could do a coordinate transformation like we did in the preceding section that would remove the effect of the gravitational field. Yet, if you are in a small laboratory in space and that laboratory is allowed to simply fall toward Earth, or toward whatever massive object you are considering, then you will think that in that laboratory there is no gravitational field.

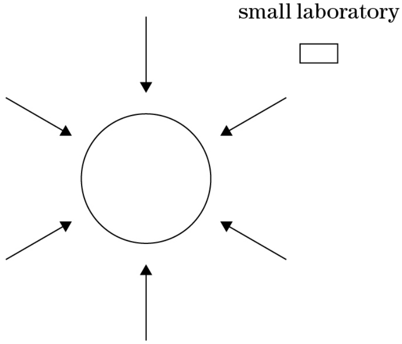

Figure 7: Gravitational field of a massive object, and small laboratory falling toward the object, experiencing inside itself no gravitation.

Exercise 1: If we are falling freely in a uniform gravitational field, prove that we feel no gravity and that things float around us like in the International Space Station.

But, again, there is no way *globally* to introduce a coordinate transformation that is going to get rid of the fact that there is a gravitational field pointing toward the center. For instance, a very simple transformation similar to equations (12) might get rid of the gravity in a small portion on one side of the Earth, but the same transformation will increase the gravitational field on the other side. Even more complex transformations would not solve the problem.

One way to understand why we can't get rid of gravity is to think of an object that is not small compared to the gravitational field. My favorite example is a 2000-mile man who is falling in the Earth's gravitational field; see figure 8. Because he is so big, different parts of his body feel different gravitational fields. Remember that the farther away you are, the weaker is the gravitational field.

His head feels a weaker gravity than his feet. His feet are being pulled harder than his head. He feels like he is being stretched, and that stretching sensation tells him that there is a gravitating object nearby. The sense of discomfort that he feels, due to the nonuniform gravitational field, cannot be removed by switching to a free-falling reference frame. Indeed, no change of mathematical description whatsoever can change this physical phenomenon.

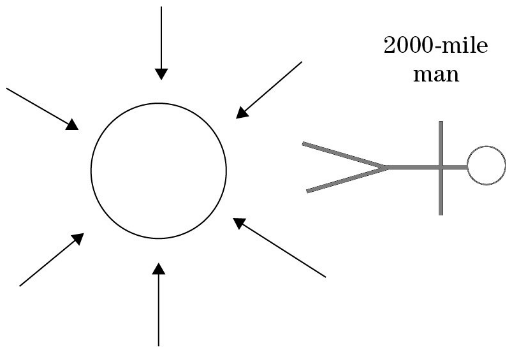

Figure 8: A 2000-mile man falling toward Earth.

The forces he feels are called *tidal forces*, because they play an important role in the phenomenon of tides, too. They cannot be removed by a coordinate transformation. Let's also see what happens if he is falling not vertically but sideways, staying perpendicular to a radius. In that case his head and his feet will be at the same distance from Earth. Both will be subjected to the same force in magnitude pointing to Earth. But since the force directions are radial, they are not parallel. The force on his head and the force on his feet will both have a component along his body. A moment's thought will convince us that the tidal forces will compress him, his feet and head being pushed toward each other. This sense of compression is again not something that we can remove by a coordinate transformation. Being stretched or shrunk, or both, by the Earth's gravitational field – if you are big enough – is an invariant fact.

In summary, it is not quite true that gravity is equivalent to going to an accelerated reference frame.

Andy: *Aha! So Einstein was wrong after all.*

Lenny: *Well, Einstein was wrong at times, but no, Andy, this was not one of those times. He just had to qualify his statement and make it a bit more precise.*

What Einstein really meant was that small objects, for a small length of time, cannot tell the difference between a gravitational field and an accelerated frame of reference.

It raises the following question: if I present you with a force field, does there exist a coordinate transformation that will make it vanish? For example, the force field inside the elevator, associated with its uniform acceleration with respect to an inertial reference frame, was just a vertical force field pointing downward and uniform everywhere. There was a transformation canceling it: simply use z- instead of z'-coordinates. It is a nonlinear coordinate transformation. Nevertheless, it gets rid of the force field.

With other kinds of coordinate transformations, you can make the gravitational field look more complicated, for example transformations that affect also the x-coordinate. They can make the gravitational field bend toward the x-axis. You might simultaneously accelerate along the z-axis while oscillating back and forth on the x-axis. What kind of gravitational field do you see? A very complicated one: it has a vertical component and it has a time-dependent oscillating component along the x-axis.

If instead of the elevator you use a merry-go-round, that is, a carousel, and instead of the $(x',\ z',\ t)$ coordinates of the elevator, you use polar coordinates $(r,\ \theta,\ t)$, an object that in the stationary frame was fixed, or had a simple motion like the light beam, may have a weird motion in the frame moving with the merry-go-round. You may think that you have discovered some repulsive gravitational field phenomenon. But no matter what, the reverse coordinate change will reveal that your apparently messy field is only the consequence of a coordinate change. By choosing funny coordinate transformations, you can create some pretty complicated fictitious, apparent, also called *effective*, gravitational fields. Nonetheless they are not genuine, in the sense that they don't result from the presence of massive objects.

If I give you the field everywhere, how do you determine whether it is fictitious or genuine, i.e., whether it is just the sort of fake gravitational field resulting from a coordinate transformation to a frame with all kinds of accelerations with respect to a simple inertial one, or it is a real gravitational field?

If we are talking about Newtonian gravity, there is an easy way. You just calculate the tidal forces. You determine whether that gravitational field will have an effect on an object that will cause it to squeeze and stretch. If calculations are not practical, you take an object, a mass, a crystal. You let it fall freely and see whether there were stresses and strains on it. If the crystal is big enough, these will be detectable phenomena. If such stresses and strains are detected, then it is a real gravitational field as opposed to only a fictitious one.

On the other hand, if you discover that the gravitational field has no such effect, that any object, wherever it is located and let freely to move, experiences no tidal force – in other words, that the field has no tendency to distort a free-falling system – then it is a field that can be eliminated by a coordinate transformation.

Einstein asked himself the question: what kind of mathematics goes into trying to answer the question of whether a field is a genuine gravitational one or not?

Non-Euclidean Geometry

After his work on special relativity, and after learning of the mathematical structure in which Minkowski[7] had recast it, Einstein knew that special relativity had a geometry associated with it. So let's take a brief rest from gravity to remind ourselves of this important idea in special relativity. Special relativity was the main subject of the third volume of TTM. Here, however, the only thing we are going to use about special relativity is that space-time has a geometry.

[7]Hermann Minkowski (1864–1909), Polish-German mathematician and theoretical physicist.

In the Minkowski geometry of special relativity, there exists a kind of distance between two points, that is, between two events in space-time; see figure 9.

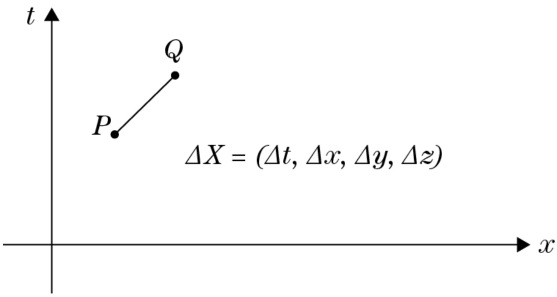

Figure 9: Minkowski geometry: a 4-vector going from P to Q.

The distance between P and Q is not the usual Euclidean distance that we could be tempted to think of. It is defined as follows. Let's call ΔX the 4-vector going from P to Q. To the pair of points P and Q we assign a quantity denoted $\Delta\tau$, defined by

$$\Delta\tau^2 = \Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2$$

Notice that $\Delta\tau$ does not satisfy the usual properties of a distance. In particular, $\Delta\tau^2$ can be positive or negative; and it can be zero for two events that are not identical. The reader is referred to volume 3 of TTM for details. Here we only give a brief refresher.

The quantity $\Delta\tau$ is called the *proper time* between P and Q. It is an invariant under Lorentz transformations. That is why it qualifies as a sort of distance, just as in three-dimensional (3D) Euclidean space the distance between two points, $\Delta x^2 + \Delta y^2 + \Delta z^2$, is invariant under isometries.

We also define a quantity Δs by

$$\Delta s^2 = -\Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2$$

We call Δs the *proper distance* between P and Q. Of course, $\Delta\tau$ and Δs are not two different concepts. They are the same – just differing by an imaginary factor i. They are just two ways to talk about the Minkowski "distance" between P and Q. Depending on which physicist is writing the equations, they will rather use $\Delta\tau$ or Δs as the distance between P and Q.

Einstein knew about this non-Euclidean geometry of special relativity. In his work to include gravity, and to investigate the consequences of the equivalence principle, he also realized that the question we asked at the end of the previous section – are there coordinate transformations that can remove the effect of forces? – was very similar to a certain mathematics problem that had been studied at great length by Riemann. It is the question of deciding whether a geometry is flat or not.

Riemannian Geometry

What is a flat geometry? Intuitively, it is the following idea: the geometry of a page is flat. The geometry of the surface of a sphere or a section of a sphere is not flat. The *intrinsic geometry* of the page remains flat even if we furl the page like in figure 10. We will expound mathematically on the idea in a moment.

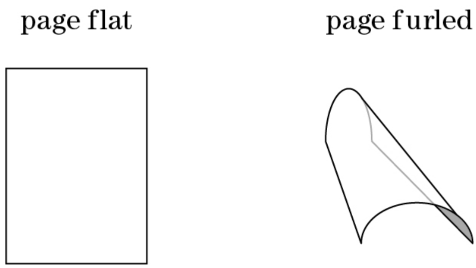

Figure 10: The intrinsic geometry of a page remains flat.

For now, let's just say that the intrinsic geometry of a surface is the geometry that a two-dimensional bug roaming on it, equipped with tiny surveying tools, would see if it were trying to establish an ordnance survey map of the surface.

If the bug worked carefully, it might see hills and valleys, bumps and troughs, if there were any, but it would not notice that the page is furled. We see it because *for us* the page is embedded in the 3D Euclidean space we live in. By unfurling the page, we can make its flatness obvious again.

Einstein realized that there was a great deal of similarity in the two questions of whether a geometry is non-flat and whether a space-time has a real gravitational field in it. Riemann had studied the first question. But Riemann had never dreamt about geometries that have a minus sign in the definition of the square of the distance. He was thinking about geometries that were non-Euclidean but were similar to Euclidean geometry – not Minkowski geometry.

Let's start with the mathematics of Riemannian geometry, that is, of spaces where the distance between two points may not be the Euclidean distance, but in which the square of the distance is always positive.[8]

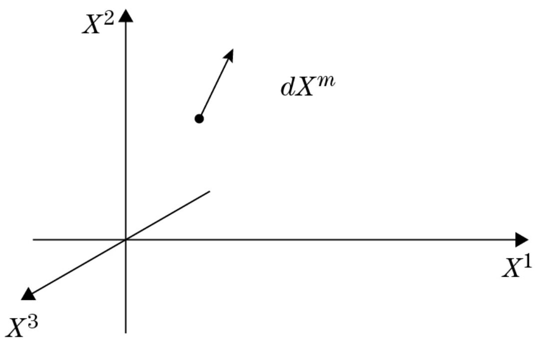

Figure 11: Small displacement between two points in a space.

We look at two points in a space; see figure 11. In our example there are three dimensions, therefore three axes, X^1, X^2, and X^3. There could be more. Thus a point has three coordinates, which we can write as X^m, where m is understood to run from 1 to 3 or to whatever number of axes there is. And a little shift between one point and another nearby has three components, which can be denoted ΔX^m or, if it is to become an infinitesimal, dX^m.

[8]In mathematics, they are called *positive definite distances*.

If this space has the usual Euclidean geometry, the square of the length of dX^m is given by Pythagoras theorem

$$dS^2 = (dX^1)^2 + (dX^2)^2 + (dX^3)^2 + \dots \tag{14}$$

If we are in three dimensions, then there are three terms in the sum. If we are in two dimensions, there are two terms. If the space is 26-dimensional, there are 26 of them and so forth. That is the formula for Euclidean distance between two points in Euclidean space.

For simplicity and ease of visualization, let's focus on a two-dimensional space. It can be the ordinary plane, or it can be a two-dimensional surface that we may visualize embedded in 3D Euclidean space, as in figure 12.

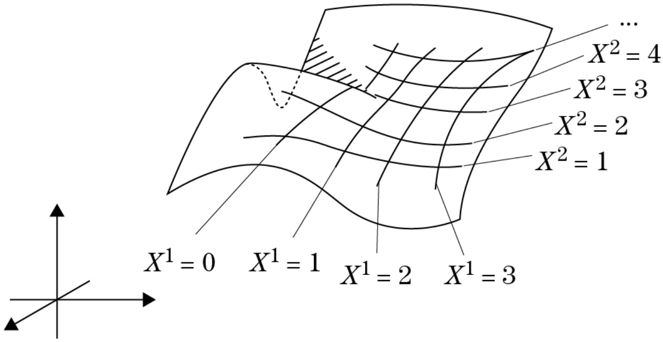

Figure 12: Two-dimensional manifold (i.e., 2D surface) and its curvilinear coordinates viewed embedded in ordinary 3D euclidean space.

There is nothing special about two dimensions for such a surface, except that it is easy to visualize. Mathematicians think of "surfaces" even when they have more dimensions. Usually they don't call them surfaces but *manifolds* or sometimes *varieties*.

Gauss had already understood that on curved surfaces the formula for the distance between two points was more complicated in general than equation (14). Indeed, we must not be confused by the fact that in figure 12 the surface is shown embedded in the usual three-dimensional Euclidean space. This is just for convenience of representation. We should think of the surface as a space in itself, equipped with a coordinate system to locate any point, with curvy lines corresponding to one coordinate being constant, etc., and where a distance has been defined. We must forget about the embedding 3D Euclidean space. The distance between two points on the surface is certainly not their distance in the embedding 3D Euclidean space, and is not even necessarily defined on the surface with the equivalent of equation (14). We will come shortly to exactly how these distances are represented mathematically.

Riemann generalized these surfaces and their metric (the way to compute distances) to any dimensions. But let's continue to use our picture with two dimensions in order to sustain intuition. And let's go slowly, so as not to miss any important detail.

The first thing we do with a surface is put some coordinates on it, which will allow us to quantify various statements involving its points. We just lay out coordinates as if drawing them with a piece of chalk. We don't worry at all about whether the coordinate axes are straight lines or not, because for all we know when the surface is a really curved surface, there probably won't even be things that we can call straight lines. We still call them X's.

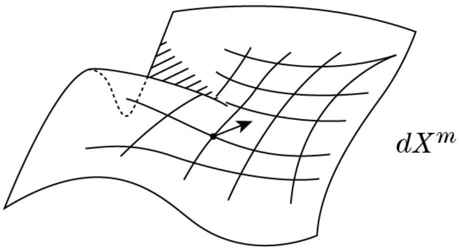

Figure 13: Two neighboring points and the shift dX^m between them.

The values of the X's are not related directly to distances. They are just numerical labels. The points ($X^1 = 0,\ X^2 = 0$) and ($X^1 = 1,\ X^2 = 0$) are not necessarily separated by a distance of one. Now we take two neighboring points; see figure 13. The two neighboring points are again related by a shift of coordinates.

But, unlike in figure 11, which was still a Euclidean space, now we are on an arbitrary curved surface with arbitrary coordinates.

Now we *define* a distance on the surface for points separated by a small shift like dX^m. It won't be as simple as equation (14), though it will have similarities with it. Here is the new definition of dS^2:

$$dS^2 = \sum_{m,\ n} g_{mn}(X)\ dX^m dX^n \tag{15}$$

The functions $g_{mn}(X)$'s, considered altogether, form what is called the *metric of the space*. It is a set of functions of the position X on the manifold under consideration.

Formula (15) is very general and applies whether the manifold is flat or curved. It is a very important formula in Riemannian geometry, and – we will soon see – even in the Minkowski geometry of relativity.

We will also see in a moment how the Einstein summation convention will enable us to rewrite formula (15) in a lighter form. The convention is explained in the section "Mathematical Interlude: Einstein Summation Convention," see infra.

Incidentally, formula (15) applies even to flat geometries equipped with curvilinear coordinates. Suppose that you take a flat geometry, like the surface of the page, but for some reason you use some curvilinear coordinates to locate points, and you ask what is the distance between two points close to each other. Then in general, the square of the distance between two points close to each other will be a quadratic form in the coordinate shifts dX^m's. A quadratic form means a sum of terms, each of which is the product of two little coordinate shifts, times a coefficient like g_{mn} that depends on X.

The surface of the Earth offers a simple example of distance on a curved manifold. Look at the distance between two nearby points characterized by longitude and latitude, as shown in figure 14. Let's denote with R the Earth radius. We take two points $(\phi,\ \theta)$ and $(\phi + d\phi,\ \theta + d\theta)$, where θ is the latitude and ϕ the longitude.

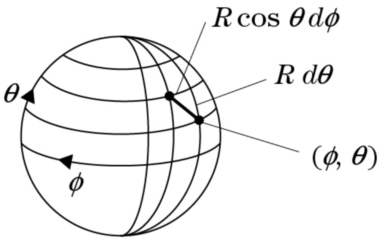

Figure 14: Formula for distance on the Earth surface. Shown are the points $(\phi,\ \theta)$ and $(\phi + d\phi,\ \theta + d\theta)$, and the segment joining them.

Apply Pythagoras theorem in a small, approximately flat, rectangular region to compute the square of the length of its diagonal. One pair of sides along a meridian have length $Rd\theta$. The other pair of sides along a parallel have length $Rd\phi$ but corrected by the cosine of the latitude. At the equator it is the full $Rd\phi$, whereas at the pole it is zero.

The general formula for the square of the distance is

$$dS^2 = R^2 \left[d\theta^2 + (\cos\theta)^2 d\phi^2\right] \tag{16}$$

It is an example of squared distance not just equal to $d\theta^2 + d\phi^2$ but having some coefficient functions in front of the differentials. In this case the interesting coefficient function is $(\cos\theta)^2$ in front of one of the terms $(dX^m)^2$. Note that the coefficient $(\cos\theta)^2$ is often written $\cos^2\theta$. Note also that in this case there are no terms of the form $d\theta\, d\phi$ because the natural curvilinear coordinates we chose on the sphere are still orthogonal at every point.[9]

In other examples – on the sphere with more involved coordinates, or on a more general curved surface like in figure 13 – where the coordinates are not necessarily locally perpendicular, the formula for dS^2 would be more complicated and comprise terms in $dX^m dX^n$. But it will still be a quadratic form. There will never be $d\theta^3$ terms. There will never be things linear. Every term will

[9] Note that spherical coordinates like we use here, which are a bit more sophisticated than Cartesian coordinates, were already much used in the sixteenth century, while Cartesian coordinates began to be used in analytic geometry only in the seventeenth century.

always be quadratic. Moreover in Riemannian geometry, at every point X, the quadratic form defining the metric locally is always positive definite.

You may wonder why we define the distance dS only for small (actually infinitesimal) displacements. The reason is that to talk about distance between two points A and B far away from each other, we must first of all define what we mean. There may be bumps and troughs in between. We could mean the shortest distance as follows: we put a peg at A and a peg at B and pull a string as tight as we can between the two points. That would define one notion of distance. Of course, there might be several paths with the same value. One might go around the hill this way. Then the other would go around the hill that way. Simply think on Earth of going from the North Pole to the South Pole.

Furthermore, even if there is only one answer, we have to know the geometry on the surface everywhere in the whole region where A and B are located, not only to calculate the distance but to know actually where to place the string. Therefore the notion of distance between any two points is more complicated than in Euclidean geometry. But between two *neighboring points* it is not so complicated. That is because locally a smooth surface can be approximated by the tangent plane and the curvilinear coordinate lines by straight lines – not necessarily perpendicular but straight.

Metric Tensor

Let's go deeper into the geometry of a curved surface and its links with equation (15), which defines the distance between two neighboring points on it. Recall the equation

$$dS^2 = \sum_{m,\ n} g_{mn}(X)\ dX^m dX^n$$

In order to get a feel about the geometry of the surface and its behavior, let's imagine that we arrange elements from a Tinkertoy Construction Set along the curved surface. For instance, they could approximately follow the coordinate lines on the surface. We would also add more rigid elements diagonally. This would create a lattice as shown in figure 15. But any reasonably dense lattice, sort of triangulating the surface, would do as well. Suppose furthermore that the Tinkertoy elements are hinged together in a way that lets them freely move in any direction from each other.

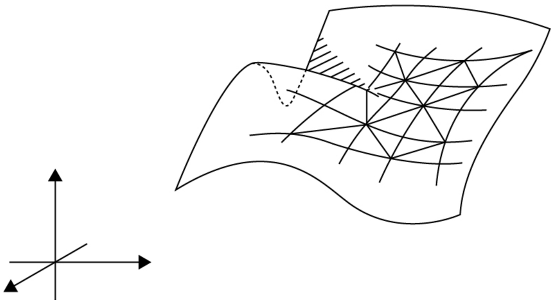

Figure 15: Lattice of rigid Tinkertoy elements arranged on the surface.

Imagine that we lift our lattice from the surface. Sometimes it will keep its shape rigidly, sometimes it won't. It will not keep its shape if it is possible to go from the initial shape to a new shape without forcing any Tinkertoy element to be stretched or compressed or bent.

In some cases it will even be possible *to lay it out flat*. It is the case, for instance, in figure 10 going from the shape on the right to the shape on the left – which is just a flat page.

Exercise 2: Is it possible to find a curved surface and a lattice of rods arranged on it that cannot be flattened out, but can change shape?

Answer: Yes. According to Gauss's Theorema Egregium, which we invite the reader to look up, a surface can be modified without stretching or compressing it as long as we preserve everywhere its Gaussian curvature. For instance, it is possible to change in such a way a section of a hyperbolic paraboloid.

We shall see that the initial surface being able to take other shapes or not corresponds to the g_{mn}'s of equation (15) having certain mathematical properties.

The collection of g_{mn}'s has a name. It is called the *metric tensor*. It is the mathematical object that enables us to compute the distance between two neighboring points on our Riemannian surface.

Mind you, the g_{mn}'s are functions of X (the points of the manifold). So, strictly speaking, we are talking about a *tensor field*. But it is customary to talk casually of the metric tensor, keeping in mind that the collection of its components depends on X.

When the lattice of Tinkertoy elements can be laid out flat, the geometry of the surface is said to be *intrinsically flat*, or just *flat*. We will define it more rigorously later.

Sometimes, on the other hand, the lattice of little rods cannot be laid out flat. For example on the sphere, if we initially lay out a lattice triangulating a large chunk of the sphere, we won't be able to lay it out on a flat plane.[10]

The question we have to address is this: if I made a lattice of little rods covering a surface, and I gave you the length of each rod, without yourself building the lattice how could you tell me whether it is a flat space or an intrinsically curved space, which cannot be flattened and laid out on a flat plane?

Let's formulate the problem more precisely and mathematically. We start from the metric tensor $g_{mn}(X)$, which is a function of position, in some set of coordinates. Keep in mind that there are many different possible sets of curvilinear coordinates on the surface, and in every set of coordinates the metric tensor will look different. It will have different components, just like the same 3-vector in ordinary 3D Euclidean space has different components depending on the basis used to represent it, but in addition the components will vary with position in different ways.

I select one set of coordinates and I give you the metric tensor of my surface. In effect I tell you the distance between every pair of neighboring points. The question is: is my surface flat or not?

[10]This is a well-known problem of cartographers, which led to the invention of various kinds of maps of the world, the most famous being the Mercator projection map invented by Flemish cartographer Gerardus Mercator (1512–1594).

To answer that question, you may think of "checking Pi." Here is the way it would go. Think of a 2D surface embedded in the usual 3D Euclidean space as shown in figure 12. You select a point and mark out a disk around it. Then you measure its radius r as well as its circumference l, and you divide l by $2r$. If you get 3.14159... you would say that the surface is flat. Otherwise you would say that it is not flat, it is intrinsically curved. Notice that this procedure is good for a two-dimensional surface, under certain conditions. Anyway it is not so great for higher-dimensional surfaces.

What is the mathematics of taking a metric tensor and asking if its space is flat? What does it mean for it to be flat? By definition, it means this:

The space is flat if we can find a coordinate transformation, that is, a different set of coordinates, in which, at any point on the surface, *the distance formula for* dS^2 *becomes just* $(dX^1)^2 + (dX^2)^2 + \ldots + (dX^n)^2$, *as it would be in Euclidean geometry.*

It is not necessary that the *initial* $g_{mn}(X)$'s form everywhere the unit matrix, with ones on the diagonal and zeroes elsewhere – as if equation (15) were just Pythagoras theorem. But we must find a coordinate transformation that brings it to that form.

In that sense, it has a vague similarity with the question of whether you can find a coordinate transformation that removes the gravitational field. In fact, it turns out not to be a vague similarity at all but a close parallel. The question is: can we find a coordinate transformation that removes the curvy character of the metric tensor g_{mn}?

To answer that geometric question, we have to do some mathematics essential to relativity. It is not possible to understand general relativity without it. The mathematics is tensor analysis plus some differential geometry. At first it looks annoying because we have to deal with all these indices floating around, and different coordinate systems, and partial derivatives of components, etc. But once we get used to it, we will see that it is simple. It was invented, as said, by Ricci-Curbastro and Levi-Civita at the end of the nineteenth century to build on works of Gauss and Riemann. It was further simplified by Einstein, who set rules for the position of indices and astutely got rid of most summation symbols.

Before explaining what is the Einstein summation convention eliminating most summation symbols, let's spend a few moments explaining the simple concept of dummy variable.

Mathematical Interlude: Dummy Variables

We are accustomed to equations where all the variables have a substantial mathematical or physical meaning. A physical example is equation (7), reproduced here:

$$L(t) = \frac{1}{2}gt^2$$

This famous equation was found by Galileo Galilei in the first half of the seventeeth century,[11] before the invention of calculus. In fact, it is one of the equations that triggered the invention of calculus by Newton and Leibniz. It describes the fall of an object: L stands for the distance of fall as a function of time, g stands for the acceleration on the surface of the Earth, and t stands for time.

Another even simpler and purely mathematical example is

$$A = ab$$

where a is the length of a rectangle, b is its width, and A is its area.

But we are also familiar with equations where one of the variables is only a handy mathematical notation without a substantial meaning. A simple example is the well-known identity expressing the value of the sum of all the squares of the integers from 1 to m

$$\frac{m(m+1)(2m+1)}{6} = \sum_{n=1}^{n=m} n^2$$

[11] We give it here in its modern form. Galileo (1564–1642) just wrote that the distance of the fall was proportional to the square of the fall duration, which, if you think of it, is a mind-blowing discovery.

Here the variable m has a substantial meaning: it is the number up to which we sum. But the variable n on the right-hand side does not have such a substantial meaning. We could rewrite the equation as

$$\frac{m(m+1)(2m+1)}{6} = \sum_{k=1}^{k=m} k^2$$

It would be exactly the same equation.

The variable n, or the variable k, is called a *dummy variable.* It is only used to conveniently express the sum.

We will meet many formulas containing one or several dummy variables, usually expressing sums, in general relativity. They are so frequent that Einstein came up with a rule to simplify them. His rule, or convention, turned out to be not only a great simplification, but also a very useful notational device to write general relativity equations, providing a guide rail as well as having a meaning on its own. The convention is the topic of the next mathematical interlude. Later in this lecture and in the rest of the book, we will discover its remarkable usefulness.

Mathematical Interlude: Einstein Summation Convention

As we go along, we will see that certain patterns keep recurring in the equations. One such pattern involves expressions in which an index such as μ is repeated in a single expression. Here is an example. For the moment it doesn't matter what it means; it's just a pattern that we will see over and over.

$$\sum_{\mu} V^{\mu} U_{\mu}$$

There are a few things to note. First of all, there is a summation over μ, which means that μ is a *dummy index.* It is just another name, in the specific context of vectors and tensors, for a dummy variable. As a consequence, what letter we use doesn't matter. The expressions with μ, as above, or with ν, as below, represent exactly the same thing, whence, as we saw, the term *dummy.*

$$\sum_{\nu} V^{\nu} U_{\nu}$$

Secondly, the dummy index appears twice in the same expression – not once, not three times, twice.

Finally, the repeated index occurs once as a *superscript* and once as a *subscript.* I often say that it appears once upstairs and once downstairs. That's the pattern: a sum over an index that appears once upstairs and once downstairs.

Einstein's famous trick – the so-called *Einstein summation convention* – was just to leave out the summation sign. The rule is: whenever we see something like $V^\mu U_\mu$, we automatically sum over the index μ.

We can readily apply the convention to formula (15) that we met earlier expressing the general form of the metric in a Riemannian space (or for that matter in a Minkowskian space as well, we shall see). It was

$$dS^2 = \sum_{m,\ n} g_{mn}(X)\ dX^m dX^n$$

With the Einstein summation convention it becomes

$$dS^2 = g_{mn}(X)\ dX^m dX^n$$

Simpler! Isn't it?

Usually, not forgetting that the g_{mn}'s components depend on X, i.e., remembering that the metric tensor is actually a tensor field, we simplify it even further to

$$dS^2 = g_{mn}\ dX^m dX^n$$

Andy: *Did it really take Einstein to invent the summation convention?*

Lenny: *I guess it did. When I was a student, I read Einstein's famous 1916 paper "The Foundation of the General Theory of Relativity." It was my habit when I learned new physics to write out the equations as I read them. At the start of the paper, the equations were written as anyone else would write them. Here's his equation 2:*

$$dX_\nu = \sum_\sigma a_{\nu\sigma} dx_\sigma$$

But then all of a sudden, right after equation 7, Einstein casually remarks that there is always a summation when indices appear twice.[12] *So from now on, he said, we'll just keep that in mind and stop writing the summation sign. It's pretty clear that he just got tired of writing them. I was pretty tired of writing them too. What a relief it was.*

End of interlude on Einstein summation convention.

Let's return to the metric and its various forms in several different coordinate systems. To find a set of coordinates that make equation (15) become equation (14) is a more involved procedure than just diagonalizing the matrix g_{mn}. The reason is that there is not one matrix. As we stressed, each component g_{mn} depends on X.

It is the same tensor field, but it has a different matrix at each point.[13] You cannot diagonalize them all at the same time. At a given point, you can indeed diagonalize $g_{mn}(X)$ even if the surface is not flat. It is equivalent to working locally in the tangent plane of the surface at X, and orthogonalizing the coordinate axes there. But you cannot say that a surface is flat because it can be made at any given point locally to look like the Euclidean plane.

Let's examine equation (14) more closely. It can be written in terms of a special matrix whose components are the *Kronecker-delta symbol* δ_{mn}, defined in the following way.[14]

First of all, δ_{mn} is zero unless $m = n$. For example, in three dimensions δ_{12}, δ_{13}, and δ_{23} are all zero, but δ_{11}, δ_{22}, and δ_{33} are nonzero. In other words, at each point the Kronecker-delta symbol is a diagonal matrix.

---

[12] Later, Einstein devised the superscript and subscript notations for the indices of tensors, and his rule henceforth applied only to pairs of indices, with the same dummy variable, one upstairs and the other downstairs.

[13] For a given set of coordinates, we have a collection of matrices – one at each point. For another set of coordinates, we will have another collection of matrices. At each point, the *components* of the tensor depend on the coordinates, but the tensor itself is an abstract object that doesn't. We already met the distinction with 3D vectors.

[14] Named after the German mathematician Leopold Kronecker (1823–1891).

Secondly, the diagonal elements are all equal to 1:

$$\delta_{11} = \delta_{22} = \delta_{33} = 1$$

Armed with the Kronecker-delta and the Einstein summation convention, we can rewrite equation (14) in the compact form,

$$dS^2 = \delta_{mn}\, dX^m dX^n \tag{17}$$

To determine if a space is flat, we look for a coordinate transformation, $X \to Y$, that turns g_{mn} into δ_{mn} everywhere. Remember that X and Y represent the *same point* P. This point P is simply located with two different reference systems, which, as we stressed, are nothing more than some geometric labeling procedure.

Later, the points P will be events in space-time, and the Kronecker-delta will be replaced by a slightly more involved diagonal matrix in Minkowski geometry (also called Minkowskian or Einsteinian geometry), but many of the ideas will remain unchanged. However let's not go too fast, and for the moment let's stay in Riemannian geometry. Riemannian geometry is everywhere locally Euclidean. It can be thought of as "Euclidean geometry on a piece of rubber."

For most metrics it is not possible to find a coordinate transformation that transforms everywhere the g_{mn} into δ_{mn}. It is only when the space is intrinsically flat that we can.

In summary, I give you the metric tensor of my surface, that is, the g_{mn} of equation (15), which we now write

$$dS^2 = g_{mn}(X)\ dX^m dX^n$$

The question I ask you is: can you, by a coordinate transformation $X \to Y$, reduce it to equation (17)? That is, in the Y system,

$$dS^2 = \delta_{mn}\, dY^m dY^n$$

There is no need to write $\delta_{mn}(Y)$, since the Kronecker-delta symbol by definition has a unique form. However, for the sake of clarity, we will sometimes still write $\delta_{mn}(Y)$ because it reminds us of which system of coordinates we are using.

If the answer is yes, the space is called *flat.* If it is no, the space is called *curved.* Of course, the space could have some portions that are flat. There could exist a set of coordinates such that in a region the metric tensor is the Kronecker-delta. But the surface is called flat only if it is everywhere flat.

This becomes a pure mathematics problem: given a tensor field $g_{mn}(X)$ on a multidimensional space (which mathematicians call a *manifold*), how do we figure out if there is a coordinate transformation that would change it into the Kronecker-delta symbol?

To answer that question, we have to understand better how things transform when we make coordinate transformations. That is the subject of *tensor analysis.* We begin to present the subject in the rest of this lecture, and will treat it in more depth in lecture 2.

The analogy between tidal forces and curvature actually is not an analogy, it is a very precise equivalence. In the general theory of relativity, the way you diagnose tidal forces (or said more accurately, their generalization) is by calculating the curvature tensor. A flat space is defined as a space where the curvature tensor is zero everywhere. Therefore it is a very precise correspondence. Simply stated:

Gravity is curvature.

But we will come to this conclusion as we get through tensor analysis. Obviously, in trying to determine whether we can transform away $g_{mn}(X)$ and turn it into the trivial $\delta_{mn}(Y)$, the first question to ask is: how does $g_{mn}(X)$ transform when we change coordinates? We have to introduce notions of tensor analysis that are rather easy.

We shall express the first tensor rule, then present a mathematical interlude spelling out some general facts on vectors and tensors, then present the second tensor rule.

We will conclude this copious lecture again with some general considerations on covariant and contravariant components of vectors and tensors.

First Tensor Rule: Contravariant Components of Vectors

Sometimes tensor notations are a bit of a nuisance because of all the indices. At first we can get confused by them. But soon we will discover that the manipulations obey strict rules and turn out to be rather simple.

We shall begin with a simpler thing than $g_{mn}(X)$. Suppose that there are two sets of coordinates on our surface: a set of coordinates X^m, and a second set that we could call X' as we did earlier. But then we would be running into horrible notations with cluttered expressions like X'^1. So we denote the second set of coordinates Y^m. To be very explicit, if we are on a space of dimension N, the same point P has coordinates

$$[\ X^1(P),\ X^2(P),\ \dots\ , X^N(P)\ ]$$

and also has coordinates

$$[\ Y^1(P),\ Y^2(P),\ \dots\ , Y^N(P)\ ]$$

The X's and Y's are related because if you know the coordinates of a point P in one set of coordinates, then in principle you know where the point is. Therefore you also know its coordinates in the other coordinate system. Thus each coordinate X^m is a function of all the coordinates Y^n. We can use whatever dummy index we want if that helps avoid confusion. We will simply write

$$X^m(Y)$$

Likewise each Y^m is assumed to be a known function of all the X^n's:

$$Y^m(X)$$

In short, we have two coordinate systems, each one a function of the other. The correspondence is one-to-one since these are coordinate systems. And we assume that the functions are nice and smooth.

Now we ask: how do the differential elements dX^m transform? The collection of differential elements dX^m is a small vector, as shown in figure 16. Remember that the vector itself is a pair of points (an origin and an end). It is independent of the coordinate system. But in order to work with it, it is useful to express it using its components dX^m.

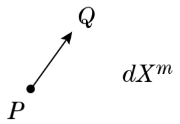

Figure 16: Small displacement expressed in the X coordinate system.

The notation dX^m is used to represent the small vector

$$dX^m = [\ dX^1,\ dX^2,\ \dots\ , dX^N\ ]$$

Said another way, when we change X a little bit, the point P moves to a nearby point Q, and the displacement is dX^m.

Let's look at the same displacement, expressed in the Y coordinate system. We want to know how dY^m can be expressed in terms of the dX^p's. It is an elementary result of calculus that

$$dY^m = \sum_p \frac{\partial Y^m}{\partial X^p}\ dX^p$$

or using the summation convention,

$$dY^m = \frac{\partial Y^m}{\partial X^p}\ dX^p \tag{18}$$

Let's spell out even more explicitly what equation (18) says: the total change of some particular component Y^m is the sum of the rate of change of Y^m when you change only X^1, times the little change in X^1, namely dX^1, plus the rate of change of Y^m when you change only X^2, times the little change in X^2, namely dX^2, and so forth up to X^N and dX^N because equation (18) means a sum over the dummy index p going from 1 to N.

We now turn to some general considerations on vectors and tensors.

So far we have used several times the term *tensor* (tensor calculus, metric tensor, curvature tensor, first tensor rule, etc.), without explaining what is a tensor! As the reader has understood, it is a fundamental mathematical tool in general relativity. You may even remember that "it extends the concept of vector." But that is certainly not a sufficient explanation to grasp what it is.

We won't go into a full fledged exposition of linear algebra and tensors – which the reader may find in any good manual on the subject. However, as I have done several times in The Theoretical Minimum series, for instance, when I dared to explain in volume 1 integral calculus or partial differentiation in brief interludes of a few pages, because we needed those tools for classical mechanics, it is time in this lecture for a third mathematical interlude presenting in some detail vectors and tensors.

Mathematical Interlude: Vectors and Tensors

Let's begin with the simplest notion of a tensor, namely a scalar. A scalar $S(X)$ is a function of position with the property that it has the same value in every coordinate system. For that reason, we could also denote it $S(P)$, but we want to insist on the coordinate system we chose to use, so we write instead $S(X)$. For the same scalar in the Y coordinate system, we will temporarily use the notation $S'(Y)$. (Later we will use $S(Y)$ and $S(X)$ for both, because it is clearer when we talk about the chain rule.)

Its transformation properties are trivial: it doesn't transform at all. An example drawn from meteorology would be the temperature at a point in space. The transformation property of a scalar reflects this triviality,

$$S'(Y) = S(X)$$

In the case of temperature, this says that the temperature at a point is just a number.[15] It does not depend on the orientation

[15] *Number* and *scalar* are two equivalent terms for the same thing. What is the reason for talking about "scalars"? Numbers are often called scalars because one number can always be obtained from another number with a change of scale: for instance, you can obtain 7 from 2, just by multiplying 2 by 3.5. You cannot do that with any pair of vectors. It is possible only with colinear vectors. Strictly speaking, the term *scalar* is reserved for real numbers. But we often also casually call complex numbers scalars.

of the coordinate system at that point. Note too that scalars do not have components, or perhaps more accurately, they have only one component: the value of the scalar itself.

Let's turn to the next simple kind of tensors, namely vectors. We shall see that there are two kinds.

We all have an intuitive idea of what a vector in a Riemannian geometry is. It is a little arrow, usually attached to a point in space. It points in a direction and it has a magnitude. An example, again from meteorology, would be the wind velocity.

In a Riemannian geometry, a vector is a thing unto itself, but given a coordinate system and a metric, it can be described by components in one of two ways: either contravariant or covariant components.

Since the terms can be a little confusing, let's stress right away that what are called the *contravariant components* of a vector are the good old components with which we construct the vector as a linear combination of the basis vectors.

We will see that we can also attach to a vector another set of numbers, called its *covariant components*. They are not its ordinary contravariant components, but something else, the geometrical meaning of which will be explained in lecture 2. The contravariant and covariant components of a vector will be simply related to each other with the help of the metric.

These components, like the components of the metric itself, will vary when the coordinates system changes. For the moment, however, let's not think of a metric, only of a system of coordinates X and a system of coordinates Y. We position ourselves at a point P. At this point, we consider a set of numbers attached to it and that depends on the coordinate system.

Disregarding any geometric interpretation, this set of numbers can be viewed as an abstract "vector." As said, we are in the case where the vector will change with the coordinate system.

In that case we will have two kinds of vectors: covariant or contravariant vectors. Notice I said covariant or contravariant *vectors* – not covariant or contravariant *components*. Later, when we have introduced a metric, we can put the two together to describe a single kind of vector (the intuitive arrow) in two ways.

What is it that makes a collection of numbers like dX^m a contravariant vector, rather than just a collection of numbers? The answer is the transformation properties under a coordinate transformation. Equation (18) defines the paradigm for the transformation of a contravariant vector.

A contravariant vector is a set of numbers V^m that transform as follows:

$$(V')^m = \frac{\partial Y^m}{\partial X^p} V^p \tag{19}$$

In this equation the variables V are the components of the vector in the X coordinate system and (V') are the components in the Y system. Looking back at equation (18), we see that the differential displacement dX^m is a contravariant vector.

There are a couple of things to note. First of all, I have used the summation convention so that the index p is summed over. Secondly, the index p in the expression $\partial Y^m / \partial X^p$ is a downstairs index. That's a convention that we have already mentioned in the interlude on Einstein summation convention and that the reader will have to remember: when an upstairs index occurs in the denominator of an expression, it counts as a downstairs index.

Generally speaking, in a "level" expression (i.e., with no denominator) or in the numerator of a fraction, a superscript index is called a *contravariant index*. And a subscript index is a called a *covariant index*. But, as we said, according to the summation convention, a superscript in the denominator of a fraction acts like a covariant index.

Let's move on to the second kind of vector – a covariant vector. If the iconic contravariant vector is the displacement dX^m, the iconic covariant vector is the gradient of a scalar $S(X)$.

Its components are given by the derivatives of the scalar along the coordinate axes:

$$\frac{\partial S(X)}{\partial X^p} \tag{20}$$

Clearly these components depend on the choice of coordinates, and will transform when the coordinates are transformed. For example, suppose we transform from the X to the Y system. To compute the components of the gradient in the Y system, we use a version of the chain rule of calculus (see lecture 2 of volume 1 of TTM, in which the chain rule is explained). We get

$$\frac{\partial S}{\partial Y^m} = \frac{\partial S}{\partial X^p} \frac{\partial X^p}{\partial Y^m} \tag{21}$$

From this we can abstract the general rule for the transformation of a covariant vector:

$$(W')_m = W_p \frac{\partial X^p}{\partial Y_m} \tag{22}$$

Thus, in equation (18), we met the *first example of transformation of a tensor*, because an ordinary vector, corresponding for instance to the position of a point, or to a displacement (in other words, a translation), or to a velocity, etc., is a contravariant vector, which is a simple kind of tensor.

Indeed, we now have the expressions, in two different coordinate systems, of the small displacement of a point on the surface (figure 16). They are dX^m and dY^m. Let's repeat that the dX^m and dY^m are two sets of components for the *same* displacement. And we know how to go from one set to the other.

Figure 17, which completes figure 16, shows the small displacement, and also locally the two sets of coordinates.

By now the reader has understood that equation (18) is simply the transformation property of the components of the displacement vector when this displacement vector (which is itself a well-defined geometric object, being defined independently of any coordinate system[16]) is expressed in the X system and in the Y system.

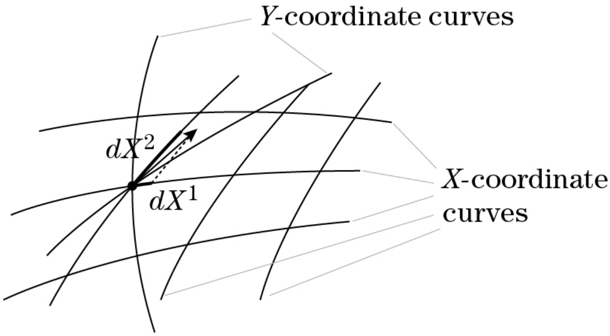

Figure 17: Small displacement, and two sets of coordinates. The small vector has components $(dX^1,\ dX^2)$ shown, but also $(dY^1,\ dY^2)$ not shown.

Note on terminology: because we will deal with vectors that can have contravariant expressions but also covariant expressions, we will prefer to speak of the contravariant components of a vector or the covariant components of a vector.

In short, the term *contravariant* comes from the fact that if we change the unit vectors in the coordinate system, for instance if we simply *divide* the length of each of them by ten, the components of a vector representing a translation will be *multiplied* by ten. Turning to the other term, *covariant* comes from the fact that, in the same kind of change of coordinates, the components of a gradient will be *divided* by ten.

[16]Notice, however, that it is difficult to speak of geometric concepts without some kind of coordinate system. The two famous American geometers Oswald Veblen (1880–1960) and John Whitehead (1904–1960), aware of the difficulty to define what is geometry, wrote in their book *The Foundations of Differential Geometry* that geometry is what experts call geometry. :-) This statement was considered outrageous by the Russian mathematician Andreï Kolmogorov (1903–1987) and his coauthors Alexander Alexandrov (1912–1999) and Mikhaïl Lavrentiev (1900–1980) in their famous book on mathematics, the English translation of which is *Mathematics: Its Content, Methods and Meaning*, MIT Press, 1969.

The interlude presented the simplest kind of tensors: tensors of rank 0, which are simply scalars; and tensors of rank 1, which are contravariant vectors and covariant vectors. The next kinds of tensors, of rank 2 or more, will be presented in the last section of this lecture.

Second Tensor Rule: Covariant Components of Vectors

Although we have already mentioned it cursorily in the preceding mathematical interlude, for the sake of symmetry, let's spell out the second tensor rule concerning the covariant components of vectors. These vectors are used to represent other things than position or translation or velocity or acceleration, etc. The reader may primarily think of gradients of scalar fields.

Examples of scalar fields are the temperature, the atmospheric pressure, the Higgs field, whatever has, at any point in the space, a value that is not multidimensional but simply a number, and that doesn't change if we change coordinates.

The wind velocity is not a scalar field because at every point it has a vector value. It is a vector field. It is important to note the following point, which should clarify things:

If we tried to consider only the first component of the vector representing the wind, we would not get a scalar field, because it would not be invariant under change of coordinates.

Thus the gradient of a scalar function is a vector (in the sense of a collection of components). But it is not an ordinary vector. Indeed, its components don't transform in the same way as do the contravariant components of ordinary vectors.

We saw earlier that an application of the chain rule gave us equation (21), which we reproduce here:

$$\frac{\partial S}{\partial Y^m} = \frac{\partial S}{\partial X^p} \frac{\partial X^p}{\partial Y^m}$$

Denoting by (W') the gradient of S with respect to the Y's, and by W its gradient with respect to the X's, it can be rewritten as equation (22), which we also reproduce, attributing it a new number:

$$(W')_m = \frac{\partial X^p}{\partial Y^m} \, W_p \tag{23}$$

Equation (23) doesn't apply only to gradients; it is the fundamental equation linking the primed and unprimed versions of the covariant components of a vector, that is, its components in the Y system and in the X system.

Notice that the indices m of W' and p of W are downstairs. The index p is a dummy index that is to be summed over as it also appears upstairs in ∂X^p. It is a nice example of the very useful Einstein summation convention and of its smooth workings.

Let's rewrite equations (19) and (23) next to each other, and relabel them:

Contravariant components

$$(V')^m = \frac{\partial Y^m}{\partial X^p} \, V^p \tag{24a}$$

Covariant components

$$(W')_m = \frac{\partial X^p}{\partial Y^m} \, W_p \tag{24b}$$

They look very much alike except that $\partial Y^m / \partial X^p$ appears in the first one, and the inverse, $\partial X^p / \partial Y^m$, in the second.

Let's recall one last time that displacements, or positions, or velocities, etc., are described with vectors having contravariant components. We saw that these change *contrary* to the basis change.

Gradients, on the other hand, are described with vectors the components of which change *like* the basis change. That is why their components are called covariant. But these vectors are different from the somewhat more intuitive contravariant vectors.

In mathematics, vectors with covariant components are sometimes viewed as vectors in the dual space of the primary vector space under consideration. They are then *dual vectors* like linear forms are. But we won't adopt this approach. For us vectors will be things that have a one-indexed collection of contravariant components and also of covariant components.

Equations (24a) and (24b) are fundamental equations for this course. The reader needs to understand them, become familiar and at ease with them, because they are absolutely central to the entire subject of general relativity. You need to know where the indices go for different kinds of objects, and how these objects transform. That is in some sense what general relativity is all about: the transformation properties of different kinds of objects.

Covariant and Contravariant Components of Vectors and Tensors

We have seen two ways to think about an *ordinary* vector. First of all, we can think of it like we have learned in high school: it is a displacement with a length and a direction, that is, *an arrow* in a space. This is geometrically well defined even before we consider any basis.

We can also think of it more abstractly as some object that has components. These components depend on the basis. If the components transform in a certain way when we change basis, namely according to equation (24a), then the object behaves exactly like our good old vectors. Therefore we can also equate the object to an ordinary vector. In tensor analysis we call them vectors whose components are contravariant.

Similarly, some other objects have components that transform according to equation (24b). They cannot be equated to our old ordinary vectors, but to other geometric things. We mentioned that mathematicians view them as dual vectors. We will just call this second type of object vectors whose components are covariant. In fact, we will see in lecture 2 that our abstract vectors have a contravariant version and a covariant version.

In tensor calculus, of which general relativity makes heavy use,[17] paradoxically for those people who have a geometric mind or intuition, it is often useful, at least at first, to forget about the geometric interpretation of the objects we manipulate, and to focus only on how collections of numbers attached to points in our space behave when we change systems coordinates.

A vector – be it with contravariant or covariant components – is a special case of a tensor. Following what we just said, we are not going to define tensors geometrically. For us, at first, tensors will be things that are defined by the way they transform. The way they transform means the way they change (or if you prefer, their components change) when we go from one set of coordinates to another. Later we will give a geometric interpretation of some tensors. We will also go deeper into contravariant and covariant components. We will see that an object with one index can have a contravariant version and a covariant version. All this will be developed in the next lecture. For the time being, let's continue to proceed step by step in our construction of the mathematical tools necessary for general relativity.

The next step, for us now, is to talk about tensors with more than one index.

The best way to approach tensors with several indices is to consider a special, very simple case to start with. Let's imagine the "product" of two vectors with contravariant components.[18] We consider the two vectors with contravariant components, V and U, and we consider the following product:

[17] Einstein developed his ideas in special relativity without using tensor calculus nor even Minkowski geometry, which Minkowski, who had been Einstein's teacher at Zurich, introduced only in 1908. Poincaré also did some preliminary work in this direction. At first Einstein did not think that this heavy mathematical recasting of the theory of relativity was useful. But he soon changed his mind. When general relativity was completed, in 1915, Einstein said it would not have been possible without abstract non-Euclidean geometry and tensor calculus. Hermann Minkowski (1864–1909) did not participate in the development of general relativity because he died in 1909. His good friend David Hilbert (1862–1943) however, did play a role in 1915; see lecture 9.

[18] It is not the dot product nor the cross product. It is going to be called the *outer product* or *tensor product*. Anyway, it is an operation that to two things associates a third thing.

$$V^m U^n$$

Without further ado, we will now always use the convention that contravariant components, or contravariant indices refering to these components, are noted upstairs.

The vectors V and U don't have to come from the same space. If the dimensionality of the space of V is M, and the dimensionality of the space of U is N, there are $M \times N$ such products. As usual, we use the notation $V^m U^n$ to denote one product as well as the collection of all of them – just like V^m denotes one component of the vector V, but is also a notation, showing explicitly the position of the index, and therefore the nature of the full vector V itself.

Let's define T^{mn} as

$$T^{mn} = V^m U^n \tag{25}$$

Notice that it matters where and in which order we write the indices of T^{mn}, because, for instance, T^{mn} is not the same as T^{nm}. The reader is invited to explain why. Soon we will also see combinations of indices upstairs and downstairs.

Product T^{mn} is a special case of tensor of rank 2. Rank 2 means that the collection of component products has two indices. It runs over two ranges: m runs from 1 to M, and n runs from 1 to N. For example, if both V and U come from a four-dimensional space, there will be 16 components $V^m U^n$. In that case T^{mn}, as we saw, represents one component but also the entire collection of 16 components.

How does T^{mn} transform?

For example V^m and U^n could be the components of the vectors V and U in the unprimed frame of reference, the reference frame using the X coordinates. Since we know how the individual components transform, when we go to the Y coordinates, we can figure out how T transforms. Let's call $(T')^{mn}$ the mn-th component of the tensor in the primed frame:

$$(T')^{mn} = (V')^m (U')^n$$

Then using equation (24a) twice, this can be rewritten as

$$(T')^{mn} = \frac{\partial Y^m}{\partial X^p} V^p \frac{\partial Y^n}{\partial X^q} U^q$$

The four terms on the right-hand side are just four numbers, so we can change their order and write it

$$(T')^{mn} = \frac{\partial Y^m}{\partial X^p} \frac{\partial Y^n}{\partial X^q} V^p U^q$$

Finally, $V^p U^q$ is just T^{pq}. So the way T transforms is

$$(T')^{mn} = \frac{\partial Y^m}{\partial X^p} \frac{\partial Y^n}{\partial X^q} T^{pq} \tag{26}$$

We found in the special case of a product of ordinary vectors how T transforms. Now this leads us to the following definition:

Anything that transforms according to equation (26) is called a tensor of rank 2 with two contravariant indices.

If there were more indices upstairs, the rule would be adapted in the obvious manner. A tensor of rank 3, all indices contravariant, would transform like this:

$$(T')^{lmn} = \frac{\partial Y^l}{\partial X^p} \frac{\partial Y^m}{\partial X^q} \frac{\partial Y^n}{\partial X^r} T^{pqr}$$

What kinds of things are tensors like that? Many things. Products of vectors are particular examples, but there are other things that are not products and still are tensors according to this definition.

We are going to see that the metric object g_{mn} is a tensor. But it is a tensor with covariant indices. So to finish this lecture, let's see how things with covariant indices transform. Equation (24b) shows how an object with only one covariant index transforms. It is a tensor of rank 1 of covariant type.

Let's begin again with the particular case of the product of two covariant vectors W and Z, or to speak less casually, two vectors with covariant components. Their product transforms as follows:

$$(W')_m(Z')_n = \frac{\partial X^p}{\partial Y^m}\,\frac{\partial X^q}{\partial Y^n}\,W_p Z_q$$

Here we have discovered a new transformation property of a thing with two covariant indices, that is, two downstairs indices.

More generally let's consider an object that we will denote T_{mn}. It is no longer simply a product of vectors but a different object. However, the letter T signals that it is something that will still be a tensor. It is a tensor with two lower indices, and it transforms according to this equation:

$$T'_{mn} = \frac{\partial X^p}{\partial Y^m}\,\frac{\partial X^q}{\partial Y^n}\,T_{pq} \tag{27}$$

Again, anything that transforms according to equation (27) is called a tensor of rank 2 with two covariant indices.

It is left to the reader to figure out how a tensor with one upper index and one lower index must transform.

In the next lecture, we will also see how the metric object g of equation (15) transforms. We will see that it is a tensor with two covariant indices.

Then the question we will ask is: given that equation (27) is the transformation property of g, can we or can we not find a coordinate transformation that will turn g_{mn} into δ_{mn}?

That is the mathematics question. It is a hard question in general. But we will find the condition.