We are now going to embark on an amazing trip through the world of General Relativity.

A full-blown mathematical exposition of all the derivations is probably well beyond most of you and, as it turns out, not very enlightening.

I will get around these difficulties by using Physics to make guesses instead doing nasty, complicated mathematical derivations. Physical thinking is very powerful as you will see.

I hope you will be able to understand all the motivations and reasoning behind everything we do.

In a few places, I will show you some of the mathematical intricacies just so you understand the important links and you will understand how to do the mathematical derivations in books.

Prepare yourself for an amazing ride.

I doubt that you have ever seen anything quite like this.

It will be tedious and roundabout at times, on purpose!

However, we will have a much better physical understanding of all the "whys".

As you will see, the underlying physics allows this subject to be fully understood

For those of you that want to see a standard mathematical set of lectures you can read the notes labeled "Regular Class Version I - Ohanian" or "Regular Class Version II - Chow" on the website. If anyone wants me to continue this class(after the end) and cover the regular approach also, let me know now and we can cover either set of notes. It will be more like a mathematics class than a physics class!

Lesson 1: The equivalence principle and tensor analysis

The equivalence principle

Einstein's revolutionary papers of 1905 on special relativity deeply clarified and extended ideas that several other physicists and mathematicians, Lorentz, 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 in acceleration with respect to one another, not just inertial ones.

Remember: 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 and with lots of help from the works of Emmy Noether and David Hilbert, Einstein published his work in November 1915.

I thinks the creators of General Relativity should be listed as **Einstein, Hilbert and Noether**.

Unlike special relativity which topped off the work of many, general relativity was thought of as essentially the work of one man according to many physicists but I do not thinks he would have succeeded so quickly without the works of Noether and Hilbert! Men have a habit of rewarding Men only especially if a woman is involved!

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, that 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 does a theory ought to have in order that the equivalence principle be true?

What are the kinds of mathematics we must use to describe it?

Most of you 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 that mean to say that "gravity equals acceleration"?

We will go through a very elementary derivation of what that means.

It is not that this derivation is important, but it is worth formalizing.

Formalizing means making equations that describe the world.

You all know that if you are in an accelerated frame of reference, an elevator accelerating upward or downward, you feel an effective gravitational field.

Children know that 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 what these mathematics really are and what in turn they can say and predict about the natural phenomenon.

Let's imagine the Einstein thought-experiment: somebody is in an elevator (figure 1).

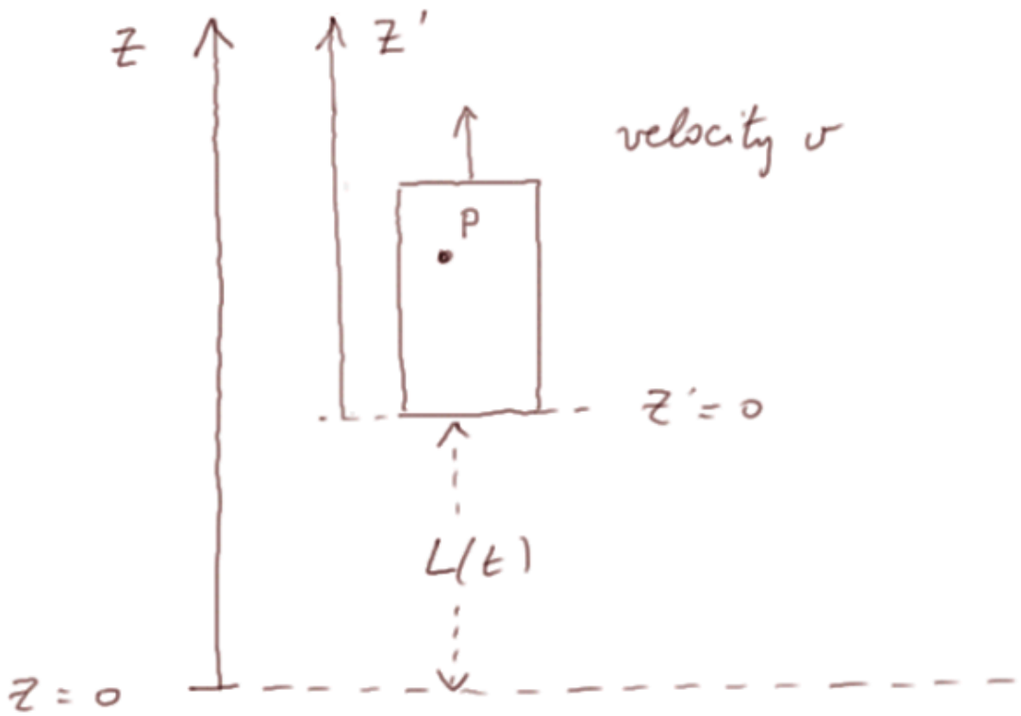


Figure 1: Elevator and two reference frames.

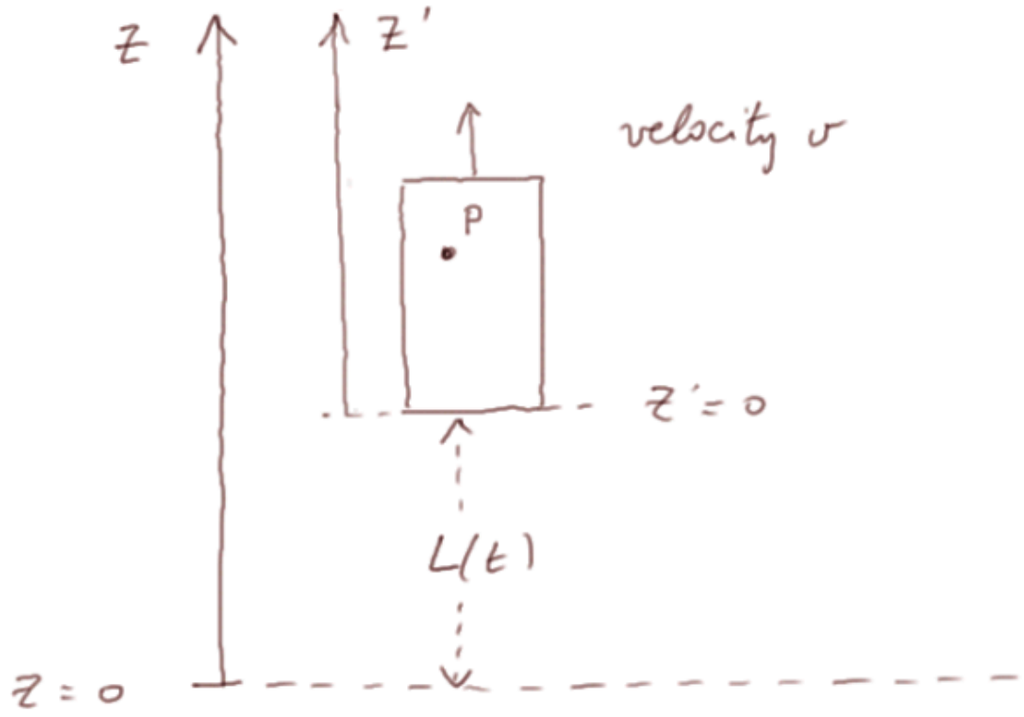
In later books, the elevator got promoted to a rocket ship.

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 the elevator is moving upward with a velocity v.

So far the problem is one-dimensional.



We are only interested in the vertical direction.

There are two reference frames: one is stationary with respect to the Earth.

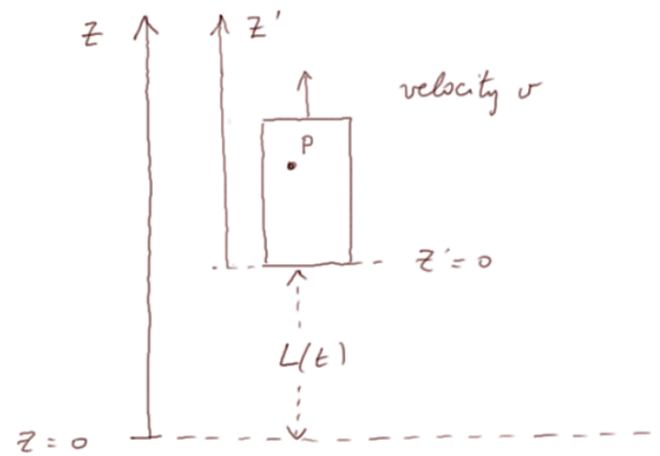
It uses the coordinate z.

And 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 $z' = 0$.



And 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 lesson: at least in 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 which are 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 get away with starting his thinking without 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 then 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.

And let’s begin with inertial reference frames.

Suppose that z′ and z are both inertial reference frames.

That means they are related by uniform velocity.

In other words

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

We have chosen the coordinates such that when t = 0, they lineup.

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 is rising and its height z is equal to vt.

So for any point we can write equation (1).

Combined with equation (2), it becomes

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

Notice that this is a coordinate transformation.

Since you are familiar with 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 for now, then we can just say 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.

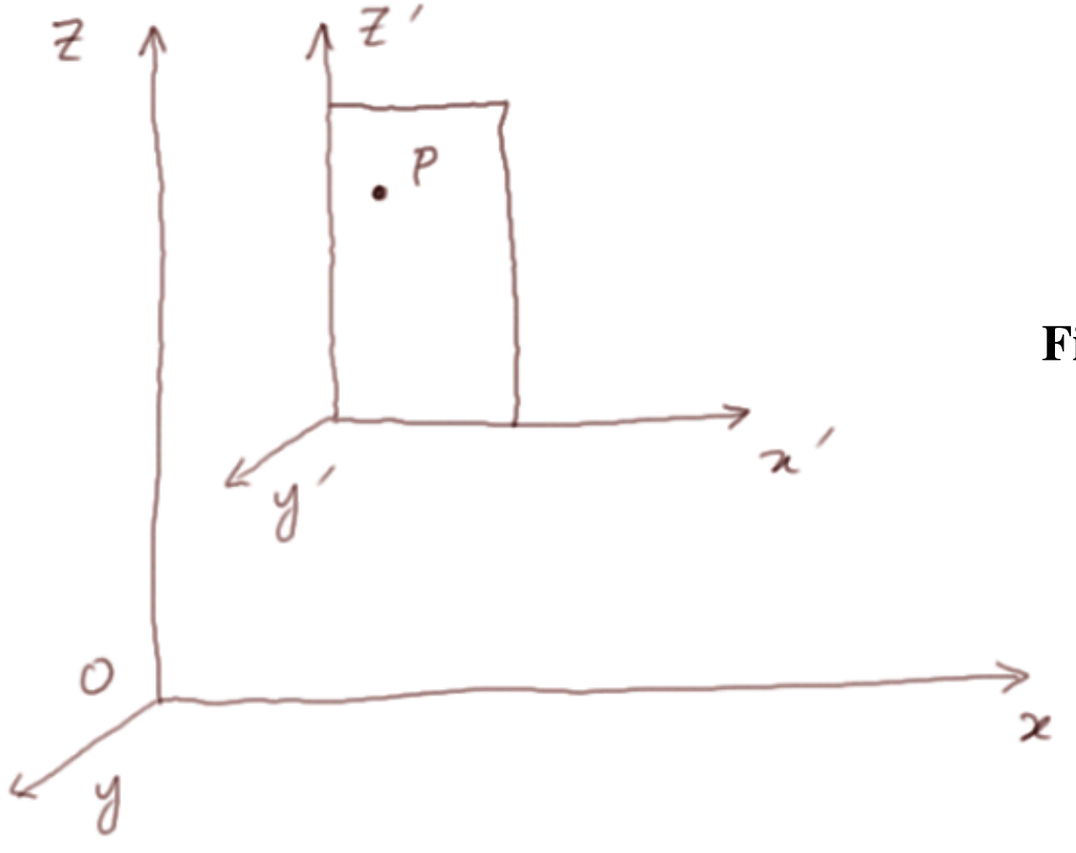


Figure 2: Elevator and two reference frames, three axes in each case.

The x coordinate will play a role in a moment with a light beam.

As long as the elevator is not sliding horizontally then x′ and x can be taken to be equal.

Same thing for y′ and y.

For the sake of clarity of the drawing, we offset the elevator a bit 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′ as being the same.

Once again, the elevator moves only vertically.

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.

Now let us look at a law of physics expressed in the stationary frame.

Let's take Newton's law of motion $F = ma$ applied to an object or a particle in the elevator.

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.

Of course F is the vertical component of force.

We forget about any x-component or y-component of force.

They are not interesting in this context.

In fact we take them equal to zero.

Whatever force is exerted, it is exerted vertically.

What could this force be due to?

There could be some charge in the elevator exerting a force on the charged particle.

Or it could just be a force due to a rope attached to the ceiling and to the particle, that pulls on it.

Any number of different kinds of forces could be acting on the particle.

And we know that the equation of motion of the particle, expressed in the original frame of reference, is given by formula (5).

What is the equation of motion expressed in the prime frame?

Well 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)

$$z' = z - vt$$

we get first of all

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

and then

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

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

All of 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 laws of physics change in going from one frame to another.

So now what can we say about the Newton's law in the prime frame of reference?

We substitute $\ddot{z}'$ for $\ddot{z}$ in equation (5), since they are equal, and we get

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

What do we find?

We find that Newton's law in the prime frame is exactly the same as Newton's law in the unprimed frame.

That is not surprising.

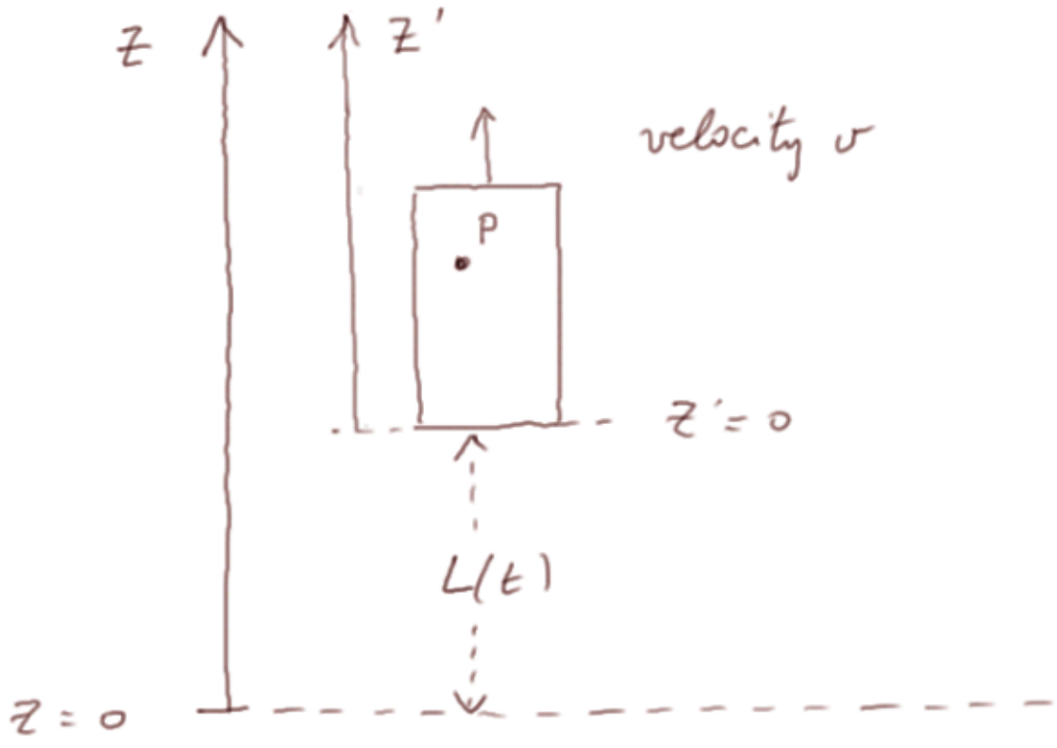
The two frames of reference are moving with uniform velocity relative to each other.

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.

So that is a sort of formalization of the argument.

Now let's turn to an accelerated reference frame.



Accelerated reference frames

Now $L(t)$ is increasing in an accelerated way (see fig 1).

The height of the elevator's floor is then given by

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

We use the letter g for the acceleration, for obvious reasons.

If you don't know what the obvious reasons are, you will find out in a minute.

We know from 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.

And after a second differentiation with respect to time, we get

$$\ddot{L} = g$$

which means that the acceleration of the elevator is constant.

The elevator is uniformly accelerated upward.

And the equations connecting the primed and unprimed coordinates are different.

For the vertical coordinates it is now

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

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

$$t' = t$$

$$x' = x$$

$$y' = y$$

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

We will continue to assume that in the z, or the 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 prime 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 prime frame of reference.

We have to do the operation of differentiating twice over again on equation (8).

We know the answer:

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

The primed acceleration and the unprimed acceleration differ by an amount g.

Now we can write Newton's equations in the prime frame of reference.

We multiply both sides of (9) by m, the particle mass, and we replace $m\ddot{z}$ by F.

We get

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

We 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 prime frame of reference.

You notice, as expected, that the force in the prime frame of reference has an extra term, a sort of *fictitious* term which is the mass of the particle times the acceleration of the elevator, with a minus sign.

What is interesting about the fictitious term $-mg$, in equation (10), is that it looks exactly like the force exerted on the particle by gravity at the surface of the Earth or the surface of any kind of large massive body.

It looks like a uniform gravitational field.

But let me spell out in what sense it looks like gravity.

The special feature of gravity is that the gravitational forces are proportional to mass – the same mass which appears in Newton's equation of motion.

We sometimes say that the gravitational mass is the *same* as the inertial mass.

That has a deep implication.

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, its motion doesn't depend on its mass.

An example would just be 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 an instant, and you know how fast it is moving, then you can predict its trajectory, without regard for what the mass of Earth is.

So equation (10) is an example of a *fictitious* force – if you want to call it that way – 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*, that extends the relativity principle which said you can juggle in the same way at rest or in a railway car in uniform motion.

In a simple example, we have equated accelerated motion and gravity.

We have begun to explain what is meant by "gravity is in some sense the same thing as acceleration".

We will have to discuss this result a bit – whether we really believe it totally or it has to be qualified.

But before we do that, let's draw some pictures of what these various coordinate transformations look like.

Curvilinear coordinate transformations

Let's first take the case where $L(t)$ is proportional to t.

That is when we have

$$z' = z - vt$$

In figure 3, every point 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.

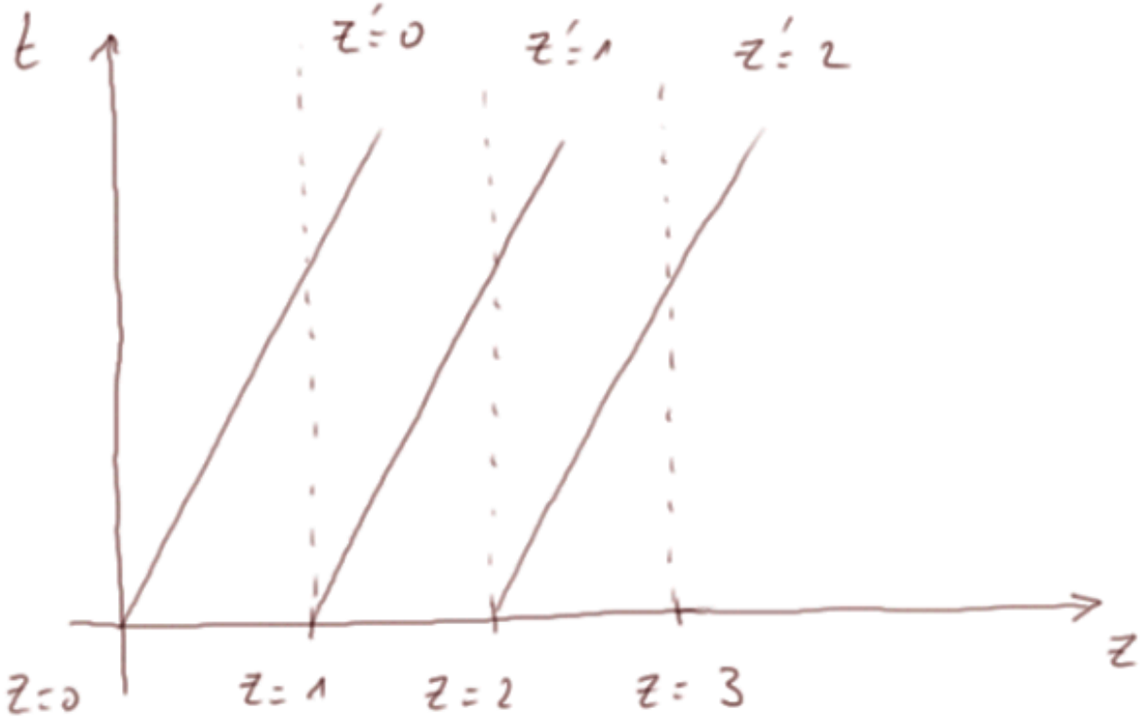
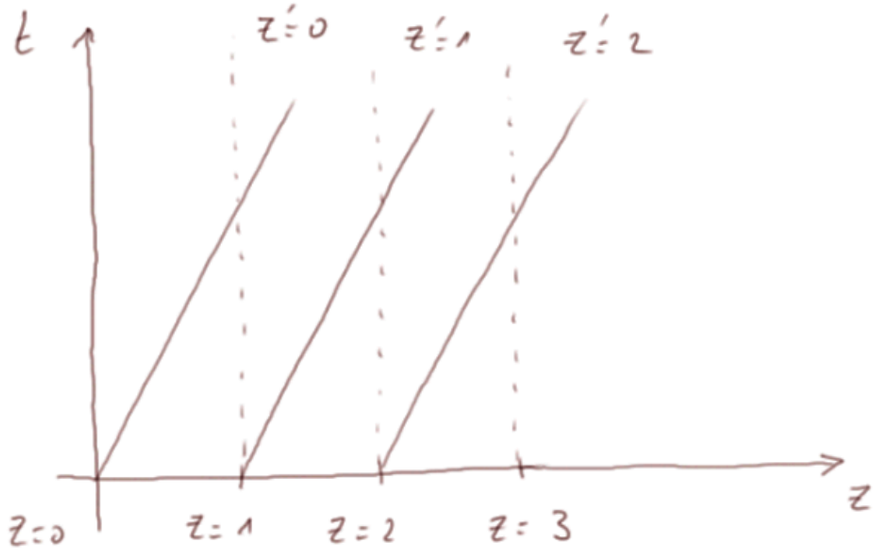


Figure 3: Linear coordinates transformation.

We represented the time trajectories of fixed z with dotted lines, and of fixed z′ with solid lines.



That is called a linear coordinates 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.

And 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, but their trajectories are straight line in spacetime – they don't change speed on their spatial straight lines.

Now let's do the same thing for the accelerated coordinate system.

Of course $t' = t$ and we left off the two other spatial coordinates x and y which don't change between the stationary frame and the elevator.

The transformation equations are now equation (8) linking z' and z, which we repeat here

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

and the other coordinates which don't change.

Again, in figure 4, every point in space-time has two pairs of coordinates (z, t) and (z′ , t′).

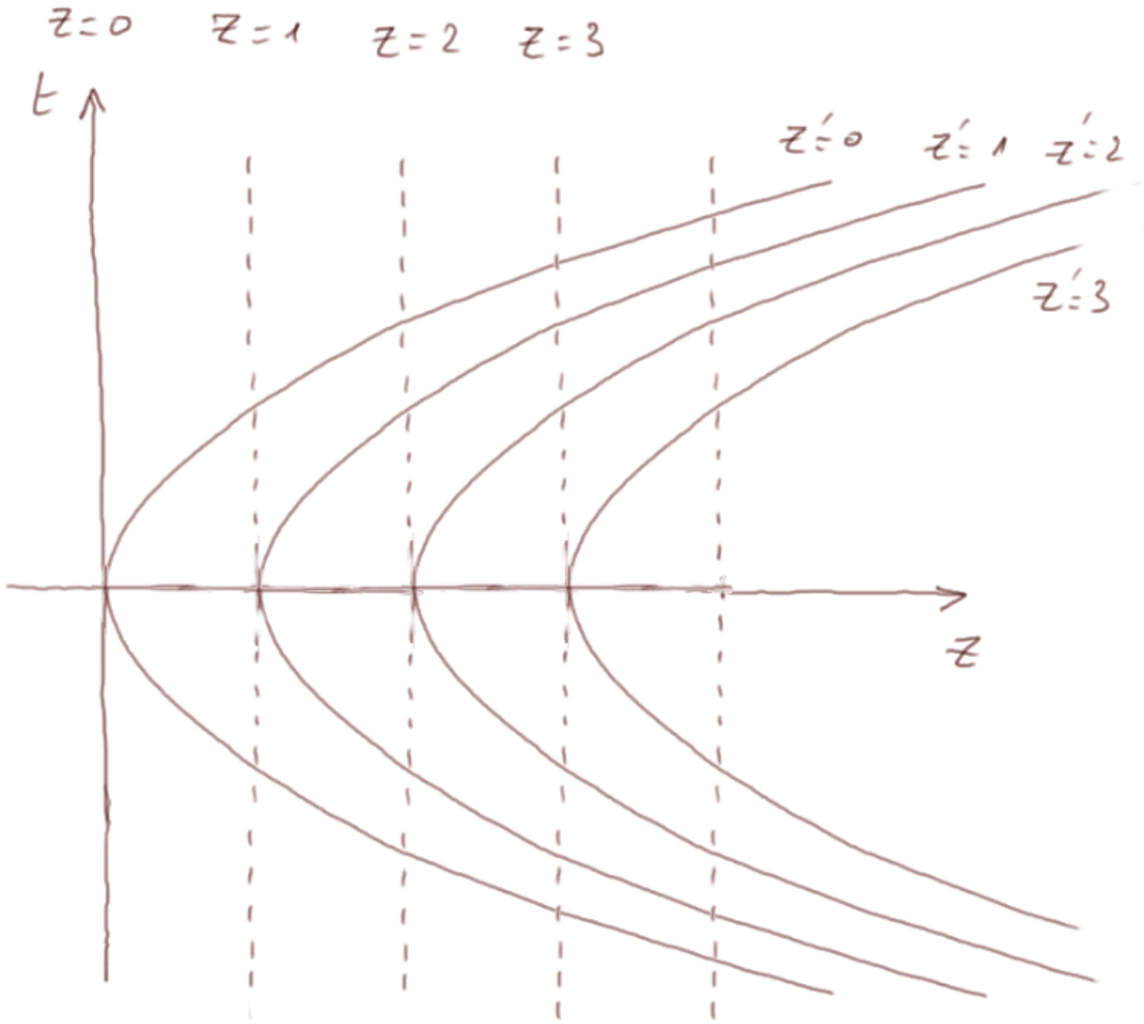


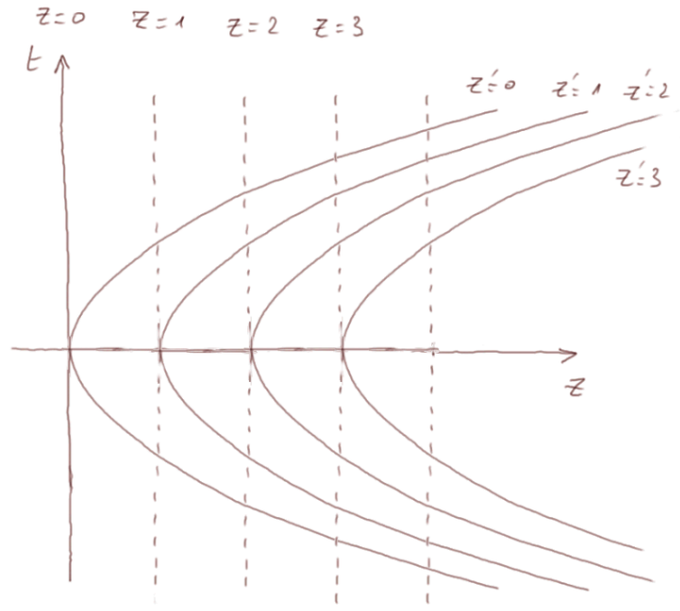
Figure 4: Curvilinear coordinates transformation.

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 going down with acceleration g.



This creates a deceleration up to, say, t = 0 and z = 0.

And 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.

The point of this figure, of course, is, not surprisingly, that straight lines in one frame are *not* straight lines in the other frame.

They become *curved* lines.

The lines of fixed t or fixed t′ of course are the same horizontal straight lines in both frames.

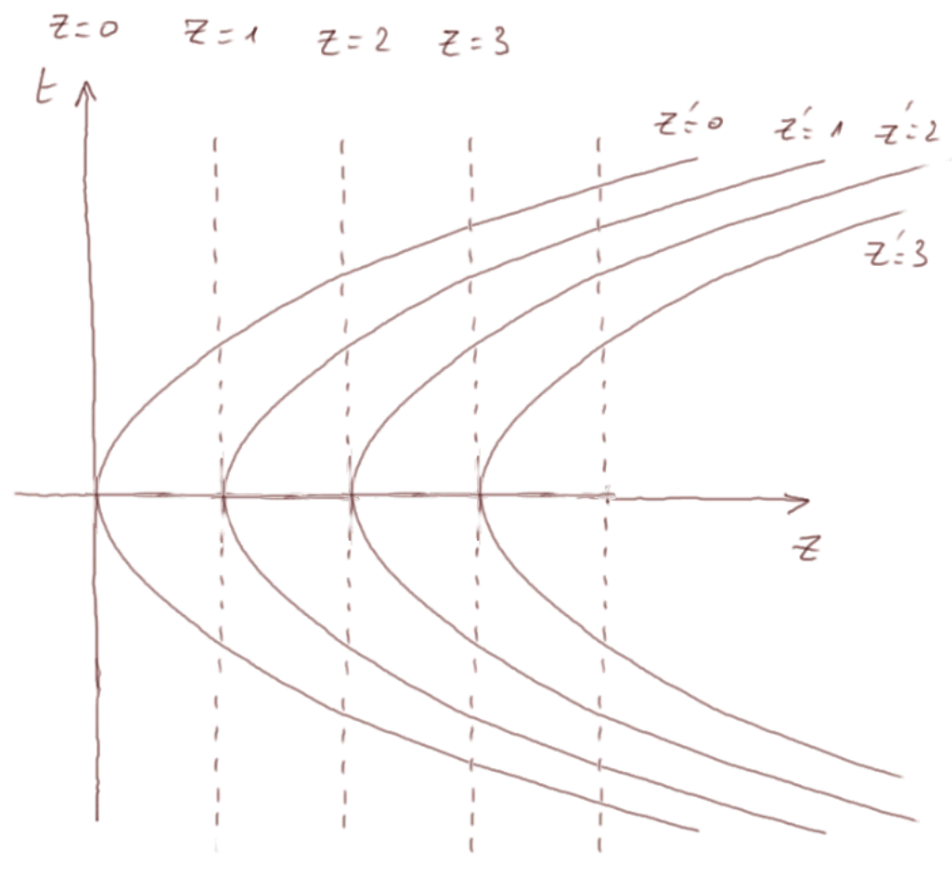
We haven’t represented them.

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 $z' =$ constant lines are actually curves, while its $t' =$ constant lines are horizontal straight lines.

So it is a curvilinear coordinates transformation.



Something Einstein understood very early is this:

> *There is a connection between gravity and curvilinear coordinates transformations of space-time*.

Special relativity was only about linear transformations, transformations which 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.

But if we want to mock-up gravitational fields with the effect of acceleration, we are really talking about transformations of the coordinates of space-time which 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 dug very deep into 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 curvature of space-time created by acceleration and therefore – if the two are the same – by gravity.

The question is: what is the influence of gravity on light?

Effect of gravity on light

After his work of 1905 on special relativity, Einstein began to think about gravity.

When he 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 creating a gravitational field 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 acceleration and gravity will affect light in the same way.

And the argument was very simple.

It was again one of these arguments which you could almost explain to a child.

Let's start out with a light beam that is moving horizontally with respect to the stationary reference frame, starting out at t = 0 (figure 5).

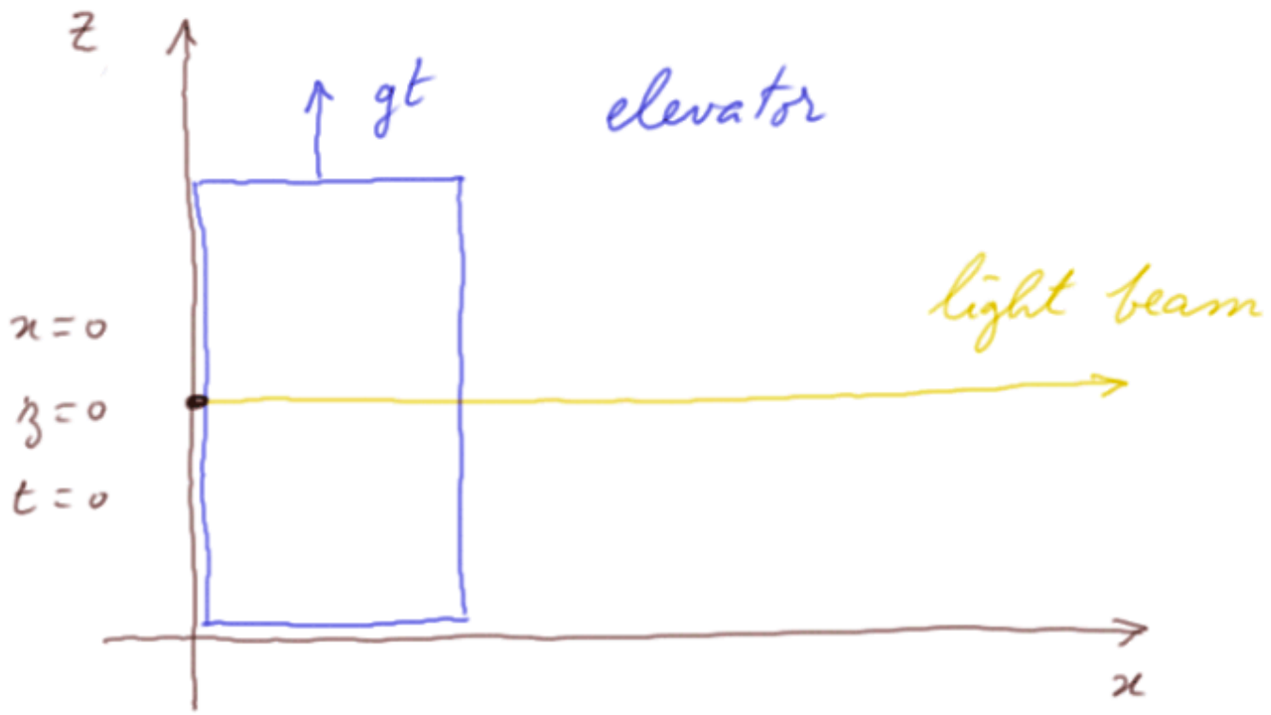


Figure 5: Trajectory of a light beam in the stationary reference frame.

And let's see what is the motion of the light beam expressed in the elevator reference frame.

At t = 0 the two frames of reference match: z′ = z for every point in space-time, see equation (8).

And they are both at rest relative to each other, that is, at t = 0 the velocity of the elevator is also zero because for any point of the elevator we have $\dot{z} = gt$.

But it begins to move up.

A flashlight sends a beam of light in the x direction, from point ($x = 0, z = 0, t = 0$) in the stationary frame.

This point in space-time has also the coordinates ($x' = 0, z' = 0, t' = 0$) in the elevator frame.

In the stationary frame of reference we know how the light beam moves.

Einstein said: in the frame of reference without any gravity the light beam goes in a straight line.

Of course there is gravity on the surface of the Earth, but let's neglect it – or transport all our experiment in outer space in an inertial frame away from any gravitational field.

The only *apparent* gravity, in the elevator, will be due to its acceleration.

The equations of motion for the light beam – think of it, for instance, as one photon – are, in the stationary frame of reference,

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

Now, let's replace the unprimed coordinates of the photon (that is, the coordinates in the stationary frame) by its primed coordinates (that is, the coordinates in the elevator frame).

Equations (12) become

$$\begin{aligned} x' &= ct \\ z' + \frac{gt^2}{2} &= 0 \end{aligned} \tag{13}$$

In the elevator, the photon trajectory is shown in figure 6.

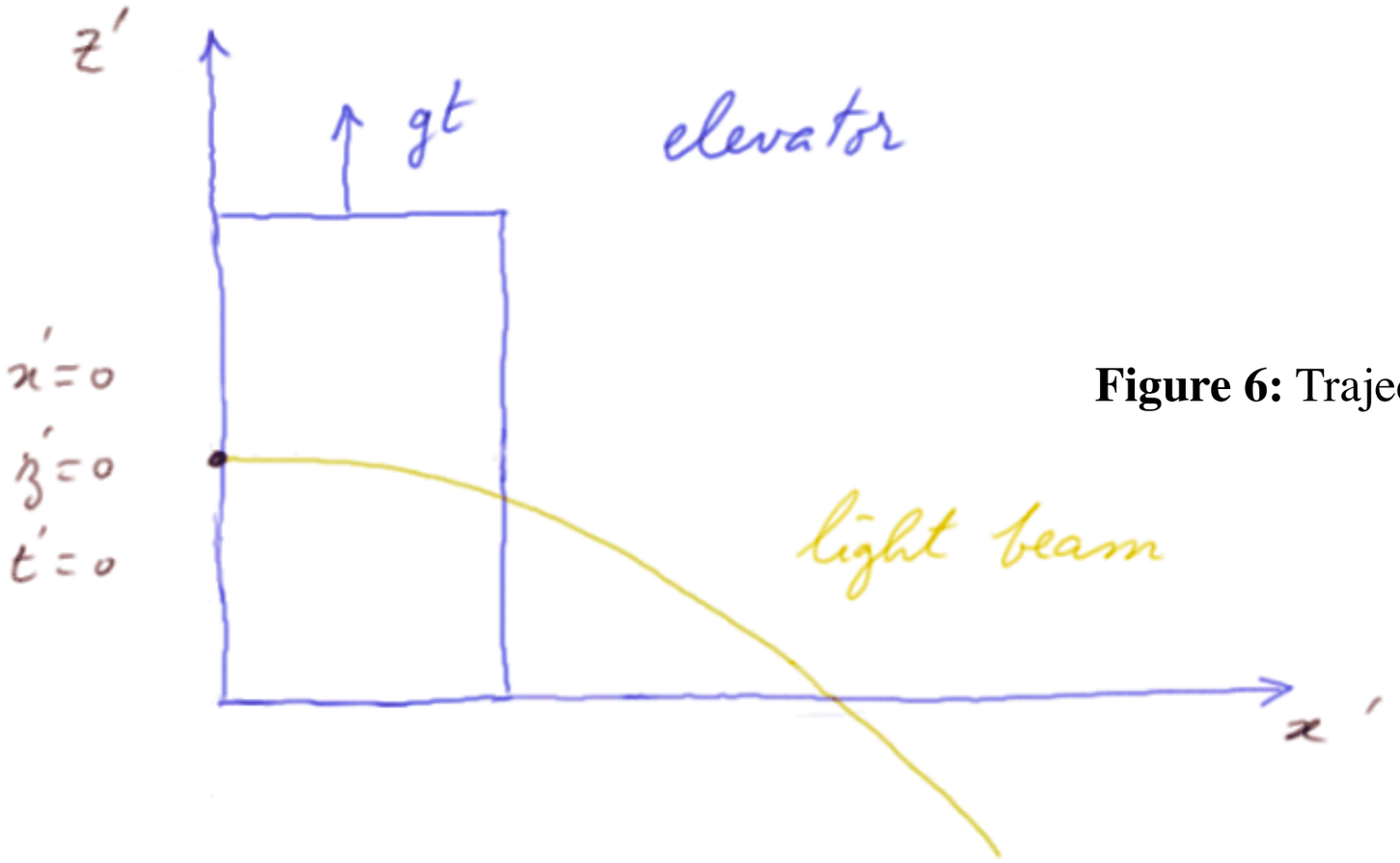


Figure 6: Trajectory of a light beam in the elevator reference frame.

It is a parabola, whose equation linking z′ and x′ we can obtain by eliminating t in equations (13).

If I am in the elevator, I say: gravity is pulling the light beam down.

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 said they are the same thing.

This proved to him that a gravitational field must bend a light ray.

And that was something that no other physicist knew at the time.

At this point what we have learned – the thing we want to abstract – is that it is interesting to think about curvilinear coordinates transformations.

When we do think about curvilinear coordinates transformations, the form of Newton's laws do change.

One of the things that happen is that apparent gravitational fields materialize, that are physically indistinguishable from ordinary gravitational fields.

Well are they really physically indistinguishable?

Not really.

So let's talk about real gravitational fields, namely gravitational field of gravitating objects like the Sun or the Earth.

Tidal forces

Say Figure 7 represents the Earth or the Sun.

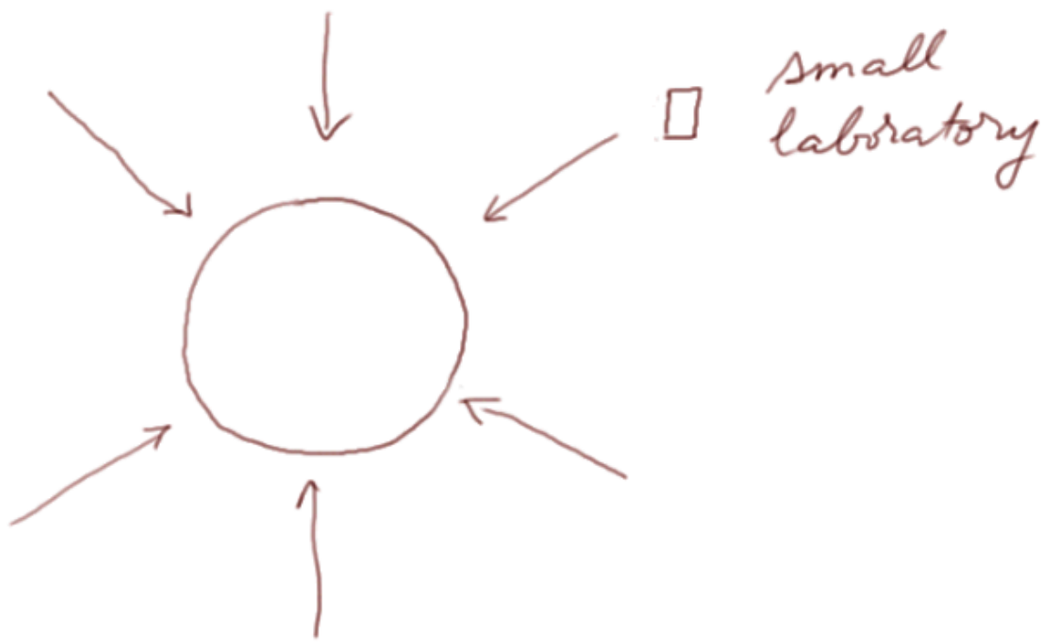


Figure 7: Gravitational field of a massive object, for instance the Earth

And the gravitational acceleration doesn't point vertically on the page.

It points toward the center.

It is pretty obvious that there is no way that you could do a coordinate transformation like we did in the preceding discussion which will remove the effect of the gravitational field.

OK, if you are in a small laboratory in space and that laboratory is allowed to simply fall toward the Earth, or whatever massive object you are considering, then you will think that in that laboratory there is no gravitational field.

This means that if we are falling freely in a uniform gravitational field, you could prove that we feel no gravity and that things float around us like in the International Space Station.

But there is no way globally to introduce a coordinate transformation which is going to get rid of the fact that there is a gravitational field pointing inward 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.

And even more complex transformations will not solve the problem.

One way to understand why we can't get rid of gravity is to think of an object which is not very very small compared to the gravitational field.

My favorite example is a 2000-mile-long-man who is falling in the Earth gravitational field, figure 8.

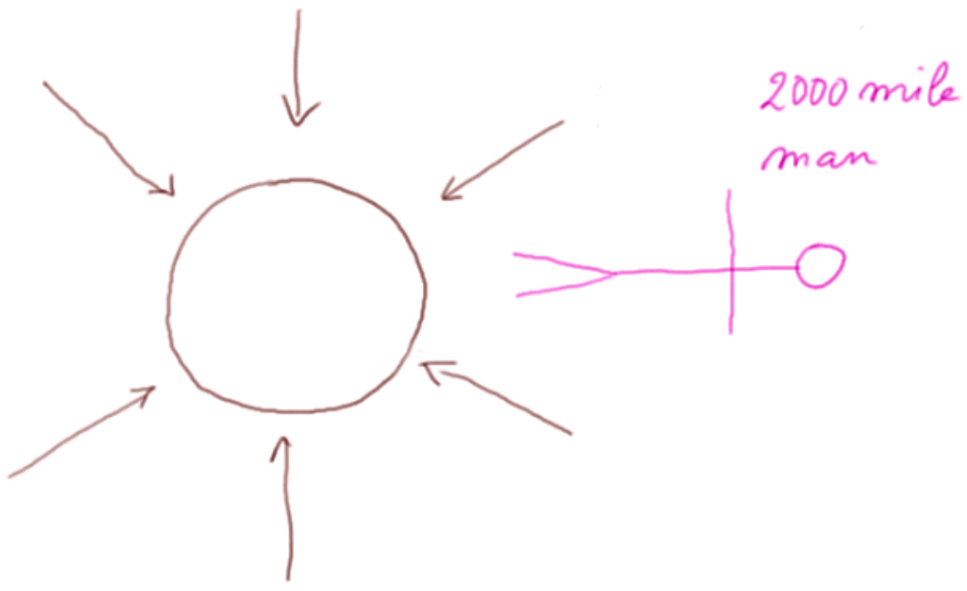


Figure 8: 2000-mile-man falling toward the Earth

Because he is large, not a point mass, the different parts of him feel different gravitational fields.

The further away you are the weaker the gravitational field is.

So his head feels a weaker gravitational field than his feet.

His feet are being pulled harder than his head.

He feels like he is being stretched.

If he is freely falling he can't distinguish this from a stretching feeling.

But he knows that there is a gravitating object there.

The sense of discomfort that he feels, due to the non-uniform gravitational field, cannot be removed by going to a falling reference frame.

And indeed no change of mathematical description whatsoever has ever changed a physical phenomenon.

The forces he feels are called *tidal* forces.

They *cannot* be removed by a coordinate transformation.

Let's also see what happens if he is not falling vertically but sideways?

In that case his head and his feet will be at the same distance of the Earth.

Both will be subjected the same force in magnitude pointing to the Earth.

But since they are radial, they are not parallel.

A part of component of the force pulling his head will be along his body, and a part of the force pulling his feet too.

So they will also *tend* to compress or shrink him.

That sense of compression is again not something that you can remove by any coordinate transformation.

Being stretched or shrunk or both by the Earth field – if you are big enough – is an *invariant* fact.

It is not something you can get rid of by doing a coordinate transformation.

So it is not true that gravity is equivalent to going to an accelerated reference frame.

Einstein of course knew this.

What he really meant is:

> *Small objects for a small length of time cannot tell the difference between a gravitational field and an accelerated frame of reference.*

That raises the question: if I present you with with a force field, does there exist a coordinate transformation which will make it disappear?

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 making it disappear: simply to use the z coordinates instead of the z'.

It is a nonlinear coordinate transformation, but nevertheless it gets rid of the force field.

With other kinds of coordinate transformations – from inertial to moving –, you can make the gravitational field look more complicated, for example transformations which 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 in the x-axis.

What kind of gravitational field do you see?

A very complicated one: it's got a vertical component and it has got a time dependent oscillating component along the x-axis.

If instead of the elevator you use a merry-go-round, and instead of the (x', z', t) coordinates of the elevator, you use polar coordinates (r, θ, t), an object that in the initial frame was fixed, or had a simple motion like the light beam, may have a weird motion.

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 appropriate coordinate transformations you can make some pretty complicated apparent gravitational fields.

They none-the-less are *not* genuine, in the sense that they don't result from the presence of a massive object.

I tell you what the gravitational field is everywhere, how do you determine whether it is just the sort of fake gravitational field coming from a coordinate transformation to a frame with various kinds of accelerations with respect to an inertial one, or it is a real genuine gravitational field?

Well, if we are talking about Newtonian gravity – not a field for instance coming from electromagnetism – there is an easy way.

You just calculate the tidal forces.

You determine whether that gravitational field will have an effect on an object which will cause it to squeeze and stretch.

If calculations are not practical, you take an object, a mass, a crystal.

You let it fall and you see whether there was stresses and strains on it.

These will be detectable things.

If you detect stresses and strains *then* it is a real gravitational field.

If you discover on the other hand that the gravitational field has no such effect, that any object, wherever it is located and let free to move, experiences no tidal force, in other words that the field has no tendency to distort a freely falling system, then it is *not* a real gravitational field.

The test object must be let free.

It should not be held in place by ropes and pulleys or anything else, be it in the elevator or on the merry-go-round.

In summary: how do we know when it is possible to make a coordinate transformation that removes the gravitational field?

How do we know when the field is just an artifact of the coordinate system, and when it is a real physical thing, due to some gravitating mass for example?

If the gravitational field is small enough that it doesn't vary between one part of the system and another, that is when the equivalence principle holds with accuracy.

Einstein knew all that.

So he 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 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.

The *only* thing we are going to use about special relativity is that space-time has a geometry.

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

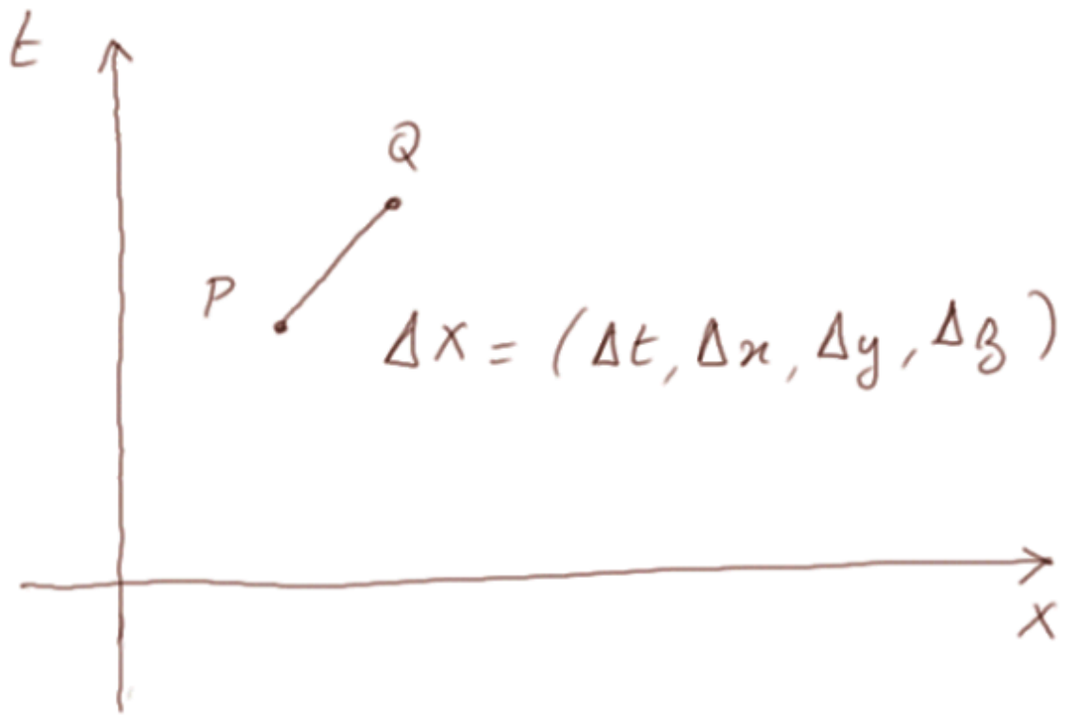


Figure 9: Minkowski geometry.

The distance between P and Q is not the usual Euclidean distance that we could be tempted to think of.

The distance, or length, between any pair of points P and Q, close to each other or not, is defined as follows.

Let's call ΔX the *4-vector* going from P to Q.

To the pair $P\ Q$ is associated the quantity denoted $\Delta\tau$, defined by

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

$\Delta\tau$ is called the *proper time* between P and Q.

It is an *invariant* under Lorentz transformations.

That is *why* it qualifies as a kind of distance, just like in Euclidean 3D space the distance of a point from the origin, $x^2 + y^2 + z^2$, is invariant under rotations.

Equivalently we may define Δs by

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

Δs is called the *proper* distance between P and Q.

$\Delta \tau$ 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, he or she 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 now 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 which 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.

And the intrinsic geometry of the page remains flat even if we furl the page like in figure 10.

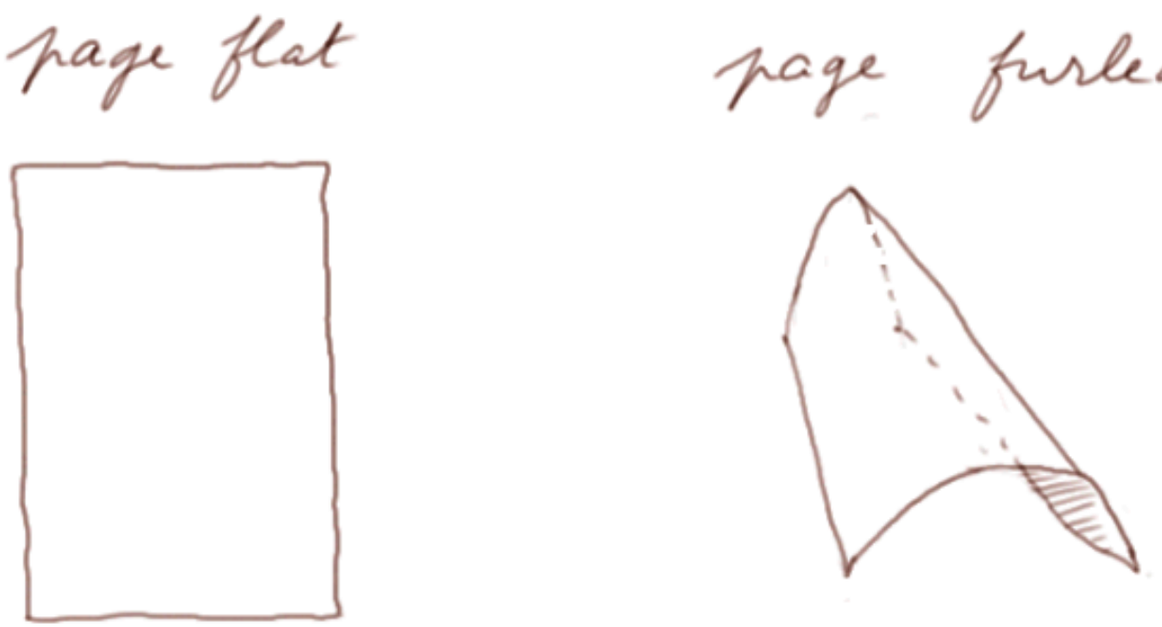


Figure 10: The intrinsic geometry of a page remains flat.

We shall expound mathematically on the idea in a moment.

For now, let's just say that the intrinsic geometry of a surface is the geometry that a 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 it worked carefully it might see hills and valleys, bumps and troughs, if there were any, but it would not determine, as we see it, that the page is furled.

We see it because for us the page is embedded in the 3D Euclidean space we live in.

And 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 which have a minus sign in the definition of the square of the distance.

He was thinking about geometries which were non-Euclidean but were *similar* to Euclidean geometry.

Minkowski geometry is non-Euclidean in two respects.

The square of the distance between two points is not just $\Delta x^2 + \Delta y^2 + \Delta z^2$.

But furthermore the square of the distance can be negative!

So we want to start with the mathematics of Riemannian geometry, that is of spaces where the distance between two points is not the Euclidean distance, but its square is always positive.

We are not going to spend a huge amount of time on it, but we shall nevertheless study a little bit the geometry of Riemannian surfaces.

We look at two points in a space (figure 11).

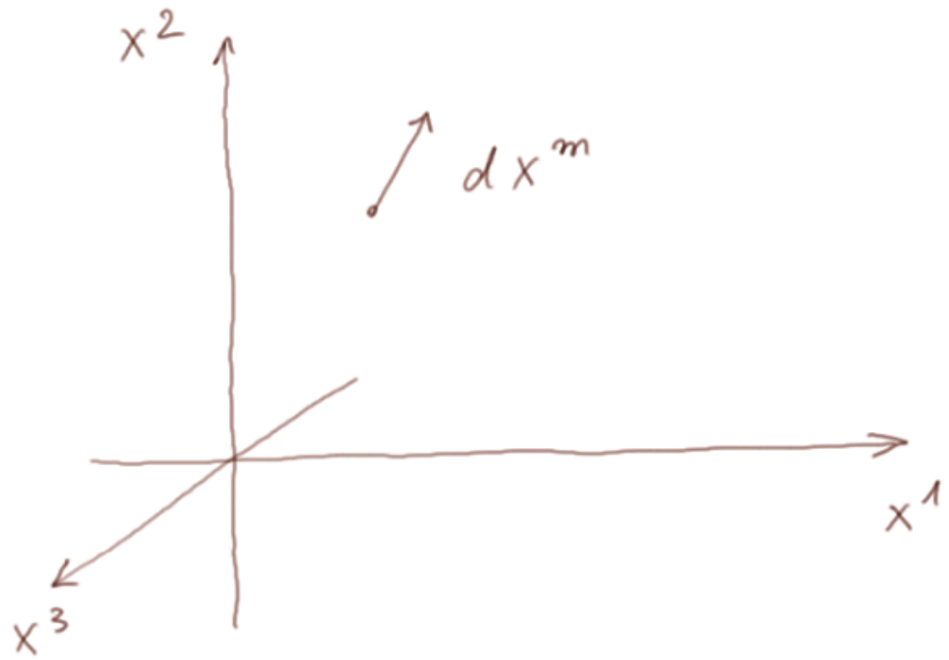


Figure 11: Distance between two points in a space.

In our picture there are three dimensions, three axes, X^1, X^2 and X^3.

There could be more.

So a point has three components, that we can write X^m, where m is understood to run from one to three or to whatever number of axes there is.

And a little shift between two points can be denoted ΔX^m or, if it is to become an infinitesimal, dX^m.

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

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

If we are in three-dimensions then there are three terms in the sum.

If we are in two dimensions, there are two of them.

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 the Euclidean space.

To pursue our study of Riemannian geometry, it is now best to think of 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 with clearly defined unit distance, because it enables us to conveniently view its curvature.

See figure 12.

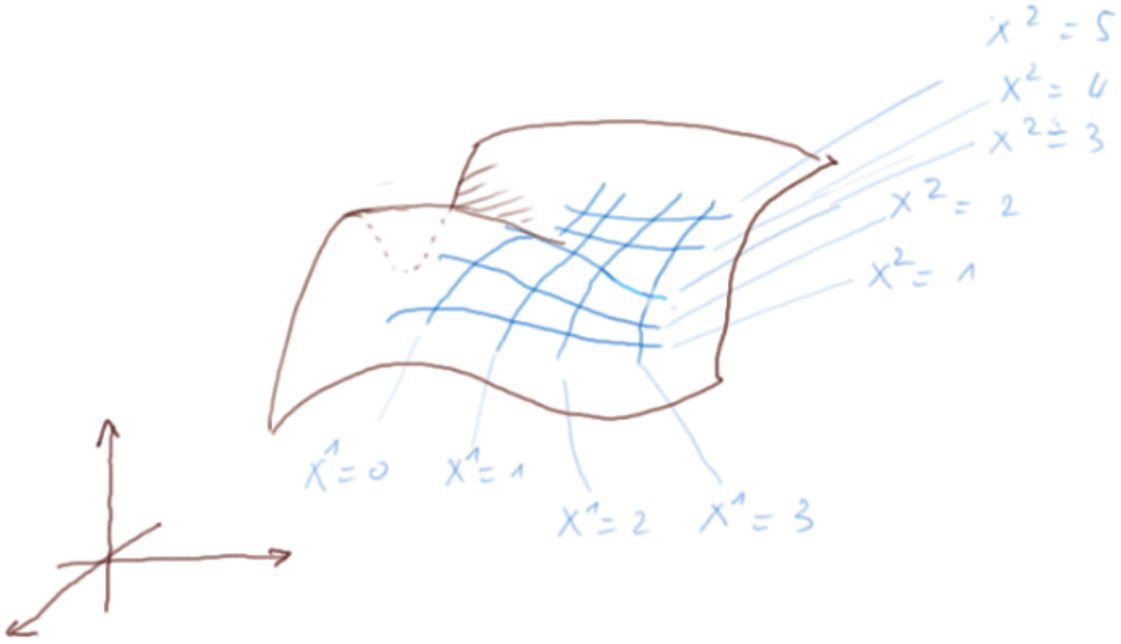


Figure 12: Two dimensional variety.

There is nothing special about two dimensions for such a surface, except that it is *easy* to visualize.

Mathematicians think of "surfaces" even when those have more dimensions.

And usually they don't call them surfaces but **varieties** - get used to the name.

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 *ought to think* of the surface, or variety, as a space in *itself*, equipped with a coordinate system to locate any point, with curvy lines corresponding to one coordinate 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).

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.

So the first thing we do with a surface is to put some co-ordinates on it.

Indeed we want to be able to quantify various statements involving its points, therefore we need to introduce some coordinates onto it.

We just lay out some kind of coordinates as if drawing them with a 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.

So we just lay out some coordinates and we still call them X's.

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 (figure 13).

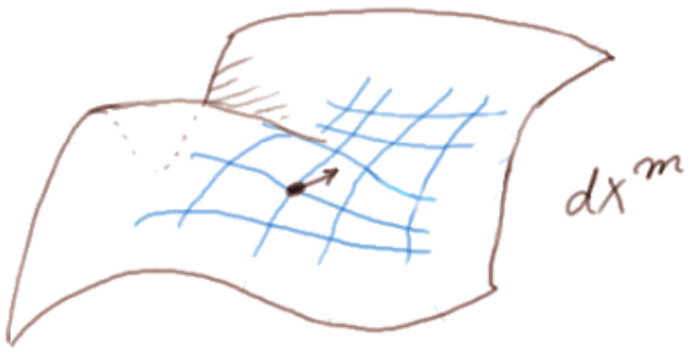


Figure 13: Two neighboring points, and the shift dX^m they form.

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).

It will have similarities with it.

Here is the new definition of dS^2

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

This is the formula in *any* geometry curved or otherwise.

Incidentally equation (15) applies even to flat geometries equipped with curvilinear coordinates.

Suppose you take a flat geometry, like the surface of the page, but 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 either the product of two little shifts, or the square of one little shift, times a coefficient like g_{mn} which depends on X.

Let's consider a *simple* example of distance on a curved surface.

See Figure 14.

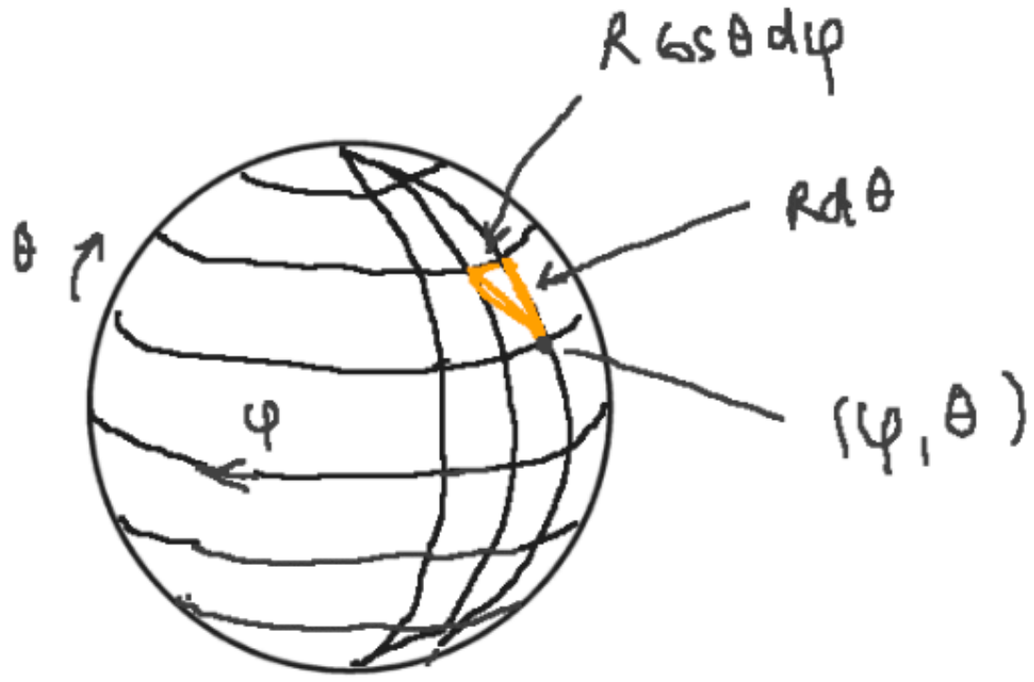


Figure 14: Distance on the surface of the Earth.

It is the distance on the surface of the Earth between two points nearby characterized by longitude and latitude.

Let's denote R the radius of the Earth.

We take two points (ϕ, θ), and $(\phi + d\phi, \theta + d\theta)$, where θ is the latitude and ϕ the longitude.

We apply Pythagoras theorem in a small approximately flat rectangular region to compute the square of the length of its diagonal.

One side along a meridian has length $Rd\theta$.

The other side along a parallel has a length $Rd\phi$ but corrected by the cosine of the latitude.

At the equator it is the full $Rd\phi$, but at the pole it is zero.

So the formula is

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

Here is an example of squared distance not just being $d\theta^2 + d\phi^2$ but having some coefficient functions in front.

In this case the interesting coefficient function is $(\cos\theta)^2$ in front of one of the terms $(dX^m)^2$.

We also note that there are no terms of the form $dX^m dX^n$ because the natural curvilinear coordinates we chose on the sphere are *still orthogonal* at every point.

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 have 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, everything will be quadratic.

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.

Well, we could mean the shortest distance as follows: we would put a peg at A, 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* one way.

And then the other would go around the hill the *other* way.

Think on the 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, not only to calculate the distance but to know actually where to place the string.

So the notion of distance between any two points is a *complicated* one.

But the notion of distance between two neighboring points is *not* so complicated.

That is because locally a smooth surface can be *approximated* by the tangent plane and the curvilinear coordinates lines by *straight* lines – not necessarily perpendicular but straight.

Metric tensor

Now we shall 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.

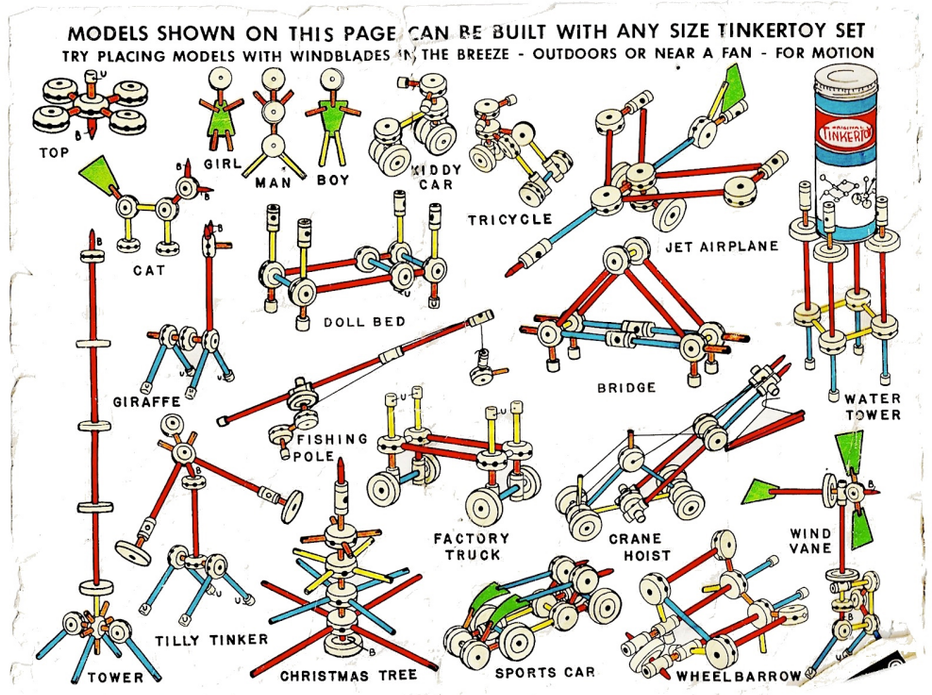
Recall the equation

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

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.

And we would also add more rigid elements diagonally.



This would create a lattice as shown in figure 15.

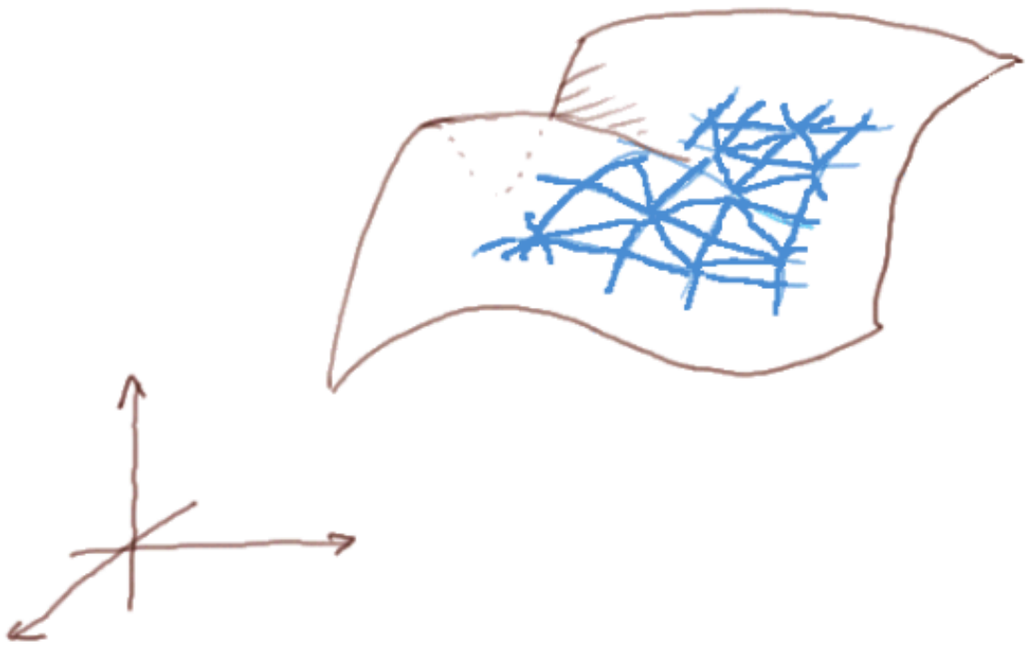


Figure 15: Lattice of rigid Tinkertoy elements arranged on the surface.

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 let them freely move in any direction from each other.

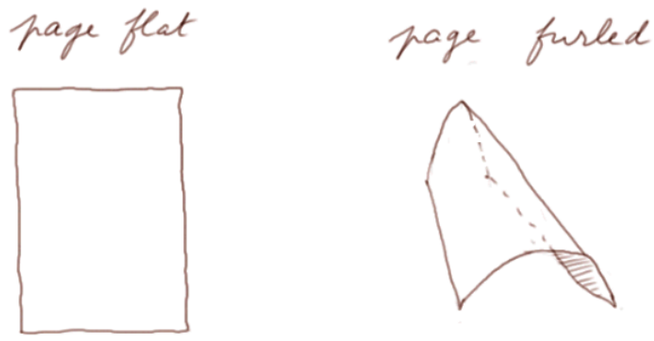
Now let's lift our lattice from the surface.

Sometimes it will keep its shape rigidly, sometimes it won't.

It will not keep its shape if going from the initial shape to a new shape doesn't force any Tinkertoy element to be stretched or compressed or bend.

And 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.



A question to keep in mind:

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

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 local distance between two neighboring points on our Riemannian surface.

When the lattice of Tinkertoy elements can be laid out flat, the geometry of the surface is said to be flat.

We will define it more carefully 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 sphere, we won't be able to remove it to lay it out on a flat plane.

The question now we have to address is this: if I had 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 it is an intrinsically curved space which cannot be flattened out and laid out on a flat plane?

Let's formulate it more precisely and mathematically: we start from the metric tensor $g_{\mathrm{mn}}(X)$ 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 may 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.

So I select one set of coordinates and I give you the metric tensor of my surface.

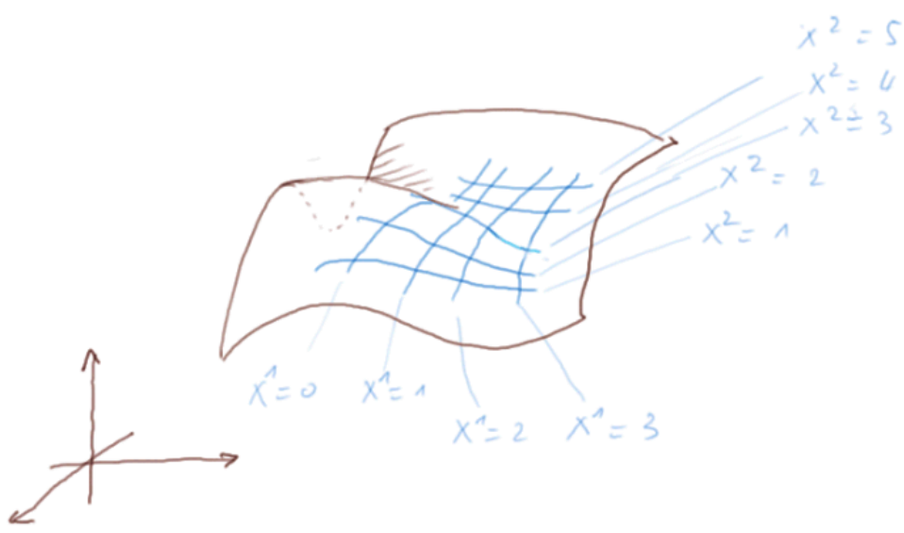
In effect I tell you the distance between every pair of neighboring points.

Now the question is: is my surface flat or is it not flat?

Of course we have to explain what we mean by flat.

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 it is flat.

Otherwise you would say that it is not flat, it is curved.

That procedure is *good* for a two-dimensional surface under certain conditions.

But anyway it is *not* so great for higher dimensional surfaces.

So 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

> *The space is flat if you can find a coordinate transformation, that is, a different set of coordinates, in which the distance formula for* dS^2 *is just* $(dX^1)^2 + (dX^2)^2 + ...(dX^n)^2$, *as it would be in Euclidean geometry.*

That the initial g_{mn}'s form the unit matrix, with ones on the diagonal and zeroes elsewhere – as if equation (15) were just Pythagoras theorem – is *not necessary*.

But we must find a coordinate transformation which can make it *look* like that.

In that sense, it has a vague similarity – it turned out not to be a vague similarity at all but a close parallel – with the question of whether you can find a coordinate transformation which removes gravitational field.

The *question* is: can you find a coordinate transformation which removes the curvy character of the metric tensor g_{mn}?

In order 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 and a little bit of differential geometry.

At first it looks annoying because we shall have to deal with all these indices floating around, and different coordinate systems, and partial derivatives of components, etc.

But once you get used to it, it is really very simple.

It was invented by Ricci-Curbastro and Levi-Civita at the end of the XIXth century to build on works of Gauss and Riemann.

And it was actually made even simpler by the famous Einstein summation convention which astutely gets rid of most summation symbols.

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.

g_{mn} depends on X.

It is the *same* tensor field, but it has a *different* matrix at each point.

For a given set of coordinates, it has a collection of matrices – *one* for each point.

And for another set of coordinates, it will have *another* collection of matrices.

And we still talk of the *same* tensor field.

Its components *depend* on the coordinates.

But the tensor itself is an *abstract object* which doesn't.

We *already* met the distinction with 3D vectors.

You *cannot* diagonalize all of them at the same time.

At a given point, you can indeed diagonalize $g_{\mathrm{mn}}(X)$ even if the surface is not flat.

It is equivalent to working locally in the tangent plane of the surface at X, and orthogonalizing out the co-ordinate 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.

So we are looking for a coordinate transformation, $X \rightarrow Y$, which *turns* g_{mn} into the Kronecker delta function δ_{mn}.

Remember, X and Y represent the *same* point on the surface.

They are its coordinates in *different* coordinate systems.

And the Kronecker delta function is

$$\delta_{mn} = 1, \text{ if } m = n$$
$$\delta_{mn} = 0, \text{ if } m \neq n$$

We want the square of the interval distance to *become*

$$dS^2 = \sum_{m,\ n} \delta_{mn}(Y)\ dY^m dY^n \tag{17}$$

You can also *think* of δ_{mn} as the unit matrix, with ones everywhere on the diagonal, and zeroes everywhere off the diagonal.

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

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

and I ask you if by a coordinate transformation $X \rightarrow Y$ you can reduce it to equation (17).

If *yes*, the space is called flat.

If *no*, the space is called curved.

Of course the space could have some *portions* which are flat.

There could exist a set of coordinates such that in a region the metric tensor be the Kronecker delta.

But it is called flat if it is *everywhere* flat.

This is now a mathematics problem: given a tensor field $g_{\mathrm{mn}}(X)$ on a multi-dimensional variety, how do we figure out if there is a coordinate transformation which would change it into the Kronecker delta function?

To *answer* that question, we have to understand better how things *transform* when you make coordinate transformations.

And that is the subject of *tensor analysis*.

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 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:

Gravity is curvature.

But we will come to it as we get through tensor analysis.

Obviously, in trying to determine whether we can transform away $g_{\mathrm{mn}}(X)$ and turn it into the trivial $\delta_{\mathrm{mn}}(Y)$, the first question to ask is *how* does $g_{\mathrm{mn}}(X)$ transform when we change coordinates?

We have to introduce *notions* of tensor analysis which are rather easy.

First tensor rule: contravariant 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 you will discover that the manipulations obey strict rules and turn out to be rather simple.

We shall *begin* with a simpler thing than $g_{\mathrm{mn}}(X)$.

Suppose there are two sets of coordinates on our surface.

There a set of coordinates X^{m}.

And I could call the other coordinates X' as we did earlier.

But then we would be running into *horrible* notation like X^{11} and X'^{11}.

So the second set of coordinates, we call Y^{m}.

To be very explicit, a same point P on the surface has coordinates $X^1(P), X^2(P), X^3(P), \ldots, X^{\mathrm{N}}(P)$ if we are on variety of dimension N.

And it also has coordinates $Y^1(P), Y^2(P), Y^3(P), \ldots, Y^{\mathrm{N}}(P)$.

But we will almost never recall the variable P.

So we have at every point a collection of X^m coordinates and a collection of Y^m coordinates.

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 know its coordinates in the other coordinate system.

So each coordinate X^m is a *function* of all the coordinates Y^n.

We can use whatever dummy index variable we want if that helps avoid confusion.

In fact we write simply

$$X^m(Y)$$

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

$$Y^m(X)$$

We have two coordinate systems, *each* one being a function of the other.

Now we ask: how the differential elements dX^m transform?

The collection of differential elements dX^m is a small vector as shown in figure 16.

Figure 16: Displacement dX^m.

Remember, 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, most of the time we *must look* at it expressed with components.

And those dX^m are the components of the little displacement under consideration, in the X coordinate system.

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

$$dX^m = [\, dX^1,\ dX^2,\ dX^3,\ \ldots\ , dX^N\ ]$$

Said another way, when we change a little bit X^1, and change a little bit X^2, and change a little bit X^3 etc., the point X moves to a nearby point, and the displacement is dX^m.

Now we look at how dY^m, for any given m, can be expressed in terms of the dX^p's.

It is an *elementary result of calculus* that

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

Remember elementary calculus: if a function f depends on two variables a and b we write

$$f(a,\ b)$$

Then the total differential of f when a and b each moves a little bit is given by the formula

$$df = \frac{\partial f}{\partial a} da + \frac{\partial f}{\partial b} db$$

Think of a function f as simple as the area of a rectangle of length a and width b: $f(a, b) = ab$.

Equation (18), reproduced here

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

says nothing more than figure 17 in a more general setting.

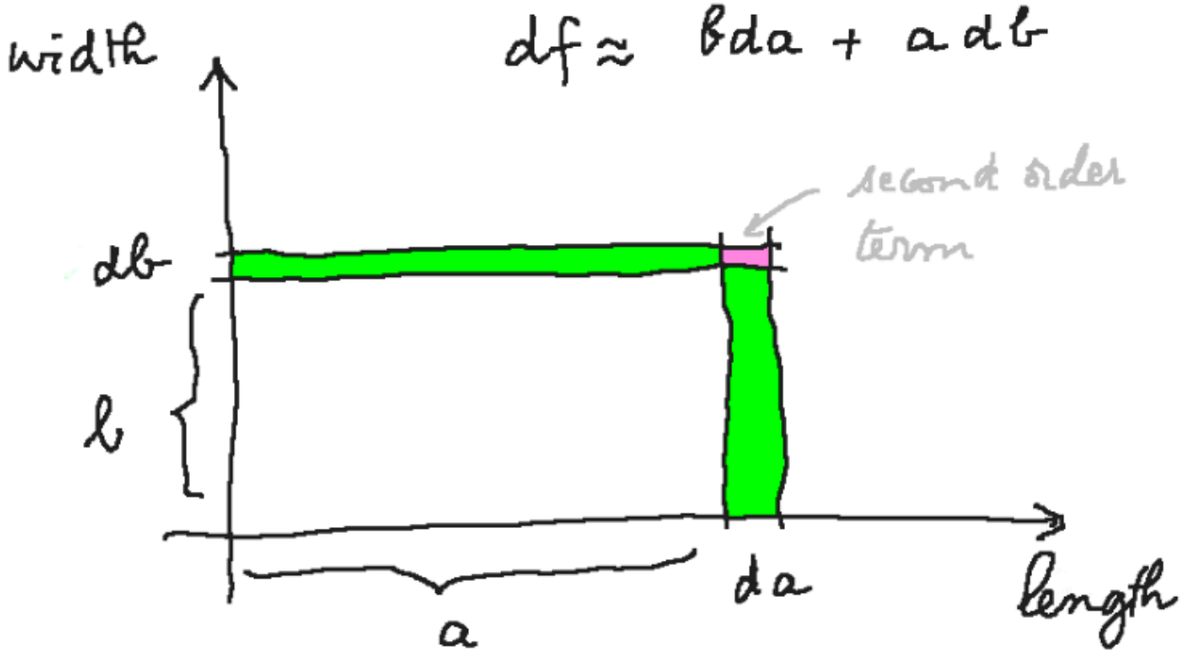


Figure 17: $df = \frac{\partial f}{\partial a} da + \frac{\partial f}{\partial b} db$ it is the green area.

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.

Now right here we have the **first example of the transformation of a tensor**.

We now have the expression of the small displacement of a point on the surface (figure 16), expressed in two different coordinate systems, dX^m and dY^m.

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.

Let's represent again the small displacement, and also locally the two sets of coordinates, figure 18:

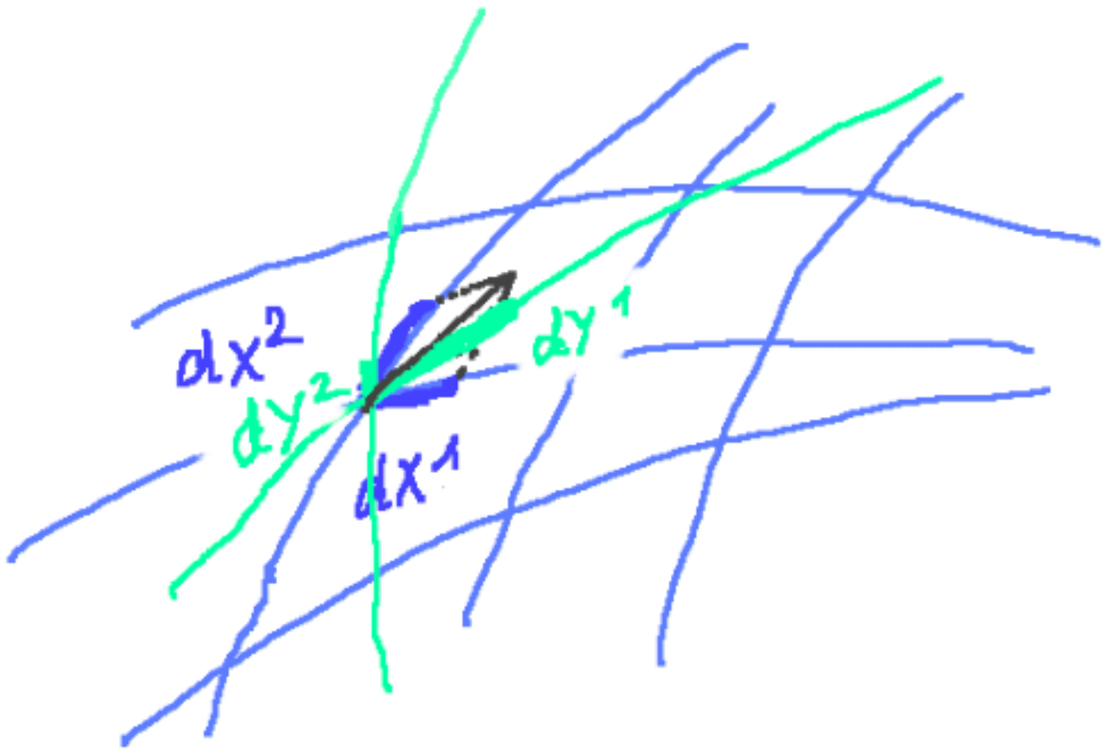


Figure 18: Small displacement, and two sets of coordinates.

Equation (18) is the transformation property of the components of the vector.

Now usual vectors, we shall call **contravariant vectors**.

So equation (18), which is really a set of equations, is the transformation properties of the components of what is called a contravariant vector.

If we leave for a moment geometric representations or visualizations, and think in more abstract mathematical terms, a contravariant vector simply means a thing which transforms as in equation (18).

Let's give the general definition of a **contravariant vector**.

It is an object, a "thing", which has *components*(upper index)

$$V^m$$

These components *depend* on the coordinate system.

V^m is the *collection* of components of the object in the *X* frame.

Let's denote $(V')^{m}$ the components of the *same* object in the *Y* frame.

For the object to be a contravariant vector its two sets of components must *satisfy*

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

Equation (19) is the *fundamental* equation of a contravariant vector.

We will soon see other types of objects called *covariant* vectors.

Remember, the **ordinary vectors** of $\mathbb{R}^3$, or $\mathbb{R}^n$, or of any vector space are **contravariant vectors**.

The position of a point is a contravariant vector.

The speed of a particle is a contravariant vector.

In tensor analysis, they are called contravariant because they change **contrary** to the change of basis.

For instance if your new basis vectors are simply the old basis vectors, all divided by 10, then V' will simply be V multiplied by 10.

There are *many things* which transform according to equation (19) and therefore are contravariant vectors.

Einstein summation convention

One of Einstein's great contributions to notations was that you can leave out the summation sign in equation (19):

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

can be rewritten

$$(V')^m = \frac{\partial Y^m}{\partial X^p} V^p \tag{20}$$

People will figure out that there is an *implicit* summation just by looking at the equation.

They will see that the left hand side of equation (20) has no index p, while the right hand side has an index p, so it *must* be summed over.

Whenever there is a *repeated* index like p, that seems not to make sense, because it appears on the right side of the equation, but disappears from the left side, it really *stands* for summation over that index.

Second tensor rule: covariant vectors

Now let's talk about covariant vectors.

They are *completely different* creatures.

The most typical covariant vectors are *gradients*.

Suppose we have a scalar function *defined* on our space, surface, variety – whatever you call it.

That is, if we want to use the terminology usually introduced, we consider a *scalar field*.

A scalar function is a function that *doesn't* have to be transformed.

It just has some *meaningful* value at every point in space.

Let's call the function S.

Its values are $S(X)$, if we use X for the coordinates of a point.

Then the gradient of S is also a kind of vector.

What is the gradient of S?

It is the set of derivatives of S with respect to each X^p

$$\frac{\partial S(X)}{\partial X^p} \tag{21}$$

Example 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 multi-dimensional but simply a number, and which 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.

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

So the gradient of a scalar function is a vector.

It is the gradient vector field.

But it **doesn't transform in the same way as contravariant vectors**.

If we compute the gradient of S with respect to the Y coordinates instead of the X coordinates, we get a different set of values, and we want to know how there are related.

It is easy to figure out.

Suppose we know the partial derivatives of S with respect to the X^{p}'s, equation (21).

We want to compute

$$\frac{\partial S}{\partial Y^m}$$

Again it is an *elementary* calculus problem.

The rate of change in S when we change Y^m a little bit, keeping all the other Y coordinates fixed, is just derivative of S with respect to X^p times the derivative of X^p with respect to Y^m – summed over all the p's.

$$\frac{\partial S}{\partial Y^m} = \frac{\partial S}{\partial X^p} \frac{\partial X^p}{\partial Y^m} \tag{22}$$

This is some version of the so-called *chain* rule.

Again the right hand side of equation (22) means sum over p.

We shall no longer write the big Σ_p sign meaning that the sum is over p.

If there is a repeated index on one side, which doesn't show up on the other side, it must mean that it is summed over.

That is Einstein summation convention.

And it makes mathematics texts on certain topics less cluttered and easier to read.

Now let's see how equation (22) *compares* with equation (20) for contravariant vectors?

Let's denote W the gradient of S with respect to the X coordinates, and W' the gradient of S with respect to the Y coordinates.

Equation (22) can then be rewritten

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

Notice that we now write the indices m of W' and p of W downstairs.

We will see why this makes nice sense in a minute.

Equation (23) is the *fundamental* equation linking the primed and unprimed version of components of a covariant vector, that is its components in the Y system and in the X system.

Let's rewrite both equations (20) and (23) next to each other, and relabel them (24a) and (24b):

Contravariant vectors

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

Covariant vectors

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

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 repeat that ordinary vectors – displacements of positions, or velocities for instance – are contravariant vectors.

We saw that they change **contrary** to the basis change.

Gradients, on the other hand, change like the basis change.

For instance *if* the new basis vectors are obtained simply by dividing the old ones by 10, then the gradient components will also be divided by 10.

That is why such objects are called **covariant**.

They are called covariant vectors, but they are *not* ordinary vectors.

In fact in mathematics, they are the so-called *dual* vectors.

Equations (24a) and (24b) are *fundamental* equations for this class.

You need to *understand* them, become familiar and at ease with them, because they are completely *central* to the entire subject of general relativity.

He or she *needs to know* where the indices go for different kinds of objects, how these objects transform.

That is in *some sense* what general relativity is all about.

It is about the *transformation properties of different kinds of objects*.

Covariant and contravariant 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.

And this is geometrically well defined even before we consider any basis.

But we can also think of it more abstractly as some object which has components.

And its 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 old vectors.

Therefore we can also equate the object to an ordinary vector.

In tensor analysis we call them *contravariant* vectors.

Similarly, some other objects have components which transform according to equation (24b).

They cannot be equated to our old ordinary vectors.

We call them *covariant* vectors.

A vector – be it contravariant or covariant – is a *special case of a tensor*.

We are *not* going to define tensors geometrically.

For us, at first, tensors will be things which are *defined* by the way that they transform.

The way they transform *means* the way they change when we go from one set of coordinates to another.

Later on we will give a *geometric interpretation* of tensors.

We will also go *deeper* into contravariant and covariant vectors.

We will see that an object with one index *can have* a contravariant version and a covariant version.

All this will be studied in the next lesson.

For the time being, let's continue to proceed step by step in our construction of the mathematical tools necessary for general relativity.

So 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.

Let's imagine the *product* of two contravariant vectors.

It is *not* the dot product nor the cross product.

It is going to be called the *outer product* or the tensor product.

Anyway, it is an operation which, given two things, associates a third thing.

We consider the two contravariant vectors V and U, and we consider the product

$$V^m U^n$$

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 MN 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 contravariant vector V, but is also a notation, showing explicitly the position of the index, 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 we write the indices of T^{mn}, because T^{mn} is not the same as T^{nm}

T^{mn} is a special case of a tensor or *rank* 2.

Rank two 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 prime frame.

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

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

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

Finally $V^{\mathrm{p}}U^{\mathrm{q}}$ 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 vectors how T transforms.

Now this leads us to the following definition:

> *Anything which transforms according to equation (26)*
> *is called a contravariant tensor of rank 2.*

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 kind of things are tensors like that?

Many things.

Products of vectors are particular examples, but there are other things which are not products and still are tensors according to the above definition.

We are going to see that the metric object g_{mn} is an tensor.

But it is a tensor with covariant indices.

So to finish this lesson 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.

Their product transforms as follows

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

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

More generally let's call it T_{mn} – a different object, but we still use T for tensor.

It is a tensor with two lower indices and it transforms according to this equation

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

You should now *be able* to figure out how a tensor with one upper and one lower index must transform.

In the next lesson, we will also see how the metric object g transforms.

And we will *discover* 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 which will turn g_{mn} into δ_{mn}.

That is the *mathematics* question.

It is a *hard* mathematics question in general.

But we *will find* the condition.

Lesson 2: Tensor mathematics

Introduction

A good notation, as we said, will carry you a long way.

When it is well conceived, it just sort of automatically tells you what to do next.

That means that you can do physics in a completely *mindless* way.

It is like having Tinkertoys.

It is pretty clear where the stick has to go into.

It has to go into the thing with the hole.

You can try putting a hole into a hole or forcing a stick into a stick.

There is only one thing you can do.

You can put the stick into the hole, and the other end of the stick can go into another hole.

Then there are more sticks and more holes you can put them into, etc.

The notation of general relativity is *much like that*.

If you follow the rules, you *almost* can't make mistakes.

But you have to *learn* the rules.

They are the rules of tensor algebra and tensor analysis.

Flat space

The *question* we are aiming at in this chapter is to understand enough about tensor algebra and analysis, and metrics, to be able to distinguish a flat geometry from a non-flat geometry.

That *seems* awfully simple.

Flat *means* like a plane.

Non-flat *means* with bumps and lumps in it.

And you would think we could tell the difference very easily.

Yet sometimes it is *not* so easy.

For example as discussed in the last lesson, if I consider this page, it is *flat*.

If I roll it or furl it, the page now looks curved but it is *not really* curved.

It is exactly the *same* page.

The relationship between the parts of the page, the distances between the letters, the angles, and so forth, *don't* change.

At least the distances between the letters measured along the page *don't* change.

So a folded page, if we don't stretch it, if we don't modify the relations between its parts, *doesn't* acquire a curvature.

Technically it introduces what is called an **extrinsic curvature**.

Extrinsic curvature has to do with the way a space – in this case the page – is *embedded* in a higher-dimensional space.

For instance whatever I do with the page is *embedded* in the three dimensional space of the room.

When the page is laid out flat on the desk, it is embedded in the embedding space in *one* way.

When it is furled like in figure 1, it is embedded in the same space in *another* way.

Figure 1: Intrinsic and extrinsic geometries and curvatures: the intrinsic geometry of the page remains flat.

The extrinsic curvature which we perceive has to do with **how** the space of the page is *embedded* in the larger space.

But it has nothing to do with its **intrinsic geometry**.

If you like, you can think of the intrinsic geometry as the geometry of a tiny little bug that moves along the surface.

It cannot look out of the surface.

It only looks around while crawling along the surface.

It may have surveying instruments with which it can measure distances along the surface.

It can draw a triangle, measure also the angles within the surface, and do all kinds of interesting geometric studies.

But it *never* sees the surface as embedded in a larger space.

Consequently the bug will *never* detect that the page might be embedded in different ways in a higher dimensional space.

It will *never* detect it if we create a furl like in figure 1, or if we remove the furl and flatten the page out again.

The bug just learns about the *intrinsic* geometry.

The intrinsic geometry of the surface means the geometry that is *independent* of the way the surface is embedded in a larger space.

General relativity and Riemannian geometry, and a lot of other geometries, *are* all about the intrinsic properties of the geometry of the space under consideration.

It doesn't have to be two dimensional.

It can have any number of dimensions.

Another *way to think* about the intrinsic geometry of a space is this.

Imagine sprinkling a bunch of points on this slide – or on a three dimensional space, but then we would have to fiddle with it in four dimensions or more...

Then draw lines between them so they triangulate the space.

And then state what the distance between every pair of neighboring points is.

Specifying those distances specifies the geometry.

Sometimes that geometry can be flattened out without changing the length of any of these little links.

In the case of a two-dimensional surface, it means laying it out flat on the desk without stretching it, tearing it, or creating any distortion.

Any small equilateral triangle has to *remain* an equilateral triangle.

Every small little square has to *remain* a square, etc.

But if the surface is *intrinsically* non-flat there will be small constructions that *cannot* be flattened out.

The other day on his motorbike a friend saw on the road the following bulge(Figure 2), probably due to pine roots, with lines drawn on it, and a warning painted on the pavement.

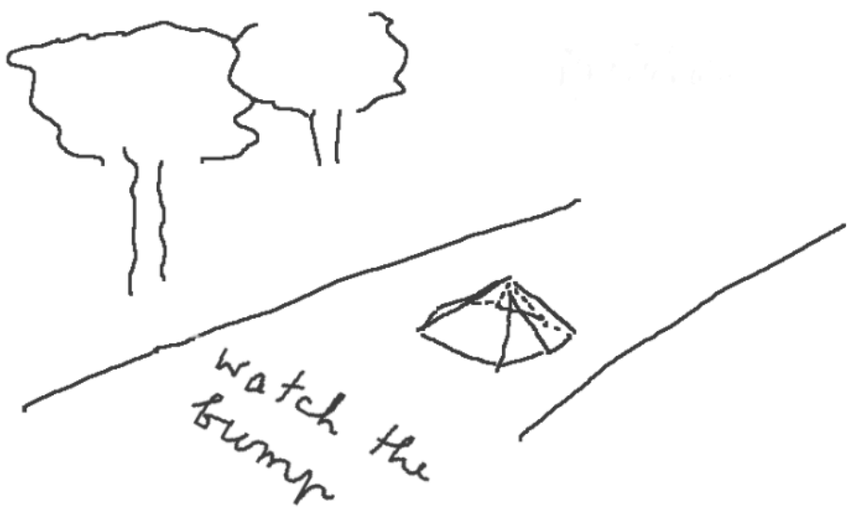


Figure 2: Watch the bump.

The road menders must have taken a course in general relativity!

Such a bump cannot be flattened out without stretching or compressing some distances.

A *curved* space is basically one which cannot be flattened out without distorting it.

It is an *intrinsic* property of the space, not *extrinsic*.

Metric tensor

We want to answer the mathematical question:

Given a space and its metric defined by the following equation

$$dS^2 = g_{mn}(X)\ dX^m dX^n \tag{1}$$

is it really flat or not?

It is important to *understand* that the space may be intrinsically curved, like the road with a bump in figure (2), or we may think that it is curved because equation (1) looks complicated, when actually it is intrinsically flat.

For instance we can draw on a flat page a bunch of funny curvilinear coordinates as in figure 3.

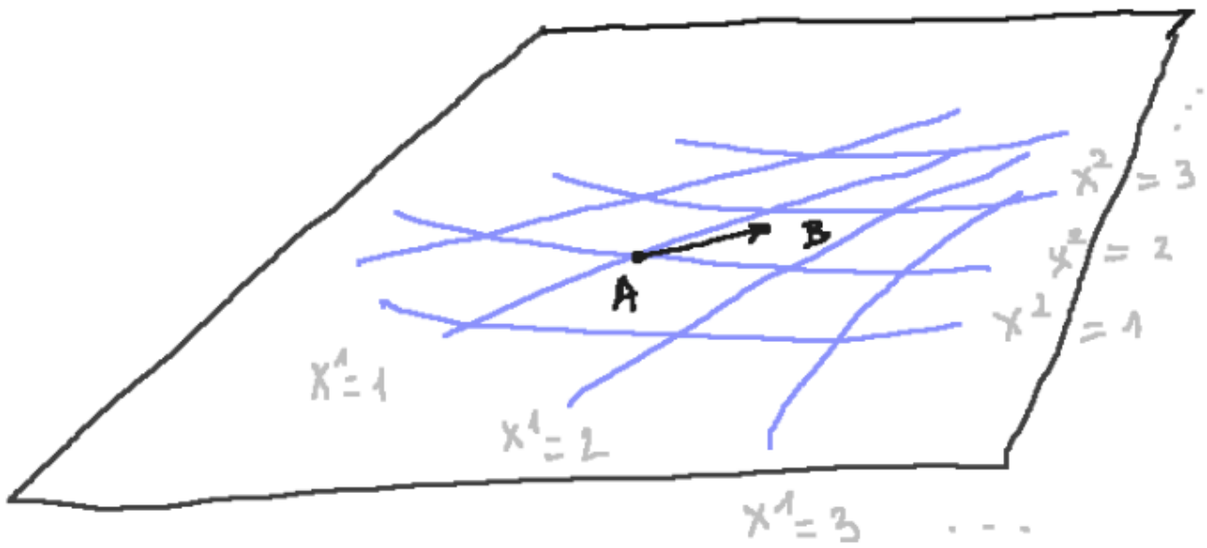


Figure 3: Curvilinear coordinates *X*'s of a flat page.

At each point *A*, if we want to compute the distance between *A* and a neighboring point i, we cannot apply Pythagoras theorem.

We have to apply the Al-Kashi theorem which generalizes Pythagoras taking into account the cosine of the angle between the coordinate axes.

And also perhaps we have to correct for units which are not unit distances on the axes.

Yet the page is intrinsically flat, be it rolled or not in the embedding 3D Euclidean space.

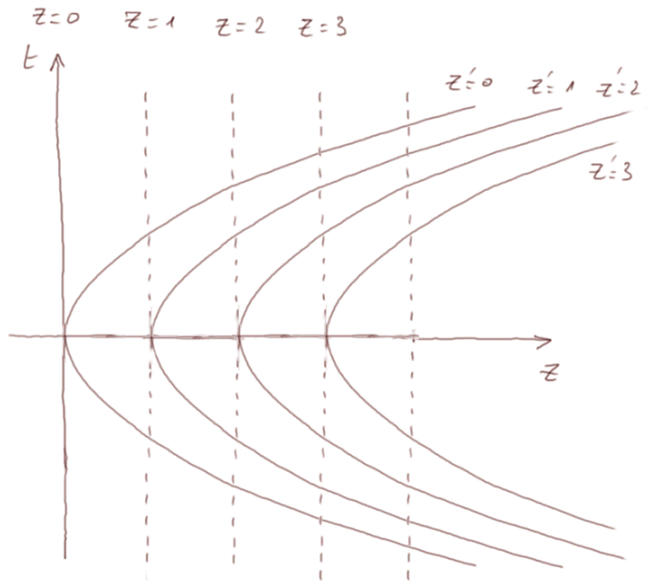
It is easy to find a set of coordinates *Y*'s which will transform equation (1) into Pythagoras theorem.

On the pages of school notebooks they are even usually shown.

And it doesn't disturb us to look at them, interpret them, and use them to locate a point, even when the page is furled.

Our mathematical goal matches closely the question we addressed in the last lesson of whether there is a real gravitational field or the apparent gravitational field is just due to an artifact of funny space-time coordinates.

For instance in figure (4) of lesson 1



the curvilinear coordinates were due to the accelerated frame we were using, not to tidal forces.

The space-time was intrinsically flat.

So we want to tackle the mathematical question.

Typically we are given the metric tensor of equation (1).

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

The mathematical question is a hard one.

It will keep us busy during the entire lesson and more.

Before we come to it, we need to get better acquainted with tensors.

We have begun to talk about them in the last lesson.

We introduced the basic contravariant and covariant transformation rules.

In this lesson, we want to give a more formal presentation of tensors.

Scalars and vectors are special cases of tensors.

Tensors are the general category of objects we are interested in.

Scalar, vector and tensor fields

So for us, tensors are collections of indexed values which depend on coordinate systems.

And they transform according to certain rules when we go from one coordinate system to another.

We are going to be interested in spaces such that at every point P in space, located by its coordinates X in some coordinate system, there may be some quantities associated with that point – what we call fields.

And those quantities will be tensors.

There will also be all kinds of quantities that will not be tensors.

But in particular we will be interested in tensor fields.

The simplest kind of tensor field is a scalar field $S(X)$.

A scalar field is a function which to every point of space associates a number – a scalar –, and everybody, no matter what coordinate system he or she uses, agrees on the value of that scalar.

So the transformation properties in going, let's say, from the X^m coordinates to the Y^m coordinates is simply that the value of S at a given point P doesn't change.

We could use extremely heavy notations to express this fact in the most unambiguous way.

But we will simply denote it

$$S'(Y) = S(X) \tag{2}$$

The right hand side and the left hand side denote the value of the same field at the same point P, one in the Y system, the other in the X system.

Y is the coordinates of P in the Y system, X is the coordinates of P in the X system.

And we add a prime to S when we talk of its value at P using the Y coordinates.

With practice, equation (2) will become clear and unambiguous.

To understand what distinguishes a scalar field from any old scalar function, notice that if we fix the coordinate system then "the first coordinate of a vector field" is a scalar function, but is not a scalar field, because it depends on the coordinate system, and it changes if we change coordinate system.

Let’s represent, on a two-dimensional variety, the X coordinate system.

And now, to avoid confusion, let’s not embed the surface in any larger Euclidean space (see Figure 4).

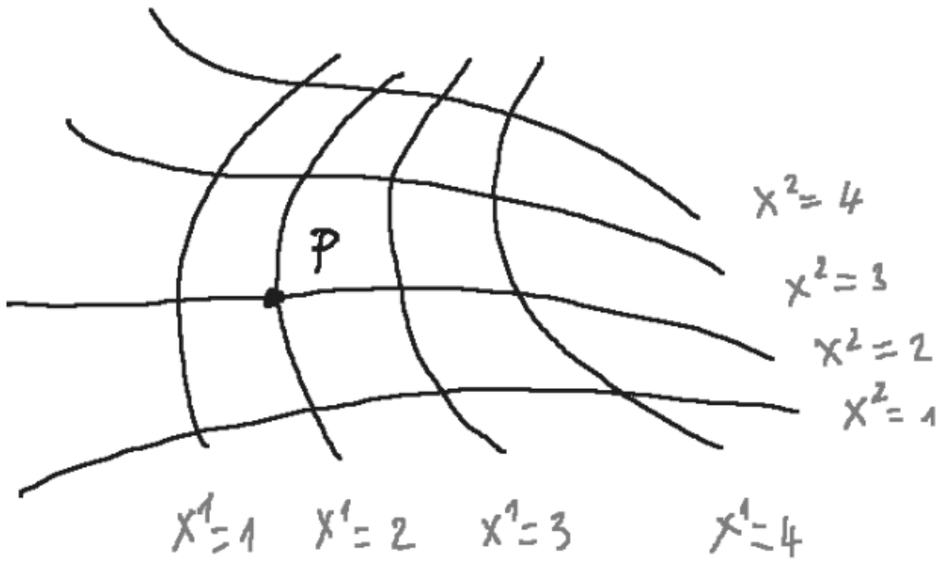


Figure 4: Curvilinear coordinates *X*’s on a two-dimensional curved variety.

Any point *P* of the space is located by the values its coordinates X^1 and X^2.

Of course we could think of a higher dimensional variety.

There would then be more coordinates.

Globally, we denote them X^m.

Now on the same space, there could be another coordinate system, Y, to locate points (see Figure 5).

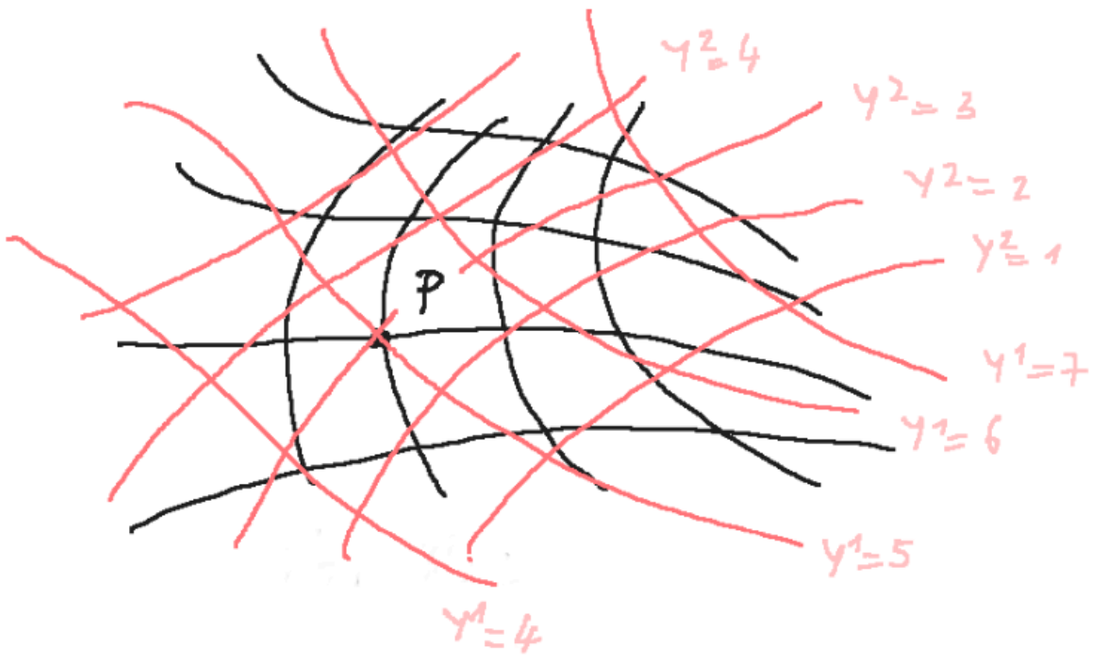


Figure 5: Second coordinate system Y on a two-dimensional curved variety.

In our figure, the point P has coordinates $(2, 2)$ in the X system, and $(5, 3)$ in the Y system.

Of course these coordinates don't have to be integers.

They can take their values in the set of real numbers, or even in other sets.

What is important to note is that at any point P, there are two collections of coordinates

$$X^m \text{ and } Y^m$$

The X^m and Y^m are related.

At any point P, each coordinate X^{m} is a function of all the Y^{m}.

And conversely.

We write it this way

$$X^m = X^m(Y) \tag{3a}$$

and

$$Y^m = Y^m(X) \tag{3b}$$

This is a coordinate transformation of some kind, and its inverse.

It can be pretty complicated.

We will only assume that functions (3a) and (3b) are continuous, and that we can differentiate them when needed.

But there is nothing more special than that.

Scalars transform trivially.

If you know the value of S at a point P, you know it no matter what coordinate system you use.

Next are vectors.

They come in two varieties: contravariant vectors which we denote with an upstairs index V^m and covariant vectors with a downstairs index V_m

We spoke about them in the last lesson.

Now we are going to see a little bit about their geometrical interpretation.

What it intuitively means to be contravariant or to be covariant?

Geometric interpretation of contravariant and covariant components of a vector

Let's consider a coordinate system, and draw its axes as straight lines because we are not interested at the moment in the fact that the coordinates may be curved and vary in direction from place to place.

We could also think of them locally, where every variety is approximately flat (a surface, locally, is like a plane) and every coordinate system locally is formed of approximately straight lines or surfaces if we are in more than two dimensions (see Figure 6).

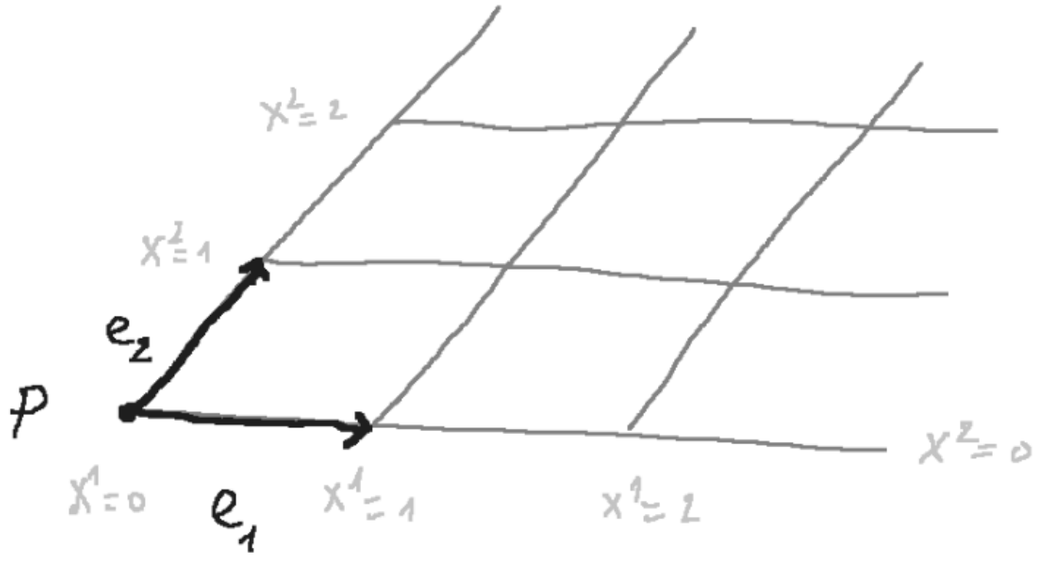


Figure 6: Coordinate system at point P.

We are mostly concerned with the fact that the coordinate axes may not be perpendicular, and with what the implications of the non perpendicularity of these coordinates are.

Furthermore the distance between two axes, say $X^1 = 0$ and $X^1 = 1$, is not necessarily 1.

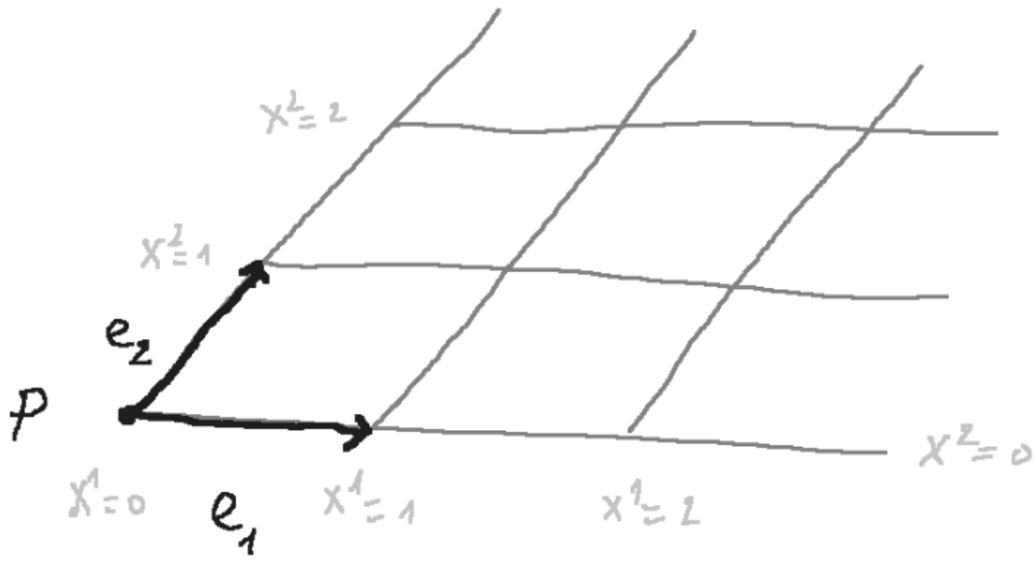
The values of the coordinates are just numerical labels, which don't correspond directly to distances.

Now let's introduce some vectors.

On our two-dimensional variety, we introduce two

$$e_1 \text{ and } e_2$$

as shown on figure 6.



If we had three dimensions, there would be a third vector e_3 sticking out of the page, possibly slanted.

We can label these vectors e_i

As i goes from 1 to the number of dimensions, the geometric vectors e_i's correspond to the various directions of the coordinate system.

Next step in the geometric explanation of contravariant and covariant vectors:

we consider an arbitrary vector V, see figure 7.

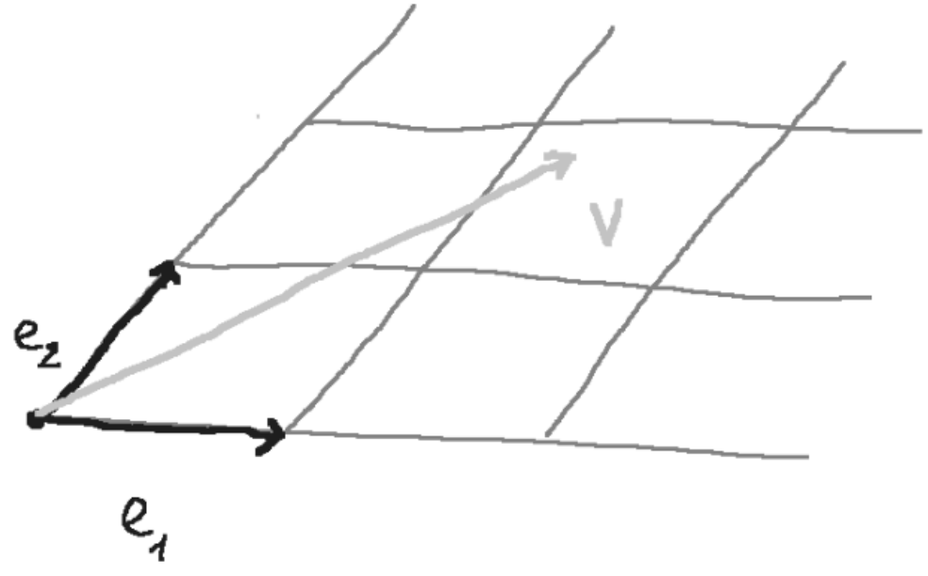


Figure 7: Vector V.

The vector V can be expanded into a linear combination of the e_i's.

We shall write V^i for the i-th coefficient, and suppose there are 3 dimensions.

Then

$$V = V^1 e_1 + V^2 e_2 + V^3 e_3 \tag{4}$$

The things which are the vectors, on the right hand side of this formula, are the e_i's.

The V^i's are actually numbers.

They are the components of the vector V in the e_i basis.

The coefficients V^i are called the contravariant components of the vector V.

It is just a name.

And there is nothing in what I did that required me to put the index 1 of e_1 downstairs and not upstairs, and the index 1 of V^1 upstairs.

It is a convention to write the expansion of V in the form of equation (4).

So, first of all, we see what the contravariant components are.

They are the expansion coefficients of V, that is, the numbers that we have to put in front of the three vectors e_1, e_2 and e_3 to express a given vector as a sum of vectors colinear to the basis.

This jives with what we have said previously: ordinary vectors are contravariant vectors.

Next step:

We look at the projection of V on the e_{i}'s using the dot product.

Let's start with e_1

$$V \cdot e_1$$

Now if we were just using conventional Cartesian coordinates, perpendicular to each other, and if the e_{i}'s really were unit vectors, that is, if the distance representing each coordinate separation was one unit of whatever the units we are dealing with, then the coefficients V^1, V^2 and V^3 would be the same as the dot products.

For instance we would have $V \cdot e_1$ equal the first contravariant component of V.

However, when we have a peculiar coordinate system with angles and with non-unit separations between the successive coordinate lines in figure 7, this is not true.

So let's see if we can work out *V* with the values $V \cdot e_1, V \cdot e_2, V \cdot e_3 \ \ldots$

Incidentally, $V \cdot e_1$ **is called** V_1, **denoted with a covariant index**.

Notice how notations fit nicely together,

We can write equation (4) as

$$V = V^m e_m \tag{5}$$

using the Einstein summation convention.

Now let's see how we can relate the contravariant components V^{m} and the covariant components V_{n}.

To reach that goal we take the dot product of each side of equation (5) with e_{n}.

We get

$$V \cdot e_n = V^m e_m \cdot e_n \tag{6}$$

And $V \cdot e_n$ is by definition V_n.

$e_m \cdot e_n$ is something new.

Let's isolate it.

It has two lower indices.

We will see that it turns out to be the metric tensor (expressed in the e_i's basis).

Let's see this connection between $e_m \cdot e_n$ and the metric tensor.

The length of a vector is the dot product of the vector with itself.

Let's calculate the length of V.

Using equation (5) twice we have

$$V \cdot V = V^m e_m \cdot V^n e_n \tag{7}$$

We must use two different indices m and n.

Recall indeed that, in the implicit summation formula $V^{\mathrm{m}}e_{\mathrm{m}}$, the symbol m is only a dummy index.

So in order not to mix things up, we use another dummy index n for the second expression of V.

If you are not yet totally at ease with Einstein summation convention, remember that, written explicitly, the right hand side of equation (7) means nothing more than

$$(V^1e_1 + V^2e_2 + V^3e_3).(V^1e_1 + V^2e_2 + V^3e_3)$$

But now the right hand side of equation (7) can also be reorganized as

$$V \cdot V = V^m V^n e_m \cdot e_n \tag{8}$$

The quantity $e_{\mathrm{m}}.e_{\mathrm{n}}$ we call g_{mn}.

So equation (8) rewrites

$$V \cdot V = V^m V^n g_{mn} \tag{9}$$

This is *characteristic* of the metric tensor.

It *tells* you how to compute the length of a vector.

The vector could be for instance a small displacement *dX*.

Then equation (9) would be the computation of the length of a little interval between two neighboring points

$$dX \cdot dX = dX^m dX^n g_{mn} \tag{10}$$

So now we have a *better understanding* of the difference between covariant and contravariant indices, that is to say covariant and contravariant components.

Contravariant components *are* the coefficients we use to construct a vector *V* out of the basis vectors.

Covariant components *are* the dot products of *V* with the basis vectors.

They are *different* geometric things.

They would, however, be the same if we were talking about ordinary Cartesian coordinates.

We inserted that discussion in order to give you some geometric idea of what covariant and contravariant means and also what the metric tensor is.

For a given collection of basis vectors e_i's and a given vector V, let's summarize all this in the following box

$$V = V^m e_m$$
$$V_n = V.e_n \tag{11}$$
$$g_{mn} = e_m.e_n$$

These relations are very important, and we will make frequent use of them in the construction of the theory of general relativity.

Let's just make one more note about the case when the coordinates axes are Cartesian coordinates.

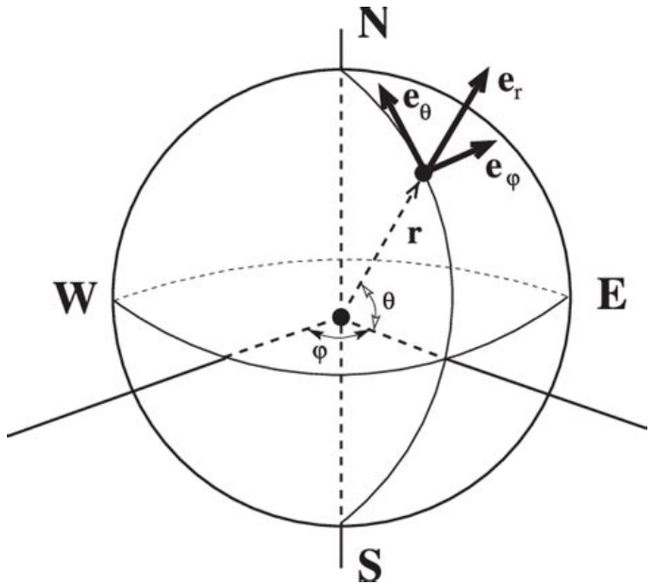
Then, as we saw, the contravariant and the covariant components of V are the same.

And the metric tensor is the unit matrix.

This means that the basis vectors are perpendicular and of unit length.

Indeed, they could be orthogonal without being of unit length.

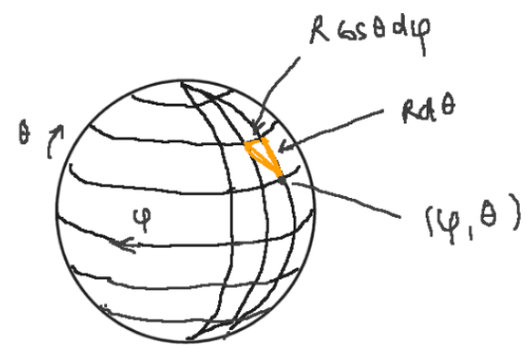
In polar coordinates (also see figure 8 below), the basis vectors at any point P on the sphere are orthogonal, but they are not all of unit length.



The longitudinal basis vector has a length which depends on the latitude.

It is equal to the cosine of the latitude.

That is why, on the sphere of radius one, to compute the square of the length of an element dS we can use Pythagoras theorem, but we must add $d\theta^2$ and $cos^2\theta\ d\phi^2$.



Also note that nothing enjoins us to represent the sphere in perspective, embedded in the usual 3D Euclidean space, like we did in figure 14 of lesson 1.

We can also represent it – or part of it – on a page.

Let's do it for a section of the Earth around one of its poles.

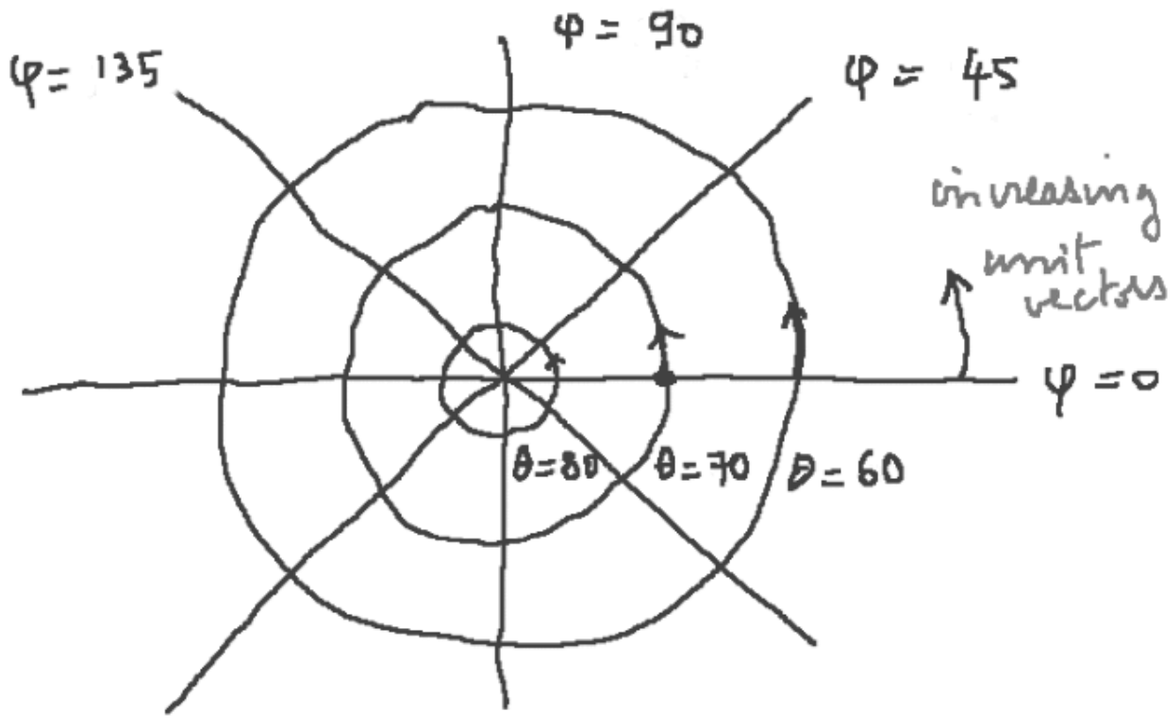


Figure 8: Map of the Earth around the North pole.

This is a representation on a page – therefore, out of necessity, flat – of a non-flat Riemannian surface with curvilinear coordinates, in this case a section of sphere in polar coordinates.

We touch here on the classical problem of cartographers: how to represent a section of sphere on a page, that is, how to make useful maps for mariners.

This is a well-known problem of cartographers, which lead 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).

This ought to clarify the fact that we can represent on a page a curved, truly non-flat, variety, and a curvilinear coordinate system on it.

This is also what is achieved by ordnance survey maps, which can show hills and valleys, slopes, distances on inclined land, gradients and things like that, see figure 9.

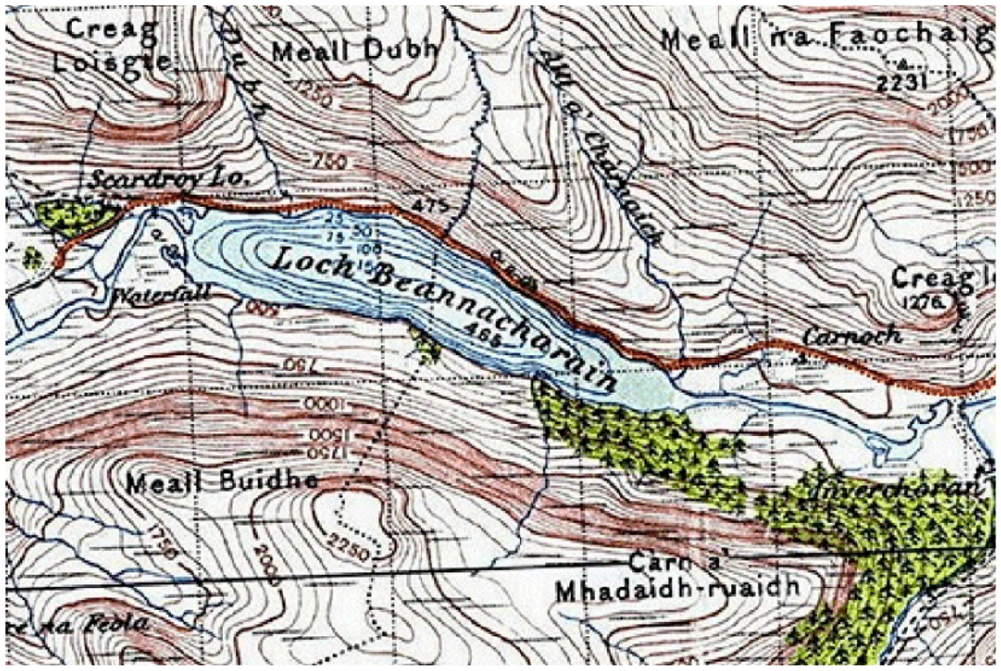


Figure 9: Ordnance survey map.

The curvy lines shown are the lines of equal height with respect to an underlying flat plane, which is a locally flat small section of the sphere on which we represent the mountainous relief.

The grid of straight lines is a coordinate system on the sphere.

We will assume that you can deal with the notions of curved surfaces, and distances on them.

So now let's come to tensor mathematics.

Tensor mathematics

Tensors are objects which are characterized by the way they transform under coordinate transformations.

We already talked a little bit about them at the end of lesson 1.

Now we want to go over again what we said and go further.

Notice that to say that tensors are characterized by the way they transform is no more strange than to say in $\mathbb{R}^3$ that (a, b, c) is a vector, or if you prefer "can be thought of as a vector", if and only if this collection of 3 numbers depends on a basis, is the expression of a thing in that basis, and it transforms in the usual way when we change basis.

Let's go over this in more detail.

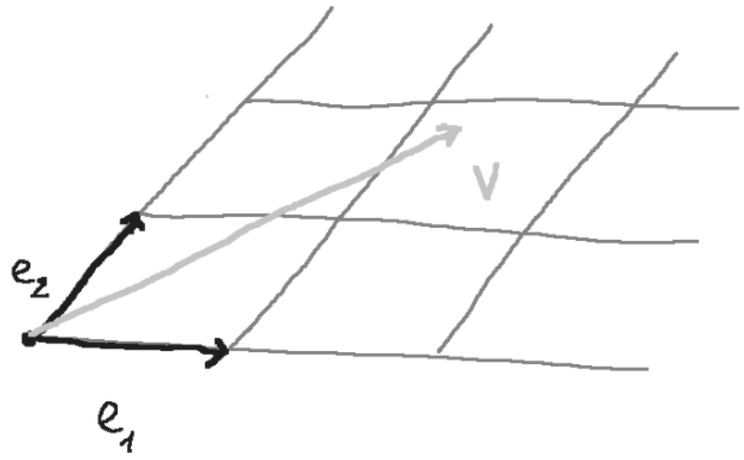
So let's start with a vector V.

It has contravariant components in the X coordinates.

We called them V^m.

And it has contravariant components in the Y coordinates, which we called $(V')^m$.

In figure 7, if we change the coordinate system, keeping the abstract geometric vector *V* fixed in the space, we are still talking about the same vector, but we will clearly change its components.



How do the contravariant components change when we change coordinates?

We have seen the rule.

Let us repeat it here.

Remember, in the formula below, prime means "in the *Y* system" and unprimed means "in the *X* system".

$$(V')^m = \frac{\partial Y^m}{\partial X^n} V^n \tag{12}$$

And now let's look at a covariant vector.

For the most typical example, we start from a scalar field *S*(*X*) which we differentiate with respect to the *X* components, and the *Y* components.

We have seen this rule as well.

The partial derivatives, which are covariant components, are related as follows

$$\frac{\partial S}{\partial Y^m} = \frac{\partial X^n}{\partial Y^m}\frac{\partial S}{\partial X^n} \tag{13}$$

Notice the difference.

And notice how the notation carries you along.

In equation (12)

- the index *m* is upstairs

- on the right hand side the proportionality factor is $\partial X^n/\partial Y^m$

- the sum is still on *n*.

If there is no index *n* on the left hand side, but an index *n* appears on the right, then an index *n* upstairs has to be balanced by an index *n* downstairs.

And we can "contract" them.

This means that they represent a sum, are only *dummy* indices, and disappear.

In both equations you can see the pattern.

And as said, the notation pretty much carries you along.

Equation (12) is the standard form for the transformation property of contravariant components.

And equation (13) is the standard form for the transformation property of covariant components, if they come from differentiating a scalar.

More generally it would be equation (14) below

$$(W')_m = \frac{\partial X^n}{\partial Y^m} W_n \tag{14}$$

Let’s go now to tensors of higher rank.

A tensor of higher rank *simply* means a tensor with more indices.

For the sake of pedagogy and completeness in this lesson 2, we overlap a bit what we did at the end of the last lesson.

We start with a tensor of rank two, with one contravariant index and one covariant index.

It is nothing more than a "thing" represented in a given basis by a collection of numbers.

These numbers are indexed with two indices.

Furthermore in another basis the same "thing" is represented by another collection of numbers and the two collections are subject to specific transformation rules related to the relationship between the two bases.

Let’s consider the tensor in a Y basis, that is to say, a Y coordinate system.

We denote it

$$(W')^{m}_{n}$$

The simplest example of such a thing would be, as we saw, just the product of two vectors, one with a contravariant index, one with a covariant index.

By "product of the vectors" we mean the collection of all the products of components.

What makes the thing a tensor is its transformation property.

So let's write it

$$(W')^m_{\ n} = \frac{\partial Y^m}{\partial X^p}\ \frac{\partial X^q}{\partial Y^n}\ W^p_{\ q} \tag{15}$$

This tells us how a tensor of rank 2, with one contravariant and one covariant index, transforms.

For each index on the left hand side, there must be a $\partial Y/\partial X$ or a $\partial X/\partial Y$ on the right hand side.

And you simply track where the indices go.

Let's do another example of a tensor of rank 2 with two covariant indices

$$(W')_{mn}$$

How does it transform?

By now you should begin to be able to write it mechanically

$$(W')_{mn} = \frac{\partial X^p}{\partial Y^m} \frac{\partial X^q}{\partial Y^n} W_{pq} \tag{16}$$

These rules are very general.

If you take a tensor with any number of indices, the pattern is always the same.

To express the transformation rules from an unprimed system X to a prime system Y, you introduce partial derivatives, in one sense or the other as we did, on the right hand side, and you sum over repeated indices.

We now know the basic notational device to express a tensor of any rank and type in one coordinate system or another.

Who invented it?

Einstein was the one who dropped the summation symbol, because he realized he didn't need it.

Gauss began to use formulas akin to equations (12) and (13) in his study of surfaces.

Riemann continued in the development of non-Euclidean geometry.

Ricci-Curbastro and Levi-Civita gave a formal presentation of tensor analysis in their fundamental work written in March 1900.

The notation is the work of many, but it is very systematic.

Notice something about tensors.

If they are zero in one frame, there are necessarily zero in any other too.

This is obvious for scalars: if a scalar is 0 in one frame, it is 0 in every frame, because its value depend only on the geometric point where it is measured, not the coordinates of that point.

Now suppose a vector V is zero in some frame – let's say the X frame.

To say that V is zero doesn't mean that some component is equal to zero, it means all of its components are equal zero.

Then equation (12) or equation (14) show that it is going to be zero in any frame.

Likewise with any tensor, if all of its components are 0 in one frame, that is, in one coordinate system, then all of its components are 0 in every frame.

As a consequence, once we have written down an equation equating two tensors in one frame, for instance

$$T^{lmn}_{pqr} = U^{lmn}_{pqr}$$

it can be rewritten

$$T^{lmn}_{pqr} - U^{lmn}_{pqr} = 0$$

So, considering that $T - U$ is still a tensor (see later for discussion on tensor algebra), we see that

if two tensors are equal in one frame, they are equal in any frame.

That is the *basic* value of tensors.

They allow you to *express* equations of various kinds, equations of motion, equations of whatever it happens to be, in a *form* where the same exact equation will be true in any coordinate system.

That is of course a deep advantage to thinking about tensors.

There are *other* objects which are not tensors.

They will have the property that they may be zero in some frames and not zero in other frames.

We are going to come across some of them.

Tensors have a certain invariance to them.

Their components are not invariant.

They change from one frame to another.

But the statement that a tensor is equal to another tensor is frame independent.

Incidentally, when you write a tensor equation, the components have to match.

It doesn't make sense to write an equation like $W_q^{\ p}$ (where p is contravariant and q covariant) equals T^{pq} (where both indices are contravariant).

Of course you can write whatever you like, but if, let's say in one coordinate system, the equation $W_q^{\ p} = T^{pq}$ happened to be true, then it would usually not be true in another.

So normally we wouldn't write equations like that.

When thinking of two vectors, if we can write $V = W$, then they are equal in all coordinates systems. Note that in Euclidean geometry, or in non-Euclidean geometry with a positive definite distance, for $V = W$ to be true it is necessary and sufficient that the magnitude of $V - W$ be equal to zero.

But this statement is not true in the Minkowski geometry of relativity, where the proper distance between two events may be zero without them being the same event.

In other words, notice that the magnitude of a vector and the vector itself are two different things.

The magnitude of a vector is a scalar, whereas the vector is a complex object.

It has components.

It points in a direction.

To say that two vectors are equal means that their magnitudes are the same and their directions are the same.

A tensor of higher rank is yet a more complicated object which points in several directions.

It has got some aspect of it that points in one direction and some aspects that point in other directions.

We are going to come to their geometry soon.

But for the moment we define them by their transformation properties.

The next topic in tensor analysis is operations on tensors.

Tensor algebra

What can we do with tensors that make new tensors?

We are not at this point interested in things that we can do to tensors which make other kinds of objects which are not tensors.

We are interested in the operations we can do with tensors which will produce new tensors.

In that way we can make a collection of things out of which we can *build* equations.

And the equations will be the *same* in every reference frame.

First of all you can multiply a tensor by a numerical number.

It is still a tensor.

That rule is obvious and we don’t need to spend time on it.

Then, we shall examine *four* operations.

Most of them are very simple.

The last one is not simple.

(1) **Addition of tensors.**

We can add two tensors of the same type, that is, of the same rank and the same numbers of contravariant and covariant indices.

And addition of course includes also subtraction.

If you multiply a tensor by a negative number and then add it, you are doing a subtraction.

(2) **Multiplication of tensors.**

We can multiply any pair of tensors to make another tensor.

(3) **Contraction of a tensor.**

From certain tensors we can produce tensors of lower rank.

(4) **Differentiation of a tensor.**

But this will not be ordinary differentiation.

It will be covariant differentiation.

We will define it and see how it works.

Those are the four basic processes that you can apply to tensors to make new tensors.

The first three are straightforward.

As said, the last one is more intricate: differentiation with respect to what?

Well, differentiation with respect to position.

These tensors are things which might vary from place to place.

They have a value at each point of the surface under consideration.

They are tensor *fields*.

At the next point on the surface they have a different value.

Learning to differentiate them is going to be fun and hard.

Not very hard, a little hard.

Furthermore it belongs, strictly speaking, to tensor analysis and will be taken up in the Lesson 3.

Adding tensors: you only add tensors if their indices match and are of the same kind.

For example if you have a tensor

$$T = T^{m\ldots}_{\ldots p}$$

with a bunch of upstairs contravariant indices, and a collection of downstairs covariant indices, and you have another tensor **of the same kind**

$$S = S^{m\ldots}_{\ldots p}$$

in other words their indices match exactly, then you are permitted to add them and construct a new sensor which we can denote

$$T + S$$

It is constructed in the obvious way: each component of the sum

$$(T + S)^{m\ldots}_{\ldots p}$$

is just the sum of the corresponding components of T and S.

And it is obvious too to check that $T + S$ transforms as a tensor with the same rules as T and S.

The same is true of $T - S$.

It is a tensor.

This is the basis for saying that tensor equations are the same in every reference frame – because $T - S = 0$ is a tensor equation.

Next, multiplication of tensors: now, unlike addition, multiplication of tensors can be done with tensors of any rank and type.

The rank of a tensor is its number of indices.

And we know that the two types, for each index, are contravariant or covariant.

We can multiply $T^{l}{}_{mn}$ by $S^{p}{}_{q}$.

The tensor multiplication being not much more than the multiplication of components and of the number of indices, we will get a tensor of the form $P^{lp}{}_{mnq}$.

Let's see a simple example: the tensor multiplication, also called **tensor product**, of two vectors.

Suppose V^{m} is a vector with a contravariant index, and let's multiply it by a vector W_{n} with a covariant index.

This produces a tensor with one upstairs index *m* and one downstairs index *n*

$$V^m \, W_n = T^m_n \tag{17}$$

A tensor is a set of values indexed by zero (in the case of a scalar), one (in the case of a vector) or several indices.

This tensor *T* of equation (17) is a set of values – depending of the coordinate system in which we look at it – indexed by two indices *m* and *n*, respectively of contravariant and covariant type.

It is tensor of rank two, contravariant for one index and covariant for the other.

We could have done the multiplication with some other vector X^n and this would have produced some other tensor

$$V^m \, X^n = U^{mn} \tag{18}$$

We sometimes use the sign ⊗ to denote the tensor product.

So equations (17) and (18) are sometimes written

$$V^m \otimes W_n = T^m_n$$

$$V^m \otimes X^n = U^{mn}$$

And this applies to the product of any tensors.

The tensor product of two vectors is not their dot product.

We will see how the dot product of two vectors is related to tensor algebra in a moment.

With the tensor product we produce a tensor of higher rank, by just juxtaposing somehow all the components of the multiplicands.

How many components does $V^m \otimes X^n$ have?

Since we are going to work mostly with 4-vectors in space-time, let's take V and X to be both 4-vectors.

Each is a tensor of rank one with a contravariant index.

Their tensor product U is a tensor of rank 2.

It has 16 independent components, **each of them the simple multiplication of two numbers**

$$U^{11} = V^1 X^1,\ U^{12} = V^1 X^2,\ U^{13} = V^1 X^3,\ \ldots$$

$$\ldots\ U^{43} = V^4 X^3,\ U^{44} = V^4 X^4$$

It is not the dot product.

The dot product has only is *one* component, not sixteen.

It is a *number*.

Sometimes the tensor product is called the *outer* product.

But we shall continue to call it the tensor product of two tensors, and it makes another tensor.

Typically the tensor product of two tensors is a tensor of different rank than either one of the multiplicands.

The only way you can make a tensor of the same rank is for one of the factors to be a scalar.

A scalar is a tensor of rank zero.

You can always multiply a tensor by a scalar.

Take any scalar S multiply it by, say, V^{m}.

You get another tensor of rank one, i.e. another vector.

It is simply *V elongated* by the value of S.

But generally you get back a tensor of higher rank with more indices obviously.

We are in the course of learning tensor algebra and tensor analysis.

It is a bit dry.

Where will these tensors come in?

We will meet them in real life soon enough.

But so far this is just a notational device.

Out of the four operations mentioned above, we already have addition and multiplication.

Let's now turn to contraction.

Contraction is also an easy algebraic process.

But in order to prove that the contraction of a tensor leads to a tensor we need a tiny little theorem.

No mathematician would call it a theorem.

They would at most call it maybe a lemma.

Here is what the lemma says.

Consider the following quantity (we begin to use also a, b, c, etc. for indices because there just aren't enough letters in the m range or the p range for our needs.)

$$\frac{\partial X^b}{\partial Y^m}\frac{\partial Y^m}{\partial X^a} \tag{19}$$

Remember that the presence of m upstairs and downstairs means implicitly that there is a sum to be performed over m.

Expression (19) is the same as

$$\sum_m \frac{\partial X^b}{\partial Y^m} \frac{\partial Y^m}{\partial X^a} \tag{20}$$

What is the object in expression (19) or (20)?

Do you recognize what it is?

It is the change in X^b when we change Y^m a little bit, times the change in Y^m when you change X^a a little bit, summed over m.

That is, we change Y^1 a little bit, then we change Y^2 a little bit, etc.

What is expression (20) supposed to be?

Let's go over it in detail.

Instead of X^b, consider any function F.

Suppose F depends on (Y^1, Y^2 ..., Y^M), and each Y^m depends on X^a.

Then, from elementary calculus,

$$\frac{\partial F}{\partial Y^m}\frac{\partial Y^m}{\partial X^a}$$

is nothing more than the partial derivative of F with respect to X^a (partial because there can be other X^n's on which the Y^m's depend).

That is

$$\frac{\partial F}{\partial Y^m}\frac{\partial Y^m}{\partial X^a} = \frac{\partial F}{\partial X^a}$$

What if what F happens to be X^b?

Well, there is nothing special in the formulas.

We get

$$\frac{\partial X^b}{\partial Y^m}\frac{\partial Y^m}{\partial X^a} = \frac{\partial X^b}{\partial X^a}$$

But what is $\partial X^b/\partial X^a$?

It looks like a stupid thing to look at.

The X^n are independent variables, so the partial derivative of one with respect to another is either 1, if they are the same, that is if we are actually looking at $\partial X^a/\partial X^a$, or 0 otherwise.

So $\partial X^b/\partial X^a$ is the Kronecker delta symbol.

We shall denote it

$$\delta_a^b$$

Notice that we use an upper index and a lower index.

We shall find out that $\delta_a{}^b$ itself happens to also be a tensor.

That is a little weird because it is just a set of numbers.

But it is a tensor with one contravariant and one covariant index.

Now that we have the little lemma we need in order to understand index contraction, let's do an example.

And then define it more generally.

Let's take a tensor which is composed out of two vectors, one with a contravariant index and the other with a covariant index,

$$T^m_n = V^m \, W_n \tag{21}$$

Now what contraction means is: **take any upper index and any lower index and set them to be the same and sum over them**.

In other words take

$$V^m \, W_m \tag{22}$$

This means $V^1W_1 + V^2W_2 + V^3W_3 + ... + V^MW_M$, if M is the dimension of the space we are working with.

We have identified an upper index with a lower index.

We are not allowed to do this with two upper indices.

We are not allowed to do with two lower indices.

But we can take an upper index and a lower index.

And let's ask how expression (22) transforms.

Let's look at the transformation rule applied first to expression (21).

We already know that it is a tensor.

Here is how it transforms

$$(V^m\ W_n)' = \frac{\partial Y^m}{\partial X^a}\ \frac{\partial X^b}{\partial Y^n}\ (V^a\ W_b) \tag{23}$$

We write $(V^{\mathrm{m}}\ W_{\mathrm{n}})'$, but we could also write $(V^{\mathrm{m}})'\ (W_{\mathrm{n}})'$, because we know that they are the same.

Indeed that is what we mean when we say that the outer product of two vectors forms a tensor: we mean that we can take the collection of products of their components in any coordinate system.

Calculated in any two systems, $(V^{\mathrm{m}})'\,(W_{\mathrm{n}})'$ and $V^{\mathrm{m}}\,W_{\mathrm{n}}$ will be related by equation (23).

Equation (23) is the transformation property of the tensor $T^{\mathrm{m}}{}_{\mathrm{n}}$ which has one index upstairs and one index downstairs.

Now let $m = n$ and contract the indices.

Remember, contracting means identifying an upper and a lower index and sum over them.

So on the left hand side we get

$$(V^m\,W_m)'$$

How many indices does it have?

Zero.

The index m is summed over.

The quantity is a scalar.

It is by definition the expression of the scalar $V^{m}W_{m}$ in the prime coordinate system, which as we know doesn't change.

So the contraction of $V^{m}W_{n}$ did create another tensor, namely a scalar.

We can check what equation (23) says.

It should confirm that $(V^{m}W_{m})'$ is the same as $V^{m}W_{m}$.

Now our little lemma comes in handy.

$$(V^m\ W_n)' = \frac{\partial Y^m}{\partial X^a}\ \frac{\partial X^b}{\partial Y^n}\ (V^a\ W_b)$$

On the right hand side of (23), when we set $m = n$ and sum over m, the sum of the products of partial derivatives is $\delta_{a}^{\ b}$.

So the right hand side is $V^a W_a$.

But a or m are only dummy indices, therefore equation (23) says indeed that

$$(V^m\ W_m)' = V^m\ W_m$$

So by contracting two indices of a tensor we make another tensor, in this case a scalar.

It is easy to prove that if you take any tensor with a bunch of indices, any number of indices upstairs and downstairs,

$$T^{nmr}_{pqs} \tag{24}$$

and you contract a pair of them (one contravariant and one covariant), say r and q, you get

$$T^{nmr}_{prs} \tag{25}$$

where the expression implicitly means a sum of components over r, and this is a new tensor.

Notice that the tensor of expression (24) has six indices, whereas the tensor of expression (25) has only four.

And notice two more things:

(a) If we looked at $V^m W^n$, we would be dealing with a tensor which cannot be contracted.

The analog of equation (23) would involve

$$\frac{\partial Y^m}{\partial X^a} \frac{\partial Y^n}{\partial X^b}$$

This quantity doesn't become the Kronecker delta symbol when we set $m = n$ and sum over it.

And $(V^m)'(W^m)'$ would not be equal to $V^m W^m$.

(b) The dot product of two vectors V and W is the contraction of the tensor $V^m W_n$.

But in that case one vector must have a contravariant index, and the other a covariant index.

In other words, contraction is the generalization of the dot product (also called inner product) of two vectors.

We are going to deal with inner products as soon as we work again with the metric tensor.

More on the metric tensor

The metric tensor plays a big role in Riemannian geometry.

We showed its construction with the basis vectors e_{m}'s, see figure 7 and after.

In the set of equations (11), we wrote

$$g_{mn} = e_m \cdot e_n$$

But let's now define it on its own terms *abstractly*.

Again these are things we have already covered before, but let's do them again now we *have* a bit more practice with tensors.

The definition of the metric tensor goes like this.

Consider a differential element dX^{m} which just represents the components of a displacement vector dX.

In other words, we are at a point P on the Riemannian surface (or Riemannian space if we are in more that two dimensions), see figure 10,

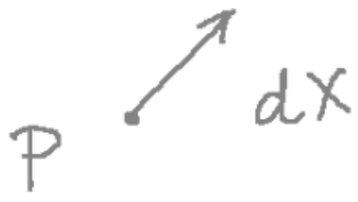


Figure 10: Displacement vector dX

and we consider an infinitesimal displacement which we call dX – even though we also attach X to a specific coordinate system.

We could call the small displacement dS but traditionally dS is a scalar representing a length.

The contravariant components of dX are the coefficients of the vector dX in the expansion given by equation (4), which we rewrite below specifically for dX, supposing furthermore to make notations simple that there are three dimensions and therefore three axes,

$$dX = dX^1 e_1 + dX^2 e_2 + dX^3 e_3 \tag{26}$$

Each dX^{m} is a contravariant component of the little displacement vector of figure 10.

Now we ask: what is the length of that displacement vector?

Well, we need to know more about the geometry of the surface (also called **variety**) to know what the length of the little vector is.

The surface or variety could be some arbitrarily shaped complicated space.

Specifying what the geometry of the variety is, in effect is specifying what the lengths of all the infinitesimal displacements are.

As said, we usually denote the length dS, and we usually work with its square.

When the variety locally is Euclidean, dS is defined with Pythagoras theorem, but when the axes locally are not orthogonal or the dX^{m} are not expressed in units of length, or both, then Pythagoras theorem takes a more complicated form.

It is still quadratic in the dX^{m}'s, but it may also involve products $\mathrm{d}X^{\mathrm{m}}dX^{\mathrm{n}}$ and there is a coefficient g_{mn} in front of each quadratic term.

The square of the length of any infinitesimal displacement is given by

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

In general the g_{mn} depend on where we are, that is, they depend on P, which we locate with its coordinates X's.

So we write more generally

$$dS^2 = g_{mn}(X)\ dX^m dX^n \tag{27}$$

We are going to stick with the case of four dimensions because we are in a class on relativity.

For the moment, however, we don't consider the Einstein-Minkowski distance whose square can be a negative number.

We are in a Riemannian geometry with four dimensions, where all distances are real and positive.

In that case how many independent components are there in the g_{mn} object?

Answer: to begin with there are 16, because g_{mn} is a 4 x 4 array.

But $dX^1 dX^2$ is exactly the same as $dX^2 dX^1$.

So there is no point in having a separate variable for g_{12} and g_{21}, because they can be made equal to each other.

So there are only 4 + 3 + 2 + 1 = 10 independent components in g_{mn} in four dimensions, see figure 11.

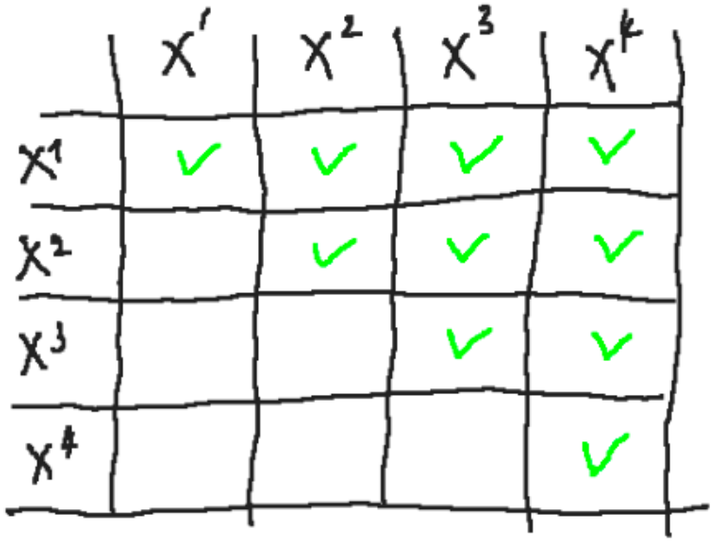


Figure 11: Independent components in g_{mn}.

Similarly in a three-dimensional space there would be 6 independent components in g_{mn}.

And in two dimensions it would be 3.

So far we haven't proved that g_{mn} is a tensor.

I called it the metric tensor, but let's now prove that it is indeed such an object.

The basic guiding principle is that the length of a vector is a scalar, and that everybody agrees on that length.

People using different coordinate systems won't agree on the components of the little vector *dX* (see figure 10), but they will agree on its length.

Let's write again the length of *dX*, or rather its square

$$dS^2 = g_{mn}(X)\ dX^m\ dX^n \tag{28}$$

And now let's go from the *X* coordinates to the *Y* coordinates.

Because dS^2 is invariant, the following holds

$$g_{mn}(X)\ dX^m\ dX^n = g'_{pq}(Y)\ dY^p\ dY^q \tag{29}$$

Now let's use this elementary calculus fact

$$dX^m = \frac{\partial X^m}{\partial Y^p}\ dY^p \tag{30}$$

And plug expression (30) for dX^m and for dX^n into equation (29).

We get

$$g_{mn}(X)\ \frac{\partial X^m}{\partial Y^p}\ \frac{\partial X^n}{\partial Y^q}\ dY^p\ dY^q = g'_{pq}(Y)\ dY^p\ dY^q \tag{31}$$

The two sides of equation (31) are expressions of the same quadratic form in the dY^p's.

That can only be true if the coefficients are the same.

Therefore we established the following transformation property

$$g'_{pq}(Y) = g_{mn}(X)\ \frac{\partial X^m}{\partial Y^p}\ \frac{\partial X^n}{\partial Y^q} \tag{32}$$

This is just *exactly* the transformation property of a tensor with two covariant indices.

So we discovered that the metric tensor is indeed *really* a tensor.

It *transforms* as a tensor.

This will have *many* applications.

The metric tensor has two lower indices because it multiplies the differential displacements dX^m's in equation (28) which have upper indices.

But the metric tensor is also just a matrix with *m n* indices.

Remembering that $g_{ij} = g_{ji}$, it is the following matrix, which we still denote g_{mn},

$$g_{mn} = \begin{pmatrix} g_{11} & g_{12} & g_{13} & g_{14} \\ g_{12} & g_{22} & g_{23} & g_{24} \\ g_{13} & g_{23} & g_{33} & g_{34} \\ g_{14} & g_{24} & g_{34} & g_{44} \end{pmatrix}$$

It is a *symmetric* matrix.

There is one more fact about this matrix, that is about the tensor g_{mn} thought of as a matrix.

It has *eigenvalues*.

And the eigenvalues are *never* zero.

The *reason* the eigenvalues are never zero is because a zero eigenvalue would correspond to a little eigenvector of zero length.

But there are *no* vectors is 0 length.

In Riemannian geometry *every* direction has a positive length associated with it.

What do we *know* about matrices which are symmetric and whose eigenvalues are all non-zero?

Answer: they *have* inverses.

The matrix of the metric tensor – both denoted g_{mn} or g for simplicity – has an inverse which in matrix algebra would be denoted g_{mn}^{-1} or simply g^{-1} and

$$g^{-1}\, g = \text{the unit matrix}$$

In tensor algebra, the inverse matrix is not denoted g_{mn}^{-1} nor g^{-1}.

It is denoted g^{mn}, with the indices upstairs.

g^{mn} is also a tensor.

Its defining property is that, as a matrix, it is the inverse of the initial matrix g_{mn} with two lower indices.

Let's write the corresponding equations.

It is the last thing we shall do in this lesson.

Let's do it slowly.

Consider two matrices A and B.

Let's say two square matrices for simplicity, one denoted with upper indices and the other with lower indices

$$A^{mn} \text{ and } B_{pq}$$

How do we multiply them?

That is, how do we compute the *m q* component of the product?

It is very simple.

If you remember matrix algebra, it is

$$(AB)^m_q = A^{mr} B_{rq}$$

So let's compute the product of the two matrices g_{mn} by g^{mn}.

By definition of g^{mn}, we have

$$g_{mn}\ g^{np} = \delta^p_m \tag{33}$$

where $\delta_m{}^p$ is the identity matrix.

Equation (33) is an equation in *matrix* algebra.

But it is also an equation in *tensor* algebra.

It is indeed elementary to show that g^{np} is also a tensor.

Its expression in a Y coordinate system is by definition $(g')^{np}$, such that

$$g'_{mn}\ (g')^{np} = \delta^p_m$$

Then there are various mathematical ways to arrive at the analog of equation (32) for the tensor g with upper indices.

As a tensor equation, equation (33) shows on its left hand side the contraction of the tensor $g_{mn} \otimes g^{qp}$.

And it says that the contraction of that product is the Kronecker delta object, which is necessarily also a tensor since it is the result of the contraction of a tensor.

g^{mn} **is called the metric tensor with two contravariant indices**.

The fact that there is a metric tensor with downstairs indices and a metric tensor with upstairs indices will play an *important* role.

So far everything we have seen on tensors was easy.

It is essentially learning and getting accustomed with the notation.

Next step: differentiation of tensors

In the next lesson we will go on to the subject of curvature, parallel transport, differentiation of tensors, etc.

The idea of a covariant derivative will be a little more complicated than tensor algebra.

Not much.

But it is essential.

We have to know how to differentiate things in a space, if we are going to do anything useful.

In particular, if we are going to study whether the space is flat, we have to know how things vary from point to point.

The question of whether a space is flat or not fundamentally has to do with derivatives of the metric tensor – the character and nature of the derivatives of the metric tensor.

So in the next lesson we will talk a little bit about tensor calculus or tensor analysis, differentiation of tensors, and especially the notion of curvature.

Lesson 3 : Flatness and curvature

Introduction

General relativity has a reputation for being very difficult.

I think the reason is that it is very difficult.

It is calculation-intensive: symbols, indices, awe-inspiring equations.

There are ways people have invented to express things in more condensed very neat notations.

But just learning them in itself is a task that that would take some time.

Things like forms, spinors and twistors and all sorts of other mathematical objects.

You could call many of them just notational devices if you like.

And they do simplify the equations.

So I sometimes feel that in presenting these things the way I do, it is sort of like Maxwell who wrote down every single equation of his set of Maxwell equations.

At first he wrote twenty altogether.

Although now we write only four.

We don't usually write all components of the equations.

We put them together into vector notation and so forth.

And if we are smart, we can even avoid the indices by inventing symbols like del, curl, laplacian, etc.

The same thing could be done to some extent for general relativity.

But as I said the computational techniques are harder.

I am by no means an expert in general relativity.

But when I have to do something involving general relativity, I go back to the indices.

I can follow if somebody else uses the fancy methods, but I can't get around to doing them myself.

It is not that I don't know how to use them.

But they are not natural to me.

I wind up going back to the same things we are doing here : indices, derivatives, the whole works.

My feeling is that it is the right way to see it for the first time.

But it is a nuisance.

It has indices all over the place, notations all over the place, too much of it.

We can't help it.

So some aspect of it you have to get over.

You have to get over that hump of tensor notation, bunches of indices upstairs and downstairs.

And then it goes easier, more smoothly... until you come to do a calculation.

Then you are right back into the mess again.

Now if you are really interested in doing the computations in general relativity, there are packages.

You just put in the metric as a function of position, the metric tensor.

And out it will spit various tensors, Riemann tensors, this kind of tensor, that kind of tensor, Einstein tensors.

And then you can say, without even looking at the results: okay, please Mr. Computer, set the Einstein tensor equal to the energy momentum tensor and tell me what comes out.

So, yes, computers can do a lot better than us.

As I said, general relativity is hard.

And part of the difficulty is just the complexity of the notations that are necessary.

By the way, the subject that we have been studying up to now is Riemannian geometry.

It is the geometry of *smooth* curved metric spaces.

They can have highly non-flat shapes, but locally they are like planes or higher dimension flat spaces, and their local geometry is like Euclidean geometry.

I suppose we have Riemann to *blame* for the complexity of the notation.

Although, actually he simplified it.

Einstein simplified it further.

But nevertheless it is notationally and computationally *intensive*.

The basic concepts are really not so hard.

But we have to go through the mess before we get to the concepts or before we can express those concepts.

General relativity in modern physics

One of the most active areas of research among theoretical physicists, at the beginning of the XXIst century, is the quantization of general relativity, the quantization of black hole physics.

The other most active area is quantum field theory.

There are also astrophysical areas where black holes are very important.

Astrophysical areas are entirely *classical* general relativity.

There is *numerical* relativity.

It means actual computations with numbers.

This involves special mathematical techniques employed with computers.

It is usually used for the purpose of solving astrophysical problems : putting in numerically various initial situations, for instance a pair of black holes orbiting around each other, and simply running the computer to find out what happens, by solving Einstein's equations.

It is a very subtle business by the way.

You don’t just put in the equations and leave the computer to grind.

It is much more subtle than that.

Numerical relativity is also a very hard subject.

The purpose for instance can be to find out how much gravitational radiations are emitted when two black holes fall into each other.

These are entirely classical radiations.

Bang!

A lot of gravitational radiation is produced, gravitational waves.

Why are we interested in them?

Among other reasons, because with the latest detectors we have a chance of detecting them.

Anything you can detect is a window onto a new natural phenomenon.

So gravitational wave detectors will be able to detect gravitational waves from the collision of two black holes.

It is estimated that there are every year a few collisions of black holes that can be seen by their gravitational radiation.

So there is a lot of different active areas nowadays related to general relativity – classical general relativity, that is involving no quantum uncertainty.

The equations of general relativity are still full of things to teach us.

Furthermore the equations of black holes have been found to be extremely useful in condensed matter physics and fluid dynamics, in other words in domains having nothing to do with black holes and gravity.

Different domains, same equations.

In short, there is a heavy amount of activity devoted to applying the equations of general relativity to other areas of physics.

It is probably one of the most active areas of research.

This preeminence of general relativity has not always been the case.

Over my career in physics it went from being real backwaters in physics to the forefront.

Nobody studied general relativity very much in the 1960s when I was a student, except people who were really interested in the subject per se.

They were then focused specifically on questions of general relativity.

But the bulk of theoretical physicists were not interested.

Then it started to change at the end of the XXth century.

It became interesting in astrophysics and eventually took over questions of fundamental theoretical physics.

Why?

Because of a clash of gravity in general relativity.

The equations lead to the hypothesis of black matter just like problems related to light lead to special and general relativity.

Riemannian geometry

This is the last lesson in which we will be studying Riemannian geometry as such, without really discussing gravity.

In the next lesson we will really get into gravity.

What do all these manipulations of tensors have to do with gravity?

We saw it a little bit in the first lesson.

The problem of finding out whether there is a real gravitational field, as opposed to just some funny coordinates with which we are locating events in space-time, leading to curved lines but not intrinsically curved trajectories (see figure 4 of lesson 1), is mathematically identical to the problem of finding out if a certain geometry – characterized by its metric tensor – is flat or not.

Let's think of a two dimensional space or variety to begin with.

That means a surface S of some sort, where each point is located with two real coordinates X^1 and X^2.

See figure 1, where it is shown embedded in the usual 3D space for convenience :

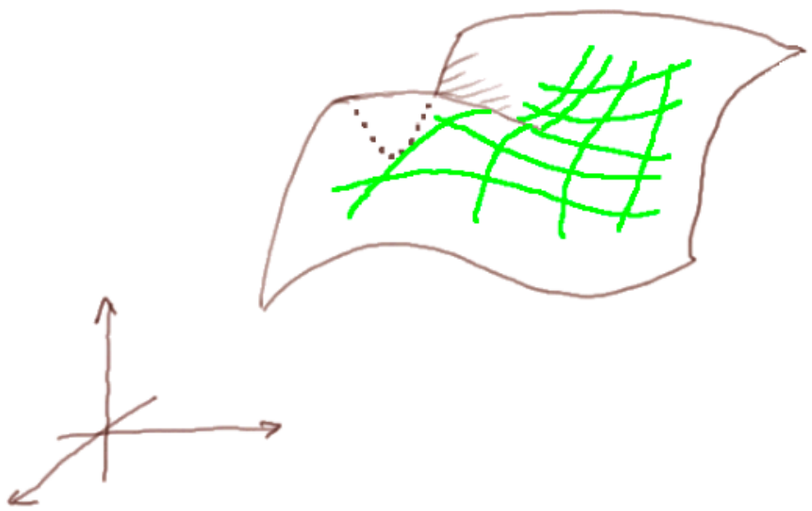


Figure 1 : Two dimensional variety with a system of coordinates on it.

For the system of coordinates, mathematicians technically talk of a mapping from $\mathbb{R}^2$ into $\mathbb{S}$, with nice properties of smoothness and differentiability.

And $\mathbb{S}$ has a metric, defined at any point P with coordinates X on it, by the now familiar equation

$$dS^2 = g_{mn}(X)\ dX^m\ dX^n \tag{1}$$

where m and n run over 1 and 2.

Notice that this metric defines the distance between two infinitesimally close points, not between two points far away.

A flat geometry is one for which Euclid's postulates are correct.

There are parallel lines, all the stuff we learned in high school Euclidean geometry.

And the surface $\mathbb{S}$ can be laid out on a plane without applying any distortion, stretching or compression, on it.

In the case of two dimensions, such a surface is also called a *developable* surface.

Any geometry is characterized by its metric tensor.

A flat one is not necessarily one whose metric, given by equation (1), has the special form of the Kronecker delta tensor.

That is not the right statement.

A flat geometry is one where we can find another system of coordinates Y such that at any point P, now located with those Y coordinates, the metric has a simple form

$$dS^2 = (dY^1)^2 + (dY^2)^2 \tag{2}$$

Notice that to find such a transformation is always possible locally at any given point P, because locally any surface is like a plane.

But it is not always possible to find such a transformation globally over the whole surface.

In other words, more generally, given an arbitrary metric tensor which varies from place to place – assuming it is smooth, that it has all the good differential properties and all – the problem of determining whether the corresponding space is flat is a somewhat difficult problem.

The bad way to approach the problem is to search through all possible coordinate systems, that is for each possible coordinates Y to transform the initial metric tensor and see whether its new form is the Kronecker tensor.

Of course, this will take an infinite amount of time.

We will never get there.

So we need a better technique.

The better technique is to search for a diagnostic quantity, that is a quantity which is built out of the metric and its derivatives, that we can calculate.

It will be a quantity computed at every point, so in reality it will be a scalar field.

If it is zero everywhere throughout the space that will say the space is flat.

If it is non-zero at some place, it will say that the space is curved there.

The diagnostic quantity that does the job is called the **curvature tensor**.

The goal of this lesson is to get to the curvature tensor of a Riemannian geometry.

It is a complicated goal – I warn you about this.

You will have to pay attention or else you will fall asleep.

It can be a very soporific subject, so we will try from time to time to tell a joke to keep you awake.

But on the other hand, if you can follow it, you will learn a lot.

And it is very interesting.

I think you will find it very interesting.

We will try not to make it too boring.

What do we start with ?

We start with a space.

And a space means first of all a number of dimensions.

In Riemannian geometry the number of dimensions can be any integer.

In principle you can even have a zero-dimensional space.

But that is just a point!

There isn't much to say about the geometry of a point.

So let's go to the next number of dimensions.

A one dimensional space is either an infinite line or a closed curve – a loop.

If it is a closed curve what is it characterized by?

It is characterized by only one thing : the total length of the curve.

Every loop is equivalent to every other loop of the same length.

In other words – just think about it for a moment – take a piece of rope that closes on itself to form a loop and has a certain length.

Wiggle it or curve it in any kind of ways, it can always be mapped or put on top of another piece of rope of exactly the same length.

It can be done in a precise way without stretching either of them.

A bug crawling successively around the two loops of the same length would not notice any difference.

So one dimensional spaces are all equivalent to each other, up to their length.

There are only two general kinds : either they are infinite, or they form a loop of a certain length L.

And every loop of length L is identical to any other loop of the same length.

And it is an *intrinsic* property.

The bug walking along the closed rope cannot distinguish the shape we give it in 3D.

All it can do is count the number of steps it takes to walk around the rope.

For instance, the bug makes an initial mark someplace, goes around the rope till it comes back to the mark, and records the number of steps it took it to do it.

That is the *only* thing the bug can say, or measure, about the closed rope.

Two dimensional spaces are where things start to be more complicated – and more interesting.

There are flat ones.

There are curved ones.

A flat one is a plane.

A curved one could be a sphere.

It could a space with lumps, the surface of the Earth including the mountains and the valleys.

It can even have a weird topology, for instance the surface of a donut – also called a torus.

You can poke another hole in the torus and make a torus with two holes, and so forth.

So by the time you get the two dimensional spaces they are pretty complicated.

Nevertheless they are characterized by the metric tensor.

We are concerned with whether they are flat.

The first question then is : find something which distinguishes whether they are flat.

If g_{mn} is not equal to δ_{mn}, that doesn't mean the space is not flat.

It just means you may be in the wrong coordinates.

You can't diagnose the space by asking whether g_{mn} is δ_{mn}.

We want to find a quantity called R – not necessarily a scalar – such that if it is zero everywhere the space is flat, and if it is not zero someplace the space is not flat there.

It would be best if it is a tensor, because if it is a tensor and it is non-zero, it is non-zero in every frame of reference.

R stands for Riemann.

He was the first person who asked this question and then answered it.

So we want to find a tensor R such that if R is equal to 0, that implies that the space is flat and that there exist coordinates such that

$$g_{mn} = \delta_{mn} \tag{3}$$

That is what we are going to search for : the **curvature tensor**.

Recall a few properties of the metric tensor and what you do with it.

First of all g_{mn} is by definition a function of position in general : $g_{\mathrm{mn}}(X)$.

If it were not a function of position, if it was a constant matrix, that is a constant set of coefficients, then it would be very easy to find coordinates in which g_{mn} is δ_{mn}.

You simply have to straighten out the angles of the coordinate axes and adjust the units so that they correspond to length.

So if g_{mn} is not dependent on position, then the space is definitely flat.

So we start with $g_{\mathrm{mn}}(X)$, where X is the set of coordinates of a point P.

What do we do with it?

We compute

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

That gives you the square of the length of the little differential displacement dX (see figure 2).

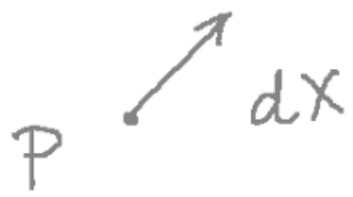


Figure 2 : Little displacement dX.

We call the displacement dX for simplicity, but we could call it dP.

It is indeed a small vector which will have different components in different coordinate systems.

In a Y system, we will call its components dY.

So the square of the length of the little displacement is

$$dS^2 = g_{mn}(X)\ dX^m\ dX^n \tag{4}$$

The metric tensor is a tensor.

We proved that in the last lesson.

We used the fact that dS^2 is an invariant quantity.

All observers, that is all coordinate systems, should agree about the length of a given little vector as shown in figure 2.

You express the vector in Y coordinates instead of X coordinates.

The length has to be the same and that leads immediately to the proof that g_{mn} really is a tensor, that is, transforms like a tensor.

Next property of g_{mn} is that it can be viewed as a matrix and it has an inverse.

For instance in four dimensions, the covariant metric tensor g_{mn} is also the symmetric matrix

$$g_{mn} = \begin{pmatrix} g_{11} & g_{12} & g_{13} & g_{14} \\ g_{12} & g_{22} & g_{23} & g_{24} \\ g_{13} & g_{23} & g_{33} & g_{34} \\ g_{14} & g_{24} & g_{34} & g_{44} \end{pmatrix}$$

The inverse of g_{mn} is denoted g^{mn}.

It is a matrix, and it is also a contravariant tensor.

It satisfies the following equation

$$g^{mn}\ g_{nr} = \delta^{m}_{r} \tag{5}$$

A remark on notations : equation (5) is a matrix equation as well as a tensor equation.

As matrix equation it says that g^{mn} times g_{mn} yields the unit matrix.

As a tensor equation it says that the tensor product of g^{mn} with g_{nr}, when it is then contracted along the index n, gives the Kronecker delta tensor.

Up to now, we haven't paid much attention to the placing of the indices of the Kronecker delta symbol, see for instance equation (3) where we wrote both of them downstairs because we equated it to g_{mn}.

The point to note is that in matrix algebra, the Kronecker delta symbol is generally denoted δ_{mn}.

While in tensor algebra of general relativity or Riemannian geometry it is most of the time denoted $\delta^{m}{}_{n}$ with one upstairs index and one downstairs index (the downstairs index slightly shifted to the right if we like neat typography) – at least when the Kronecker delta symbol is treated as a tensor.

This is a thing to prove : if you take the Kronecker delta symbol and you transform it pretending it is a tensor of whichever type, you just get back the Kronecker delta symbol of the same type.

There is another important point to mention : in Riemannian geometry, as opposed to Einsteinian geometry, all lengths and their squares are positive.

This says something about the properties of the metric tensor : there are no directions along which if you evaluate dS^2 you get anything but a positive number.

That is equivalent to a statement about the eigenvalues of g_{mn}.

It says that all of the eigenvalues of the matrix are *positive*.

To be complete, the following statements are *equivalent* :

(1) dS^2, of equation (4), is always positive

(2) the quadratic form $g_{mn}(X)\, dX^m\, dX^n$ is positive definite

(3) the matrix g_{mn} has only positive eigenvalues

(4) the matrix g_{mn} is positive definite

That will *change* when we go to the theory of real gravity and have to use the Einsteinian geometry.

The metric tensor allows you to do something: it *allows* you to raise and lower indices.

Consider a vector with contravariant components V^m

You can put it in correspondence with another vector with covariant components.

It is really the same vector.

We talked about this a little bit in the last lesson.

We saw the difference between contravariant and covariant components.

Remember, suppose V^m and V_n denote the *same* geometric vector V.

The V^m's are the components of V in the basis of e_i's, see figure 7 of lesson 2.

The components V_n's are defined by equations (11) of lesson 2.

To construct the covariant components of a vector, given its contravariant components, we just multiply it by the covariant metric tensor

$$V_n = V^m \, g_{mn} \tag{6}$$

Likewise, it can be shown that the following holds

$$V^m = V_n \, g^{mn} \tag{7}$$

So if you know the upper indices, that is the contravariant components of a vector V, equation (6) is the way you build its covariant components.

And if you know the covariant components of *V*, equation (7) is the way you build its contravariant components.

The proof is not difficult.

Gaussian normal coordinates

To what *extent* can you force the metric to look like a flat metric?

Now we mean not only that the space is flat, but that we are in a coordinate system such that g_{mn} is *equal* to δ_{mn}.

If that is the case, the space is definitely flat – that is the *definition* of flatness.

But to what extent given any space can you find coordinates such that g_{mn} might be the Kronecker delta over some limited region.

Here is a *theorem* : at any given point P on the surface, we can find a system of coordinates which are *good* in the immediate vicinity of the point, see figure 3.

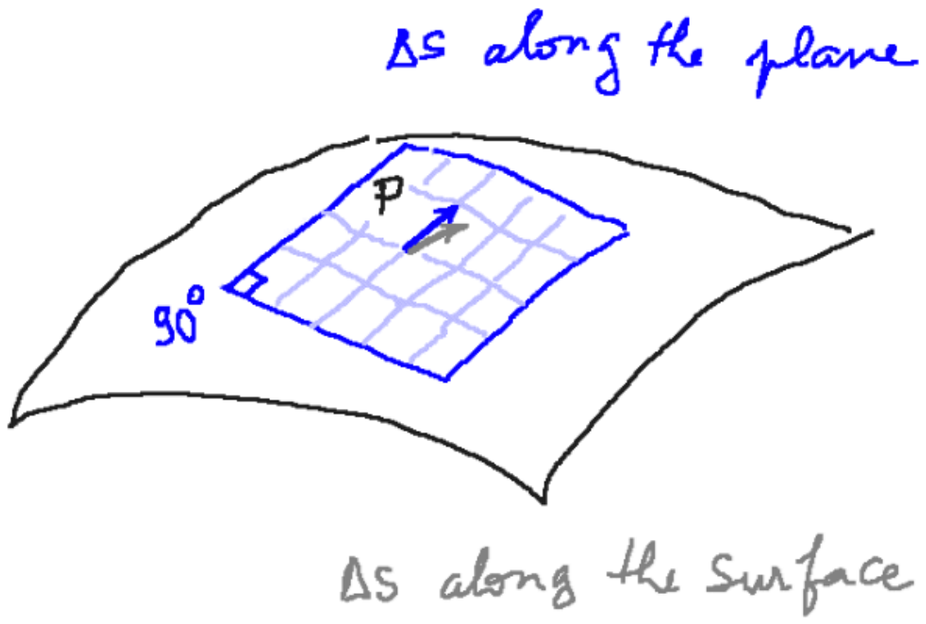


Figure 3 : Displacement of length ΔS along the surface and along the tangent plane in the same direction.

The coordinates are represented on the **tangent** plane.

We could also have represented them – slightly curved – on the surface itself.

They will *not* be globally flat.

Equation (4) will *not* reduce over the whole variety to Pythagoras theorem.

But in the immediate vicinity of point P, with a suitable choice of coordinates the metric tensor g_{mn} *can* be made to be approximately like the Kronecker delta.

Such coordinates are called **Gaussian normal coordinates**.

Here is the way we proceed.

We position ourselves at point P and we move along any first direction as straight as we possibly can.

Later we will learn what is meant by "as straight as we possibly can".

It will *mean* along a geodesic.

So you make as straight a curve as you can.

Say, you are a little bug.

You move on the surface following your nose going straight ahead.

That defines one coordinate axis.

Then you come *back* to point P.

You have some surveying tools to figure out which other directions make a right *angle* with the fist line.

On a two dimensional surface there is only one other direction (in one sense or the other).

In three dimensions, there is a whole plane.

And you go off in an orthogonal direction as straight as you can.

You build that way a complete set of coordinates based on those directions.

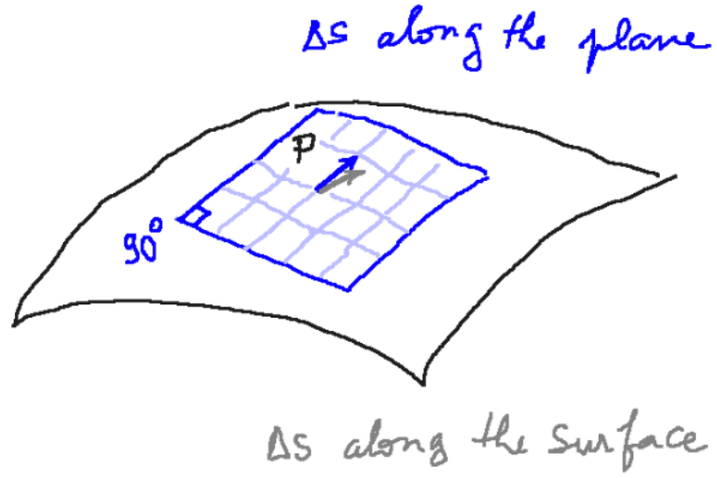
The theorem says that at every point P of the surface you can choose Gaussian normal coordinates such that at that point, whose coordinates are, say, X_0

$$g_{mn}(X_0) = \delta_{mn} \tag{8}$$

You can do that in *more* than one way.

If you found coordinates with which equation (8) is true, you can rotate the coordinates.

This will produce a different set of axes such that equation (8) in the new set is still true.



In figure 3, think of pivoting the coordinate system around P.

The theorem says furthermore that, at point *P*, once you have chosen the directions, you can also choose the *X*'s such that the **derivative** of any element of the metric tensor $g_{mn}(X)$ at that point with respect to any direction in space, X^r, be set equal to zero

$$\frac{\partial g_{mn}}{\partial X^r} = 0 \tag{9}$$

Equation (9) will be true, at a given point, *only* for the first derivatives.

Unless the space is flat, the derivatives of higher order at that point won't be zero

$$\frac{\partial^2 g_{mn}}{\partial X^r \partial X^s} \neq 0 \tag{10}$$

So, at a point, there is no content really in saying that the metric can be chosen to be, let's say, flat-like.

Up to the first derivatives included that can always be done.

It is in the second derivatives of the metric tensor that the flatness or non-flatness of the space somehow starts to show up.

How do we *prove* it?

Let's take up the point of interest that we called P , of coordinates X_0, to be the origin.

$$X_0 = 0$$

Now suppose we have some general metric and some coordinates Y in which the metric has some form, which does not satisfy equations (9).

Let's look for some X's which will be functions of the Y's, and choose them in the following way : at the place where $X = 0$, in other words at the origin, let's also *assume* that $Y = 0$.

So the two sets of coordinates have the *same* origin.

That means that X will *start* out just equal to Y plus something quadratic in Y

$$X^m = Y^m + C^m_{nr} Y^n Y^r \tag{11}$$

plus some more *complicated* terms.

We are simply *expanding* each X^m in powers of Y^1, Y^2,... Y^N, where N is the number of dimensions of the variety.

How *many* such $C^m{}_{nr}$ are there?

Suppose we work in *four* dimensions.

Then there 10 *distinct* combinations $Y^n Y^r$, because $Y^n Y^r = Y^r Y^n$.

And for each n and r we have four $C^m{}_{nr}$, when m runs from 1 to 4.

That means there are 40 *nr independent* coefficients.

Now how many independent components of g are there?

Answer: ten.

So there are 40 equations of type (9).

Finally we reached 40 equations to solve for 40 unknowns.

That allows us to be sure, at point P, that not only $g_{mn}(X) = \delta_{mn}$, but also that the derivatives of g_{mn} and δ_{mn} will match up to quadratic order.

That means that we will be able to solve the forty equations (9), but that we will *fail* to set the left hand sides of equations (10) to be equal to zero.

To summarize, at any point P, a smooth variety *is* locally flat.

We can approximate it by its tangent space, see figure 3.

And we can construct coordinates X's such that P is located at the origin, and the metric tensor has the form

$$g_{mn}(X) = \delta_{mn} + o(X) \tag{12}$$

where $o(X)$ represent terms of second order and higher.

We also *interpret* equation (12) as saying that the metric is locally Euclidean up to second order.

The reason we introduced this theorem is because we want to give some meaning to derivatives of tensors.

We have talked about various operations we can do on tensors to make new tensors : we can add them, subtract them, multiply them, contract them…

We already can do all sorts of things.

Can we differentiate them?

In particular, if we have a tensor which is a function of position – that means all of its components are functions of position –, can we construct another tensor which is in some appropriate sense the derivative with respect to position of the tensor that we have ?

The answer is *yes*.

We shall see how to do it and why it is a little bit tricky.

Covariant derivatives

To differentiate a tensor with respect to position, we could think :

> *"Okay. Let's take the components of the tensor – for instance contravariant components – and just differentiate them."*

That would yield a new set of components, with one more index, which would be simply the derivatives of the components of the tensor.

But there is the following problem.

Think for instance of the derivatives of a vector.

We could differentiate each component with respect to each direction.

We sure can do that.

This would produce a two dimensional collection of values.

But it would not be a tensor.

A tensor is an abstract object which has, in each coordinate system, a multi-indexed collection of components.

And these collections of components must transform from one system to another according to equations (14) to (16) of lesson 2.

If we differentiated a vector field, at the same point, using two different coordinate systems, we would obtain two collections of double-indexed values.

But they would not be linked by a tensor equation like those mentioned, which are necessary to define a tensor.

Let us see why.

Consider a surface and a point P on it, figure 4.

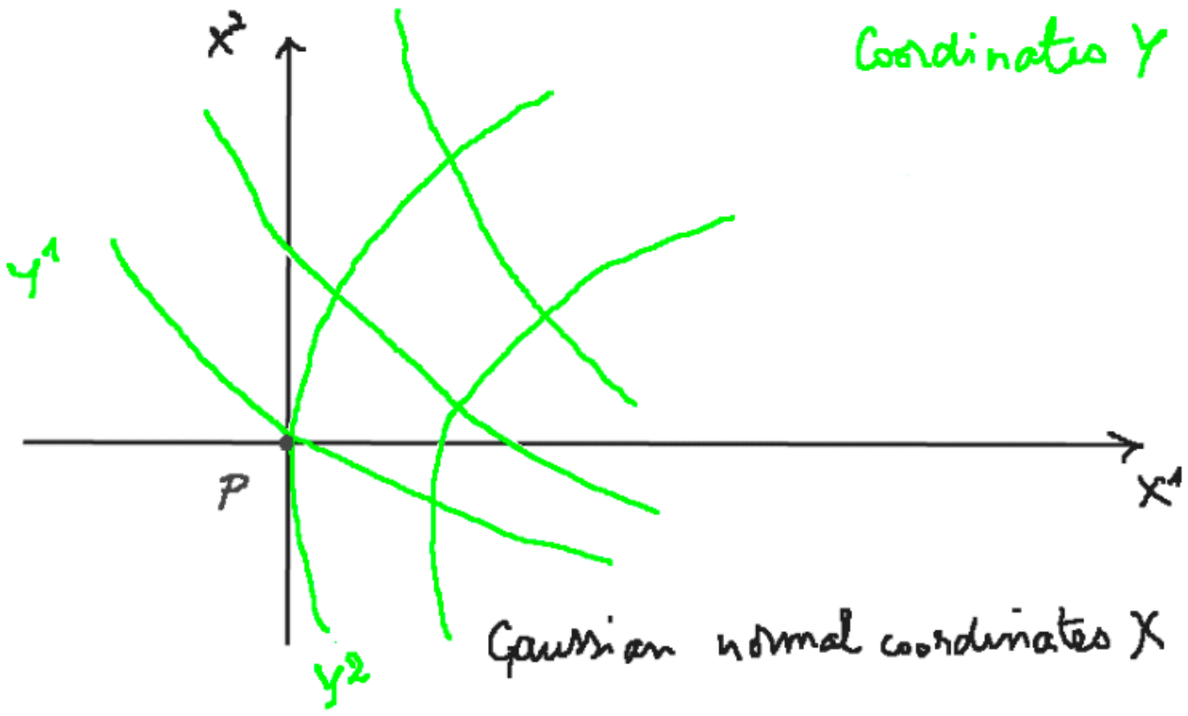


Figure 4 : Surface viewed at P with Gaussian normal coordinates X, and any coordinates Y.

We have two sets of coordinates on the surface, coordinates X and coordinates Y.

If the space is flat, for X we just use flat ordinary Cartesian coordinates.

Or if the space is curved, we use a set of Gaussian normal coordinates at P, that is coordinates X that are locally, at P, as straight and orthogonal as possible so to speak.

And there is another set of coordinates Y, for instance the initial ones.

For convenience, at P we chose X such that the X^2-axis is tangent to the Y^2-axis.

Think of a vector field defined over the surface.

The vector field is made of a different vector at every point.

In order not to clutter the picture, these vectors are not shown on figure 4 yet.

There is one at P and there are plenty around P.

Before we ask how to differentiate a vector field, let's ask what it would mean for the vector field to be constant in space.

We run into the following difficulty: because the space is curved it becomes hard to compare the vector at one point with the vector at another point.

The coordinates X cannot be chosen to be everywhere flat.

Then what exactly do we mean by saying that a vector at one point is equal to a vector at another point?

It does not really mean anything because to compare a vector at P with a vector at Q, unless we have nice flat coordinates over the whole surface, there is no unique way to do that.

So let's look at it for a moment.

If the space is really flat, then we know what it means for a vector at one point P of the surface to be the same or to be equal to a vector at another point Q of the surface, see figure 5.

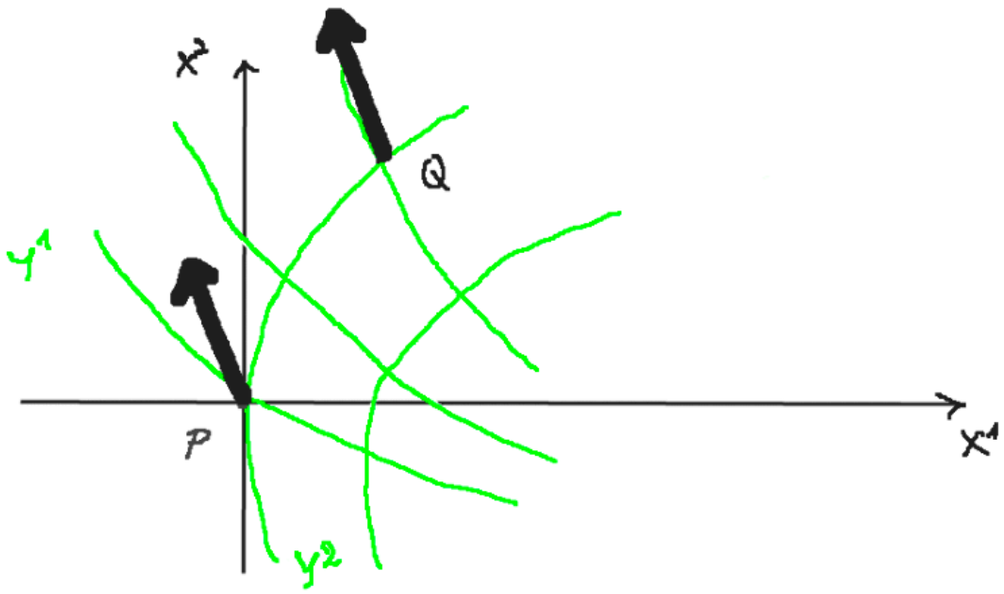


Figure 5 : Two equal vectors, at P and at Q.

It means they point in the same direction and have the same length.

Therefore in the X coordinates they have the same components.

But what about their components in the Y coordinate system?

In fact let's take a special case.

Suppose both vectors point vertically in the X axes, figure 6.

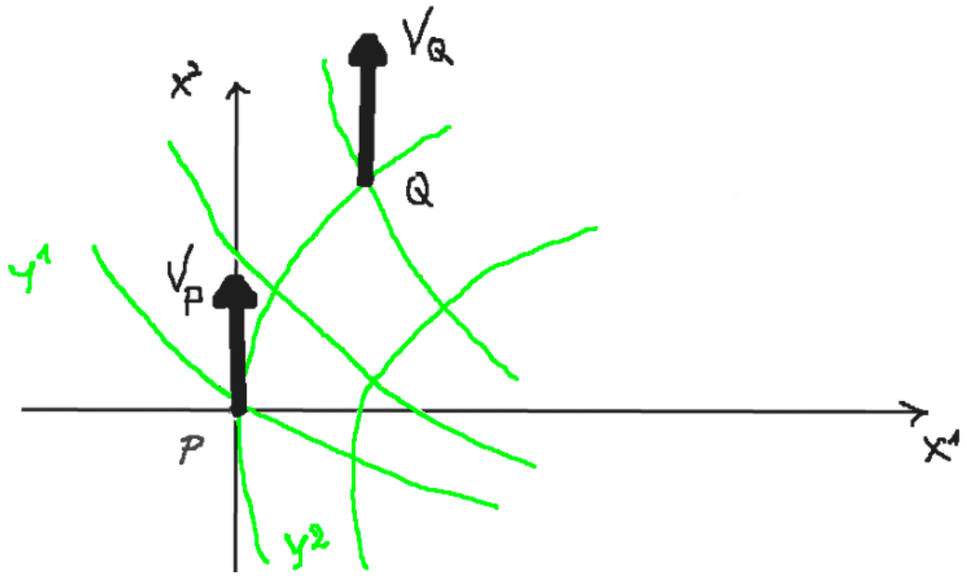


Figure 6 : Two equal vertical vectors, at P and at Q.

In that case, V_P only has an X^2 component in the flat coordinates.

And so does V_Q.

They have the same components in the X axes.

Are the components of V_P and V_Q in the Y coordinates the same?

Answer : No.

Along the Y axes, V_Q has a Y^1 component and a Y^2 component, while V_P only has a Y^2 component.

So it's clear that even though the vectors V_P and V_Q are the same, their components are not the same in the Y coordinate system.

This will be true, be it for covariant vectors or for contravariant vectors: when we have curvilinear coordinates, we *cannot* judge the equality of two vectors.

And in general we will be dealing with curvilinear coordinates.

Another way to say the same thing is that the derivative of the *m*-th component of *V*, with respect to the *r*-th direction of the coordinate system, might be 0 in one coordinate system and not 0 in the other coordinate system.

$$\frac{\partial V_m}{\partial X^r} = 0$$

$$\frac{\partial V'_m}{\partial Y^r} \neq 0$$

It might even be the case – as in figure 5 or 6 – that all of the derivatives of V_m are 0 in one coordinate system and not 0 in the other coordinate system.

That will be because the coordinates are shifting, not because the vector is changing.

You see what we are driving at: the equation saying that the derivatives of a vector are all equal to 0 may true in one reference frame, but untrue in another.

Therefore it cannot be a tensor equation.

In other words

the ordinary derivatives of the components of a vector with respect to the coordinates do not form themselves a tensor.

If they were components of a tensor we would think of this quantity

$$T_{mr} = \frac{\partial V_m}{\partial X^r}$$

as a rank 2 tensor with an m index and an r index.

But it were a tensor, the following fact would have to be true: if T_{mr} is zero in one frame or one coordinate system, it is zero in every coordinate system.

And it is simply not true with that T_{mn} – not because the vector may change from point to point, but because the orientation of the coordinates changes.

So we need a better definition of the derivative of a vector than just differentiating its components.

We need something which if it is zero in one frame is zero in every frame.

Here is how we will define the derivative of a vector.

As a preliminary, notice that to define the derivative at a point P we only need to look at points in the vicinity of P.

The first thing we do is to construct a set of Gaussian normal coordinates at point P.

Remember: Gaussian normal coordinates are as *straight as possible* near P.

They are well defined over the whole variety, and they make up an approximately Euclidean system of coordinates in the vicinity of P.

So we re-express all the vectors of the vector field in the new coordinates X that are locally flat and Euclidean.

To follow the procedure geometrically, let's look again at two vectors, one at P and the other at Q nearby, and for clarity make the second also slightly different from the first - see figure 7.

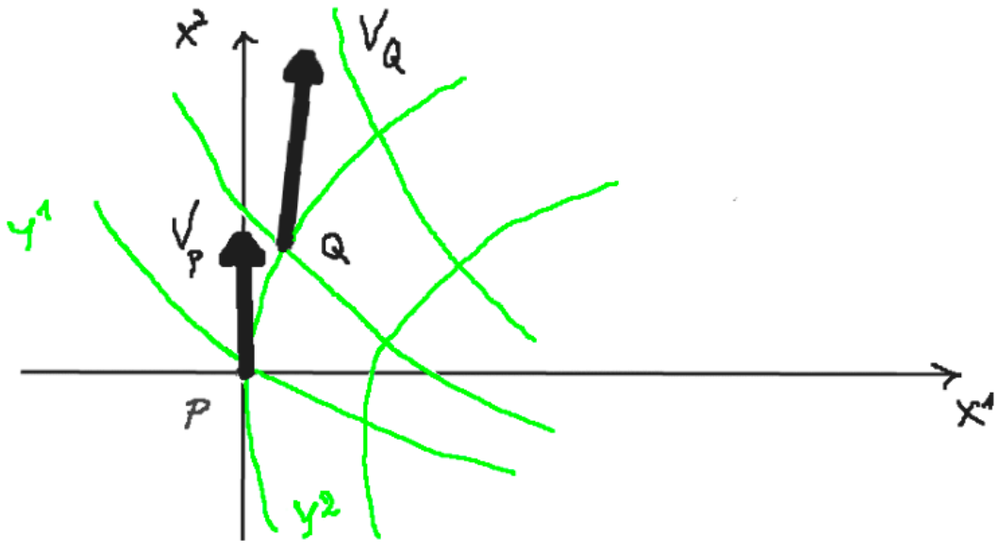


Figure 7 : Two vectors nearby.

Then pretend that the Gaussian normal coordinates are really nice flat coordinates over the whole vicinity of P.

Visually translate the vector V_{Q} so that its origin be the same as V_{P}'s, and look at the difference between V_{Q} and V_{P} .

The components, in Gaussian normal coordinates, of the difference between V_{Q} and V_{P} are the kind of elements that we will use to define the derivatives of the vector field at P.

For instance if we were interested in the derivatives along the PQ direction, it would approximately be $\overrightarrow{V_Q - V_P}$ divided by the small distance between P and Q.

But we shall be interested in the derivatives of V along the X axes.

Now the derivatives $\partial V_{\mathrm{m}}/\partial X^{\mathrm{r}}$, in the Gaussian normal coordinates, define the derivatives of V at P.

Finally, if we want to work with our initial Y coordinates, we take the double-indexed collection of partial derivatives produced by the differentiation in Gaussian normal coordinates at P.

We treat it as a tensor, that is, we consider this collection as the components of a tensor in the X coordinates.

And we transform it back into the Y coordinates using the tensor equations linking X and Y.

This necessarily produces a tensor, since when transformed from any Y coordinates into the X coordinates, it gives the same thing.

Now when we look at the derivatives of a vector in the general coordinates Y, we will have the addition of two terms: one term because the vector may be changing and the other term because the coordinates may be changing.

As we saw, the coordinates may shift, may be rotating out from under you, even if the vector doesn't change (see figures 5 and 6).

But before we look into these two terms, let's repeat the prescription as a methodical procedure.

> You have a vector field V on a variety, equipped with any coordinate system Y. You want to calculate the derivative of V at P. Then follow these steps :
>
> 1. Change coordinates to use Gaussian normal coordinates at P, let's call them X (notice that they are valid over the whole surface, and approximately flat at P).
> 2. Differentiate V at P in the usual way, using the X coordinates.
> 3. Consider the collection of partial derivatives you got as the components of a tensor of rank 2 in the X coordinates.
> 4. Switch back to your original coordinate system Y, and re-express in that original system the tensor you got, using the tensor equations linking X and Y.

Let's look at what we get and then comment on each term.

We find that, with this new definition, the derivatives at P are the old ones corrected by the addition of another term :

$$D_r V_m = \partial_r V_m - \Gamma^t_{rm}\, V_t \tag{13}$$

Here is how to read equation (13) : $D_r V_m = \partial_r V_m - \Gamma^t_{rm}\, V_t$

(1) The notation $D_\mathrm{r} V_\mathrm{m}$ is by definition the partial derivative of V_r with respect to the r-th direction in X obtained from the above procedure, that is, in the Gaussian normal coordinates X, and then re-expressed in the Y coordinates.

(2) $\partial_\mathrm{r} V_\mathrm{m}$ is the ordinary partial derivative of V_m with respect to the r-th direction in Y calculated directly in the Y coordinates; ∂ is a short hand for $\frac{\partial}{\partial Y^r}$.

(3) – $\Gamma^t_{rm} V_\mathrm{t}$ is the additional term due to the fact that the coordinates Y themselves evolve in the vicinity of P.

The minus sign is a pure convention.

This whole second term on the right hand side of equation (13) must clearly be proportional to V_t , i.e., if you double the size of V_t it must be twice as big.

And the coefficient Γ^t_{rm} in front of V_t is a new mathematical object appearing in the differentiation procedure.

We shall talk about it.

The term $\Gamma^t_{rm} V_{\mathrm{t}}$ does not have a derivative related to the vector because if doesn't come from the fact that the vector is changing.

It comes from the fact that the coordinates are changing in the vicinity of *P*.

The right hand side of equation (13) is what you will get if you take a vector, differentiate it in Gaussian normal coordinates and then transform the double-indexed collection of derivatives to other coordinates as a tensor.

In any other coordinate system what you will get is the usual derivative in that coordinate system minus an object times the components of *V* themselves.

As usual $\Gamma^t_{rm} V_{\mathrm{t}}$ means of course the sum over *t*.

Equation (13) holds in any arbitrary coordinate system.

Of course the V_{m}'s or V_{t}'s and the Γ^t_{rm}'s depend on the coordinate system.

Notice that there is not only one Γ^t_{rm} : they form a *three*-indexed collection.

$D_r V_m$ is called the covariant derivative of V_m .

This is not because the index m is downstairs.

The terminology has another origin.

If we had differentiated V^m, we would have obtained a formula analogous to equation (13).

And it would also be called the covariant derivative of V^m.

And, by its very construction, $D_r V_m$ is a tensor.

Christoffel symbols

The coefficients Γ^t_{rm} have two names : **connection coefficients** and **Christoffel symbols**.

The name **connection coefficients** comes from the fact that they connect neighboring points and tell us how to calculate the rate of change of a vector field from one point to another, even though the coordinate system may be changing.

They are also named **Christoffel symbols** after Elwin Bruno Christoffel.

They have been known on occasion as the "Christ awful" symbols because they seem complicated.

With some practice, however, you will discover that they are not that complicated.

They are just an extra linear term.

But I grant you that they are complicated and unlikeable enough.

Let's investigate what follows from the definition of the covariant derivative and the Christoffel symbols.

We are not going to prove every single fact we state, because there are just too many little pieces.

It follows from the definition of the covariant differentiation – namely, to differentiate a vector V at a point P, go to a set of Gaussian normal coordinates at P, differentiate the vector in the ordinary manner, treat the object you obtain as a tensor with two indices, change coordinates, etc. – that the Christoffel symbols have a symmetry :

$$\Gamma^t_{rm} = \Gamma^t_{mr} \tag{14}$$

The geometry of general relativity is the Minkowski-Einstein geometry which is an extension of Riemannian geometry with a non-positive definite metric and the Christoffel symbols we will use will be symmetric in the sense of equation (14).

A remark to build our physical intuition : doing the derivative in Gaussian normal coordinates which are almost flat, or as flat as can be, and then treating what we obtain as an object in its own right, is very similar to what we do in gravitational theory when we evaluate something in a freely falling frame.

For example in a freely falling frame we calculated how light moved across an elevator and then we transformed it to the frame of reference in which the elevator was accelerating.

That is closely related to the operations we have been doing in this lesson: we calculate something because we know how to do it in coordinates which are as flat as possible.

That would be a freely falling frame in general relativity.

And then we transform it in any coordinate we like, accelerated coordinates or anything we like, and we translate the statement from one coordinate system to another.

In the construction of the covariant derivative, the calculation of the variation of a vector from point to point is done first in Gaussian normal coordinates, and then it is transformed in any coordinate system.

Equation (13), reproduced below,

$$D_r V_m = \partial_r V_m - \Gamma^t_{rm}\ V_t \tag{15}$$

is the form that you get for the corresponding collection of components.

It is a tensor.

However, $\partial_r V_m$ is not a tensor.

Therefore $\Gamma^t_{rm} V_t$ cannot be a tensor.

And Γ^t_{rm} cannot be a tensor either.

We will see that the Γ^t_{rm}'s are build-up out of the derivatives of the metric $\partial_r g_{mn}$.

In fact in a coordinate system in which the derivatives of the metric are 0, the Christoffel symbols are 0.

But a tensor, if it is zero in one coordinate system, it is 0 in every coordinate system.

So that is another way to see that they can't be tensors.

Let's look now at the covariant derivative of higher rank tensors, because we will need this for curvature.

Suppose we have a tensor with more than one index, say

$$T_{mn}$$

and we want to differentiate it covariantly along the *r*-th axis.

We denote the resulting tensor

$$D_r T_{mn}$$

Its expression is the analog of equation (15), except that for every index in the tensor T_{mn} there will be a term like $\Gamma^t_{rm} V_{\mathrm{t}}$.

Let's see you how it works :

We start by working only on the *m* index, letting n be passive, and writing the equivalent of (15)

$$D_r T_{mn} = \partial_r T_{mn} - \Gamma^t_{rm} T_{tn} - \ldots$$

But we are not finished.

We have to do exactly the same with the *n* index, letting this time m be passive.

$$D_r T_{mn} = \partial_r T_{mn} - \Gamma^t_{rm} T_{tn} - \Gamma^t_{rn} T_{mt} \tag{16}$$

That is the form of the covariant derivative at point *P* of the tensor T_{mn}.

The rule is the same: we switch to Gaussian normal coordinates at *P*, we do the ordinary differentiation of the tensor with respect to each direction X^{r}.

This adds one more index to the collection of components that formed T_{mn}.

Notice that in order not to clutter equations we no longer use prime signs for one system and un-prime for the other.

The last time we did it in this lesson was in our comments following figure 6.

And we re-express the new tensor in the original coordinate system with the usual tensor equations (equations (15) and (16) of lesson 2 and their generalizations).

This allows us to to differentiate any tensor.

At the moment we are only dealing with tensors with covariant indices.

We will come in a moment to tensors with contravariant indices.

You may wonder : what is all this intricate business of covariant differentiation of tensors for?

Answer: it is for comparing things at different points.

We want to be able to talk about rates of variation of things along coordinate lines, with objects which have an existence irrespective of the system of coordinates we work with.

Remember that a vector in ordinary 3D has an existence irrespective of the basis we are using.

For certain work and calculations with it – not all of them – we need a representation of the vector in a basis.

The collection of components to represent it and work with it is different from one basis to another, but the vector we are talking about is the same.

It is even like the difference between things and their notations.

When we talk about 12 in the decimal system or 1100 in the binary system, we are actually talking about the same thing: the number twelve – which some other people call douze...

Where are we going to use covariant derivatives?

Answer : in field equations.

Field equations are going to be differential equations which represent how a field changes from one place to another.

But we want them to be the same equations in every reference frame.

We don’t want to write down equations which are special to some peculiar frame.

We want them to be valid in general.

That is, if they are true one frame they will be true in all frames.

That means they have to be tensors equations.

So we have to know how to differentiate tensors to get other tensors.

Another point worth stressing: the Christoffel coefficients will be present in equation (16) even in a flat space – like a plane, or this page, or the ordinary 3D Euclidean space – if you chose funny coordinates (see figure 3 of lesson 2).

That is an important point: terms like $\Gamma^t_{rm} T_{\text{tn}}$ are there even in flat space if you are using funny coordinates.

In fact if you choose any coordinates in which the derivatives of the g_{mn}'s are not zero, that is, in which the coordinates vary from point to point sinuously (viewed from an embedding space for instance), then $\Gamma^t_{rm} T_{\text{tn}}$ will be there.

The presence of terms like $\Gamma^t_{rm} T_{\text{tn}}$ in the covariant derivative of a tensor is not a characteristic feature of curved spaces, it is a feature of curved coordinates.

To begin to use our new tool, let's apply equation (16) to the metric tensor itself.

There is something special, however, about the metric tensor: in Gaussian normal coordinates, its derivatives are all zero.

That means that the covariant derivative of the metric tensor is zero.

That simple observation is what is going to allow us to compute the Christoffel symbols.

Let's write equation (16) with the metric tensor.

We are in a coordinate system that is general, curvilinear or not, flat of not, whatever:

$$D_r g_{mn} = \partial_r g_{mn} - \Gamma^t_{rm} g_{tn} - \Gamma^t_{rn} g_{mt}$$

But we know that this is zero.

Why ?

Because the ordinary derivative of the metric tensor in Gaussian coordinates is zero.

So, in any coordinate system, we have

$$\partial_r g_{mn} - \Gamma^t_{rm} g_{tn} - \Gamma^t_{rn} g_{mt} = 0 \tag{17a}$$

Now let's write the same equation, except permuting the indices.

It is a little trick to get as much information about the Christoffel symbols as we can, and eventually, via some nice eliminations, to be able to isolate one Christoffel symbol and express it in terms of the ordinary partial derivatives of *g* with respect to the axes in any coordinate system.

So equation (17a) becomes

$$\partial_m g_{rn} - \Gamma^t_{mr} g_{tn} - \Gamma^t_{mn} g_{rt} = 0$$

But the middle term, by symmetry, can be rewritten interchanging *m* and *r*

$$\partial_m g_{rn} - \Gamma^t_{rm} g_{tn} - \Gamma^t_{mn} g_{rt} = 0 \tag{17b}$$

And for exactly the same reasons we can write

$$\partial_n g_{rm} - \Gamma^t_{rn} g_{tm} - \Gamma^t_{mn} g_{rt} = 0 \tag{17c}$$

Let's write the three equations of interest next to each other, to look at them more conveniently :

$$\partial_r g_{mn} - \Gamma^t_{rm} g_{tn} - \Gamma^t_{rn} g_{mt} = 0 \qquad \text{(a)}$$

$$\partial_m g_{rn} - \Gamma^t_{rm} g_{tn} - \Gamma^t_{mn} g_{rt} = 0 \qquad \text{(b)}$$

$$\partial_n g_{rm} - \Gamma^t_{rn} g_{tm} - \Gamma^t_{mn} g_{rt} = 0 \qquad \text{(c)}$$

How can we add them, or subtract them, or do something clever, that will isolate only one of the terms with a gamma ?

Let's add equation (b) to (c) and subtract (a).

Of course we will get $\partial_n g_{rm} + \partial_m g_{rn} - \partial_r g_{mn}$ plus some other terms.

But the middle term of (a), $\Gamma^{t}_{rm}\, g_{tn}$, will disappear, and so will the last term of (a), $\Gamma^{t}_{rn}\, g_{mt}$.

And we will be left with twice the same last term with a gamma, $\Gamma^t_{mn} g_{rt}$.

So we are in luck : (b) + (c) – (a) yields

$$\partial_n g_{rm} + \partial_m g_{rn} - \partial_r g_{mn} = 2\Gamma^t_{mn} g_{rt} \qquad (18)$$

We are still not done.

We would like to get Γ^{t}_{mn} by itself.

Our goal, indeed, is to find out what the Christoffel symbols are in terms of derivatives of the metric.

We are almost there.

Fist of all, equation (18) shows that if all the derivatives of the metric are zero, the Christoffel symbols must be 0.

But how are we going to get rid of the g_{rt} on the right hand side of equation (18)?

The answer comes from recalling that g_{rt} has an inverse.

We saw that in the form of matrix equations, as well as in the form of tensor equations, see equation (33) of lesson 2.

We multiply both sides of equation (18) by the inverse tensor, and move also the factor 2, and we get

$$\Gamma^t_{mn} = \frac{1}{2}\, g^{rt}\ [\, \partial_n g_{rm} + \partial_m g_{rn} - \partial_r g_{mn}\,] \tag{19}$$

This is the expression of the Christoffel symbols in terms of the ordinary derivatives of the metric tensor.

It looks pretty simple.

The indices m and n are symmetric.

You can interchange them, the Christoffel symbol won't change.

There is one negative term and two positive terms.

It is not very complicated.

The problem is that there is a boatload of them.

When you think about a four dimensional space and let all the coefficients range from 1 to 4, there is just a lot of Christoffel symbols.

That is what makes doing calculations in general relativity a very tedious business.

Intrinsically there is nothing hard about it.

But in fact doing a calculation in a general relativity context usually fills page after page of nothing more complicated than just computing these derivatives and assembling them together.

Equation (19) holds for any coordinate system and any metric tensor.

Notice that all our calculations are at one point P.

Whatever coordinate system our variety is equipped with, we position ourselves at a point on it, consider the metric tensor g_{mn} there, calculate the gammas there with equation (19).

The use of Gaussian normal coordinates at P was just for intermediate reasoning, calculation and proof purposes.

We are now back in the initial coordinate system of our space.

The g_{mn}, g^{mn} and Γ^{t}_{mn} depend on P.

But equation (19) is general.

At every point, it expresses the connection coefficients – the other name of the Christoffel symbols – in terms of the derivatives of g.

These connection coefficients enable us to figure out how any vector or tensor varies, in our space, when we move a little bit along the coordinate lines.

The problem with the Christoffel symbols is that they are not tensors.

They can be zero in one frame of reference, and not zero in another.

For example, in a set of Gaussian normal coordinates at point P, all of the Γ^{t}_{mn} are equal to zero.

This can be seen in many ways.

Since the metric tensor in that case is constant (even equal to the Kronecker delta tensor, but that is not necessary), equation (19) tells us that $\Gamma^{t}_{mn} = 0$.

But in some other coordinate systems the Christoffel symbols are not.

Remember that even in an intrinsically flat space, we can have coordinates such that the metric tensor is not constant.

Then the Chritoffel symbols won't be zero.

As said, they are related to the coordinate system not to the intrinsic geometry of the space.

A sphere is intrinsically non flat.

In the polar coordinates θ and ϕ (see lesson 1, figure 14), the components of g are not constant, therefore the Christoffel symbols are not zero in that system of coordinates.

Even on a sphere, however, at any given point we can build a set of Gaussian normal coordinates (like maps do – you just tinker with the longitude), and then the Christoffel symbols at that point will be zero.

All this is conceptually tricky when we meet it for the first time.

But at the end of the day the rule is simple : calculate the Christoffel symbols and, in many contexts, replace ordinary derivatives with covariant derivatives.

You could write your equations in Gaussian normal coordinates.

Then they would just involve ordinary derivatives and we would not have to wade through a river of Christoffel symbols.

But if you want the same equations in general coordinates then replace the ordinary derivatives by covariant derivatives.

That is the procedure.

It will require you to think about it.

You will have to sit down, to follow carefully the reasonings, to do calculations.

Then what we are doing will become clear.

Curvature tensor

What is curvature?

It is easiest to start with two-dimensional curvature.

Intuitively it is easy to understand: it is a characteristic of something that is round and cannot be flattened out.

But we are going to give it some more mathematical definition.

How do we probe for curvature ?

Let's begin by drawing a space which is curved.

The sphere is obviously curved.

I don't want, however, to deal with a sphere.

I want to look at a cone.

But it is going to be a cone with a round summit, figure 8.

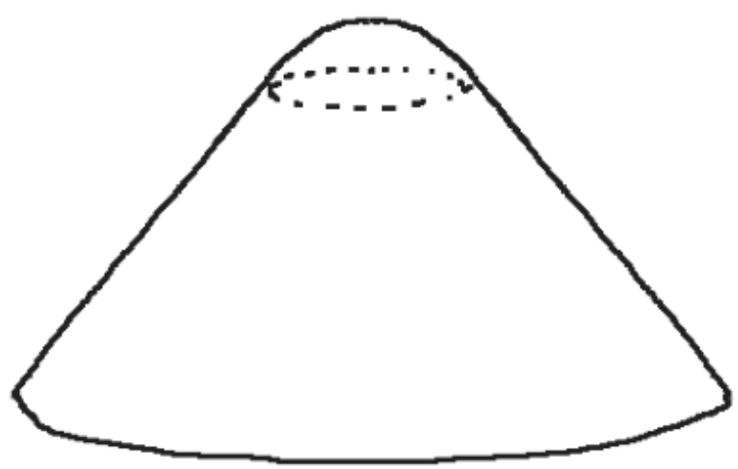


Figure 8 : Cone with a rounded summit.

Think of the top of a mountain the sides of which are nice and flat like those of a volcano, and the top is round.

So if you are away from the top of the mountain, below the dotted line, around you the surface is flat.

It may not look flat because, like the furled page in figure 1 of lesson 2, we represented the mountain embedded in 3D Euclidean space.

But the surface is what mathematicians call **developable**: any section with no hole in it, cut from the side of the mountain, can be flattened onto a plane without distortion.

The rounded cone only differs from a flat space in this vicinity of the summit.

To see that just take the same space below the dotted line but continue it so that it really does form a genuine cone as in figure 9.

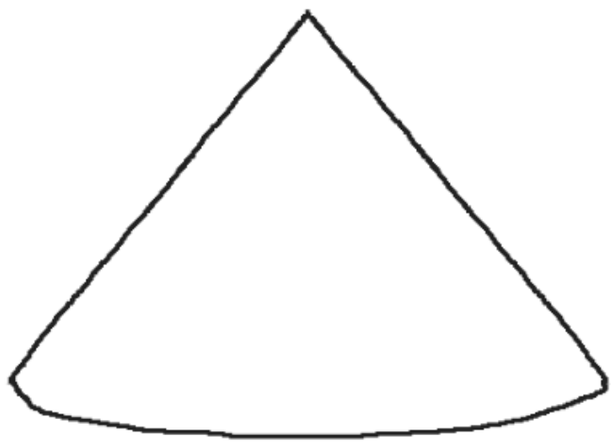


Figure 9 : Genuine cone.

Then slice it along a generatrix, that is, a straight line on the cone going to the top.

And open it up.

You can lay out flat on a plane the shape that you get.

It is a disk with a missing piece, see figure 10.

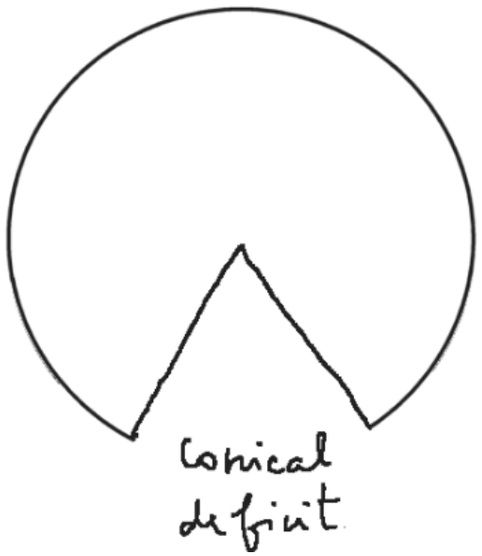


Figure 10 : Cone opened up and laid flat (smaller scale than figure 9).

The missing piece is called the deficit angle, or the **conical deficit**.

We can see that the bigger the conical deficit is, the pointier the cone will be.

Now, on the flat surface of figure 10, let's consider a collection of identical vectors arranged around the shape as shown in figure 11.

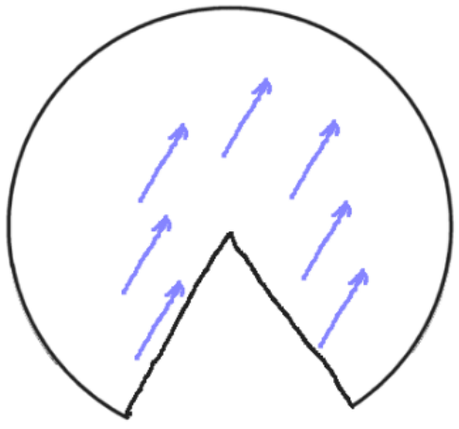


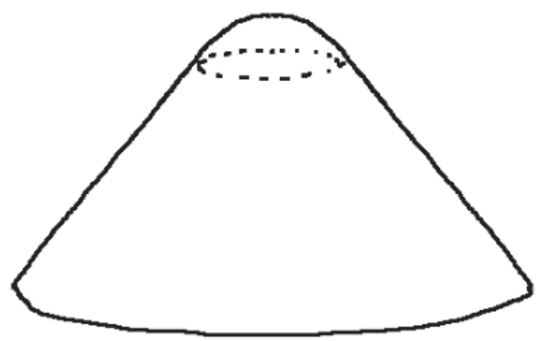
Figure 11 : Identical vectors.

On the flat surface, all the vectors point in the same direction.

But when we fold again the shape to form the cone, we see that the vectors no longer point in the same direction.

Think of them as very small so that they don't have to be bent. The first one on the left is along a generatrix, but the last one on the right points in another direction.

So the tip of the cone is such that if you carry a vector around (in the same direction on the flattened version) when you get around to the other side it has been rotated.



The rotation is due to the conical deficit.

Exactly the same is true on the rounded cone of figure 8: if we take a vector on the flat side, below the dotted line, and we carry it around the mountain in such a way that, when it is opened up and laid on a plane, it is always pointing in the same direction, by the time we get back to the other side it will be pointing in some other direction.

That is because the summit of the shape is curved.

That is the *effect* of curvature.

A region of curvature has the property that if you take a vector pointing in the same direction as much as you can do, that is locally parallel to itself, and take it around the region of curvature, then the vector will rotate.

The vector will rotate when it goes around any region of curvature even though you took pains to make sure that at every point you were moving it parallel to itself.

In other words you took pains to make sure that the covariant derivative of the vector was always zero.

But nevertheless, when you came around to the other side, it is shifted.

There is another way to say this, which is equivalent and actually more useful.

Consider a curved space with some curvature at point P, figure 12.

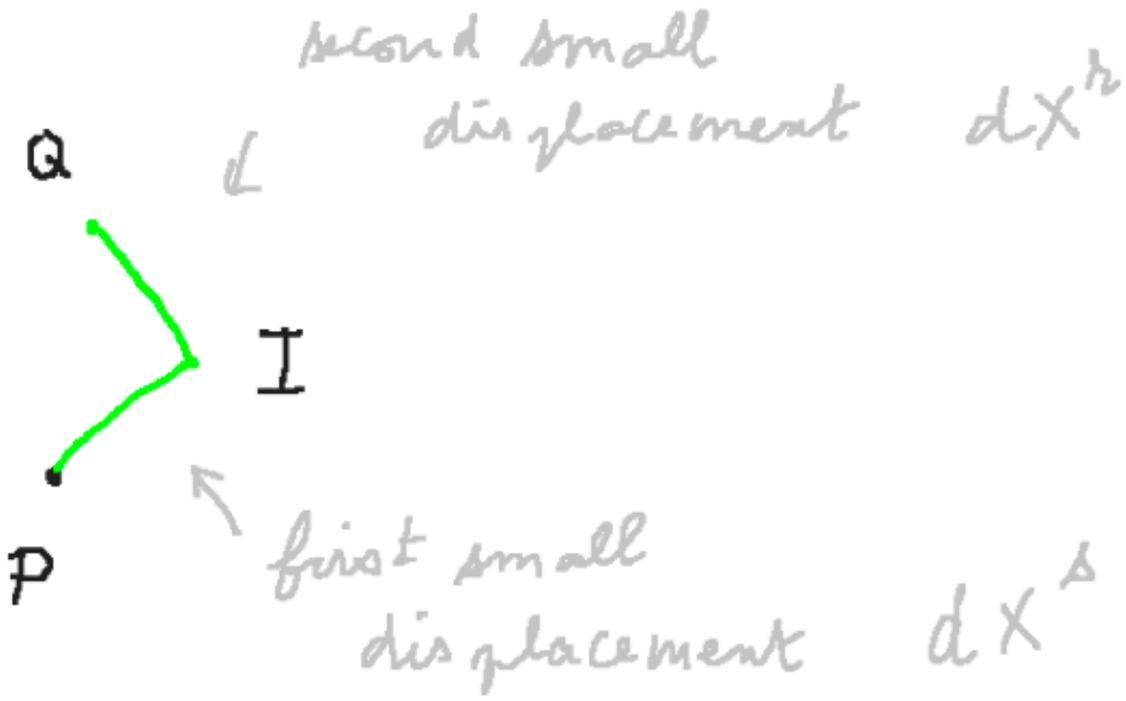


Figure 12 : Displacements to differentiate a vector field along two axes.

Take a vector field and differentiate it along one axis (first displacement in figure 12).

Then differentiate it along the second axis (second displacement in figure 12).

That is, you consider the vector field at P.

Then you move a bit along one axis, and consider the new value of the vector field at I.

Then you move another bit along the second axis, and consider the value of the vector field at Q.

Over each displacement, the vector will change.

How will it change?

It will change typically by differentiating the vector along the two axes in sequence.

We first differentiate the vector along one axis and then differentiate it along the second axis.

And this will produce a small change in the vector due to the two derivatives.

The total change in the vector consists of two changes.

And that total change is proportional to a second derivative.

To get a feel for this, think of a function f of two independent variables a and b.

View it in 3D.

Suppose f is nice and smooth everywhere, and in particular, at the point (0,0), is tangent to the plane formed by the a and b axes.

Suppose we are interested in $f(a, b) - f(0,0)$.

If a and b are small, this variation is approximately equal to $$ab\frac{\partial^2 f}{\partial a \partial b}(0.0)$$

That is true in any coordinates : if, to compare the vectors at Q and at P, you compared the vector at I with the vector at P, and then compared the vector at Q with the vector at I, what you would be calculating is the second partial derivative of the vector with respect to the two directions.

In figure 12, if the first displacement is along the direction X^s and the second displacement along the direction X^r, then the variation of the m-th component of the vector V – let's say it has covariant indices – would be

$$D_r D_s V_m \tag{20}$$

This expression is calculated covariantly.

And in Gaussian normal coordinates, expression (20) would just contain ordinary derivatives.

We could have also gone in the other direction, figure 13.

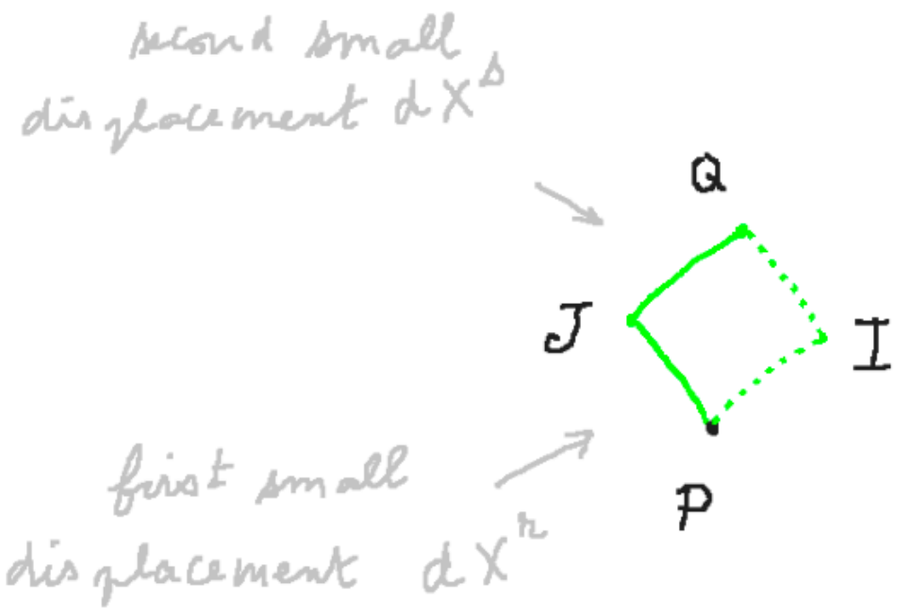


Figure 13 : Displacements in the other order.

That is to say we could have gone first in the *r* direction and then in the *s* direction and calculated the way the vector changes from *P* to *J* and then from *J* to *Q*.

The variation of *V* would then be

$$D_s D_r V_m \tag{21}$$

Ordinarily, and in flat space in general, expression (20) and expression (21) are equal to each other.

That is

$$D_r D_s V_m = D_s D_r V_m \tag{22}$$

This is just a version of the fact, in calculus, that the partial derivatives of a function of several variables can be taken in the order you like.

But equation (22) is not true in curved space.

In curved space the difference

$$D_r D_s V_m - D_s D_r V_m$$

can be thought of as taking the vector around the closed loop

$$P \to I \to Q \to J \to P$$

Let's go back to our cone, either the genuine cone (figure 9), or the cone with a rounded top but looking at the part below the dotted line (figure 8).

Consider a vector field which when the cone is opened and laid flat is constant.

Fold the flat shape to form the cone.

We discovered that if we follow the vector field on a closed loop around the top, we don't get back to the same vector we started with.

This is due to the following fact, that is important enough to note with italics.

In curved space covariant derivatives are not interchangeable.
In flat space they are interchangeable.

That is the way we will test whether the space is flat or not.

We will test whether differentiating tensors, and in particular vectors, in opposite order gives the same result.

- If the answer is yes everywhere in the space for any vector, then the space is flat.

- If we discover that there are places in the space where the order of differentiation gives different answers, then we know the space has some kind of defect in it (like the point of the genuine cone) or has curvature (like the summit of the rounded cone).

So all we have to do is compute the second covariant derivatives of a vector in opposite order and compare them.

In principle it is not complicated.

In practice it will be a little complicated, but will remain manageable.

We have all the tools to do it.

Now it is a mechanical operation, consisting of pure plug-ins.

We *sketch* the steps below, and then give the answer.

We start with a vector expressed with covariant components V_n

We compute its covariant derivative in the *r* direction $D_r V_n$

And then we differentiate this, still covariantly, in the *s* direction $D_s D_r V_n$

Later on, we will interchange *s* and *r* and subtract.

(23)

Let's replace the first covariant derivative of V_{n}, with respect to *r*, by its expression given in equation (13).

We get

$$D_s D_r V_n = D_s \left[\partial_r V_n - \Gamma^t_{rn} V_t \right]$$

But notice that $\partial_r V_n - \Gamma^t{}_{rn} V_t$ is a tensor.

So we know how to differentiate it : use equation (16).

Continue to crank mechanically the calculations.

In the end, the difference between the two second order covariant derivatives yields a tensor, denoted $\mathfrak{R}^t{}_{srn}$, multiplied by V_t

$$D_s D_r V_n - D_r D_s V_n = \mathcal{R}_{srn}^{\ \ t} V_t \tag{24}$$

And here is $\mathfrak{R}$

$$\mathcal{R}_{srn}^{\ \ t} = \partial_r \Gamma^t_{sn} - \partial_s \Gamma^t_{rn} + \Gamma^p_{sn}\Gamma^t_{pr} - \Gamma^p_{rn}\Gamma^t_{ps} \tag{25}$$

There are two terms involving derivatives of Christoffel symbols, and two terms which are sums over p of products of Christoffel symbols.

$\mathfrak{R}^t{}_{srn}$ *is the curvature tensor, also called Riemann curvature tensor or Riemann–Christoffel tensor.*

It has a complicated expression.

And it is even more complicated when you remember that the Christoffel symbols are given by equation (19), which we reproduce below

$$\Gamma^t_{mn} = \frac{1}{2}\, g^{rt}\ [\, \partial_n g_{rm} + \partial_m g_{rn} - \partial_r g_{mn}\,]$$

Let's see what are the elements in the curvature tensor given by equation (25).

The Christoffel symbols involve derivatives of g.

So differentiating again produces second derivatives of g.

Remember: the second derivatives of g are the things that we cannot generally set equal to zero.

For the first derivatives of g, we saw that we can find a frame of reference where they are equal to zero.

But for the second derivatives of g we can't.

So by the time we are finished calculating the curvature tensor, the second derivatives of g have come into it.

The second derivatives are testing and probing out the geometry of the surface a little more thoroughly than just the first derivatives.

In a similar way, in the theory of functions, when at a point x you know $f(x)$ and $f'(x)$ and $f''(x)$, you are better off than with just the value of f and of its first derivative.

So the curvature tensor has second derivatives of the metric g, and it has squares or quadratic things involving first derivatives of g.

It is a complicated thing.

If we were to actually write it in terms of the metric, or we were to try to calculate it for a given metric, it could rapidly fill up pages.

But conceptually what it is doing is simply calculating the difference in a vector if you transport it around the loop in figure 13, keeping it parallel to itself, as much as you can locally at every point, until you have come all the way around.

It calculates the little change in a vector in parallel transport going around a loop.

The curvature tensor has a complicated formula, but we can calculate it.

We can put the metric tensor into a computer and ask the computer: "Is the curvature tensor 0 ?"

It is even better if you have a software that can do algebra.

If you have the metric in some algebraic form, you can do all the operations of equations (19) and (25) and then test out whether the curvature tensor is zero everywhere.

If the curvature tensor is zero everywhere, that is all its components are zero everywhere, then the space is flat.

We shall study the curvature tensor a little more.

As we said, it is a complicated thing.

Its main use is to tell us whether the space is flat.

And, if not, how unflat it is.

It is closely related to a quantity in gravitational physics.

Can you guess which one?

A local quantity which tells you that the space is not flat.

It must be something telling you whether there is really a gravitational field present or not.

Answer : tidal forces.

It is exactly related to tidal forces, those things which in a gravitational field squeeze bodies one way and stretch them another way.

Tidal forces are represented by the curvature tensor.

Here is another way to get a feel for what the curvature tensor is.

Imagine a surface which is flat away from a point in the center where it has a bulge.

It doesn't have to be a rounded cone.

It can simply be a plane with a bulge, as in figure 14.

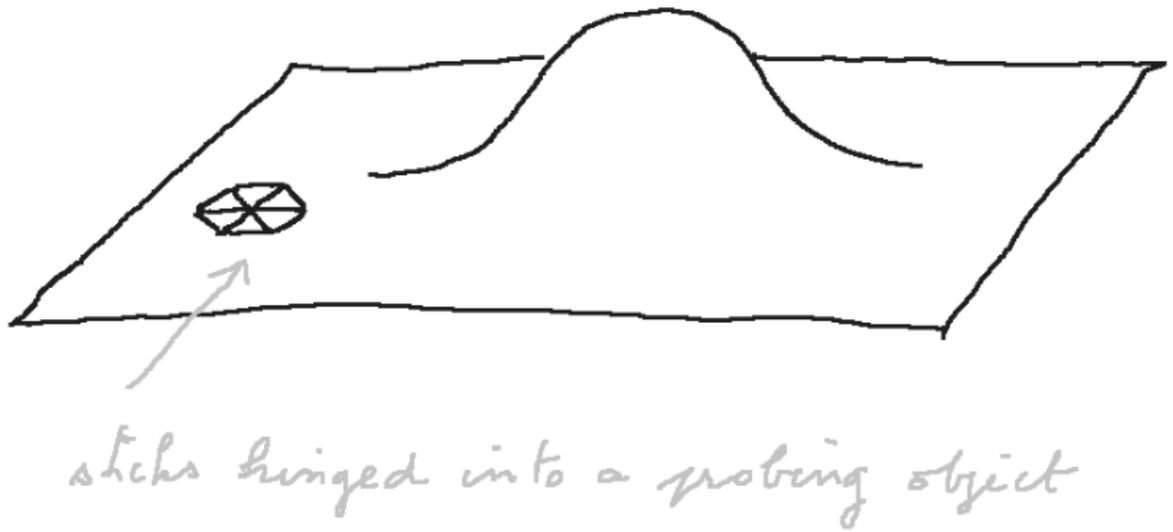


Figure 14 : Probing the curvature with a little structure of Tinkertoy sticks.

And you have a small structure of Tinkertoy sticks, all hinged at their extremities, so that their directions can move freely from each other, while remaining attached.

At first, the probing structure lays flat, without stress or distortion, in a flat part of the surface, because the probe is itself flat.

And then you start moving the probe.

While you move it in the flat region, nothing happens to it.

It remains perfectly happy.

It doesn’t get stretched, it doesn’t get distorted or deformed.

This would have also been the case on the side of the rounded cone away from the summit, by the way.

But now what happens when you try to move the probe into the curved region?

When you are trying to move it into the curved region, it simply can’t follow the curvature without having to stretch or compress some of its lengths.

It has to follow the metric properties of the curved space.

In particular if you go around the probe, what you are doing somehow is sampling the double covariant derivatives of equation (20) or (21).

You are going to find out that various angles between sticks change from their value in flat space.

The lengths of the sticks shift too, they get stressed, they get deformed.

The measure of how much the probe gets stressed locally is given by the curvature tensor.

The curvature is an important property because, if you are in region where there is curvature, you can feel it, either with tidal forces in a gravitational field, or with the probe in the experiment of figure 14.

Uniform gravitational fields don't have curvature.

That is why in free-fall, in a perfectly uniform gravitational field, you simply feel nothing.

Indeed they don't cause tidal forces.

Of course perfectly uniform gravitational fields don't really exist in nature.

You can simulate one with acceleration, but you cannot see one in nature.

They exist only approximately on the surface of big massive objects, if you limit yourself to a small solid angle.

This leads us to a last remark.

Tidal forces, or curvature on a surface, have a bigger effect on bigger objects.

The 2000-mile man in free-fall toward the Earth will feel tidal forces more strongly than a free-falling bacteria.

Similarly, in figure 14, if the probe is small compared to the bulge, it won't be much deformed when it goes over it.

Whereas if it were a bigger Tinkertoy structure, made for instance of many more hexagons, covering a larger region of the plane, like floor tiles, but still hinged so that any two connected sticks can change their directions from each other, then the probe would feel the curvature more strongly.

At the end of this third lesson we have reached the curvature tensor.

It is complicated.

Its expression is given by equation (25).

I wish we could do without the sea of symbols and indices.

But we can compute it.

Often you will be presented with the metric tensor in some analytic form.

There will be a formula for it.

With the formula you can do differentiations.

Everything consists in analytic functions that you can calculate.

So we have finally reached our initial aim.

Remember that our aim was to find a method to determine whether a space is flat.

By definition, it is flat if there exist a set of coordinates in which the metric tensor is everywhere equal to the Kronecker delta tensor.

But trying out every possible set of coordinates, and checking them at every point, was not a practical solution.

So we found the curvature tensor.

If it is zero everywhere, then we can find a set of coordinates such the metric tensor is everywhere equal to the Kronecker delta tensor.

You just position yourself at any fixed point and start to build Euclidean coordinates, like we did when we built Gaussian normal coordinates.

Now no curvature will limit us to a small vicinity.

In summary, *the curvature tensor has a complicated form but it is practical tool*.

We are finished with our mathematical study of Riemannian geometry, metrics, tensors, curvature, etc.

As far as we are concerned, our new tools will now be put to use.

In the next lesson, we will enter into gravity land.

We will see what has to change to go from Riemann geometry to Einstein geometry.

Then we will study a particular example: the Schwarzschild geometry.

It is the geometry of a black hole, a star, or any gravitating mass.

Lesson 4 : Geodesics and gravity

Introduction

Before we enter into the heart of the subject – geodesics and gravity – let's go through a quick review of some of the important points and equations we have established so far.

We saw various operations we can do on vectors and tensors: addition, tensor product, contraction, not to mention expression of their components in various coordinate systems.

Most importantly, we talked about the operation of differentiation.

We have a surface or variety, with a tensor defined at every position.

How to calculate the variation of the tensor when we move a bit from X to $X + dX$?

If we did it "naively" we would produce something which would depend not only on the tensor itself, but also on the change of the curvilinear coordinate system from X to $X + dX$.

And it would not be a tensor.

It could not be used conveniently in frame-independent equations.

Notice that the idea of a change not due to the tensor itself is more delicate than it seems.

It requires to be defined, and that is precisely the purpose of considering locally what is going on in Gaussian normal coordinates.

We defined therefore a special kind of differentiation called **covariant differentiation**.

At any given point P on the surface, we change from the current coordinate system to a set of local Gaussian normal coordinates – that is coordinates which are *closest* possible to Cartesian coordinates.

This can be done in *several* ways, because once we have one set of Gaussian normal coordinates we can rotate it and that produces another set.

But considering how we use it it does not matter which set we use.

So we choose one set of Gaussian normal coordinates at P, and we calculate the ordinary derivatives of the components of the tensor with respect to each direction in that system.

This produces a multi-indexed collection of components with one more index downstairs.

And we treat this collection as the components of a tensor expressed in that Gaussian normal coordinate system.

We can then express the derivative tensor back into the initial curvilinear coordinates if need be.

But remember : a tensor exists and is well defined irrespective of the basis we are using – just like vectors exist and are well defined in a vector space irrespective of any basis.

Bases are used to work on vectors when we need to have them in component form.

The tensor obtained by the above differentiation process is called the *covariant* derivative of the tensor we started with.

At any particular point P, if we happen to be in Gaussian normal coordinates, then the covariant derivative is *just* the ordinary derivative.

But in other coordinates, it takes on a more complicated form.

This is a *consequence* of the fact that we want it to be a tensor.

For the simplest example of a tensor, that is a covariant vector, the formula is

$$D_r V_m = \partial_r V_m - \Gamma^t_{rm} V_t \tag{1}$$

where Γ^t_{rm} is called a *Christoffel* symbol.

For a tensor with more covariant indices, the formula is a simple generalization of equation (1), carrying an extra term with a Christoffel symbol for each index.

For instance

$$D_r T_{mn} = \partial_r T_{mn} - \Gamma^t_{rm} T_{tn} - \Gamma^t_{rn} T_{mt} \tag{2}$$

Equations (1) and (2) are valid in any coordinate system.

But let's *stress again* that, at any given point, if we are locally using a coordinate system which is as close as possible to Cartesian – instead of Gaussian normal we will sometimes call it a "best local coordinate system" – then the Christoffel symbols are zero, and the right hand sides reduce to their first terms, that is to ordinary derivatives.

Having recalled all this, let's turn to a specific tensor : the metric tensor.

Cartesian coordinates are by definition a coordinate system in which the metric is everywhere constant, moreover is equal to the Kronecker delta tensor.

And a space in which such a system can be found is called *flat*.

Similarly, locally, a Gaussian normal coordinate system is one in which the metric tensor is locally the Kronecker delta tensor up to second order (that is, still behaving like the Kronecker tensor in the first order but not in the second).

Therefore, in a set of Gaussian normal coordinates, at a given point P, the ordinary partial derivatives of the components of the metric tensor are zero :

$$\partial_r \ g_{mn} = 0 \tag{3}$$

This is true *only* in a set of Gaussian normal coordinates.

However, as a consequence, the covariant derivative of the metric tensor, which is itself a tensor, **in any coordinate system**, at any point *P* on the surface, is equal to zero:

$$D_r \ g_{mn} = 0 \tag{4}$$

Looking again at the Christoffel symbols appearing in equations (1) and (2), we saw in many ways why they are *not* tensors.

Unlike tensors, they can be zero in one coordinate system and not zero in another.

We calculated their value, in any given coordinate system, in terms of the ordinary partial derivatives of the components of the metric tensor :

$$\Gamma^t_{mn} = \frac{1}{2}\, g^{rt}\ [\, \partial_n g_{rm} + \partial_m g_{nr} - \partial_r g_{mn}\,] \tag{5}$$

We see again that if the ordinary partial derivatives of the metric tensor components are zero, as is the case in a best local coordinate system, *then* the Christoffel symbols are zero in that coordinate system.

If they were tensors they would have to be zero in any coordinate system, but they are not.

Equation (5) should be memorized.

It is not too hard.

Remember that tensor indices must display a coherent structure, and follow the summation convention.

On the left hand side, we have a Christoffel symbol. It has two lower indices, *m* and *n* for instance, and one upper index, *t* for instance.

Its expression will involve a sum – expressed not with a big Σ but with the summation convention – over an index, say, r.

The formula begins with a factor 1/2 followed by the inverse metric tensor g^{rt} with upper indices.

These are the t of the Christoffel symbol and an r to operate a sum.

Then, between brackets, the formula contains three partial derivatives of the form $\partial_i g_{jk}$, where ijk cycle over mnr, two terms with a plus sign and the term with g_{mn} with a minus sign.

That is the Christoffel symbol, with one upper index and two lower indices.

Equation (5) looks slightly different from equation (19) of lesson 3, because between the brackets we wrote, in the middle term, g_{nr} instead of g_{rn}.

But remember that g is symmetric.

These Christoffel symbols go into equations (1) and (2).

That tends to make covariant derivative rather complicated objects.

If we plugged the expression of the Christoffel symbols in equations (1) or (2), and even wrote out the analytic form of the metric tensor that we might have at our disposal, it would be a nuisance.

But that is what covariant derivative means.

Equation (1) was the covariant derivative of a vector with covariant components.

Let's talk now about the covariant derivative of a vector with **contravariant** components.

We denote it

$$D_r V^m$$

As always it starts out with an ordinary partial derivative, and there is another term.

The calculations are exactly the same as what we did to calculate the covariant derivative of a covariant vector.

For the proof use the following little trick.

We can write

$$V^m = g^{mp} V_p$$

Then we take the covariant derivative of each side.

Since, in a best set of coordinates, the covariant derivative is a standard derivative, it is easy to *see* that it will satisfy the rule of differentiation of a product

$$D_r V^m = (D_r g^{mp}) V_p + g^{mp} (D_r V_p) \tag{6}$$

Now what do we know about the first term on the right hand side?

It involves the covariant derivative of the inverse metric.

We know that the covariant derivative of the metric is zero – that was equation (4).

That means, as we saw, that in a best set of coordinates the metric is constant up to second order.

Then its inverse must also be constant up to second order – go into matrix representation if necessary to convince yourself.

Therefore the first term on the right hand side of equation (6) disappears, and the equation rewrites

$$D_r V^m = g^{mp} (D_r V_p) \tag{7}$$

Now we know how to calculate the covariant derivative of a vector with lower indices.

That is given by equation (1).

If you plug that in equation (7) and do a little algebraic manipulation, you will find out the formula for covariantly differentiating a vector with a contravariant index, that is with an upper index.

Here is the result

$$D_r V^m = \partial_r V^m + \Gamma^m_{rt} V^t \tag{8}$$

As before the formula begins with just a simple derivative.

Then it has a term which would be zero in a set of best coordinates, because the covariant derivatives would simply be the ordinary ones.

But they are not zero in general coordinates.

In the term with the Christoffel symbol, there is a sum over t.

Generally speaking, in equation (8), we see that all the indices are in place as expected.

The only peculiarity is the plus sign instead of the minus sign that appeared in the covariant derivative of a vector with a lower index.

That minus sign was a convention.

Here too, but it must be the other sign.

If we have memorized the covariant derivative of a covariant vector, to remember the formula for the derivative of a contravariant derivative, it is the only thing we have to remember.

The place of the indices follow consistent rules.

There is no way to be mistaken.

The Christoffel symbols have one upper index and two lower indices.

The upper index must the same as on the left hand side of equation (8).

The second term on the right hand side is a sum, so there must be a dummy index t downstairs in the Christoffel symbol, and a corresponding t upstairs on V^t.

Just as we generalized the covariant derivative of a covariant vector to tensors with covariant indices, going from equation (1) to equation (2), we can now generalize the covariant derivative to a tensor with any collection of lower and upper indices.

A lower index will entail an extra term with a Christoffel symbol with a minus sign, an upper index will entail an extra term with a Christoffel symbol with a plus sign.

Now we come to the idea of parallel transport.

We have already touched upon it in the previous lesson.

But let's now spell it out in detail.

Parallel transport

Suppose we have a curved surface – or a higher dimensional curved space – and some vector field defined on it.

That is, at every point of our space, there is attached a vector.

And, in what follows, the vectors of the vector field will always be in the tangent plane – or in the higher tangent flat space – to the space.

We are interested in knowing, when we move along a curve on the space, see figure 1, whether the field stays parallel to itself.

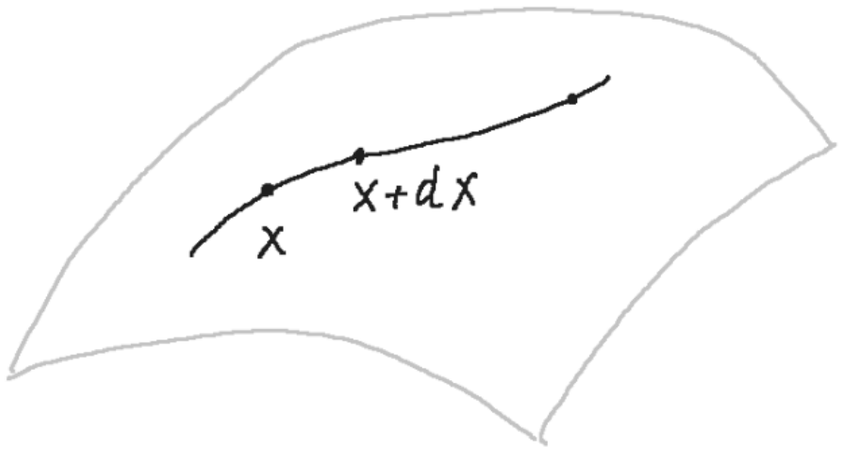


Figure 1 : Vector field and curve on a variety.

In figure 1 we have represented the space and the curve, but neither the vectors of the vector field nor the curvilinear coordinates on the surface.

At each point of the curve, imagine there is a vector.

Let’s move along the curve.

What we want to know is whether the field stays parallel to itself.

"Parallel to itself" between X and $X + dX$ on the curve *means* that the covariant derivative of the vector is 0 in the direction of the curve at that point X.

The covariant derivative is the difference between the vectors at $X + dX$ and at X, as they are written in best local coordinates, divided by the components of dX.

Let's write again the tensor which is the covariant derivative of a contravariant vector

$$D_m V^n = \frac{\partial V^n}{\partial X^m} + \Gamma^n_{mr} V^r \tag{9}$$

Now we want to consider the derivative along the trajectory or curve.

How does the vector change from point to point?

That simply corresponds to taking the covariant derivative $D_m V^n$ and multiplying it by dX^m.

Hence, the small change in the vector is

$$D_m V^n dX^m \tag{10}$$

This formula eliminates the fact that the coordinates themselves may evolve as we go from point to point.

That is the essence of covariant derivative.

Expression (10) is the small change in the vector *V* in going from one point to its neighbor, measured by the change of its components in a set of best coordinates, and then considered abstractly in any coordinate system.

Let's give it a name

$$DV^n = D_m V^n dX^m \tag{11}$$

It is the **covariant change** in the vector going from one point to a neighboring point on the trajectory.

Let's *express* this covariant change with the building blocks we have.

We *multiply* the right hand side of equation (9) by dX^m and get

$$DV^n = \frac{\partial V^n}{\partial X^m} dX^m + \Gamma^n_{mr} V^r dX^m \tag{12}$$

The *first* term on the right hand side has a simple interpretation.

It is the *ordinary* differential change in *V* disregarding anything related to a possible change in coordinates.

We *denote* it dV^n.

Equation (12) becomes

$$DV^n = dV^n + \Gamma^n_{mr} V^r dX^m \qquad (13)$$

It reads as follows : the covariant change in V is equal to the ordinary change in V plus a term equal to a Christoffel symbol multiplied by V^r and by dX^m.

This second term is of course a *double* sum according to summation convention.

Equation (13) is the formula which tells you how a vector changes from point to point.

Now suppose we are interested in finding a vector which is parallel to itself as we move along the curve.

Parallel to itself means that it doesn't change as we move from X to $X + dX$.

At each point X, we erect some best coordinates, and in those coordinates we test whether the vector is changing.

If it doesn't change in the first order – that is, its first derivative is zero –, we say : good, the vector is constant along the little segment.

We go to the next little segment, erect best coordinates at the new point, test again.

We do that along the whole curve.

It the tests say that the vector never changes in the first order, the vector is said to be parallel to itself along the curve.

There is a notion of calculus to remember: consider a function f continuous and differentiable over a segment (a, b).

Suppose $f(a) = 0$, and when we go from x to $x+\varepsilon$, $f(x+\varepsilon)-f(x) < \mathrm{O}(\varepsilon)$.

The notation $\mathrm{O}(\varepsilon)$ means something an order of magnitude smaller than ε.

That is, when ε goes to zero, $O(\varepsilon)/\varepsilon$ goes to zero too.

Then the function f is equal to zero along the whole segment (a, b).

In other words, if over any small displacement the variation of f is an order of magnitude smaller than the displacement, f' is always zero, and f stays constant.

In summary, if all along the curve the vector V satisfies

$$dV^n + \Gamma^n_{mr} V^r dX^m = 0 \tag{14}$$

then the vector maintains a relationship of being parallel to itself.

Taking a vector from one point and transporting it like this along a given curve, in such a way that it stays parallel to itself, is called **parallel transport**.

Making up a benign new word, we say that we "parallel-transport" the vector.

An *important* point about parallel transport on a curved space is that it is **trajectory-dependent**.

On the surface in figure 2, if we start at point A, take a vector V there, which lives in the tangent plane, and parallel-transport it to point B, the vector we end up with at B will depend on the path we followed from A to B:

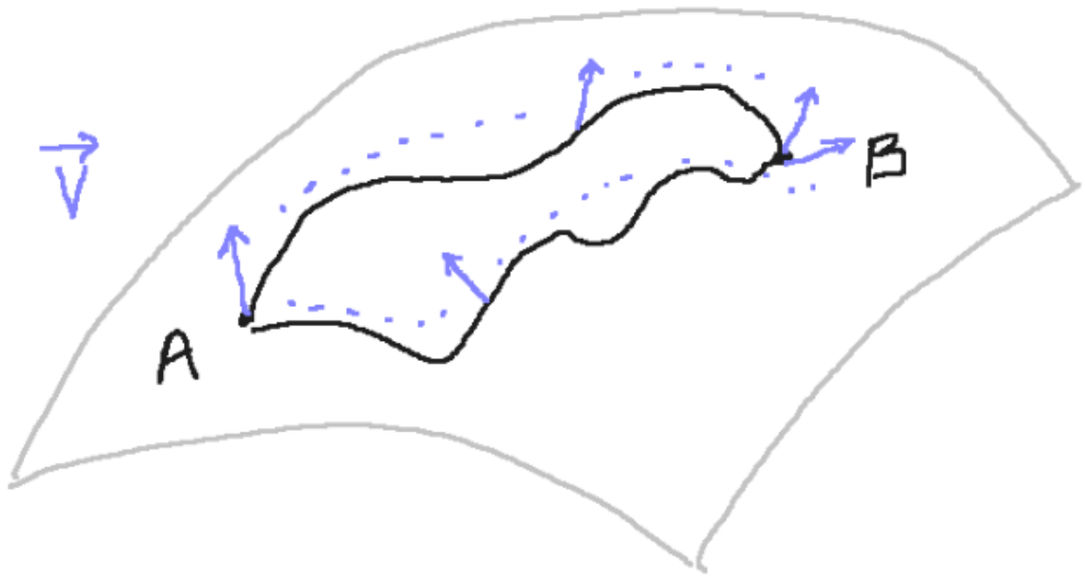


Figure 2 : Parallel-transporting V from A to B.

In figure 2, we represented the vector V (or $\vec{V}$) at A and suggested its evolution along two paths.

We did *not* represent any coordinate system.

Indeed, it is important to understand that parallel transport is *dependent* on the trajectory, but is *independent* of any coordinate system used to locate points on the surface.

At each point, anyway, we use a set of best local coordinates to do the infinitesimal parallel transport of the vector there.

When we arrive at B, the final vector we end up with depends not only on V or course, but also on the *path* we followed.

The final vector depends on the bumps and troughs we encountered along the path, that is on the *local* curvatures along the path.

Even if we came back to the same point A, depending on the loop we followed, we would end up with one or another vector.

If there *exists* a flat **connected** region – that is with no hole – and we follow a loop entirely in that region we will end up with the same vector V.

We already saw this phenomenon on the cone – pointy or rounded, it doesn't matter – in the previous lesson.

When we started with a vector on the side of the cone, and parallel-transported it around the cone we did *not* end up with the same vector.

An *alternative* path would be not to go around the top of the cone, in which case we would end up with the same vector – if we were careful to stay on the flat part.

So we see that *two* paths don't lead to same result, figure 3.

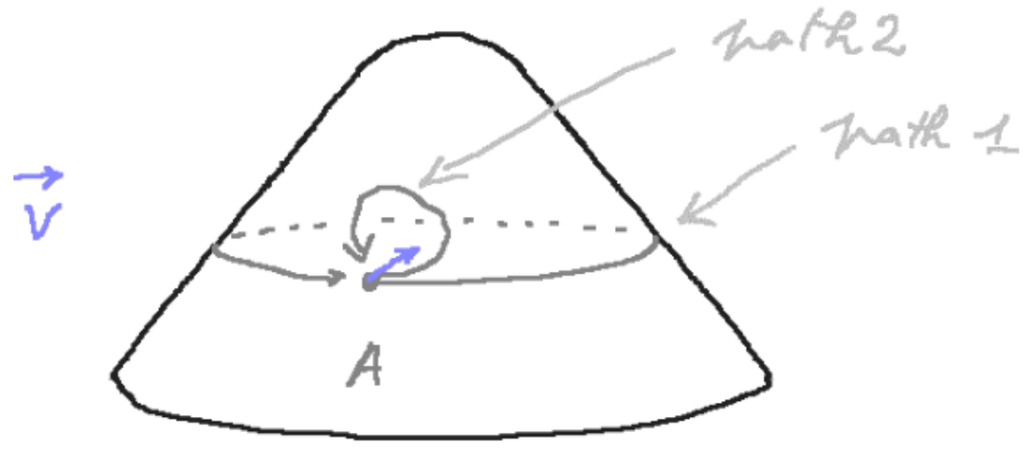


Figure 3 : Parallel-transport on a cone along two different paths.

Remember that the side of a cone is flat according to our definition, even though we see it embedded in 3D and in ordinary language it is not flat.

The side of a cone is intrinsically flat, because any section of it with no hole can be laid out on a plane without exerting any distortion on it.

More mathematically, any connected section of the side is flat *because* there exists a coordinate system the metric of which is the Kronecker delta tensor over the whole section.

Parallel-transporting a vector, that is moving it on the surface while making sure its covariant derivative remains null, also preserves its length.

It can be shown as a consequence of equation (14).

The *next* topic will concern tangent vectors to a curve, and *whether* the tangent vector stays constant or not.

When the tangent vector stays parallel to itself we will see that the curve is a *geodesic*.

For instance on the cone of figure 3, if we go from A to A around the cone, we don't follow a geodesic because obviously the tangent vector changes direction.

Even when we go from A to B as in figure 4, this is not a geodesic.

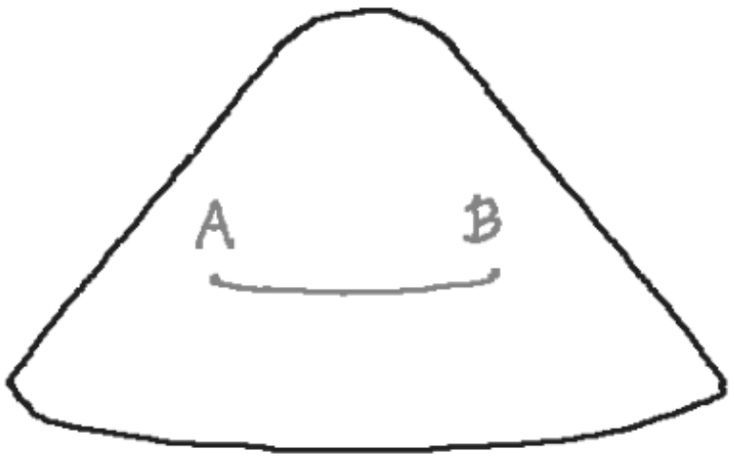


Figure 4 : Going from A to B on a cone.

On the Earth boats and airplanes *try* to follow geodesics for reasons we shall see.

When, while we are sitting on a plane, going from Paris to New York, the crew shows on a screen our trajectory, we are often *surprised* to discover that we do not follow "a straight line".

The *quotes* are because the notion of straight line on the Earth unless we are in a small region is, as we know, something to handle with care.

Figure 5 : On a plane from Paris to New York

Tangent vectors and geodesics

We now arrive at the notion of tangent vector to a curve and of geodesic.

On a surface where we consider two points A and B, a *geodesic* between A and B is a curve with certain properties.

One can *define* a geodesic in several ways :

1. The curve with the shortest distance between A and B is a geodesic.
2. A curve whose length is stationary when you wiggle it is a geodesic.
3. A third better definition looks at what happens locally along the curve. A curve which at each point is as straight as possible is a geodesic.

Of course this last definition is more intuitive than mathematical.

So let's make it more precise.

If at each point along the curve, the covariant derivative of the tangent vector is zero, that is if the tangent vector doesn't change, *then* the curve is as straight as possible.

When we talk about the tangent vector without further specification, it is of length one.

Let's try to build *more* intuition about it before turning to the mathematics.

Suppose you have a curved terrain, as in figure 6.

Figure 6 : Driving a car straight ahead, on a curved terrain.

For *convenience*, it is a two dimensional example, but there is *nothing* special about two-dimensional spaces in defining the notion of geodesic.

Secondly suppose we have a car that we drive on this terrain.

And *assume* that the size of a car, in particular the distance between the front wheels, is small by comparison with any curvature.

In other words, the car is *very small* compared to the hills and the valleys.

We start from A in some direction, and driving *straight* in the above sense – *never* turning the steering wheel –, we end up at B.

Our trajectory will wind between the hills – or we may even start from the top of a hill, that is, from a point with clear curvature.

The curve that we will execute with our car in the space, keeping the steering wheel straight, *will* nevertheless be as straight as possible.

It will be a *geodesic* in the space.

Another way to *characterize* a geodesic is to say that the tangent vector along the curve is constant.

We have an *intuitive* perception of what the tangent vector is.

But let's define it more precisely.

Consider a curve, and a point at coordinate X on it.

And take a *neighboring* point, figure 7.

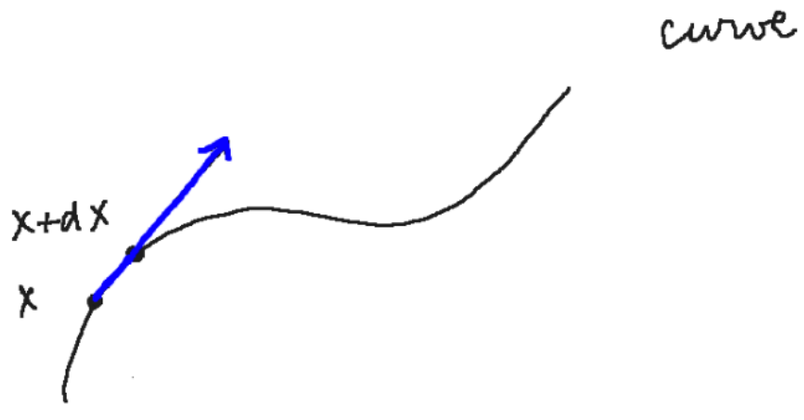


Figure 7 : Construction of the tangent vector at a point.

The points X and $X + dX$ are separated by dX which we can also denote, in tensor style, dX^{m}.

And *consider* a vector starting a X, going through $X + dX$, and of length one.

Then take the limit when the second point $X + dX$ approaches the first point X.

The resulting vector is called the *tangent* vector at X.

Consider the distance dS between the two points X and $X + dX$, as you remember, it is defined by

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

The way we *construct* the tangent vector in the X coordinate system is very simple.

The m-th component of the tangent vector is

$$t^m = \frac{dX^m}{dS} \tag{16}$$

It can be *proved* that equation (16) produces a vector of length one.

There is *one* such vector at each point along the curve.

It is called the *tangent* vector.

It *points* in the direction between two neighboring points and its length is one.

Let's now turn our *attention* to curves where the tangent vector is constant?

If we *plug-in* the tangent vector in equation (14), these curves satisfy the following equation :

$$dt^n + \Gamma^n_{mr} t^r dX^m = 0 \qquad (17)$$

This equation *holds* because once you have set your steering wheel dead ahead, you are moving in as straight a line as you can.

So the covariant change of the tangent vector is *zero*.

See the example below to *build* your intuition.

Example of calculations with Christoffel symbols

In order to *build* our intuition about geodesics, particularly where a surface has curvature – like a rounded hill – let's see what an example looks like.

Consider a point *P* on the surface of a sphere.

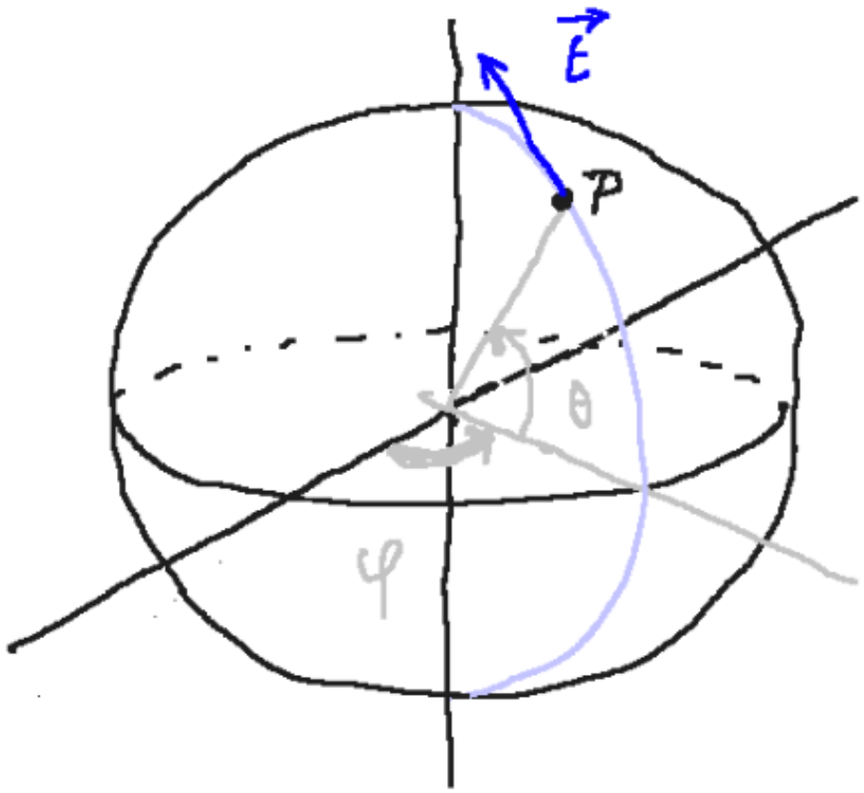


Figure 8 : 2-sphere with polar coordinates.

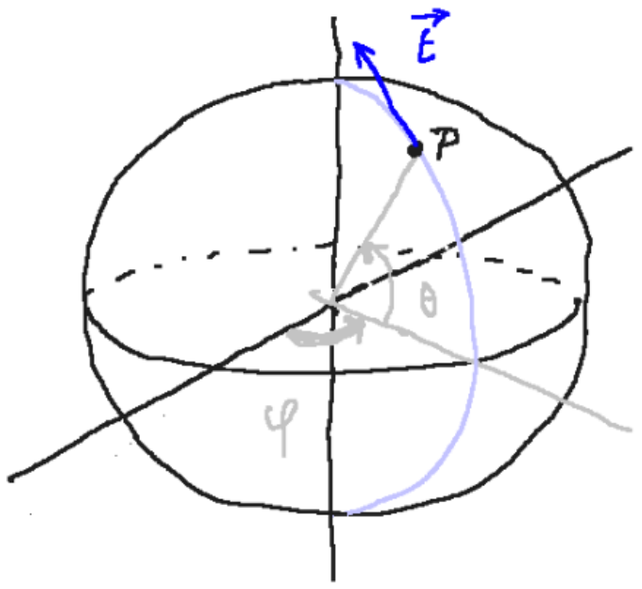
Mathematicians call such a surface a 2-sphere, *because* its points are located with two coordinates.

Let's *use* the ordinary latitude θ and longitude i, and the ordinary distance we are familiar with, for instance on the Earth.

The objective of the exercise is to show that a meridian is a geodesic.

In other words, when we follow a meridian the tangent vector **doesn't change**.

We are on a 2-sphere with polar coordinates θ and ϕ, see figure 8.



(1) The *metric tensor* of the ordinary distance is

$$\begin{pmatrix} 1 & 0 \\ 0 & cos^2\theta \end{pmatrix}$$

(2) There are eight *Christoffel* symbols using this metric. Six are *zero* and the other two are

$$\Gamma^1_{22} = sin\theta\ cos\theta$$
$$\Gamma^2_{12} = \Gamma^2_{21} = -tan\theta$$

(3) We can then *show* that the tangent vector to a meridian has everywhere components $t^1 = 1$ and $t^2 = 0$.

(4) The tensor which is the *covariant* derivative of this tangent vector is

$$\begin{pmatrix} 0 & 0 \\ 0 & -tan\theta \end{pmatrix}$$

(5) Finally if we *follow* a meridian, the covariant change of the tangent vector is always zero.

If we *carry* out all the steps in detail, then it will show you that doing actual calculations with Christoffel symbols, even on a simple example, quickly fills pages.

It will also show you that *even* on a surface with curvature there are paths where the tangent vector doesn't change.

These are the *geodesics*.

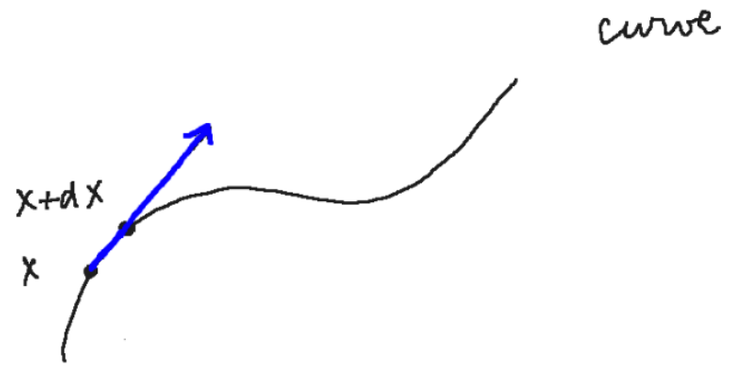
Above we looked at a meridian, because the polar coordinates make it simple to study, but of course by symmetry *any* great circle is a geodesic.

We might feel, in figure 8, that the tangent vector changes when we move along a meridian, but it is *because* we look at the 2-sphere embedded in 3D Euclidean space.

If we turned our steering wheel, however, and in the tangent plane swerved from our straight path, then the tangent vector of our trajectory would change.

More on geodesics

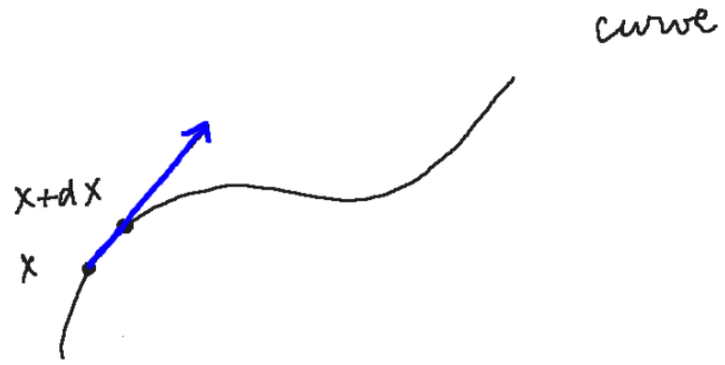
We can write equation (17) of a geodesic in a slightly neater form.



Let's *divide* both sides of the equation by dS, that is, by the little distance between two neighboring points with coordinates X and $X + dX$, see figure 7.

Equation (17) becomes

$$\frac{dt^n}{dS} = -\Gamma^n_{mr}\ t^r\ \frac{dX^m}{dS}$$



But $\frac{dX^m}{dS}$ is t^m, so we can *rewrite* equation (17) as

$$\frac{dt^n}{dS} = -\Gamma^n_{mr}\ t^r\ t^m \qquad (18)$$

This equation *only involves* the tangent vector.

It also of course involves the Christoffel symbols, but let's *suppose* we are given them.

Then equation (18) is the equation of motion on a geodesic.

The gammas are made up out of the metric.

So if we know the metric, we know what to put on the right hand side.

And we can say one more thing: since the tangent vector *t* itself is a derivative, we can write the left hand side as a *second* derivative.

$$\frac{d^2X^n}{dS^2} = -\Gamma^n_{mr}\ t^r\ t^m \qquad (19)$$

Does this look like anything *familiar*?

Maybe not.

But if we were to think of S as some measure of time as we moved along the curve, then the second derivative of position on the left would be *acceleration*.

So if *S* were like time, or if *S* were increasing uniformly with time, then we would have the following fact : an acceleration is equal to something that depends on the metric and on *t*.

We will deal with $t^r\ t^m$ *later*.

Let's observe for the time being, that equation (19) has the look of a Newton equation: *acceleration* is equal to something that depends on the gravitational field, because, as we will see, the *metric* is the gravitational field.

We will see that equation (19) *replaces* Newton's equation for the motion of a particle in a gravitational field.

In other words, in some sense a particle in a gravitational field *moves* along the *straightest* possible trajectory.

But it moves along the straightest possible trajectory **through space-time** not just through space.

Space-time

So now we have to *come* to space-time.

So far we have *been* studying the mathematics of curved spaces as Riemann would have understood them.

The German mathematician Bernhard Riemann (1826-1866) is the one who *invented* most of the mathematics for curved spaces.

But these were, so to speak, *ordinary* curved spaces in which distance was governed by the Pythagorean theorem.

In Riemann spaces, the square of the distance is *always* positive.

A Minkowski space is a *space-time* which *also* has a natural measure of distance along curves or between points – also called events in the space-time –, be they neighboring points or distant points.

As in Riemannian geometry, in Minkowski geometry the distance is *generally* expressed through its square.

But in Minkowski space this square can be *negative*.

What is the *name* of the distance in a Minkowski space-time ?

Answer : **the proper time**.

It *applies* to any pair of events, far away, close, or infinitesimally close.

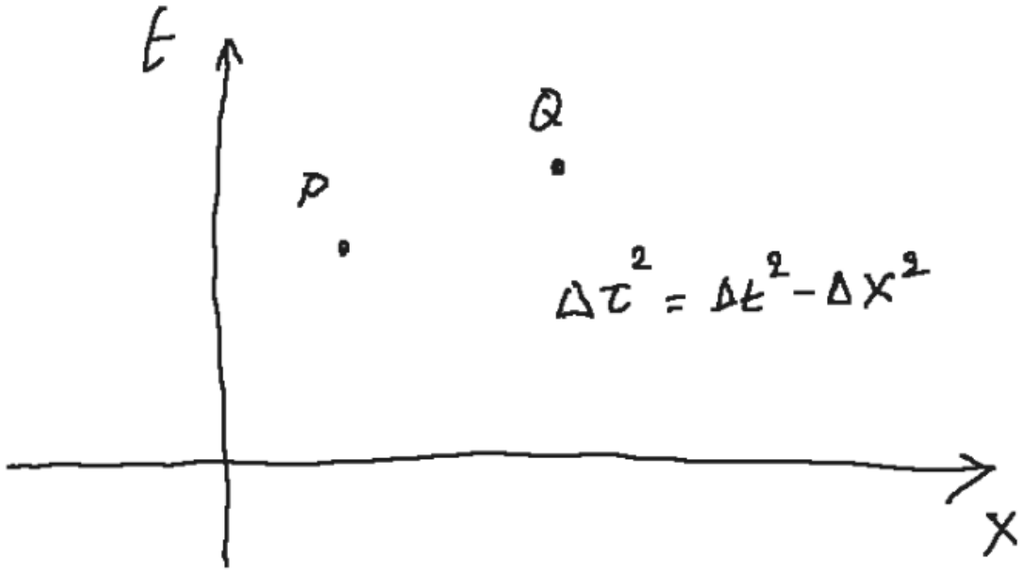


Figure 9 : Proper time between two events *P* and *Q*.

Given two points P and Q in space-time, the space-time distance between them – and now let's not call it the space- time distance, let's call it the proper time – is equal to Δt^2 not plus ΔX^2, but **minus** ΔX^2.

$$\Delta\tau^2 = \Delta t^2 - \Delta X^2$$

When P and Q are infinitesimally close, it becomes

$$d\tau^2 = dt^2 - dX^2 \tag{20}$$

And X may stand for three spatial coordinates (x, y, z).

With a Greek index it may even stands for the four coordinates, see X^μ below.

It is conventional to rewrite equation (20) as

$$dt^2 - dX^2 = -dS^2$$

where S is called the *proper* distance.

Some authors *prefer* to work with τ, some prefer to work with S.

In these lessons, we *will* mostly use the proper time τ.

Another convention it to write equation (20) as

$$dS^2 = g_{\mu\nu}\, dX^\mu\, dX^\nu \tag{21a}$$

or

$$d\tau^2 = -g_{\mu\nu}\, dX^\mu\, dX^\nu \tag{21b}$$

where anyone familiar with special relativity is familiar with the notation X^μ with a Greek index :

$$X^\mu = \begin{pmatrix} t \\ x \\ y \\ z \end{pmatrix}$$

According to standard convention this is also sometimes noted

$$X^\mu = \begin{pmatrix} X^0 \\ X^1 \\ X^2 \\ X^3 \end{pmatrix}$$

where the Greek index μ runs over 0 to 3.

When we use a Latin index, we mean only the three spatial coordinates, that is, if you read X^i, this means that i runs over 1, 2 and 3, or in other words X^i runs only over the spatial coordinates.

Let's comment on equation (21a). $$dS^2 = g_{\mu\nu}\, dX^\mu\, dX^\nu$$

The indices μ and ν run over 0 to 3.

The equation has exactly the *same* form as the usual equation for the distance in Riemannian geometry that we have already often used, see equation (1) of lesson 3 for instance.

The only *new* thing in the Minkowski geometry is the *metric* tensor $g_{\mu\nu}$ or the corresponding matrix.

It is still *diagonal*, but it has a *minus* 1 corresponding to the time axis, and *three* +1 for the space axes.

Since it plays a *central* role, it has a name.

We use the Greek letter η (written and pronounced "eta").

It is called $\eta_{\mu\nu}$ (pronounced "eta mu nu").

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{22}$$

With this metric tensor, we can check that equation (21b) expressing the proper time is the same equation as

$$d\tau^2 = dt^2 - dx^2 - dy^2 - dz^2$$

What do all the *zeros* mean in the metric tensor?

They mean that there are *no* cross terms like $dt\,dx$, or $dy\,dz$ in the definition of the proper time.

The coordinates (t, x, y, z) are coordinates specific to *Minkowski* geometry.

They are the *analog* of Cartesian coordinates in Euclidean geometry.

In special relativity, that is *all* there is to the coordinates and the metric tensor.

In general relativity, the metric tensor *becomes* a function of space and time.

We then call it $g_{\mu\nu}(X)$ (where X *clearly* stands for an event with four coordinates).

And equation (21b) becomes

$$d\tau^2 = -g_{\mu\nu}(X)\ dX^\mu\ dX^\nu \tag{23}$$

There is one more *important* thing, to mention at the outset, on the metric tensor in *relativity*.

What is the *difference* between the matrix of equation (22) and the identity matrix of Euclidean metric?

Well, it has a *minus* 1 in first position.

But more *importantly* there is an invariant concept about $g_{\mu\nu}(X)$: it has one negative eigenvalue and three positive eigenvalues.

And this will *always* be the case in general relativity.

We are not going to spend much time dealing with the mathematical fact.

The equations will always automatically take care of all of that.

But what does it mean that there is one negative eigenvalue and three positive?

It *means* there is one dimension of time and three dimensions of space.

We *could* write a metric with two minus signs on the diagonal.

It would *correspond* to a crazy space with two time directions and two space directions.

That is *not* allowed in relativity.

At the *beginning* of his work on general relativity, Einstein realized that the metric of space-time should always have one negative and three positive eigenvalues – *locally* at any point in space-time.

But, as said, we *won't* have to worry about it.

It will be *taken care* of by the equations themselves.

Other than that, all that we have done in Riemannian geometry, all the equations involving metrics, covariant derivatives, curvature, geodesics, etc. *will be exactly the same* in the space-time geometry of general relativity.

What does flat *mean* in space-time ?

Flat *no longer means* that there is a coordinate system in which the metric is the Kronecker delta.

Now it means that there is a coordinate system in which the metric has the form of $\eta_{\mu\nu}$, that is, the Kronecker delta *except* with one negative sign.

Remember that in Riemannian geometry, the condition for flatness was not that the metric we were dealing with was the unit matrix, *but* that there existed coordinates – other coordinates that we could find – in which it would be the unit matrix.

Similarly, in Minkowski geometry, the condition for flatness in space-time is *not* that the metric be $\eta_{\mu\nu}$, it is that there exist coordinates in which the metric *can be brought* to the very simple form of $\eta_{\mu\nu}$.

Mathematically speaking this condition is equivalent to saying that the metric must have one negative eigenvalue and three positive eigenvalues.

That is the *notion* of a flat space-time.

And if the space-time *cannot* be brought to the form of $\eta_{\mu\nu}$, then it is *curved*.

How do we *check* whether the space-time is curved?

We do all *exactly* the same things that we did in Riemannian geometry.

Here is a *recap* of the analogies that we have already made so far, as well as those we shall see :

Flat spaces
- Euclidean geometry → Minkowski geometry
- Kronecker δ tensor → η tensor
- Newton physics → special relativity

Non flat spaces (always locally flat)
- Curved metric → gravitational field
- Riemannian geometry → Einstein general relativity

Before going into spaces whose curvature is due to real gravitational fields (i.e. to the presence of massive bodies), we shall spend some time with "flat" spaces with Minkowski geometry.

Standard terminology, however, *doesn't* call them "flat", it simply *calls* them "spaces endowed with the Minkowski metric".

So for a *while* we are going to do special relativity.

We will *deal* with a space with the $\eta_{\mu\nu}$ metric.

And we are going to wind up looking at it in polar coordinates – not ordinary polar coordinates but *hyperbolic* polar coordinates.

The name is awe-inspiring, but the concept is *simple and well adapted* to space-time and particles moving in it, particularly particles accelerating in it.

Since we *know* from lesson one that there is a link between gravity and acceleration, and our ultimate *objective* is to describe relativistic motion of particles in gravitational fields, it is natural to *start* with studying particles accelerating in the framework of special relativity.

Special relativity

We are in the *space-time* of special relativity, which we can also call a Minkowski space.

Its metric is *defined* by the tensor

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{24}$$

Our *goal* is to define the notion of a uniformly accelerated reference frame.

We have *already* met it in lesson one when we illustrated the principle of equivalence with a the gravitational field of the Earth (in a very small region where it can be viewed as uniform) and the apparent field we experience in an elevator being uniformly accelerated.

But in lesson one, we *stayed* at an elementary somewhat casual level.

In fact, in special relativity there is a *difficulty* with the notion of uniformly accelerated reference frame.

Consider a bunch of points, separated by a fixed distance, as in figure 10.

Figure 10 : Points in space with a fixed separation.

We try to move them "uniformly".

We think of them as forming a *frame*.

Suppose we *start* accelerating them along the X-axis, and let each have the *same* constant acceleration.

We would think that they would *maintain* the same distance between them – that they would keep forming a frame.

But distances, time and simultaneity *do odd things* when we start moving sizable objects.

In fact *if* we gave all the points the same trajectory, in particular the same acceleration, we would *discover* that, in the *rest frame* of the first point for instance, they would *neither* stay fixed, *nor* evenly distributed.

The *distance* between them would grow larger and larger.

That *means* for example, if there were strings between them, that as they started moving, in an attempt to keep their distance uniform, the strings would *stretch*, and eventually *break*.

That is *not* what we would think of a uniformly accelerated reference frame, as we are *accustomed* to from non-relativistic physics.

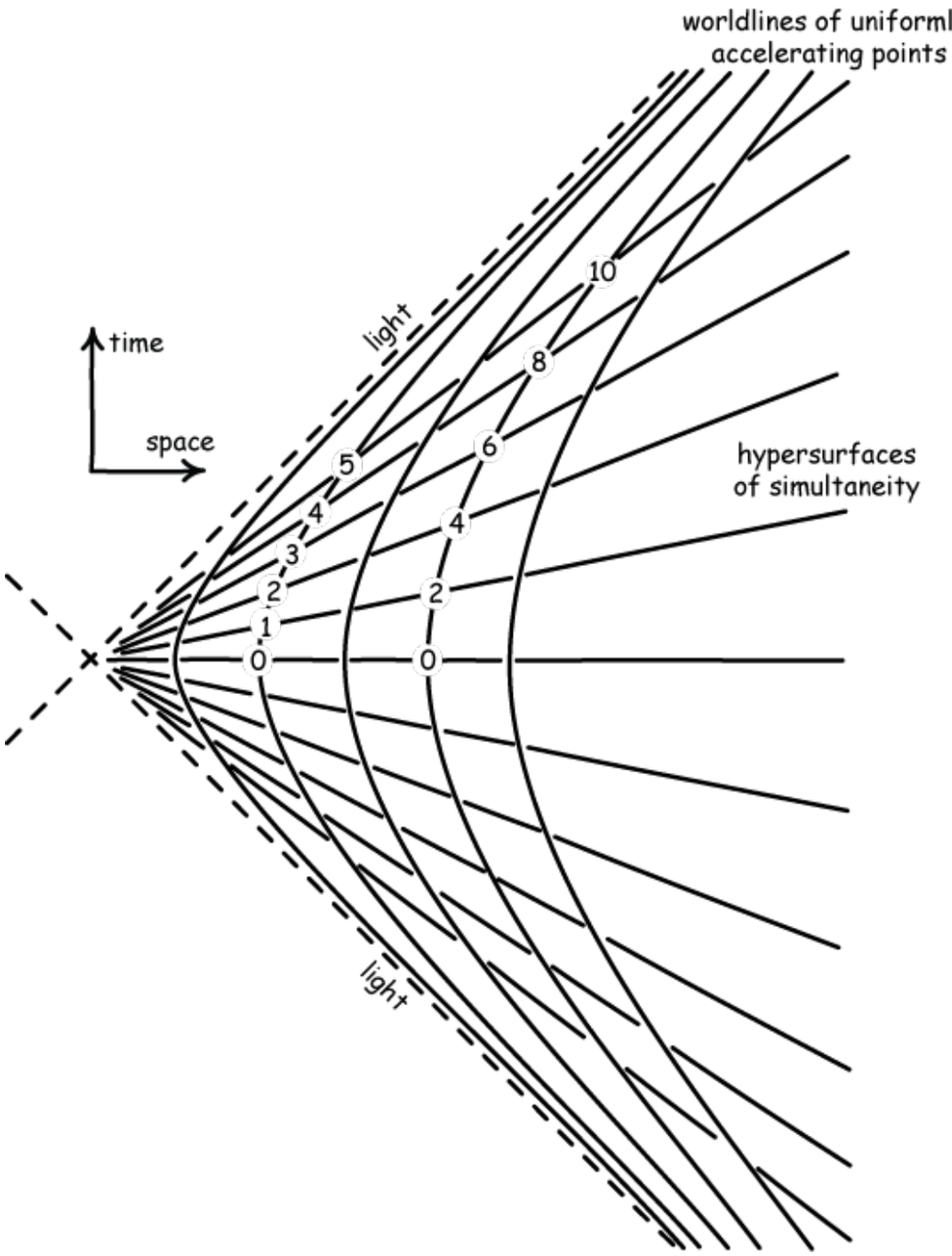
What is *nice*, in *non-relativistic* physics, about a uniformly accelerated reference frame, is that it *keeps* the same structure, the same shape.

The distances between points stay the *same*.

If you had strings connecting the points they *wouldn't* get stretched.

There is a *second* difficulty about the simplistic idea of uniformly accelerating a particle in special relativity: if we *waited* long enough, in the stationary frame, the particle would eventually *exceed* the speed of light.

So uniform acceleration, to the extent that it exists and makes good physical sense, is *not as simple* as just moving the points in figure 10 all with the same accelerated trajectory (see figure below).



We will show this result in the following discussion.

So we are going to *construct* what a relativist would call a uniformly accelerated reference frame.

But to do so, we need to go back one step, to Euclidean space and talk about polar coordinates, fig. 11, because – *surprisingly enough* – the uniformly accelerated coordinate system is the *analog* of polar coordinates.

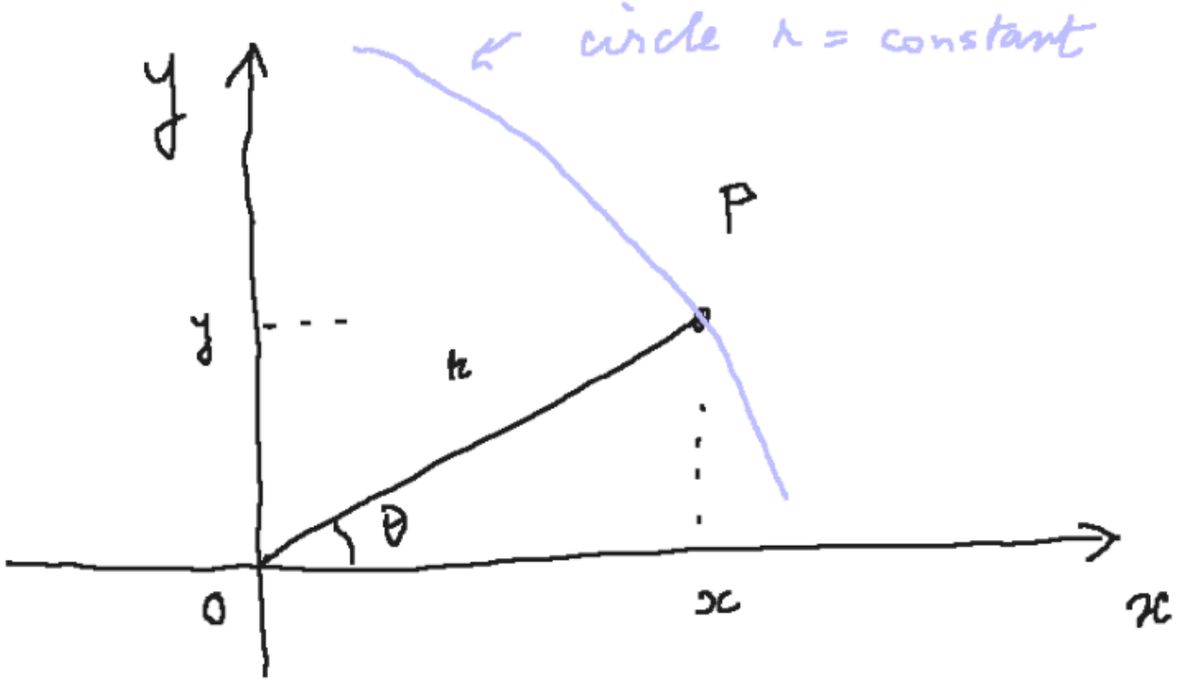


Figure 11 : Euclidean and polar coordinates in the plane.

Here are some *equations*, expressing the coordinate transformation from polar to Cartesian coordinates, which the reader should be familiar with

$$\begin{aligned} x &= r\cos\theta \\ y &= r\sin\theta \end{aligned} \tag{25}$$

We also have

$$\cos^2\theta + \sin^2\theta = 1 \tag{26}$$

which is the same as saying that

$$x^2 + y^2 = r^2 \tag{27}$$

Finally there are two more equations to remember

$$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2}$$

$$\sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i} \tag{28}$$

Equations (25) to (28) are the basic equations governing ordinary polar coordinates.

What is the equation of a circle around the origin?

It is just

$$r = \text{constant}$$

Imagine a point moving around the circle with uniform velocity, *therefore* uniform angular velocity.

Then the magnitude of the acceleration of that point is constant around the circle.

The vector acceleration constantly *points* toward the center of the circle.

What does it have to do with relativity?

In relativity we write *basically* the same equations to define a uniformly accelerated point.

We turn to the basic representation of space-time, which is the *analog* in special relativity of figure 11.

In figure 12, we see the *light cone* : the two diagonal straight lines in grey.

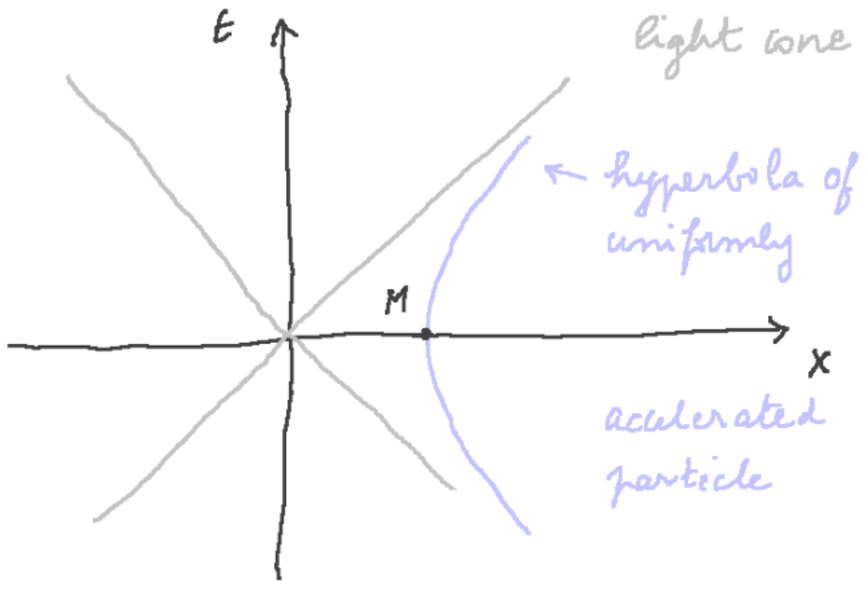


Figure 12 : Light cone

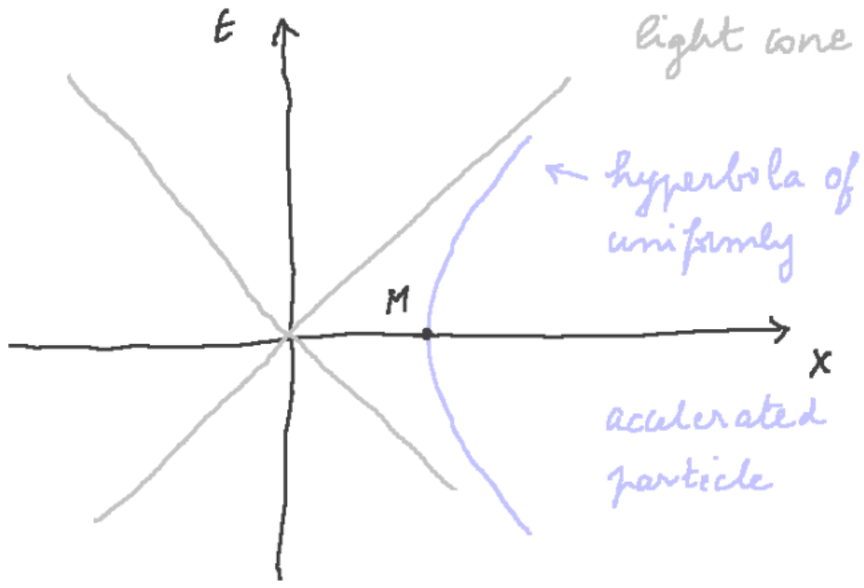
(it would be a cone if we had two spatial coordinates).

Having studied special relativity, we know that they *represent* the trajectory of a light ray starting at time 0 from the origin and going either to the right or to the left.

Let's see now what is the *analog* of the uniformly accelerated point on the circle of figure 11.

In other words, what is a *uniformly accelerated point* in special relativity?

It is defined as a point moving on a *hyperbola* as in figure 12.



The point is *clearly* accelerated.

It is *not* moving with constant velocity.

Constant velocity would correspond to a *straight* line with some slope higher the 45°.

In fig. 12, we see that from the past until time 0, the point, or particle, *moves*, spatially on the X-axis, to the left to a minimum point M.

At M its velocity is *zero*.

Therefore, in the (X, t) diagram, at M the tangent to the trajectory is *vertical*.

Then, *after* point M, the particle is moving again to the right.

In the (X, t) diagram, as the point moves up and up, the tangent to the trajectory *gets* closer and closer to 45°, that is the particle *gets* closer and closer to the speed of light, but *never* exceeds it.

Next step is the following mathematical question: what is for a hyperbola corresponding to equation for a circle in ordinary polar coordinates?

Answer : there *exists* a system of coordinates in which the hyperbola of fig. 12 has a very simple equation akin to equation (25) for the circle in polar coordinates.

It uses *hyperbolic* polar coordinates.

In equation (25), expressing the transformation of coordinates between Cartesian and ordinary polar, we shall *replace* the sine and cosine functions by their hyperbolic version.

$$x = r\cos\theta$$
$$y = r\sin\theta$$

Let's also replace the angle θ by a *parameter* ω.

The *correspondence* is

$$\cos\theta \to \cosh\omega$$
$$\sin\theta \to \sinh\omega$$

The mathematical definitions of the hyperbolic sine and cosine functions are very similar to those of ordinary sine and cosine.

But there is no more $i = \sqrt{-1}$ coefficient.

$$\begin{aligned} \cosh\omega &= \frac{e^{\omega} + e^{-\omega}}{2} \\ \sinh\omega &= \frac{e^{\omega} - e^{-\omega}}{2} \end{aligned} \tag{29}$$

Analogously to equation (26), we have that

$$\cosh^2\omega - \sinh^2\omega = 1 \tag{30}$$

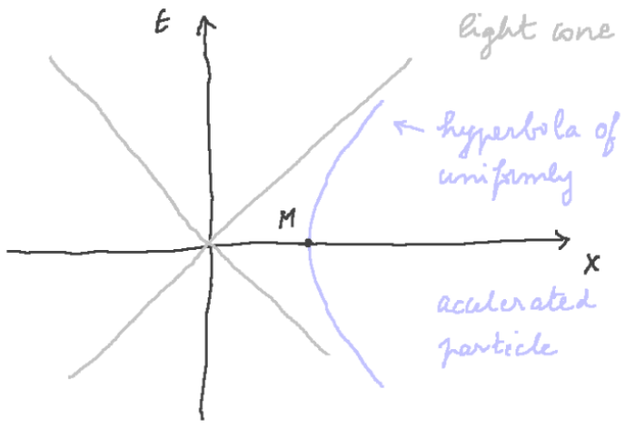
The coordinates of a point P in the (X, t) diagram are now

$$\begin{aligned} X &= r\cosh\omega \\ t &= r\sinh\omega \end{aligned} \tag{31}$$

Equations (31) *define* r and ω from X and t.

The parameter ω is *not* a geometric angle.

But *when* we move along a hyperbola with the light-ray trajectories as asymptotes, see figure 12, it is what increases from - ∞ to + ∞ – just like θ was the parameter that changed as we moved along a circle centered at the origin.



And on such a hyperbola, *r doesn't* change.

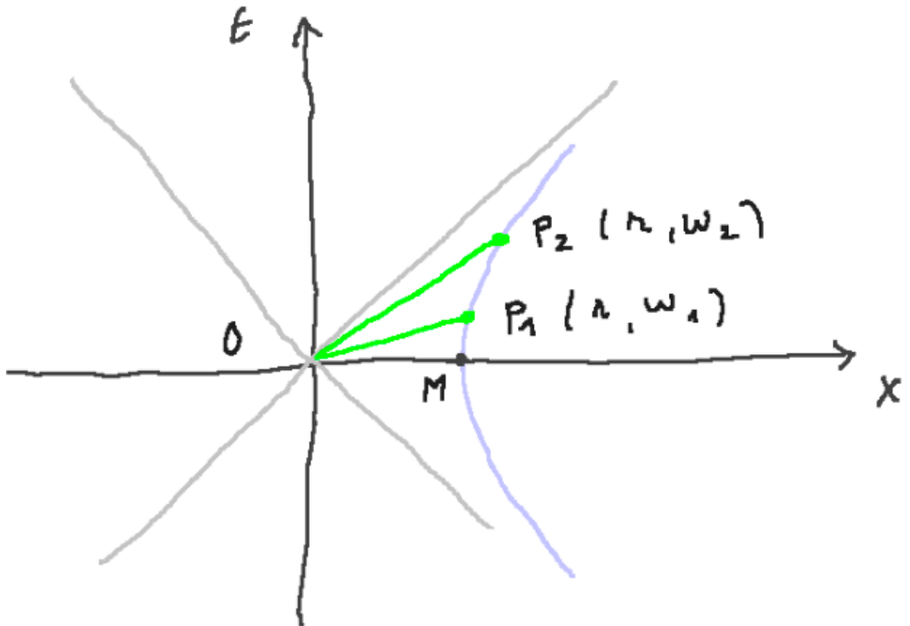
The parameter ω *plays* on the hyperbola the role of the angle on the circle.

It is sometimes *called* the hyperbolic angle.

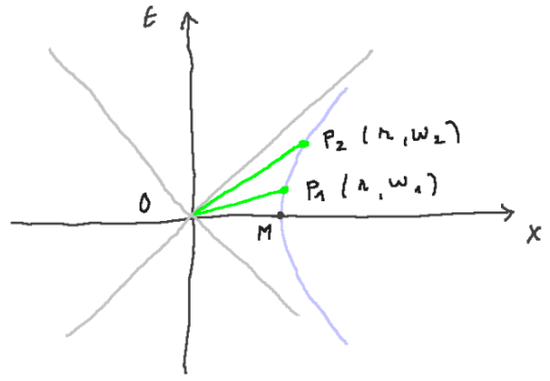
As before, equations (31) express nothing more than a coordinate transformation between the Minkowski coordinates (X, t) and the hyperbolic coordinates (r, ω).

Any point P in space-time can be *located* by its Minkowski coordinates (X, t) or by its hyperbolic coordinates (r, ω).

Figure 13 : Hyperbola in hyperbolic coordinates.



In figure 13, all the points on the hyperbola have the *same* hyperbolic radius r.



Its *value* is the distance between O and M.

And it is a *characteristic* of the curve.

On the other hand, the hyperbolic angle ω *increases* up to infinity as we move on the hyperbola closer and closer to its asymptote, that is, as the particle *moves* spatially farther and farther away on the X-axis.

So we have in Minkowski geometry the *analog* to the circle in Euclidean geometry: the hyperbola, as in figures 12 or 13, in hyperbolic polar coordinates given by equations (32), corresponds to a constant value r, and the parameter ω going from $-\infty$ to $+\infty$.

This will be *handy* to study a uniformly accelerated particle because by definition it moves along such a trajectory.

Uniform acceleration

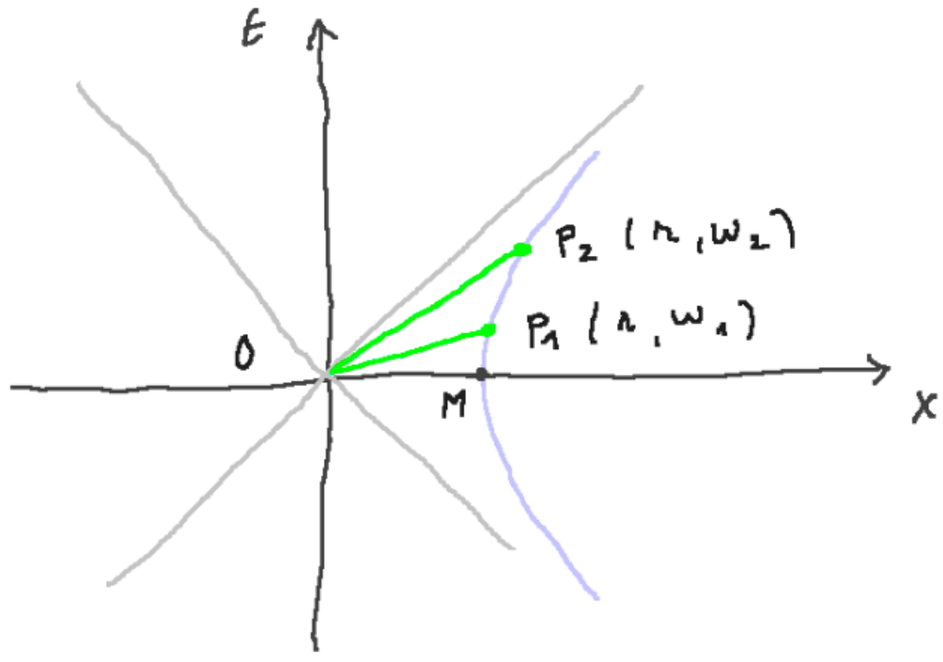
Now that we have a good understanding of what r and ω are in the hyperbolic coordinate system, *which* do you think is like time ?

On the X-axis, $\cosh \omega = 1$.

So if we move to the right on this axis, we just *increase* r.

Therefore r is like a *space* coordinate.

If from point M we travel *upward* on the hyperbola of figure 13, r stays *fixed* and we *increase* ω.



Going upward is the *analogy* with traveling around the circle in figure 11.

Therefore ω is *like* a time coordinate.

Just as there was a uniformity to the circle – at any point you could define the radius r and it was constant –, there is an *analog* uniformity to the hyperbola : the hyperbolic radius r is *constant* on a hyperbola.

In figure 14, we can see hyperbolas for *different* values of r.

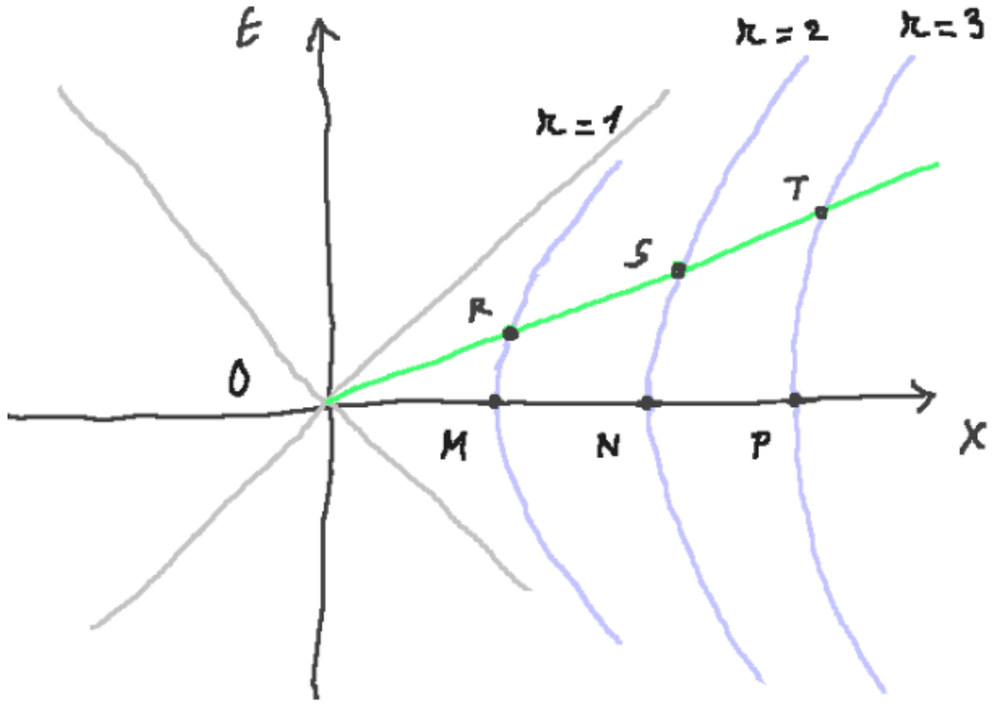


Figure 14 : Hyperbolas for different values of r.

The *analog* of equation (27) on a circle becomes on a hyperbola

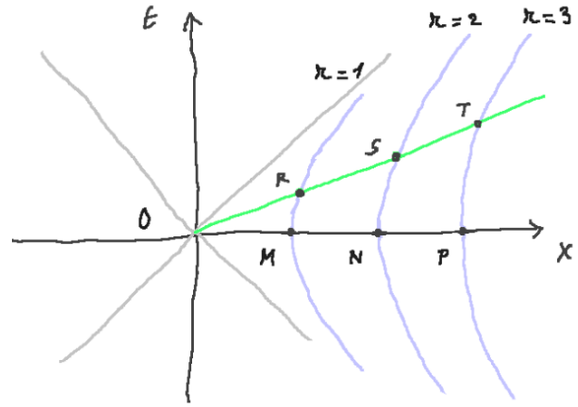
$$X^2 - t^2 = r^2 \tag{32}$$

The hyperbolic polar coordinates are just new coordinates for *every* point in the standard Minkowski diagram (t, X) of figure 14.

Those coordinates in special relativity are the *closest* thing that exists to uniformly accelerated coordinates.

For a given ω the distances between the points corresponding to $r = 1, r = 2, r = 3$ etc. *stay* the same.

In other words, in figure 14, *not only* $MN = NP$ but also $MN = NP = RS = ST$.



That can be *checked* with with your understanding of special relativity.

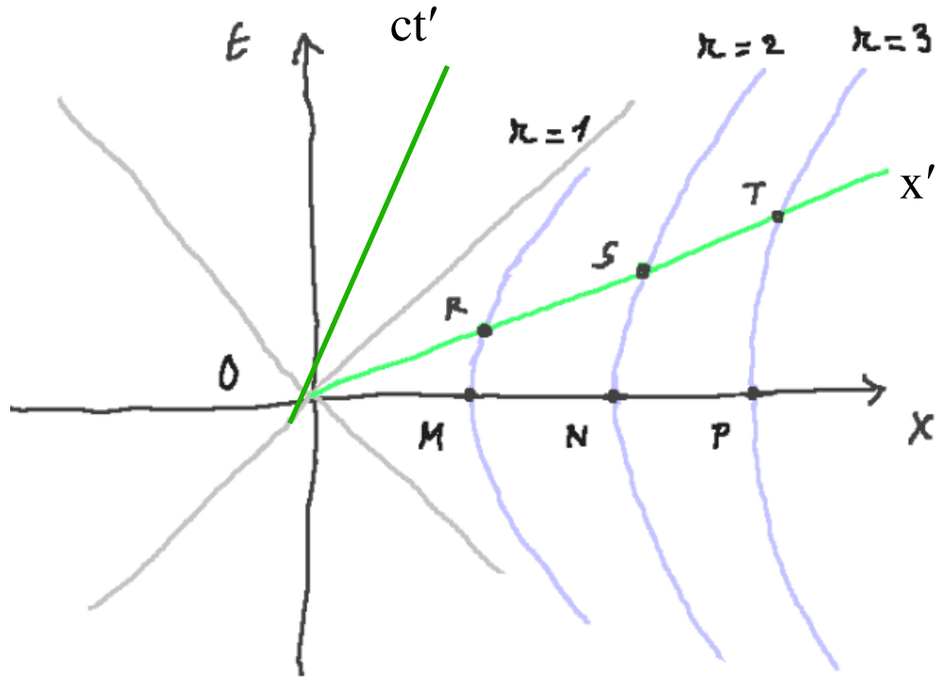
That means that as the Lorentz frame of reference *accelerates* the distance between neighboring particles just stay the *same*.

Notice also that points O, M, N and P are *simultaneous*, with $t = 0$.

And points O, R, S and T are also *simultaneous for another observer* which is moving with a certain speed relative to O.

In figure 14, we can *see* what the speed is, relative to the stationary frame, of the observer who sees R, S and T simultaneous ?

What is unusual here, and different than an ordinary accelerated frame of reference, is that the accelerations along the different trajectories corresponding to $r = 1, r = 2, r = 3,$ etc. are different.



speed is given by slope of ct′ axis

Let's look at a trajectory in the same collection as the hyperbolas of fig 14, but with a very *small r*, figure 15.

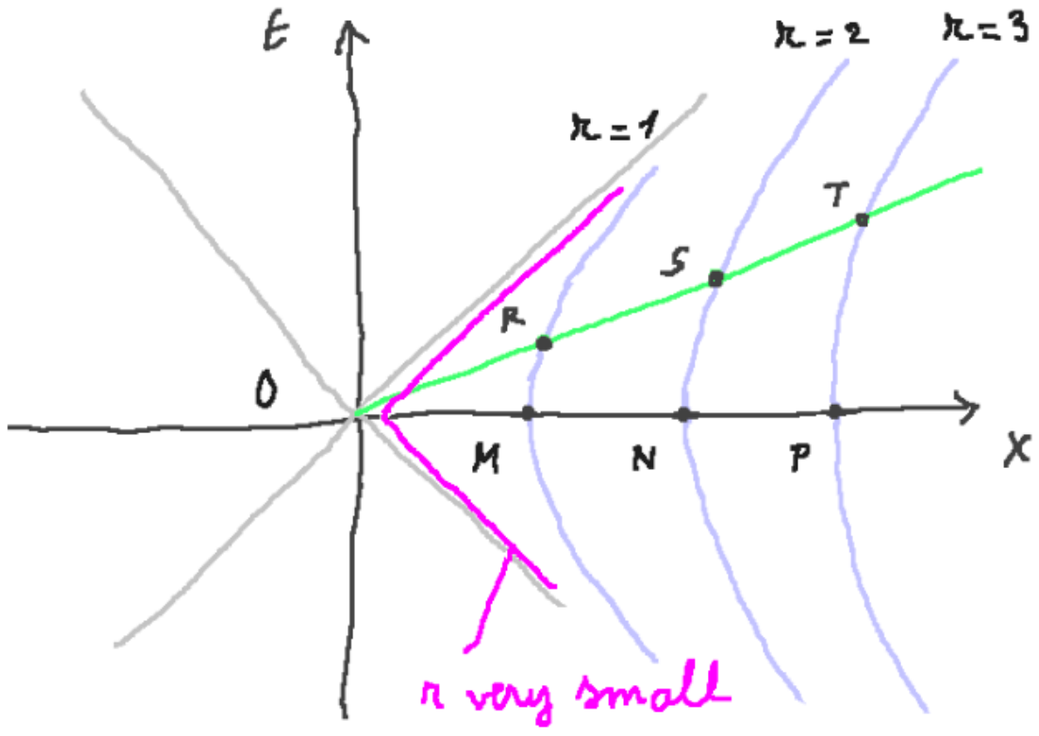


Figure 15 : Hyperbola with a very high acceleration.

On the hyperbola with a very *small r*, the particle makes a sudden change of direction when it comes close to the origin, and then speeds off to the right again.

That indicates that its trajectory has a *much larger acceleration*.

Another important fact : on a given hyperbola, the proper acceleration, that is, the actual acceleration – the actual push – that an observer would feel, is *uniform* along the trajectory.

This is the analog of the fact, on a given circle in fig 11, that at any point on the circle an observer feels the same acceleration in magnitude toward the center.

On the other hand, the accelerations on two different hyperbolas are *different*.

Again that is the *analog* of the fact that on a circle with a larger radius the centripetal acceleration is smaller.

Conversely the tinier the radius the bigger the acceleration.

The message to *remember* is this : if you want to define uniform acceleration, you *ought* to have acceleration which is constant in time.

That is okay.

But then you are *necessarily* going to have *different* accelerations at *different* points of space, that is, on trajectories with *different r*'s.

The proper acceleration of a particle with coordinates (t, X) - here X stands for (X^1, X^2, X^3) or equivalently (x, y, z) - is defined by $d^2X^n/d\tau^2$.

It is the *rate of change* of the proper velocity, with respect to proper time.

The *proper* velocity being itself $dX^n/d\tau$.

We have seen *many* times now the definition of proper time.

Remember that it *has* a physical meaning.

It is the *time recorded and displayed* by a watch attached on the wrist of an observer traveling in space-time.

Figure 10 : Points in space with a fixed separation. We try to move them “uniformly".

So, if we consider the collection of particles of figure 10, fixed with respect to each other, the further out the particle is in space, that is, the further out to the right on the X-axis, the smaller its acceleration is.

In other words, the *further* to the right, the *straighter* is its trajectory.

In fact we can write an *equation* for the proper acceleration.

The proper acceleration, A, of a particle on one of the trajectories in figure 15 *depends* only on r.

Let’s *choose* one of them, and call it R.

And let's work in units where the speed of light c is equal to 1.

Then along the hyperbola with parameter R, the proper acceleration is *given* by

$$A = \frac{1}{R} \tag{33}$$

The *bigger* the R, the *smaller* the acceleration.

It is the *same* thing that was going on on the circle in figure 11: the smaller the radius of the circle was, the bigger the acceleration had to be for the particle to stay on the circle.

Equation (33) is does not look *consistent* unit-wise.

What is the *unit* of acceleration?

It is length divided by time divided by time, *i.e.* l/t^2.

Let's *rewrite* this dimension as

$$\frac{1}{l}\,\frac{l^2}{t^2}$$

In other words, acceleration *has* unit 1 over a length times a velocity squared.

So in equation (33) – we are now familiar with this kind of reasoning on dimensions – there is a 1 *actually* with the dimension of the square of a velocity.

It *is* the speed of light, with value 1 in the units chosen.

If we worked in general units where the speed of light has *value* c, equation (33) would be

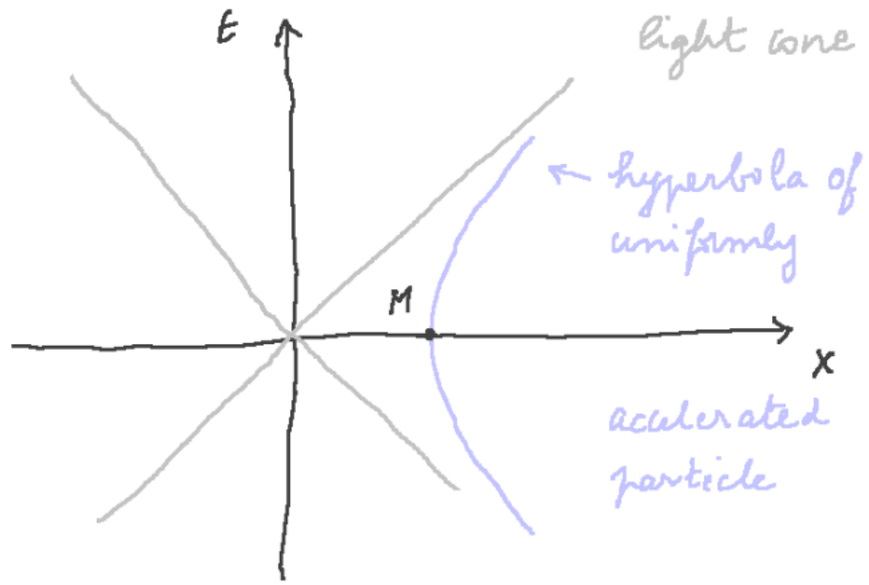
$$A = \frac{c^2}{R} \tag{34}$$

This *means* that for a fixed R, at human scale, the acceleration of a particle, in the collection of particles of figure 10, is very *large*.

We have to go to a very large R *before* we get to particles with a moderate acceleration.

We usually work in these lessons with units in which c = 1, but you should *keep in mind* that the acceleration of a particle, whose coordinates are given by equation (31) with r fixed, is very big unless r is very big.

By the way, the acceleration on a given trajectory in figure 12, for example the acceleration on the hyperbola $r = 2$, is the *ordinary* acceleration at point M.



And it is the constant acceleration that we would *experience* all along the trajectory if we rode a particle.

Notice that the uniformly accelerated reference frame we have been studying is *defined* relative to a particular origin.

The hyperbolic coordinates r and ω *depend* on the stationary frame selected.

We could *change* for another stationary frame.

This would *produce* different r and ω.

But *no real physics* depends on the choice of the stationary frame.

Uniform gravitational field

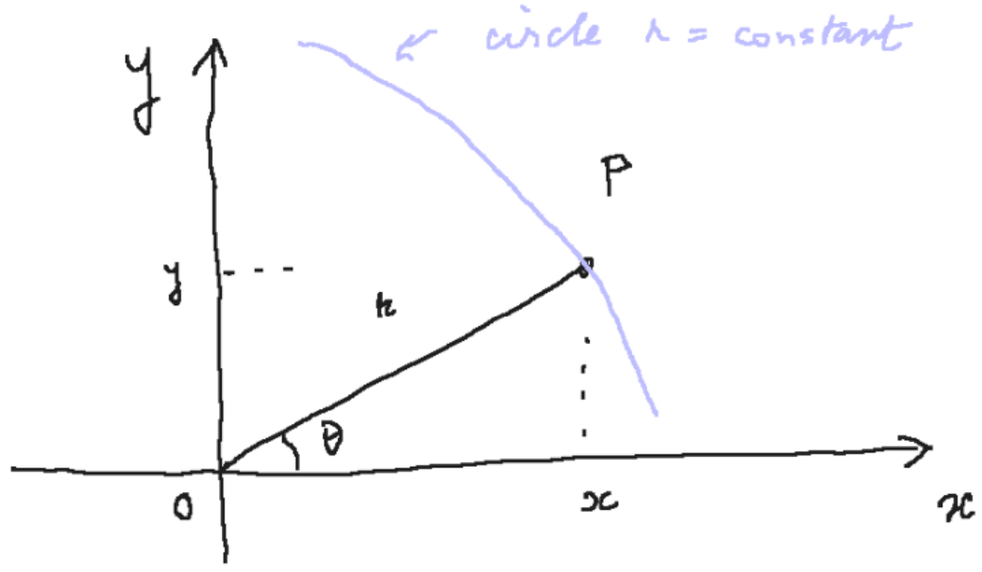
We have introduced a somewhat *arbitrary* set of coordinates for our stationary frame.

It just *has* to be Galilean, that is, it has to be such that in it Newton's equation has the *simple* form F = ma.

In that set of coordinates we are *now going to write* the equation of motion along a geodesic.

And we are going to see that the equation of motion along a geodesic looks like a particle falling in a uniform gravitational field.

Let's *first* talk about the metric of the Euclidean plane in ordinary polar coordinates as we did in figure 11



$$dS^2 = r^2 d\theta^2 + dr^2 \tag{35}$$

This just *says* that the distance between a point P, with coordinates (r, θ), and a close neighbor on the same circle, with coordinates $(r, \theta + d\theta)$, see fig. 11, for a fixed $d\theta$, *increases* linearly with r

This is the *metric* of the plane in polar coordinates.

Notice that the metric is *not* the Kronecker delta.

It *has* dr^2.

The matrix of this *two-dimensional metric* looks like this

$$g_{mn} = \begin{pmatrix} r^2 & 0 \\ 0 & 1 \end{pmatrix} \tag{36}$$

Why is it not the Kronecker delta?.

Not because the space is curved, but *because* the coordinates are curvilinear.

The space *itself* is flat.

Indeed it is the plane, and we *can go back* to Cartesian coordinates (x, y) in which the metric is the Kronecker delta.

Staying in the flat plane, the *analog* with the hyperbolic coordinates (r, ω) is

$$d\tau^2 = r^2 d\omega^2 - dr^2 \tag{37}$$

We are considering *only* two dimensions, time and one spatial coordinate.

We express the metric in terms of the proper time τ *rather* than proper distance S.

This makes *no* essential difference.

It is the *same* geometry.

When we want specifically to talk about a distance we *will* be careful to change the sign of the right hand side of equation (37).

For the moment we will ignore y and z, because we *don't* need them.

We will be *thinking* about a particle falling in a gravitational field along a vertical axis denoted X or x.

The coordinates y and z would be the other spatial coordinates and they *don't* matter for the problem we are going to discuss.

So the two coordinates that we will be *interested* in are ω and r, that is respectively, time and distance from the origin.

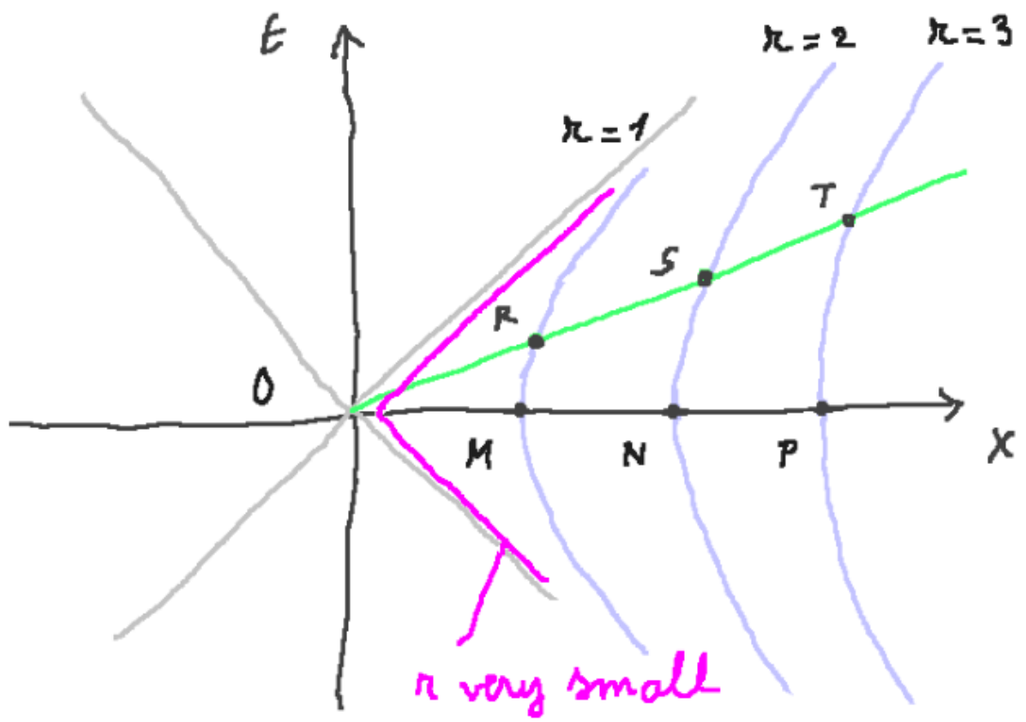
And equation (37) is the *metric*.

We want to *see* what this metric has to do with gravitation

Gravitation is *supposed* to have something to do with the metric.

So consider again figure 15, which *represents* a bunch of particles fixed with respect to each other, *in* a frame that is uniformly accelerated relative to the stationary frame.

The spatial x-axis, *even though* we said that it is a vertical axis along which a particle is falling, is represented as usual horizontally.



As we explained in the earlier, a uniformly accelerated frame is a *more subtle idea* than it seems.

In fact each particle *keeps a constant acceleration* on its trajectory.

But the accelerations of *different* particles are *different*.

As we *move* to the right, with *r increasing*, the acceleration of the corresponding particle is *smaller* and smaller, fig. 15.

If we go very *far away*, we can *find* a particle, corresponding to a hyperbolic radius R, and whose acceleration is g equal to *10 meters per second per second*.

Remember that the formula for its acceleration is $\frac{c^2}{R}$

We *set* this equal to g.

It *gives*

$$R = \frac{c^2}{g}$$

We *have* to go out this distance from O to find a particle with acceleration g.

The speed of light c is 3 times 10^8 meters per second, so c^2 is approximately 10^{17}, and R *about* 10^{16} meters.

Therefore we have to go out *10 000 billion kilometers* to find a particle with approximately the acceleration of the Earth gravitational field on its surface.

So let's go *there*.

Moreover, while there, if we *don't* move too much along the r direction, the acceleration $g = c^2/R$ *won't* change much.

It is *similar* to moving on a vertical axis near the surface of the Earth: the gravitational field doesn't change much.

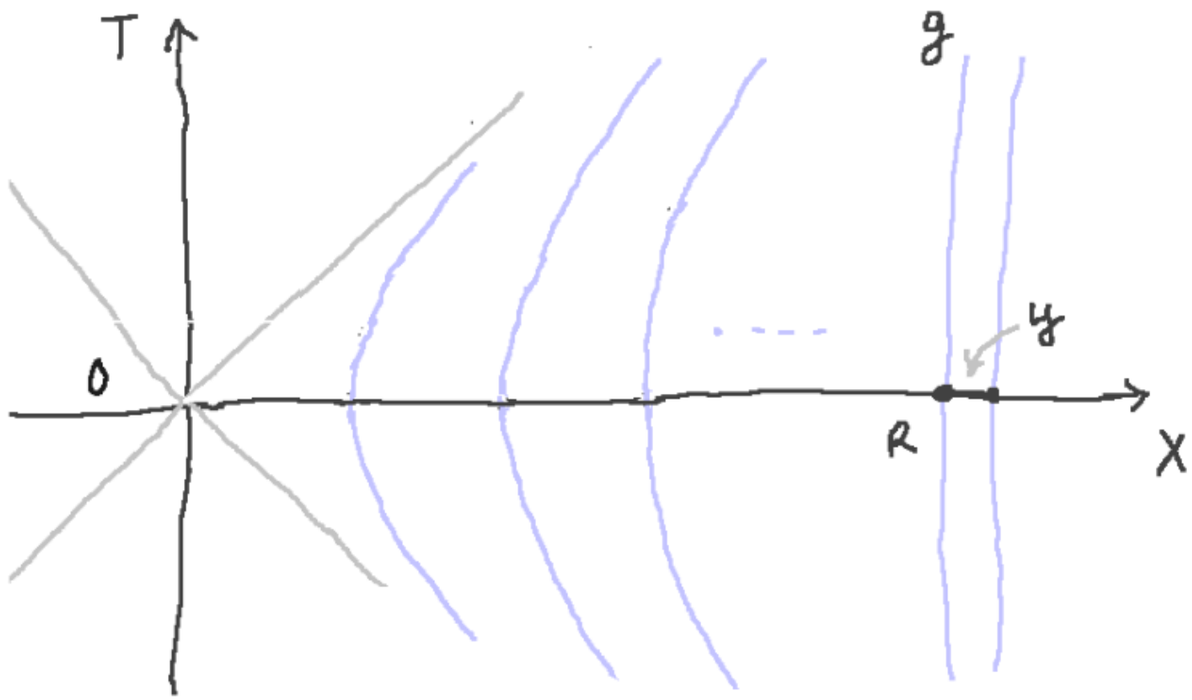
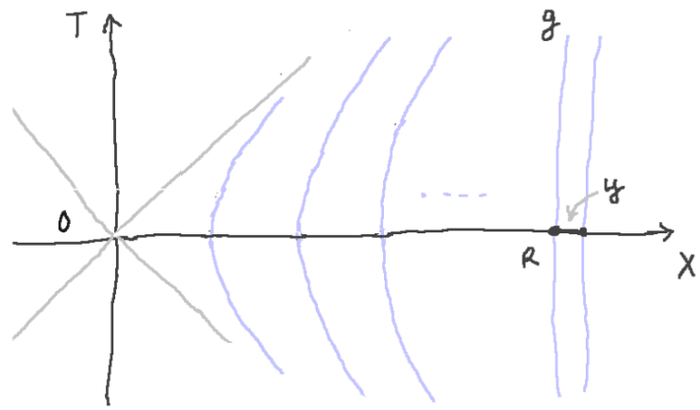


Figure 16 : Hyperbolas at R and at $R + y$.

Think of the *X*-axis as the axis of fall of a particle.

Falling would be going to the *left*.



And for the heck of it, we call *y* the *distance* from *R* on that axis.

Also for convenience, we *no longer* call the time, *t*.

We *call* it *T* .

y is a new first coordinate in the hyperbolic polar coordinates (r, ω).

$$y = r - R \tag{38}$$

All the particles with a moderate *y* have the same acceleration *g*.

Then we *rewrite* the metric of equation (37) using the new local coordinate *y*.

$$d\tau^2 = (R^2 + 2Ry + y^2)d\omega^2 - dy^2 \tag{39}$$

This *creates* an even more complicated looking metric, but not much.

And we shall *simplify* it, because we are going to focus on a limited region around R.

Notice that when we talk about r, R, y or ω, we no longer talk about the Minkowski coordinates (x, t), but about the hyperbolic polar coordinates (r, ω).

There is, however, a *simple* correspondence between r and x : when $\omega = 0$, they are equal.

That is why we can talk casually about R on the X-axis, while also meaning all the points on the hyperbola intercepting the horizontal axis at that point, *because* they all have hyperbolic radius R.

We rewrite equation (39) as $$d\tau^2 = (1 + \frac{2y}{R} + \frac{y^2}{R^2})\ R^2\ d\omega^2 - dy^2 \qquad (40)$$

One more step concerning coordinates: we give $R\omega$ a new name, we call it t.

Before rewriting again equation (40), let's notice that y/R is very small. y is measured in meters or kilometers – for instance the stunt man Felix Baumgartner, in 2012, dived from 39 km – while R is 10^{16} meters.

And y^2/R^2 is even much smaller.

We are going to keep y/R and *neglect* y^2/R^2.

Equation (40) becomes

$$d\tau^2 = (1 + \frac{2y}{R})\ dt^2 - dy^2 \tag{41}$$

We end up with a metric which apart from the term $2y/R$, which is small, just *looks* like the good old Minkowski metric $dt^2 - dr^2$.

It is space and time in a more or less ordinary way, but with a *little* correction.

The little correction $2y/R$ is what *accounts* for gravitation in the accelerated reference frame.

Keep in mind, however, that we are *really* talking about flat space.

So far we have *not introduced* any curvature.

The space we are working in is *really* a flat space.

With a change of coordinates we *can go back* to the Minkowski metric.

Therefore any gravitation that we find is in a sense the same *fake* gravitation that we found in the accelerated elevator of lesson 1.

We are *studying* physics in an accelerated coordinate system.

It is the elevator being pulled toward the *right*.

And what do we *expect* to find?

We expect to find that in that elevator there is an *effective gravitational field*.

We can also call it a *fictitious* gravitational field.

It is this field that is *associated* with the term $2y/R$ in equation (41).

To get a better understanding of the connection, let's now study the motion of a particle in a metric given by equation (41). $d\tau^2 = (1 + \frac{2y}{R})\ dt^2 - dy^2$

In units where c is equal to 1, we have $g = 1/R$.

The metric can be *rewritten*

$$d\tau^2 = (1 + 2gy)\ dt^2 - dy^2 \tag{42}$$

Have you ever seen the expression gy in studying gravitation in a uniform field?

If we introduce the mass m of a the particle, mgy is simply the *potential* energy.

The term gy is called the *gravitational potential*.

And the *term* $(1 + 2gy)$ is one plus twice the gravitational potential.

It is *extremely* general.

In any kind of gravitational field, as long as it is more or less constant with time, and not doing anything too radically relativistic, the *coefficient* in front of dt^2 in the metric is *always* one plus twice the gravitational potential.

Why do we call gy the gravitational potential other than that it just looks like it?

Answer: because if we work out the equation of motion of a particle in the metric given by equation (42), we will *find* that, as long as the particle is moving slowly, as long as we can make a good Newtonian approximation, as long as things are not too relativistic, the equation of motion that we will find is the same as that of a particle falling along the y-axis in a uniform gravitational field, as we have calculated it in classical mechanics.

What we *mean* by the y-axis of course is still the unique spatial axis, but near point R.

The *whole point* of the preceding section on uniform acceleration was to *explain* that a uniformly accelerated reference frame is something *more subtle* than just accelerating a thirty meter long steel beam along the X-axis.

We *had* to define what we meant by a uniformly accelerated frame.

It *leads* to this funny construction *where* points at different distances from O, measured for instance at time 0, each have a *fixed* acceleration, but the acceleration *differs* from point to point.

On the other hand we *can check* with a Lorentz transformation, in the frame of one of the moving points at velocity v, that the distances measured simultaneously by P between the points *don't change*.

And it also leads to *hyperbolic coordinates* and to figure 14.

Aside from that, it is really *ordinary* physics.

There *is* an accelerated elevator at M, see figure 17, there is another accelerated elevator at N, there is another one at P, there is one at R, etc.

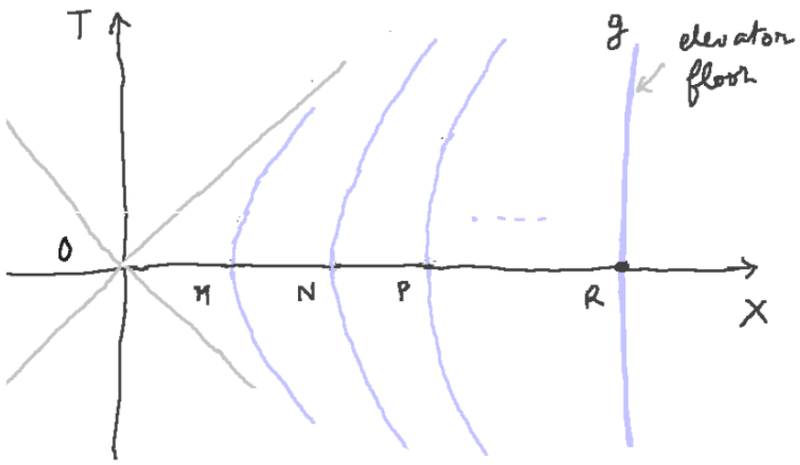


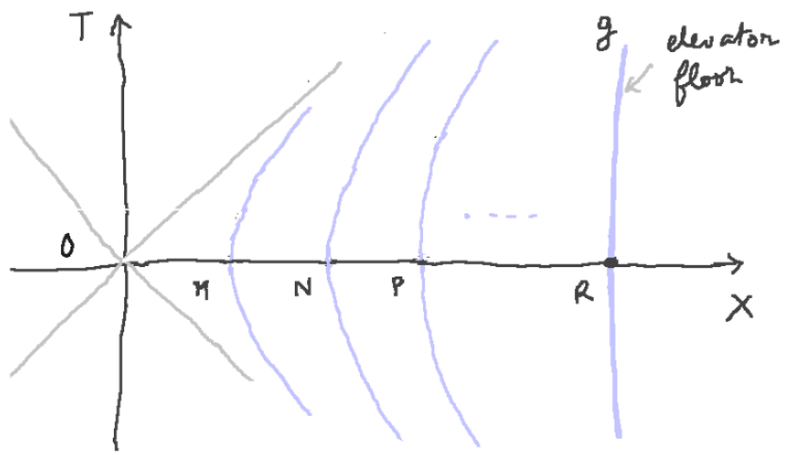
Figure 17 : Floor of the elevator.

In other words, it is just a bunch of elevators at different positions each being accelerated.

The only specific idea is that, for them to form a uniformly accelerated frame, they each must have a *different* acceleration.

They have to be *accelerated*, according to equation (34), with *acceleration* c^2/R.

We are interested, in figure 17, in the accelerated elevator *at* hyperbolic radius R.



Of course the elevator should be *imagined* on its side.

The bottom of the elevator, to be precise, *follows* the trajectory with hyperbolic radius R, that is, the thick trajectory in figure 17.

It is *important to understand* that the trajectory in space-time of the floor of the elevator is not a little horizontal segment going to the right on the X-axis – *although* it is true that if we look only at the space coordinate X, the elevator does move very slightly to the right.

Its trajectory in space-time is the *almost* vertical line, in figure 17, because it has only acceleration g, which is rather *small*.

And in the hyperbolic coordinates (r, ω), **the hyperbolic radius of the elevator remains** R.

To the right of R there are other *trajectories* – trajectories of points inside the elevator –, and they *also* have an acceleration very close to g.

The local coordinate y – it is hyperbolic radius *shifted* by R – is a coordinate to *locate* things inside the elevator, that is, it is the height of a point or a particle above the floor.

The *metric* of space-time in the *coordinate system* (t, y), which we built to locate things inside the uniformly accelerated elevator, is given by equation (42), which we *reproduce* below

$$d\tau^2 = (1 + 2gy)\ dt^2 - dy^2$$

We had to do a little bit of work to arrive at this equation, but now for us it is a *given*.

It is the *metric* tensor, in the *coordinate* system (t, y), at points in the *vicinity* of a point *moving* with acceleration g.

And in order to show the equivalence of a uniform acceleration and a gravitational field, we *chose* for the acceleration g the same value as the gravitational field of the Earth.

What is the *rule* to figure out how a particle moves ?

The rule about particle motion is that particles *move on geodesics* – not geodesics of space but geodesics of *space-time*.

In other words, we take the metric of space-time, whatever it is, and we go through exactly these same operations.

The metric of space-time that we are given is equation (42).

And the equation of motion of a particle is equation (19), which we reproduce below, with a change of sign and a new numbering,

$$-\frac{d^2X^n}{dS^2} = \Gamma^n_{mr}\ t^r\ t^m \tag{43}$$

It is the equation of motion which says : *it is a geodesic*.

Go *straight* ahead.

But not straight ahead in space now, straight ahead *in space-time*.

Otherwise the equations are the *same*.

We shall write it slightly *differently*, using $d\tau$ instead of dS.

Remember that $d\tau^2$ is just the *opposite* of dS^2.

So the *left* hand side of (43) becomes

$$\frac{d^2X^n}{d\tau^2} \tag{44}$$

This is called the *proper* acceleration.

As long as the elevator is moving *slowly*, in other words if it hasn't been accelerating long enough to get up near the speed of light, then the proper time and the ordinary time are *essentially* the same.

And expression (44) is just the *ordinary* acceleration.

We choose X to be y.

We want the y component of acceleration.

Then expression (44) is *simply*

$$\frac{d^2y}{d\tau^2}$$

Now we turn to the right hand side of equation (43). $$-\frac{d^2X^n}{dS^2} = \Gamma^n_{mr}\ t^r\ t^m$$

The *n*-th component of *X stands* for *y*.

What is t^r ?

It is dX^r/dS, *because* we are on a geodesic.

And dS is $id\tau$.

So equation (43) *becomes*

$$\frac{d^2y}{d\tau^2} = -\Gamma^y_{mr}\ \frac{dX^r}{d\tau}\ \frac{dX^m}{d\tau} \tag{45}$$

Since *m* and *r* each run over four coordinates (to be consistent with standard notations in relativity, it would be *better* to use μ and ν.

We *leave* to you to make the change in these dummy variables), the right hand side has a whole bunch of terms – ten of them to be precise because the gammas are symmetric in *m* and *r*.

Fortunately most of them are extremely small as long as the elevator is moving slowly, and as long as the movement of the object we are interested in, namely the particle with coordinate *y*, is *slow*.

Under these conditions, *only* one of the combinations $\Gamma^{y}_{mr}\ \frac{dX^r}{d\tau}\ \frac{dX^m}{d\tau}$ is significant.

What is the value of $dt/d\tau$ for slow motion?

It is essentially 1, because time and proper time in that case are almost the same.

On the right hand side of equation (45) the differential elements are the components of the 4-velocity of the particle.

We just saw that $\frac{dX^0}{d\tau}$ is essentially 1.

What are the derivatives of the space components with respect to τ ?

They are *proportional* to their actual ordinary spatial velocity.

We are assuming the spatial velocity is small compared to the speed of light so the *only* important contribution on the right hand side of (45) comes when *r* and *m* are time indices.

Let's use *t* instead of 0 for the time index.

Equation (45) *reduces* to

$$\frac{d^2y}{d\tau^2} = -\Gamma^y_{tt} \tag{46}$$

The right hand side *must* be the gravitational force.

It *must* be the derivative of the gravitational potential energy.

Let's *go back* to the expression of the Christoffel symbols in terms of the metric

$$\Gamma^p_{rs} = \frac{1}{2}\, g^{pn} \left[\frac{\partial g_{nr}}{\partial X^s} + \frac{\partial g_{ns}}{\partial X^r} - \frac{\partial g_{rs}}{\partial X^n} \right] \tag{47}$$

We *need* the one with two time covariant indices and one space contravariant index.

The space index is y.

Among the terms g^{yn}, the *only* term which is not negligible is g^{yy} and it is 1.

Since X^t is just what we *denote* t and X^y is what we *denote* y, we *get*

$$\Gamma^y_{tt} = \frac{1}{2} \left(\frac{\partial g_{yt}}{\partial t} + \frac{\partial g_{yt}}{\partial t} - \frac{\partial g_{tt}}{\partial y} \right)$$

The terms $\frac{\partial g_{yt}}{\partial t}$ and $\frac{\partial g_{yt}}{\partial t}$, which are equal, are *both* zero.

So *finally*

$$\Gamma^{y}_{tt} = -\frac{1}{2}\frac{\partial g_{tt}}{\partial y}$$

And equation (46) can be *rewritten*

$$\frac{d^2 y}{d\tau^2} = \frac{1}{2}\frac{\partial g_{tt}}{\partial y} \tag{48}$$

An equation like (48), where the second derivative of a spatial variable y with respect to time is proportional to the first derivative of some quantity with respect to y, *reminds* us of an equation of motion with potential energy.

Somehow one half of g_{tt} *must be* the opposite of a potential energy.

But we *saw* that: it is minus a potential energy with $m = 1$, also *called* gravitational potential.

In equation (42), which we *rewrite* below,

$$d\tau^2 = (1 + 2gy)dt^2 - dy^2$$

g_{tt} is the coefficient $-(1 + 2gy)$ in the metric defining dS^2, which is the *same* as $d\tau^2$ with a minus sign.

So one half the derivative of g_{tt} with respect to y is $-g$.

Equation (48) finally *becomes*

$$\frac{d^2y}{d\tau^2} = -g \tag{49}$$

That is the equation of motion of a particle in a *uniform gravitational field*.

We *went through* a rather complicated derivation to reach it, but so doing *we learned* the following points :

1. Space-time has a metric. In arbitrary coordinates, the metric can have a fairly complicated structure. In uniformly accelerated coordinates, however, it is almost the Minkowski metric, but with the extra term $2gy$ in equation (42).
2. The equation of motion along a geodesic in space-time – at least as long as things are going slowly, that is as long as Newtonian approximation is valid – is just Newton's equation in a uniform gravitational field.

The word *valid* here means *setting* the terms where c appears in the denominator to zero.

Uniform gravitational field, constant acceleration... : it is what we *expected*.

But to do it properly, using the metric, the Christoffel symbols, the geodesics and so forth, is a fairly *complicated* procedure.

Einstein *guessed* it : the hypothesis that a particle moves along a geodesic in space-time was his *starting point*, and he went in the *opposite* direction.

He *knew* about uniformly accelerated coordinate systems, but he *didn't know* about Christoffel symbols.

Somewhere along our own derivations is where he started.

And for a *uniform acceleration* – with Newtonian approximation – the metric is simply given by equation (42).

So we have come around *full circle* from the fist lesson where we talked about accelerated elevators giving rise to gravitation.

We *have shown* that in the Minkowski-Einstein space-time a uniformly accelerated reference frame *does* give rise to an effective gravitational field.

But so far we *haven't gotten to real gravitational fields*.

The gravitational field we got is *not a real* gravitational field because it *really corresponds to a flat space*.

If we were to take the metric of equation (42) and calculate the curvature tensor it would be exactly *zero*, indicating that there *does exist* coordinates where the metric has the simple form $dt^2 - dX^2$.

So the gravitation that we are *experiencing* is really exactly the gravitation due to an *accelerated* frame of reference, *not* due to real gravitating matter.

Now we *can guess* what the effect of real gravitating matter would be.

Instead of $2gy$ in equation (42), *what is* the gravitational potential due to a gravitating object ?

It *is* $-G/y$.

Here, by *convention*, for a uniform gravitational field, the gravitational potential is taken to be zero at ground level and increases to $+\infty$ when the height increases, while for the radial field created by an object, the gravitational potential is taken to be zero infinitely far away and goes to $-\infty$ when the radius goes to zero.

That is why y is now in the *denominator*, and there is a *minus* sign.

So we can *expect*, when we study the metric of a real gravitational field, that we will have something like

$$d\tau^2 = (1 - \frac{2GM}{y})\ dt^2 - dy^2 \tag{50}$$

where G is Newton's constant, and M is the mass of the gravitating object.

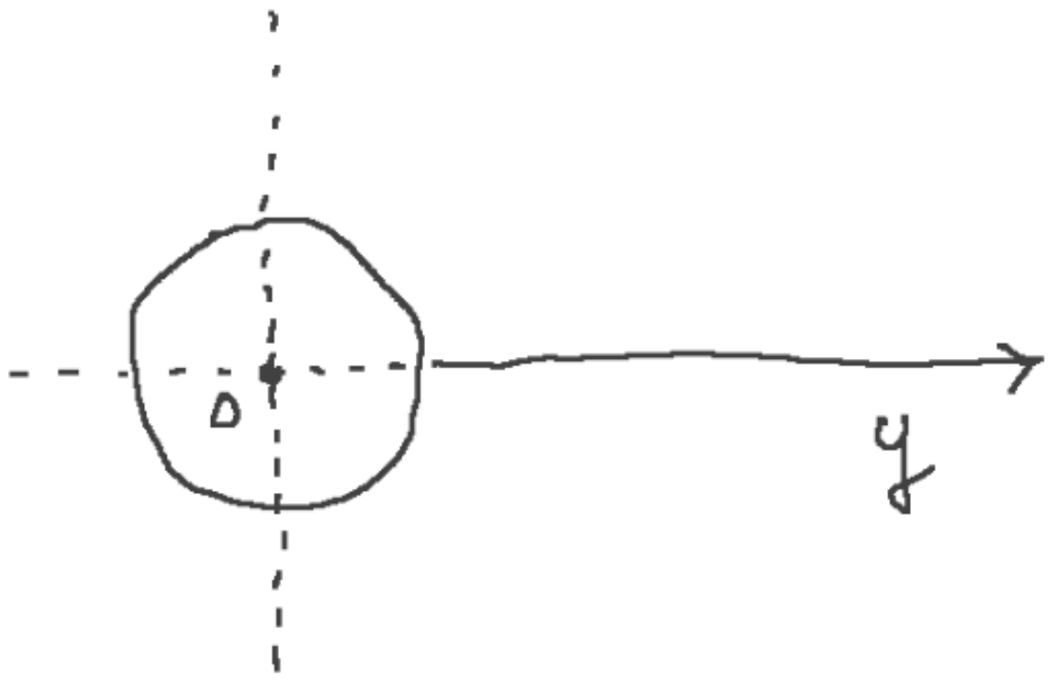


Figure 18 : Gravitating object of mass M and gravitational potential $-G/y$.

That is *almost* the Schwarzschild metric, but not quite.

We *will work out* what is the Schwarzschild metric of a gravitating object.

Equation (50) will lead to a *weird* phenomenon. $$d\tau^2 = (1 - \frac{2GM}{y})\ dt^2 - dy^2$$

When y is *large*, the term $2GM/y$ is *small*.

That is *good* because $(1 - 2GM/y)$ is *positive*.

But something *crazy* happens at the point where y is *equal* to $2GM$.

The *coefficient* in front of dt^2 becomes zero.

That point y where the coefficient changes sign is called the *horizon of the black hole*.

For the Earth, supposing all its mass was almost point-like, the *horizon* would be 9 millimeters.

Real gravitational fields and the Schwarzschild metric will be the *subject* of next lesson.

We are *not* going to derive the metric entirely from what we already know.

To derive it, we *need* fields equations.

We haven't discussed them yet, and we won't do it until *lesson 9*.

So far we have *only* discussed geometry, flatness, curvature, geodesics, etc.

And in this lesson, when we finally arrived at the space-time of relativity and its peculiar geometry, we *ended* up with a little demonstration of how, in a uniformly accelerated reference frame in space-time, movement along a geodesic *gives* rise to Newton's equation.

Lesson 5 : Metric for a gravitational field

Space-like, time-like, and light-like intervals, and light cones

Let's begin with time-like, space-like, and light-like intervals.

For that we go back to special relativity to spell out what that means.

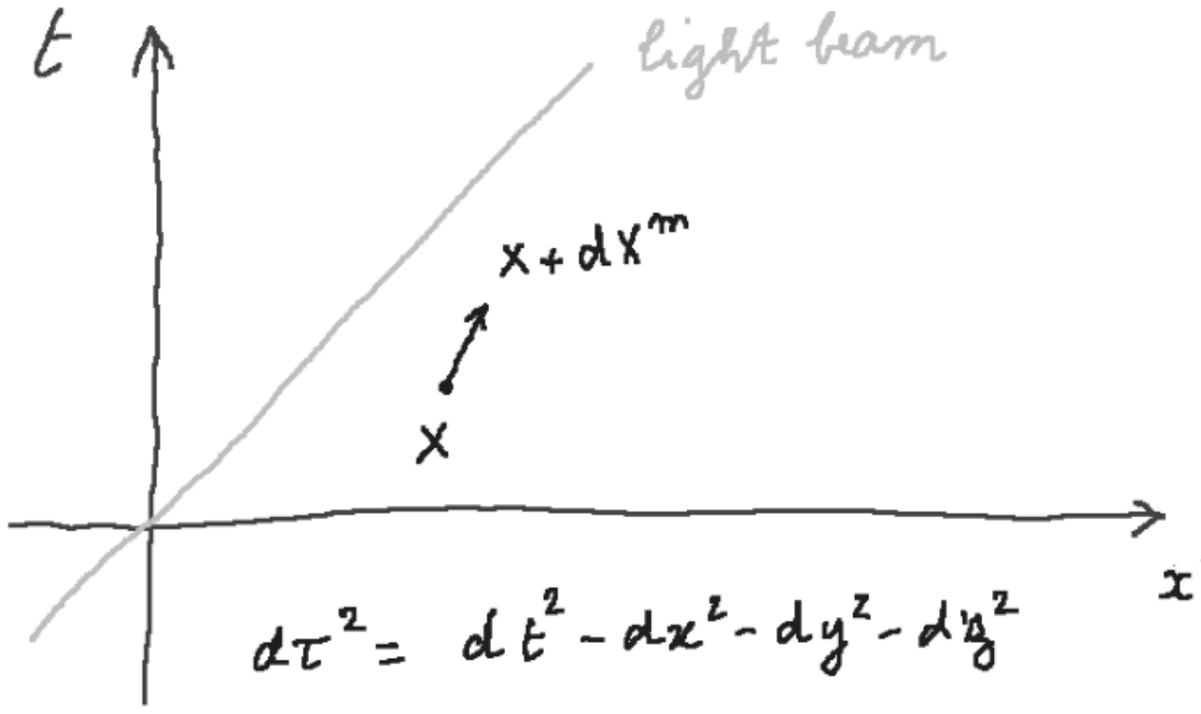


Figure 1 : Space-time and Minkowski metric.

We have discussed the metric many times.

We call it proper time

$$d\tau^2 = dt^2 - dx^2 - dy^2 - dz^2 \tag{1}$$

Sometimes we want to put in explicitly the speed of light.

It then becomes

$$d\tau^2 = dt^2 - \frac{dx^2}{c^2} - \frac{dy^2}{c^2} - \frac{dz^2}{c^2} \tag{2}$$

The reason to introduce the speed of light is to keep track of what is small and what is big under certain circumstances.

For example if we want to go to the non-relativistic limit, that is, the limit where everything is moving slowly, it is good to put back c because it reminds us that it is much bigger than any other velocities in the problem.

And it makes it easy to see which terms can be neglected and which terms cannot.

So as before we will sometimes put it in and sometimes take it out depending on circumstances.

Let's look at the sign of $d\tau^2$.

Of course, when we look at real numbers, their square is always positive.

But $d\tau^2$ is not defined as the square of a real number, it is defined by equation (2).

And it can be positive or negative, depending on whether $dx^2 + dy^2 + dz^2$ is smaller than dt^2 or bigger than $d\tau^2$.

Of course when we say $dx^2 + dy^2 + dz^2$, we mean $(dx^2 + dy^2 + dz^2)/c^2$.

If $d\tau > 0$, then the little element dX^{m} in figure 1 is said to be *time-like*.

It has more time than it has space so to speak.

Its vertical part is bigger that its horizontal part.

Its slope in figure 1 is greater than 45°.

It can be described in terms of a light cone, figure 2.

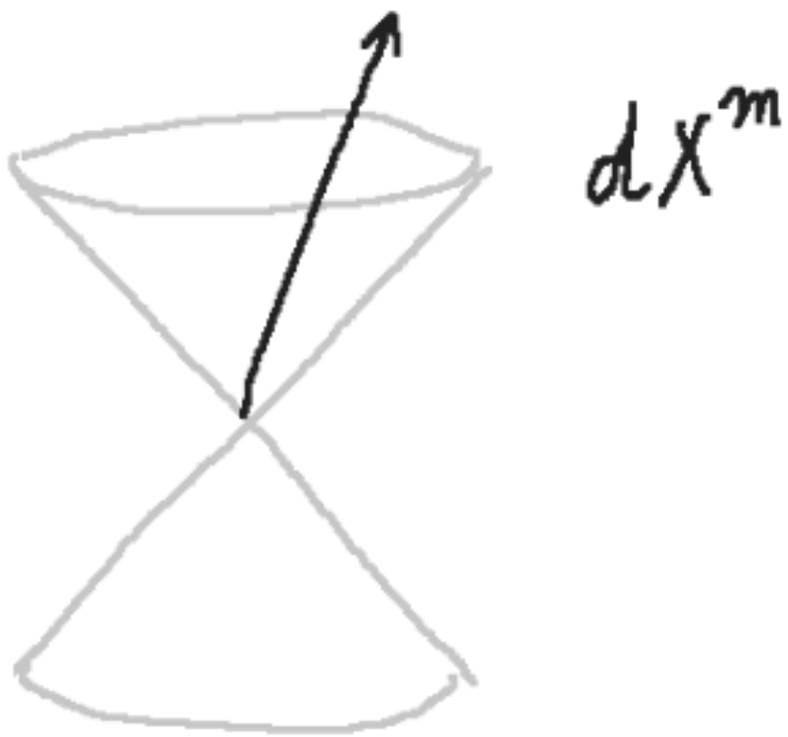


Figure 2 : Time-like interval.

If we represent two spatial coordinates x and y, in addition to the time coordinate t, and a light cone whose center is at X, then it means the little vector dX^m lies in the interior of the cone.

It could also lie in the backward direction in the same picture, pointing to the past.

In either case, it is called a *time-like* interval.

Space-like is exactly the opposite of time-like.

It corresponds to $d\tau^2 < 0$, or equivalently $dx^2 + dy^2 + dz^2$ greater than dt^2.

In that case we usually define another quantity dS^2 which is just the negative of $d\tau^2$.

By definition it is

$$dS^2 = dx^2 + dy^2 + dz^2 - dt^2 \tag{3}$$

It is the *same* object as $d\tau^2$ except for the minus sign.

Space-like vectors are those for which $dS^2 > 0$. And if we represent the cone at X as in figure 2, we get figure 3.

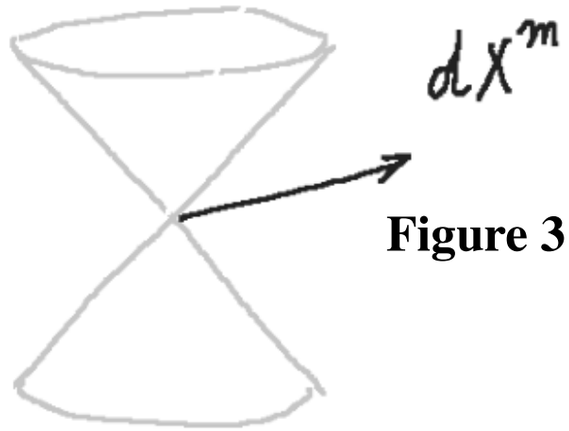


Figure 3 : Space-like interval.

Finally there are **light-like** vectors.

They are those for which both $d\tau^2 = 0$ and therefore $dS^2 = 0$.

Their *slope*, in the standard diagram, is at 45°.

They are *trajectories* of light- rays, and they lie on the *surface* of the cone of figure 3.

Those are the three kinds of vectors we can have.

Just for a moment consider what it would mean, if there were two positive signs and two negative signs instead of one positive sign and three negative signs in the definition of the metric in equation (1).

Somehow this would correspond to two time directions.

It *doesn't* mean anything in physics.

There are *never* two time directions.

There are always one time and three space.

Can you *imagine* a world with two times?

I can't.

Frankly I can't.

I cannot imagine what it would mean to have two different time directions.

And we will simply take the view that that is *not* an option.

There is always one time-like direction in the metric of equation (1) and three sp

But that *doesn't mean* that there is a *unique* direction which is time-like.

There are *many* time-like directions pointing *within* the light cone of figure 2.

The *invariant property*, corresponding to the fact that at any point there is one time and three space directions, concerns the *metric* tensor.

We can *write* the metric tensor in the following expression

$$d\tau^2 = -g_{\mu\nu}\, dX^\mu\, dX^\nu \tag{4}$$

The *minus* sign in front of $g_{\mu\nu}$ is a *convention*, so that dS^2, the square of the infinitesimal proper distance, is given by the same expression but without the minus sign.

Let's recall this important point about the proper time $d\tau$, defined by equation (4): it is the time *recorded* by a clock *accompanying* the particle along its trajectory.

In other words, it *has* a physical practical meaning which is often useful to remember.

Of course, as we know, for particles going slowly – and by slowly we mean up to thousands of miles per second – the proper time is *essentially* the same as the standard time of the stationary observer in the stationary frame of figure 1.

This is easily seen from equation (2), because c is *very big* compared to ordinary velocity, or to the spatial components of the 4-velocity.

Similarly, proper distance, dS, along the trajectory of a particle is distance *measured* by a meter stick carried along by the particle

In summary, proper distance, $\sqrt{dS^2}$, *really* is a distance.

And proper time, $\sqrt{d\tau^2}$, *really* is a time.

Let's keep that in mind.

Equation (3) is the definition of the metric with the coordinates (t, x, y, z).

It *can* always be written in terms of a matrix.

In this case it is the matrix η *as shown*

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{5}$$

Remember that it is the analog in the Minkowski space of the Kronecker delta, that is simply the unit matrix, in Euclidean space.

The story is not that there is only one time direction, but that there is *always* only one negative eigenvalue in the metric.

In special relativity and in general relativity, that *will still* be the case whatever the metric is and whatever the coordinate system is.

And since the metric, in general, depends on the point in space-time where we look at it, this invariance statement will be *true* at any point.

A metric which would have two negative eigenvalues or three negative eigenvalues, would have more than one time.

And we just *don't* even think about that.

That is something that physics does not seem to have *made use of*, several time axes.

The concepts of time-like, space-like and light-like displacements are *not restricted* to special relativity.

They apply generally whatever the metric, and whatever the point which we consider.

In the preceding discussion, we were in the *flat space of special relativity*, but the concepts *apply* in general relativity where the space is *intrinsically* not flat.

Now we shall consider a metric *more general* than $\eta_{\mu\nu}$.

We *consider* $g_{\mu\nu}(X)$.

At every point, that is, at every X, there is a matrix.

And the matrix *must have* one negative eigenvalue and three positive eigenvalues.

In other words, wherever you are – standing there – you should *experience* a world with one time direction and three space directions, or *more exactly* one negative eigenvalue and three positive.

That *means* that every point in space has a light cone *associated* with it, figure 4.

Figure 4 : Light cone at each point.

These light cones *can be* tilted and change shape depending of the curvy aspect of the coordinates at each point.

But at *each point* the metric *has* three positive and only one negative eigenvalue.

And at *each point* there is the *notion* of time-like displacement, space-like displacement and light-like displacement.

So we have to *make sure*, when we write a metric, that it does have this property of the right number of positive and negative eigenvalues.

The property of having a certain number of eigenvalues positive and a certain number negative is called the *signature* of the metric.

What was the signature of the metric of ordinary flat space?

Not ordinary space-time, just the page you are reading.

It is + +.

The signature of Minkowski space in special relativity with three spatial coordinates is − + + +.

When somebody gives us a metric, or wherever we get it from – we might get it as a present in the mail, we might calculate it from some equations of motion, from some field equations – we should make sure that that metric has *signature* $-+++$.

If it doesn't, it means we did something *wrong*.

And not only should we have that signature at some point, but we should have it at *every* point.

Notice that the shape of the light cone, in particular its angle of openness, is a *pure coordinate issue*, see figure 4 and figure 2.

In particular, if in the standard Minkowski metric and representation of figure 2, we chose units such that the speed of light is not 1 but the huge number c, the cone would be *extremely flat*, and the picture not very useful.

So much on the signature of the metric.

Let's now *move on* to the metric of a *massive* body, and more *specifically* of a black hole.

But first of all let's *revisit* geodesics, *deriving* them in *another* way.

Geodesics and Euler-Lagrange equations

We *learned*, in lesson 4, what is the *definition* of a geodesic.

And we used the corresponding equation – equation (19) of lesson 4 – in the example of a free particle in a uniformly accelerated reference frame.

A geodesic is a curve whose tangent vector *stays parallel* to itself all along the curve.

In other words, it is a *trajectory where we always go straight*.

In this lesson we are going to learn a *different* definition, which in many ways is *more useful*.

But let us first *recall* the original definition,

$$\frac{d^2X^m}{d\tau^2} = -\Gamma^{\mu}_{\sigma\rho}\,\frac{dX^\sigma}{d\tau}\,\frac{dX^\rho}{d\tau} \tag{6}$$

The *left hand side* is the derivative of the tangent vector along some curve, which all along the curve *should be equal* to the double sum involving Christoffel symbols of the right hand side.

That is the *standard definition* of a geodesic.

But remember that in lesson 4, we also mentioned another: it is the analog of the definition of a geodesic in ordinary space, that is, the *shortest distance* between two points.

Or better yet, it is the *curve between two points whose length is stationary*.

So another way of writing equation (6) is to say that we "*extremelize*" – or we *make minimum* – the length of the curve between two points.

Let's *start* with ordinary space, the page of any book, or a curved version of the page with hills and valleys.

We take *two points* in the space and any curve between them.

We *calculate* the distance along the curve.

Then we *search* for the curve which minimizes its length.

How to calculate it?

Let's *spell out* the logic.

We *start*, as we said, with any curve C between A and B.

There are plenty of them *shown* in grey in figure 5.

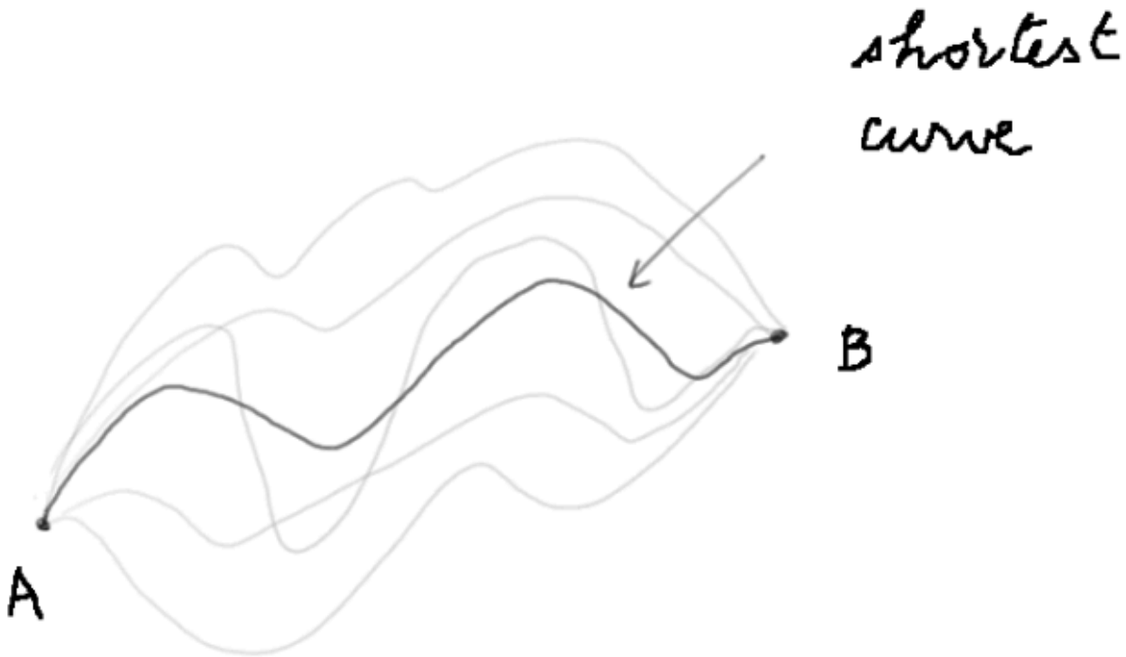


Figure 5 : Determination of a geodesic between A and B.

When the space is flat it is a straight line.

On the given curve C, for each little segment on the curve, we have

$$dS^2 = g_{mn}(X)\ dX^m\ dX^n \tag{7}$$

We are now in *ordinary Riemannian space, not in space-time*.

We will *come back to space-time* in a moment.

Equation (7) can be rewritten

$$dS = \sqrt{g_{mn}(X)\ dX^m\ dX^n} \tag{8}$$

This is just *Pythagoras theorem* applied to a little segment on curve C.

Then we *add them all up*.

This gives the *distance* along curve C in figure 5.

$$S = \int_{along\ curve\ C} \sqrt{g_{mn}(X)\ dX^m\ dX^n} \tag{9}$$

Finally we *look* for the curve *C* which makes *S minimum or extremum*.

That's the logic.

The *mathematics* which *implement* this logic comes *from* classical mechanics.

It is a problem in *calculus of variations*, *analogous* to minimizing the action of a particle along a trajectory.

In other words, we can think of equation (9) as *expressing* the action of a particle moving from *A* to *B* along curve *C*.

$$S = \int_{along\ curve\ C} \sqrt{g_{mn}(X)\ dX^m\ dX^n}$$

Then the rule for calculating the geodesic is to "extremelize" this quantity, or to *make it stationary*.

The equation that tells us *how* to minimize a quantity like *S* in equation (9) is called the *Euler-Lagrange equation*.

Digression (D) on Calculus of Variations and Lagrange Equations

Many problems in Newtonian mechanics are *more easily analyzed* by means of *alternative* statements of the laws, including *Lagrange's equation*.

As a prelude to these techniques, we now consider *general principles* of the techniques of the calculus of variations.

Emphasis will be placed on those *aspects* of the theory of variations that have a *direct bearing* on classical systems, *omitting some existence proofs*.

Our primary interest here is in *determining* the path that gives extremum solutions, for example, the shortest distance (or time) between two points.

Statement of the Problem

The *basic* problem of the calculus of variations is to determine the function $y(x)$ *such that* the integral

$$J = \int_{x_1}^{x_2} f\{y(x), y'(x); x\}\, dx \tag{D-1}$$

is an *extremum* (i.e., either a maximum or a minimum).

In Equation D-1, $y'(x) = dy/dx$, and the *semicolon* separates the independent variable x from the dependent variable $y(x)$ and its derivative y'(x).

The functional J depends on the function $y(x)$, and the limits of integration are *fixed*.

The function $y(x)$ is then to be *varied until* an extreme value of J is found.

By this we *mean* that if a function $y = y(x)$ gives the integral J a minimum value, then any neighboring function, no matter how close to $y(x)$, must make J *increase*.

The *definition* of a neighboring function may be made as follows.

We give all possible functions y a *parametric* representation $y = y(\alpha, x)$ such that, for $\alpha = 0$, $y = y(0, x) = y(x)$ is the function that *yields* an extremum for J.

We can then write

$$y(\alpha, x) = y(0, x) + \alpha\eta(x) \tag{D-2}$$

where $\eta(x)$ is some *function* of x that has a continuous first derivative and that vanishes at x_1 and x_2, because the varied function $y(\alpha, x)$ must be *identical* with $y(x)$ at the endpoints of the path:

$$\eta(x_1) = \eta(x_2) = 0$$

The situation is *depicted* schematically in Figure D-1.

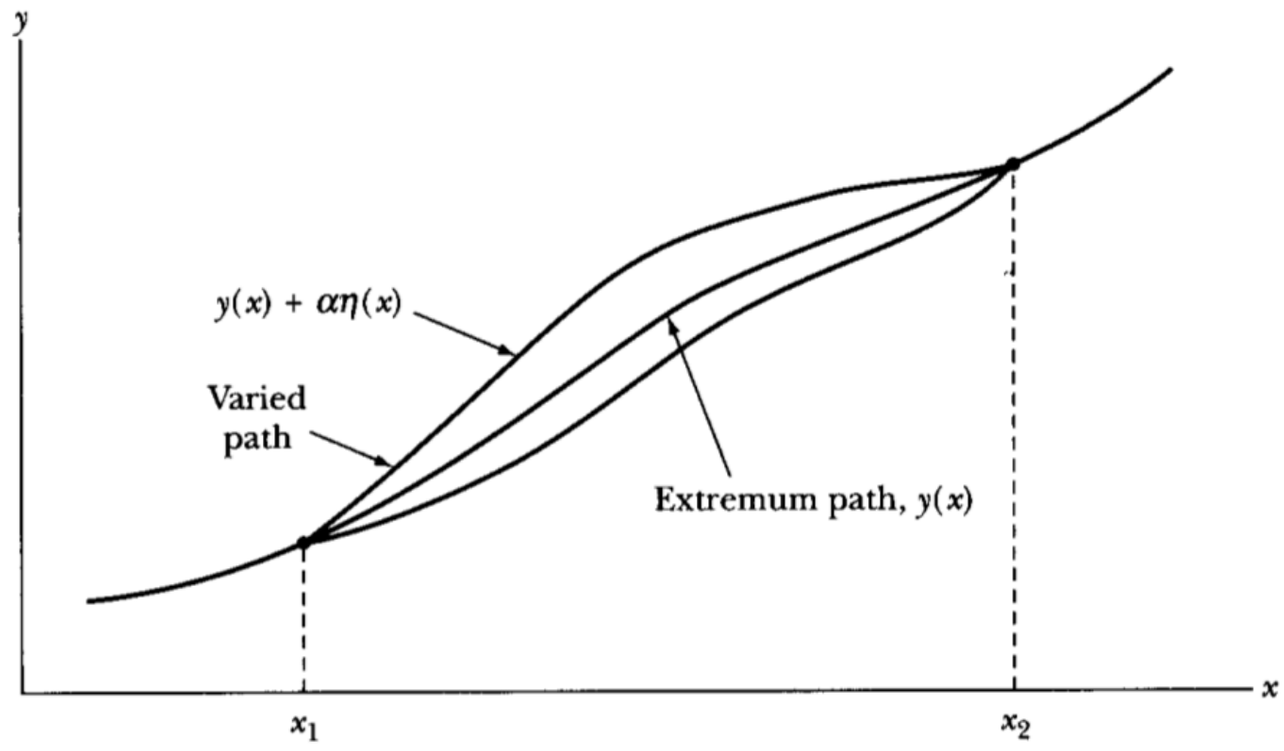


FIGURE D-1 The function $y(x)$ is the path that makes the functional J an extremum. The neighboring functions $y(x) + \alpha\eta(x)$ vanish at the endpoints and may be close to $y(x)$, but are not the extremum.

If functions of the type given by Equation D-2 are considered, the integral J becomes a *functional* of the parameter α :

$$J(\alpha) = \int_{x_1}^{x_2} f\{y(\alpha, x), y'(\alpha, x); x\}\, dx \tag{D-3}$$

The *condition* that the integral have a stationary value (i.e., that an extremum results) is that J be *independent* of α in first order along the path giving the extremum ($\alpha = 0$), or, *equivalently*, that

$$\left.\frac{\partial J}{\partial \alpha}\right|_{\alpha=0} = 0 \tag{D-4}$$

for *all* functions $\eta(x)$.

This is only a *necessary* condition; it is *not sufficient*.

EXAMPLE 1

Consider the function $f = (dy/dx)^2$, where $y(x) = x$.

Add to $y(x)$ the function $\eta(x) = \sin x$, and *find* $J(\alpha)$ between the limits of $x = 0$ and $x = 2\pi$.

Show that the stationary value of $J(\alpha)$ occurs for $\alpha = 0$.

Solution. We may construct neighboring varied paths by *adding* to $y(x)$,

$$y(x) = x \tag{D-5}$$

the *sinusoidal* variation $\alpha \sin x$,

$$y(\alpha, x) = x + \alpha \sin x \tag{D-6}$$

These paths are *illustrated* in Figure D-2 for $\alpha = 0$ and for *two* different nonvanishing values of α .

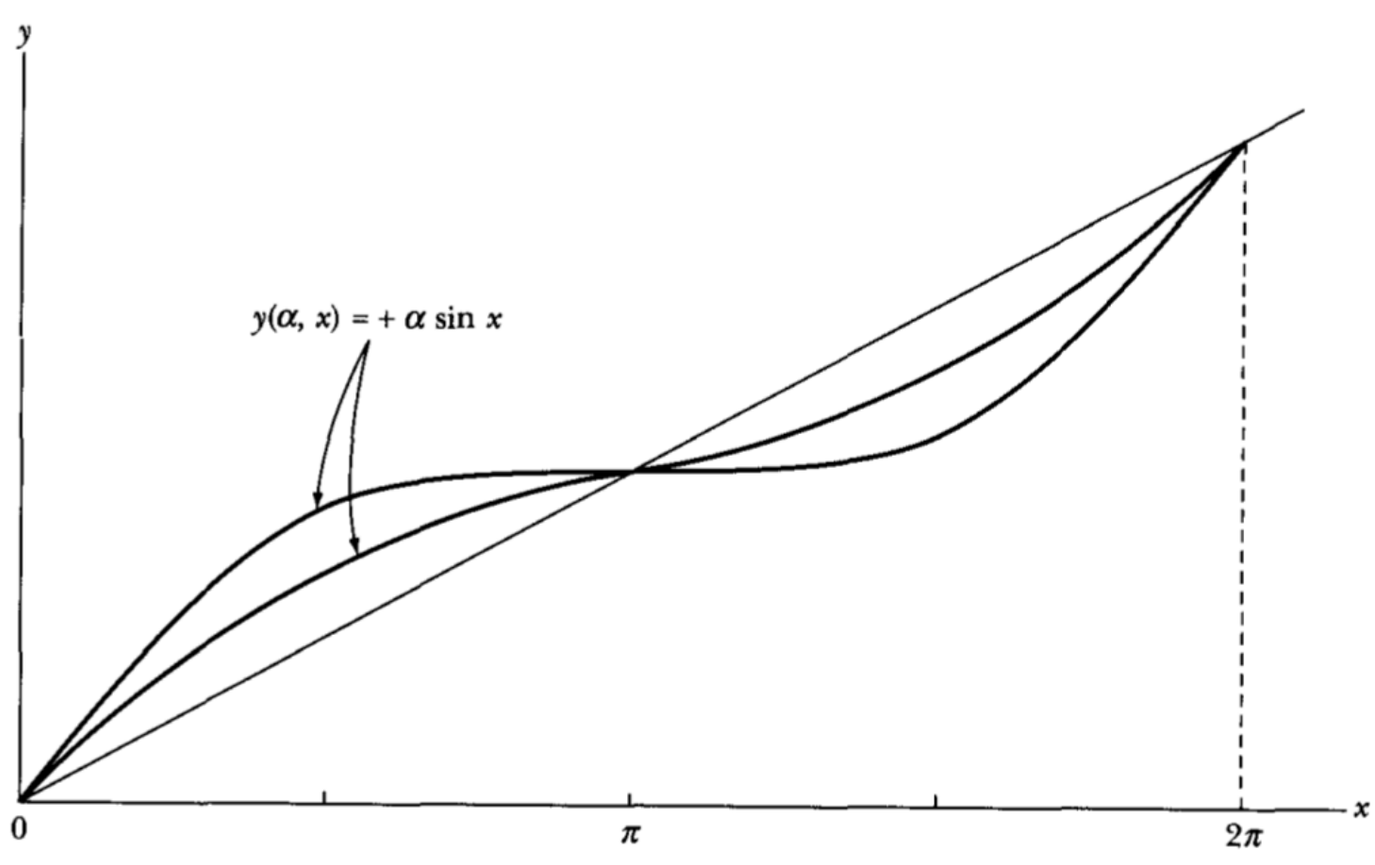


FIGURE D-2 Example 6.1. The various paths $y(\alpha, x) = x + \alpha \sin x$. The extremum path occurs for $\alpha = 0$.

Clearly, the function $\eta\,(x) = \sin x$ obeys the endpoint conditions, that is, $\eta(0) = 0 = \eta(2\pi)$.

To *determine* $f(y, y'; x)$ we first determine.

$$\frac{dy(\alpha, x)}{dx} = 1 + \alpha \cos x \tag{D-7}$$

then

$$f = \left(\frac{dy(\alpha, x)}{dx}\right)^2 = 1 + 2\alpha \cos x + \alpha^2 \cos^2 x \tag{D-8}$$

Equation D-3 now *becomes*

$$J(\alpha) = \int_0^{2\pi} (1 + 2\alpha \cos x + \alpha^2 \cos^2 x)\, dx \tag{D-9}$$

$$= 2\pi + \alpha^2 \pi \tag{D-10}$$

Thus we *see* the value of $J(\alpha)$ is *always greater* than $J(0)$, no matter what value (positive or negative) we choose for α .

The condition of Equation D-4 is *also satisfied*.

Euler's Equation

To *determine the result* of the condition expressed by Equation D-4, we *perform* the indicated differentiation in Equation D-3:

$$\frac{\partial J}{\partial \alpha} = \frac{\partial}{\partial \alpha} \int_{x_1}^{x_2} f\{y, y'; x\}\, dx \tag{D-11}$$

Because the limits of integration are fixed, the differential operation affects *only* the integrand.

Hence,

$$\frac{\partial J}{\partial \alpha} = \int_{x_1}^{x_2} \left(\frac{\partial f}{\partial y} \frac{\partial y}{\partial \alpha} + \frac{\partial f}{\partial y'} \frac{\partial y'}{\partial \alpha} \right) dx \tag{D-12}$$

From Equation D-2, we *have*

$$\frac{\partial y}{\partial \alpha} = \eta(x); \qquad \frac{\partial y'}{\partial \alpha} = \frac{d\eta}{dx} \tag{D-13}$$

Equation D-12 *becomes*

$$\frac{\partial J}{\partial \alpha} = \int_{x_1}^{x_2} \left(\frac{\partial f}{\partial y} \eta(x) + \frac{\partial f}{\partial y'} \frac{d\eta}{dx} \right) dx \tag{D-14}$$

The second term in the integrand can be *integrated* by parts:

$$\int u\,dv = uv - \int v\,du \tag{D-15}$$

$$\int_{x_1}^{x_2} \frac{\partial f}{\partial y'}\frac{d\eta}{dx}\,dx = \frac{\partial f}{\partial y'}\eta(x)\Big|_{x_1}^{x_2} - \int_{x_1}^{x_2} \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right)\eta(x)\,dx \tag{D-16}$$

The integrated term *vanishes* because $\eta(x_1) = \eta(x_2) = 0$.

Therefore, Equation D-12 *becomes*

$$\begin{aligned}\frac{\partial J}{\partial \alpha} &= \int_{x_1}^{x_2}\left[\frac{\partial f}{\partial y}\eta(x) - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right)\eta(x)\right]dx \\ &= \int_{x_1}^{x_2}\left(\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'}\right)\eta(x)\,dx\end{aligned} \tag{D-17}$$

The integral in Equation D-17 now appears to be *independent* of α .

But the functions y and y' with respect to which the derivatives of f are taken are still functions of α .

Because $(\partial J/\partial\alpha)|_{\alpha=0}$ must vanish for the extremum value and because $\eta(x)$ is an arbitrary function (subject to the conditions already stated), the integrand in Equation D-17 must *itself* vanish for $\alpha = 0$:

we get

$$\boxed{\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'} = 0} \qquad \textbf{Euler's equation} \qquad \text{(D-18)}$$

EXAMPLE 2

We can use the calculus of variations to solve a *classic* problem in the history of physics: the *brachistochrone*.

Consider a particle moving in a constant force field starting at rest from some point (x_1,y_1) to some lower point (x_2, y_2).

Find the path that allows the particle to accomplish the transit in the *least possible time*.

Solution. The coordinate system may be *chosen* so that the point (x_1,y_1) is at the origin.

Further, let the force field be *directed* along the positive x-axis as in Figure D-3.

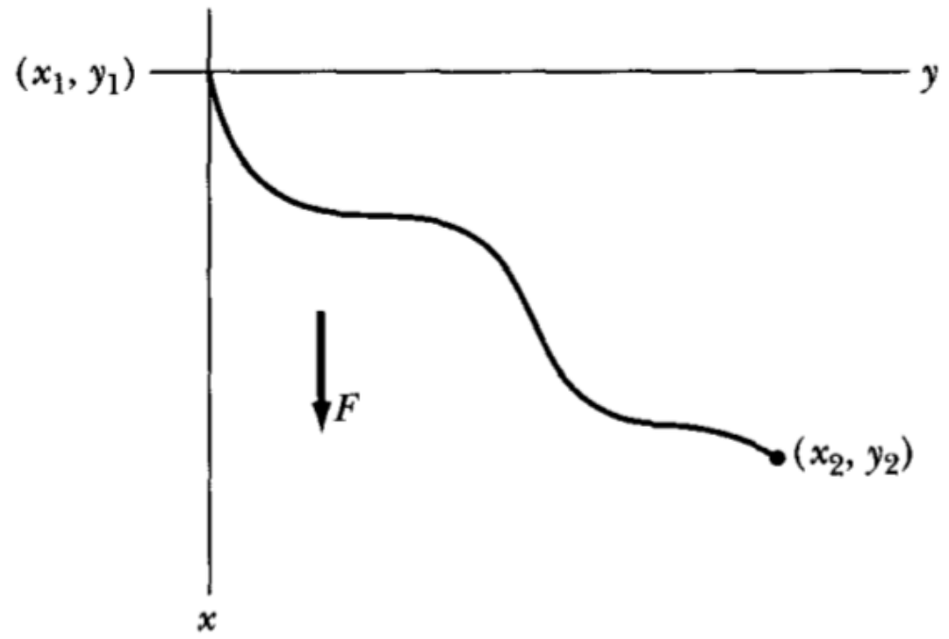


FIGURE D-3 Example 2 The *brachistochrone* problem is to find the path of a particle moving from (x_1, y_1) to (x_2, y_2) that occurs in the least possible time. The force field acting on the particle is F, which is down and constant.

Because the force on the particle is constant - and if we *ignore* the possibility of friction - the field is conservative, and the total energy of the particle is $T + U = constant$.

If we measure the potential from the point $x = 0$ [i.e., $U(x = 0) = 0$], then, *because* the particle starts from rest, $T+U= 0$.

The kinetic energy is $T = \frac{1}{2} mv^2$, and the potential energy is $U = -F\text{x} = -mgx$, where g is the acceleration imparted by the force.

Thus

$$v = \sqrt{2gx} \tag{D-19}$$

The *time* required for the particle to make the transit from the origin to (x_2, y_2) is

$$t = \int_{(x_1,y_1)}^{(x_2,y_2)} \frac{ds}{v} = \int \frac{(dx^2 + dy^2)^{1/2}}{(2gx)^{1/2}}$$
$$= \int_{x_1=0}^{x_2} \left(\frac{1 + y'^2}{2gx}\right)^{1/2} dx \qquad \text{(D-20)}$$

The time of transit is the quantity for which a *minimum* is desired.

Because the constant $(2g)^{-1/2}$ does not affect the final equation, the function f may be identified as

$$f = \left(\frac{1 + y'^2}{x}\right)^{1/2} \qquad \text{(D-21)}$$

And, *because* $\partial f/\partial y = 0$, the Euler equation (Equation D-18) *becomes*

$$\frac{d}{dx}\frac{\partial f}{\partial y'} = 0$$

or

$$\frac{\partial f}{\partial y'} = \text{constant} \equiv (2a)^{-1/2}$$

where a is a *new* constant.

Performing the differentiation $\partial f/\partial y'$ on Equation D-21 and squaring the result, we *have*

$$\frac{y'^2}{x(1 + y'^2)} = \frac{1}{2a} \tag{D-22}$$

This may be *put* in the form

$$y = \int \frac{x\,dx}{(2ax - x^2)^{1/2}} \tag{D-23}$$

We now make the following change of variable:

$$\begin{aligned} x &= a(1 - \cos\theta) \\ dx &= a \sin\theta\, d\theta \end{aligned} \tag{D-24}$$

The integral in Equation D-23 then *becomes*

$$y = \int a(1 - \cos\theta)\, d\theta$$

and

$$y = a(\theta - \sin\theta) + \text{constant} \tag{D-25}$$

The *parametric* equations for a cycloid passing through the origin are

$$\left.\begin{aligned} x &= a(1 - \cos\theta) \\ y &= a(\theta - \sin\theta) \end{aligned}\right\} \tag{D-26}$$

which is just the solution *found*, with the constant of integration set equal to zero to conform with the requirement that $(0, 0)$ is the starting point of the motion.

The path is then as *shown* in Figure D-4, and the constant a must be *adjusted* to allow the cycloid to pass through the specified point (x_2, y_2).

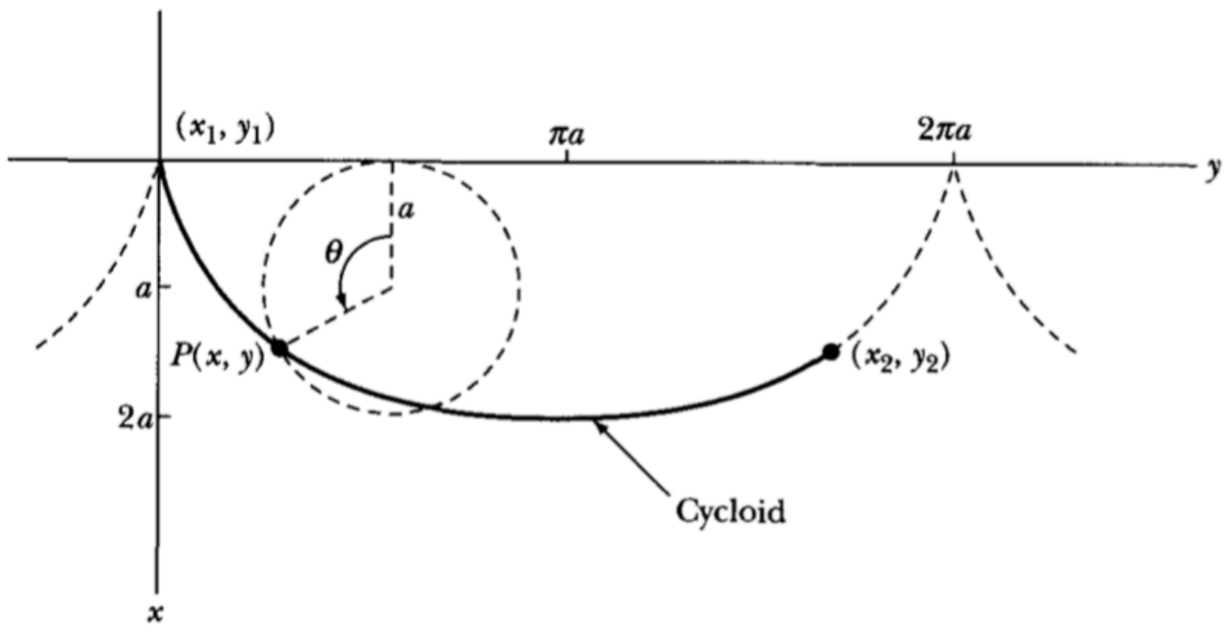


FIGURE D-4 Example 2. The solution of the brachistochrone problem is a cycloid.

Solving the problem of the brachistochrone does indeed *yield* a path the particle traverses in a *minimum* time.

But the procedures of variational calculus are designed only to produce an *extremum* - either a minimum or a maximum.

It is almost *always* the case in dynamics that we desire (and find) a minimum for the problem.

Lagrangian Dynamics

Hamilton's Principle

Minimal principles in physics have a long and interesting history.

The search for such principles is *predicated* on the notion that nature always minimizes certain important quantities when a physical process takes place.

The first such minimum principles were developed in the field of optics.

Hero of Alexandria, in the second century B.C., found that the law governing the reflection of light could be obtained by asserting that a light ray, traveling from one point to another by a reflection from a plane mirror, always takes the shortest possible path.

A simple geometric construction *verifies* that this minimum principle does indeed lead to the equality of the angles of incidence and reflection for a light ray reflected from a plane mirror.

Hero's principle of the shortest path *cannot*, however, yield a correct law for refraction.

In 1657, Fermat *reformulated* the principle by postulating that a light ray always travels from one point to another in a medium by a path that requires the *least time*.

Fermat's principle of least time *leads* immediately, not only to the correct law of reflection, but *also* to Snell's law of refraction.

Minimum principles continued to be sought, and in the latter part of the seventeenth century the beginnings of the calculus of variations were developed by Newton, Leibniz, and the Bernoullis when such problems as the brachistochrone (as we saw) and the shape of a hanging chain (a catenary) were solved.

The *first* application of a general minimum principle in mechanics was made in 1747 by Maupertuis, who asserted that dynamical motion *takes place* with minimum action.

Maupertuis's principle of least action was based on *theological* grounds (action is minimized through the "wisdom of God"), and his concept of "action" was rather *vague*.

Action, as we will see, is a quantity with the dimensions of *length* x *momentum* or *energy* x *time*.

Only *later* was a firm mathematic foundation of the principle given by Lagrange.

Although it is a *useful* form from which to make the transition from classical mechanics to optics and to quantum mechanics, the principle of least action is *less general* than Hamilton's Principle and, indeed, can be derived from it.

In 1828, Gauss developed a method of treating mechanics by his principle of **least constraint**; a modification was *later* made by Hertz and embodied in his principle of **least curvature**.

These principles are closely related to Hamilton's Principle and add nothing to the content of Hamilton's more general formulation; their mention only *emphasizes* the continual concern with minimal principles in physics.

In two papers published in 1834 and 1835, Hamilton *announced* the dynamical principle on which it is possible to base *all* of mechanics and, indeed, *most* of classical physics.

Hamilton's Principle may be stated as follows:

> *Of all the possible paths along which a dynamical system may move from one point to another within a specified time interval (consistent with any constraints), the actual path followed is that which minimizes the time integral of the difference between the kinetic and potential energies.*

In *terms* of the calculus of variations, Hamilton's Principle *becomes*

$$\delta \int_{t_1}^{t_2} (T - U)\, dt = 0 \qquad \text{(D-27)}$$

where the symbol δ is a *shorthand* notation to describe the variation discussed earlier.

This variational statement of the principle requires only that the integral of T - U be an *extremum*, *not* necessarily a *minimum*.

But in almost *all* important applications in dynamics, the *minimum* condition occurs.

The kinetic energy of a particle expressed in fixed, rectangular coordinates is a function only of the $\dot{x}_i$, and if the particle moves in a conservative force field, the potential energy is a function only of the x_i:

$$T = T(\dot{x}_i), \quad U = U(x_i)$$

If we *define* the difference of these quantities to be

$$L \equiv T - U = L(x_i, \dot{x}_i)$$

then Equation D-27 *becomes*

$$\boxed{\delta \int_{t_1}^{t_2} L(x_i, \dot{x}_i)\, dt = 0} \tag{D-28}$$

The function L appearing in this expression may be *identified* with the function f of the variational integral (see earlier discussion),

$$\delta \int_{x_1}^{x_2} f\{y_i(x), y_i'(x); x\}\, dx \tag{D-29}$$

if we make the transformations

$$x \rightarrow t$$
$$y_i(x) \rightarrow x_i(t)$$
$$y_i'(x) \rightarrow \dot{x}_i(t)$$
$$f\{y_i(x), y_i'(x); x\} \rightarrow L(x_i, \dot{x}_i)$$

The Euler-Lagrange equations *corresponding* to Equation D-29 are therefore

$$\boxed{\frac{\partial L}{\partial x_i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}_i} = 0, \quad i = 1, 2, 3} \qquad \text{Lagrange equations of motion} \qquad \text{(D-30)}$$

These are the **Lagrange equations of motion** for the particle, and the quantity L is called the **Lagrange function** or **Lagrangian** for the particle.

After all that intense mathematics, I will now present a discussion of Lagrangian Methods from the Mechanics class I taught to some of you in 2020.

An Alternative Path - The Lagrangian Method

Let us restrict our attention to one-dimensional systems.

We then have

$$E = K + V$$

where

$$K = \frac{1}{2}mv^2 = \frac{1}{2}m\dot{x}^2 = \text{ kinetic energy}$$
$$V = V(x) = \text{ potential energy}$$

Now Newton's second law takes the form

$$F_x = -\frac{\partial V(x)}{\partial x} = ma_x = m\ddot{x}$$

Some algebra gives

$$m\ddot{x} = \frac{d}{dt}(m\dot{x}) = \frac{d}{dt}\left(\frac{\partial}{\partial \dot{x}}\right)\left(\frac{1}{2}m\dot{x}^2\right) = \frac{d}{dt}\left(\frac{\partial K}{\partial \dot{x}}\right)$$

Combining everything we then have

$$-\frac{\partial V}{\partial x} = \frac{d}{dt}\left(\frac{\partial K}{\partial \dot{x}}\right)$$

assuming that

$$K = K(\dot{x}) \rightarrow \frac{\partial K}{\partial x} = 0$$ no dependence on position

$$V = V(x) \rightarrow \frac{\partial V}{\partial \dot{x}} = 0$$ no dependence on velocity

we have

$$\frac{\partial (K - V)}{\partial x} = \frac{d}{dt}\left(\frac{\partial (K - V)}{\partial \dot{x}}\right)$$

or

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) - \frac{\partial L}{\partial x}$$

where we define

$$L = K - V = \text{ Lagrangian}$$

and this equation is called **Lagrange's equation**.

It is important to realize that *I have not introduced* any new physical content into the theory at this point.

I have only found an *alternative* way to write Newton's second law!

$$\int L(\vec{r},\vec{v},t)dt \quad \text{when done over path of motion} \;=\; \text{minimum}$$

The derivation actually goes *other* way

—> Newton's Laws This the principle of "least action" ; It is related to Fermat's principle of least time…. as we saw earlier.

Examples

(1) For an object falling in a gravitational field, we have

$$K = \frac{1}{2}m\dot{y}^2$$
$$V = mgy$$
$$L = K - V = \frac{1}{2}m\dot{y}^2 - mgy$$

We then have

$$\frac{\partial L}{\partial \dot{y}} = m\dot{y} \rightarrow \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{y}}\right) = m\ddot{y}$$
$$\frac{\partial L}{\partial y} = -mg$$

so that equation of motion is

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{y}}\right) - \frac{\partial L}{\partial y} = 0$$
$$m\ddot{y} = -mg \rightarrow \ddot{y} = -g$$

which agrees with the standard result.

(2) For an object attached to a spring, we have

$$K = \frac{1}{2}m\dot{x}^2$$

$$V = \frac{1}{2}kx^2$$

$$L = K - V = \frac{1}{2}m\dot{x}^2 - \frac{1}{2}m\dot{x}^2$$

We then have

$$\frac{\partial L}{\partial \dot{x}} = m\dot{x} \rightarrow \frac{d}{dt}\left(\frac{\partial L}{\partial yjx}\right) = m\ddot{x}$$

$$\frac{\partial L}{\partial x} = -kx$$

so that the equation of motion is

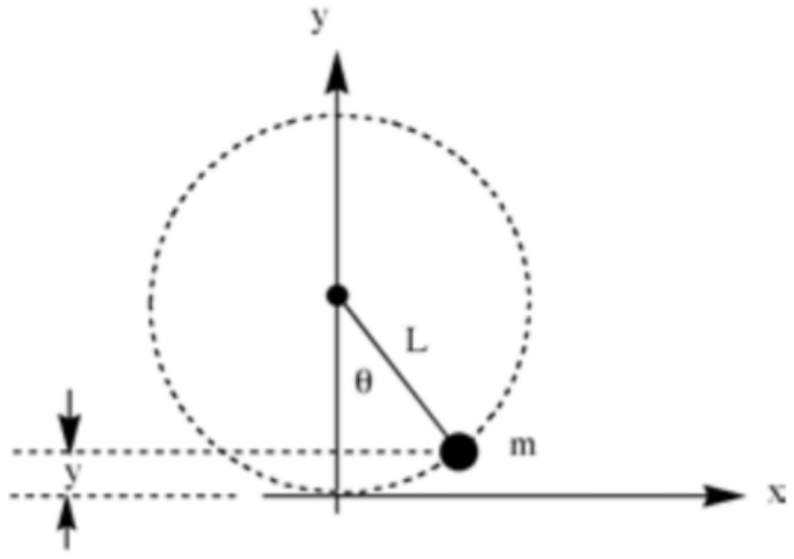
$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) - \frac{\partial L}{\partial x} = 0$$

$$m\ddot{x} = -kx \rightarrow \ddot{x} = -\frac{k}{m}x = -\omega_0^2 x$$

which agrees with the standard result.

Extra work for simple problems…….

(3) For an object attached to a string (simple pendulum) (see figure)



we have

$$L\dot{\theta} = v_\theta$$

$$K = \frac{1}{2}mL^2\dot{\theta}^2$$

$$V = mgy = mgL(1-\cos\theta)$$

$$L = K - V = \frac{1}{2}mL^2\dot{\theta}^2 - mgL(1-\cos\theta)$$

We then have

$$\frac{\partial L}{\partial \dot{\theta}} = mL^2\dot{\theta} \rightarrow \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\theta}}\right) = mL^2\ddot{\theta}$$

$$\frac{\partial L}{\partial \theta} = -mgL\sin\theta$$

so that the equation of motion is

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\theta}}\right) - \frac{\partial L}{\partial \theta} = 0$$

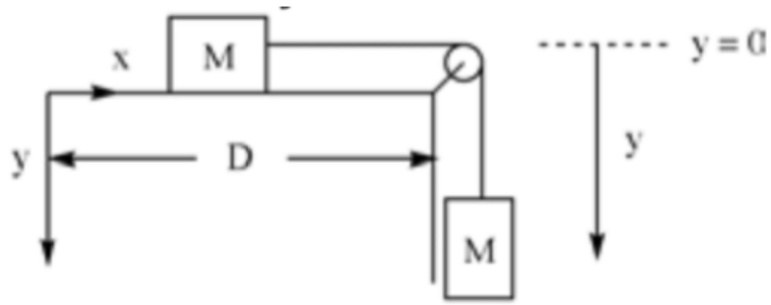
$$mL^2\ddot{\theta} = -mgL\sin\theta \rightarrow \ddot{\theta} + \frac{g}{L}\sin\theta = 0$$

For small angles we have $\sin\theta \approx \theta$, so that our equations reduces to

$$\ddot{\theta} + \frac{g}{L}\theta = \ddot{\theta} + \omega_0^2\theta = 0$$

which agrees with the standard result.

(4) Now consider the sliding blocks shown in figure



We have the following constraint:

$$\ell = \text{ length of string } = (D - x) + y$$
$$\rightarrow \dot{\ell} = 0 = -\dot{x} + \dot{y} \rightarrow \dot{x} = \dot{y}$$

$$K = \frac{1}{2}M\dot{x}^2 + \frac{1}{2}M\dot{y}^2 = M\dot{y}^2$$
$$V = -Mgy$$
$$L = K - V = M\dot{y}^2 + Mgy$$

If we ignore the string mass (first approximation) we have

We then have

$$\frac{\partial L}{\partial \dot{y}} = 2M\dot{y} \rightarrow \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{y}}\right) = 2M\ddot{y}$$
$$\frac{\partial L}{\partial y} = Mg$$

so that the equation of motion is

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{y}}\right) - \frac{\partial L}{\partial y} = 0$$
$$2M\ddot{y} = Mg \rightarrow \ddot{y} = \frac{g}{2}$$

which has the solution

$$y(t) = \frac{g}{4}t^2$$

which agrees with the standard result if we remember that y is increasing downwards.

Now what happens if the string has a mass m. We then have

$$K = \frac{1}{2}M\dot{x}^2 + \frac{1}{2}M\dot{y}^2 + \frac{1}{2}m\dot{y}^2 = \left(M + \frac{1}{2}m\right)\dot{y}^2$$
$$V = -Mgy + U_{string} == Mgy - g\int_0^y ydm$$
$$= -Mgy - \int_0^y y\frac{m}{\ell}dy = -Mgy - \frac{mg}{2\ell}y^2$$
$$L = K - V = \left(M + \frac{1}{2}m\right)\dot{y}^2 + Mgy + \frac{mg}{2\ell}y^2$$

We then have

$$\frac{\partial L}{\partial \dot{y}} = (2M + m)\dot{y} \rightarrow \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{y}}\right) = (2M + m)\ddot{y}$$
$$\frac{\partial L}{\partial y} = Mg + \frac{mg}{\ell}y$$

so that the equation of motion now becomes

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{y}}\right) - \frac{\partial L}{\partial y} = 0$$
$$(2M + m)\ddot{y} = Mg + \frac{mg}{\ell}y \rightarrow \ddot{y} = \frac{M}{2M + m}g + \frac{mg}{(2M + m)\ell}y$$

which has the solution

$$y(t) = \frac{M}{m}\ell(1 - \cosh(\gamma t)) \quad , \quad \gamma = \sqrt{\frac{mg}{(2M+m)\ell}}$$

This the standard result .

Note there is the same amount of work independent of complexity of problem

Consider the limit as m $\rightarrow$ 0. We have

$$y(t) = \lim_{m\to 0} \frac{M}{m}\ell\left(1 - \cosh\sqrt{\frac{mg}{(2M+m)\ell}}t\right)$$

$$= \lim_{m\to 0} \frac{M}{m}\ell\left(1 - \frac{1}{2}\left(e^{\sqrt{\frac{mg}{(2M+m)\ell}}t} + e^{-\sqrt{\frac{mg}{(2M+m)\ell}}t}\right)\right)$$

$$= \lim_{m\to 0} \frac{M}{m}\ell\left(1 - \frac{1}{2}\left(1 + \sqrt{\frac{mg}{(2M+m)\ell}}t + \frac{1}{2}\frac{mg}{(2M+m)\ell}t^2\right.\right.$$

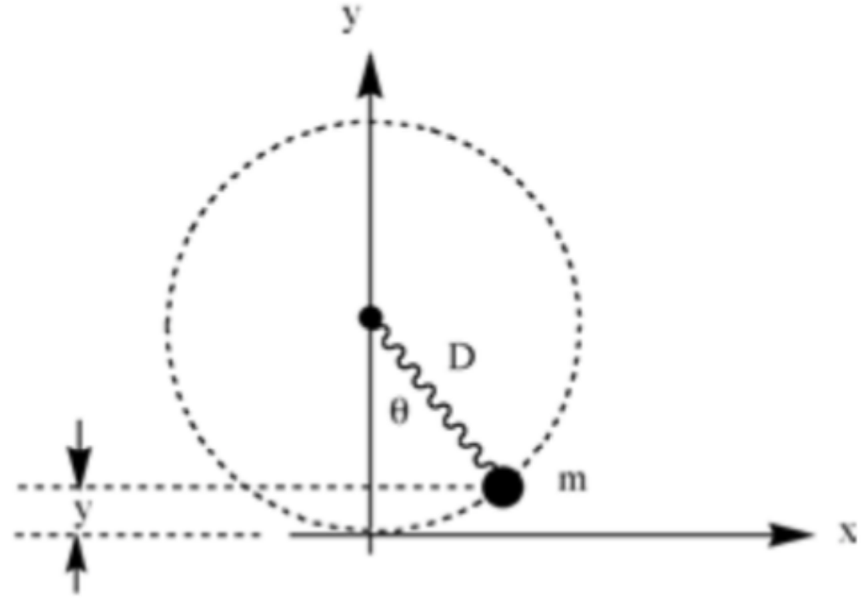
$$\left.\left. +1 - \sqrt{\frac{mg}{(2M+m)\ell}}t + \frac{1}{2}\frac{mg}{(2M+m)\ell}t^2\right)\right)$$

or

$$y(t) = \lim_{m\to 0} \frac{M}{m}\ell\frac{mg}{2(2M_m)\ell}t^2 = \frac{g}{4}t^2$$

which is the standard result for a massless string = 0.

(5) Now let us modify this pendulum problem and turn the string into a spring of spring constant k and unstretched length b. See figure



Now D is the variable length of the spring.

We have

$$K = \frac{1}{2}mv^2 = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2) = \frac{1}{2}m(\dot{D}^2 + D^2\dot{\theta}^2)$$

$$V = mgD(1-\cos\theta) + \frac{1}{2}k(D-b)^2$$

$$L = K - V = \frac{1}{2}m(\dot{D}^2 + D^2\dot{\theta}^2) - mgD(1-\cos\theta) - \frac{1}{2}k(D-b)^2$$

We then have one equation for each variable(the two-dimensional problem generates two equations)

$$\frac{\partial L}{\partial \dot{\theta}} = mL^2\dot{\theta} \rightarrow \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\theta}}\right) = mL^2\ddot{\theta}$$

$$\frac{\partial L}{\partial \theta} = -mgL\sin\theta$$

$$\frac{\partial L}{\partial \dot{D}} = m\dot{D} \rightarrow \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{D}}\right) = m\ddot{D}$$

$$\frac{\partial L}{\partial D} = -mg(1-\cos\theta) - k(D-b)$$

so that the equations of motion are

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\theta}}\right) - \frac{\partial L}{\partial \theta} = 0$$

$$mL^2\ddot{\theta} = -mgL\sin\theta \rightarrow \ddot{\theta} + \frac{g}{L}\sin\theta = 0$$

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{D}}\right) - \frac{\partial L}{\partial D} = 0$$

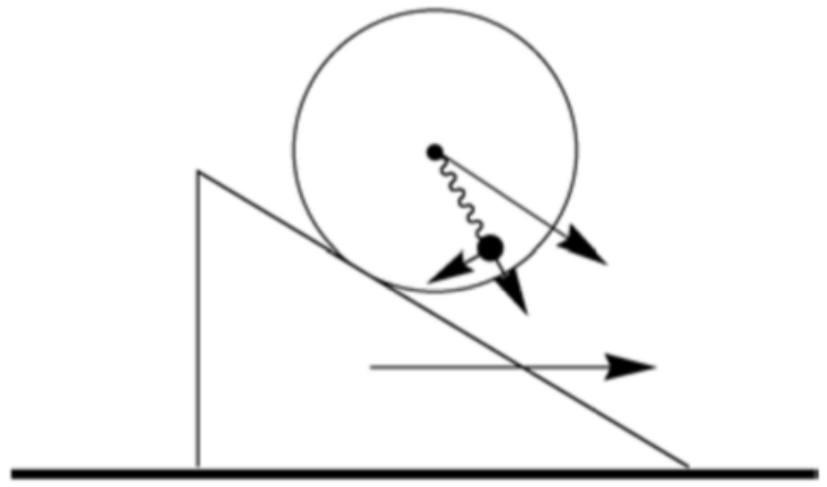
$$m\ddot{D} = -mg(1-\cos\theta) - k(D-b) + mD\dot{\theta}^2$$

$$\rightarrow \ddot{D} + \frac{k}{m}(D-b) - D\dot{\theta}^2 + g(1-\cos\theta) = 0$$

Clearly this is a much more *complicated* problem.

In both last examples, you could have solved problems using force diagrams without new methods.

What *happens*, however, in the case shown in figure



where we have a cylinder with an attached springy pendulum rolling down an inclined plane that can slide along the table with all the motions shown allowed!

This *usually* requires an advanced mechanics course.

Try it you will see that the Lagrangian method is not so difficult!

One final idea I need to introduce is the **Hamiltonian** (or the energy).

Conservation of Energy

The Lagrangian that describes a closed system (i.e., a system not interacting with anything outside the system) cannot depend explicitly on time, that is,

$$\frac{\partial L}{\partial t} = 0$$

so that the total derivative (using coordinates q_i) of the Lagrangian $L(q_i, \dot{q}_i)$ becomes

$$\frac{dL}{dt} = \sum_j \frac{\partial L}{\partial q_j}\dot{q}_j + \sum_j \frac{\partial L}{\partial \dot{q}_j}\ddot{q}_j$$

where the usual term, $\partial L/\partial t$, does not now appear.

But Lagrange's equations are

$$\frac{\partial L}{\partial q_j} = \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_j}$$

Substituting this result for $\partial L/\partial q_j$ above, we have

$$\frac{dL}{dt} = \sum_j \dot{q}_j \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_j} + \sum_j \frac{\partial L}{\partial \dot{q}_j}\ddot{q}_j$$

or

$$\frac{dL}{dt} - \sum_j \frac{d}{dt}\left(\dot{q}_j \frac{\partial L}{\partial \dot{q}_j}\right) = 0$$

so that

$$\frac{d}{dt}\left(L - \sum_j \dot{q}_j \frac{\partial L}{\partial \dot{q}_j}\right) = 0$$

The quantity in the parentheses is therefore constant in time; denote this constant by -*H*:

$$L - \sum_j \dot{q}_j \frac{\partial L}{\partial \dot{q}_j} = -H = \text{constant}$$

If the potential energy *U* does not depend explicitly on the velocities $\dot{x}_{\alpha,i}$ or the time *t*, then *U*= *U*($x_{\alpha,i}$).

The relations connecting the rectangular coordinates and the generalized coordinates are of the form $x_{\alpha,i} = x_{\alpha,i}(q_j)$ or $q_j = q_j(x_{\alpha,i})$, where we exclude the possibility of an explicit time dependence in the transformation equations.

Therefore, $U = U(q_j)$, and $\partial U/\partial \dot{q}_j = 0$.

Thus

$$\frac{\partial L}{\partial \dot{q}_j} = \frac{\partial (T - U)}{\partial \dot{q}_j} = \frac{\partial T}{\partial \dot{q}_j}$$

The equation above defining *H* can then be written as

$$(T - U) - \sum_j \dot{q}_j \frac{\partial T}{\partial \dot{q}_j} = -H$$

One can also show that

$$\sum_l \dot{q}_l \frac{\partial T}{\partial \dot{q}_l} = 2\sum_{j,k} a_{jk}\dot{q}_j\dot{q}_k = 2T$$

Therefore, we have

$$(T - U) - 2T = -H$$

or

$$T + U = E = H = \text{constant}$$

The total energy E is a *constant* of the motion for this case.

The function H, called the **Hamiltonian** of the system

Now we return to our discussion of General Relativity.

When you go from the principle of least action to the Euler-Lagrange equation, the principle of least action *turns* into a differential equation involving a Lagrangian as we discussed.

Typically, when rewritten as explicitly as possible, the Euler-Lagrange equation becomes an equation of the type $F = ma$.

Going from minimizing the quantity in equation (9) to equation (6) is exactly that operation.

In fact equation (6) looks like equating an acceleration to something.

And the thing is a kind of force.

As we will see, sometimes the integral we will be minimizing will be a physical quantity called the **Action** - hence the principle is often *called* the **Principle of Least Action**.

Now let's come back to relativity and to our actual problem of geodesic, where we are not concerned with ordinary distance but with proper time.

If we want to *express* the quantity to be minimized, which involves proper time, we deal with almost exactly the same expression as in equation (9) *except* for a minus sign in front of the metric.

From equation (4) we *get* that the proper time between point 1 and point 2 in space-time is *given* by

$$\tau = \int_1^2 \sqrt{-g_{\mu\nu}(X)\ dX^\mu\ dX^\nu} \tag{10}$$

This is the expression that we will *want* to minimize.

Let's suppose the expression defined by equation (10) really corresponds to the motion of a particle that starts at the space-time point 1 and ends at the space-time point 2, see figure 6.

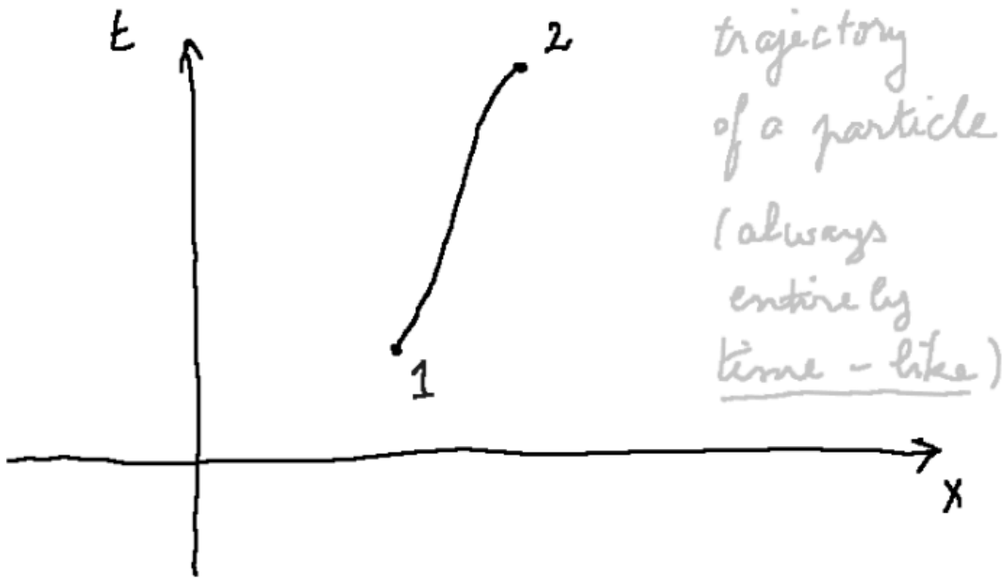


Figure 6 : Trajectory of a particle : geodesic in space-time.

The *action* we are interested in *depends* on one more quantity.

It depends on the *mass* m of the particle.

The *actual* action then is

$$A = -m \int_1^2 \sqrt{-g_{\mu\nu}(X)\ dX^\mu\ dX^\nu} \tag{11}$$

This is a *definition* of the mass.

We will find out that *putting* a coefficient called mass here is important for thinking about energy and so forth.

And the minus sign is *strictly* a convention in the definition of mass.

We want to make this action *A stationary*.

What do we do with the right hand side of equation (11)?

A priori it is a *completely unrecognizable* object to work on with our mathematical toolbox.

Okay, it is an integral.

That is, it is just a *sum* of infinitely many infinitely small elements.

Mathematically speaking, it is more precisely *defined* as the limit, when the number N of elements *goes* to infinity and the *sizes* of the elements go to zero, of finite sums of N elements of the form $f(u_n)\Delta u_n$.

Such a definition is also *due* to Riemann.

But *where* is the differential element ?

What is the variable to integrate over ? Etc.

Usually for us an integral which we know how to calculate, or at least manipulate, has the form

$$\int F(\ some\ variable\ )\ d\ some\ variable$$

where the variable may be some spatial quantity or may be time, or some other clearly identified physical quantity.

Normally we *don't* see integrals where beneath the integral there is a square root, and *inside* the square root there is a product of differentials like $dX^\mu\ dX^\nu$.

We are going to solve it in the standard way, introducing *dt* inside the integral, and even the square root, to arrive at a more familiar expression.

To start with, let's break up the trajectory of the particle into little time segments Δt, figure 7.

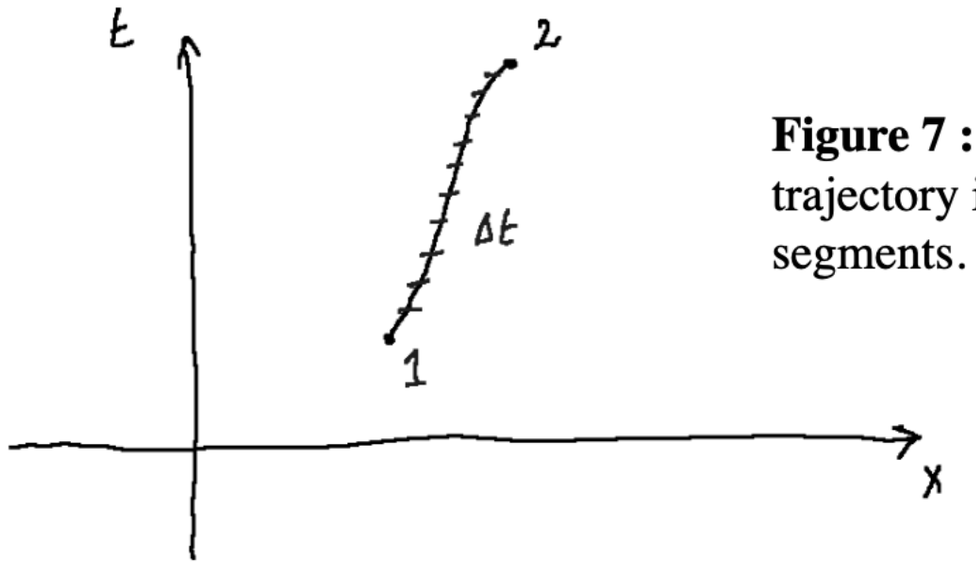


Figure 7 : Breaking the trajectory into little time segments.

Let's skip the finicky maths and make them directly *dt*, because, after a good basic course on calculus, we know that in our circumstances it is allowed.

Equation (11) *becomes*

$$A = -m \int_1^2 \sqrt{-g_{\mu\nu}(X) \, \frac{dX^\mu}{dt} \, \frac{dX^\nu}{dt} \, dt^2} \tag{12}$$

Some of the differentials dX^μ are in fact *dt*, *because t* is one of the four coordinates in $X^\mu = (t, x, y, z)$.

What *happens* when we have *dt*/*dt*?

That is *just* 1.

And what *happens* when we have dx/dt or the analog with y or z ?

It is just the *ordinary* velocity.

We can also *pull* the dt^2 out of the square root, and obtain a standard differential element dt under the integral sign :

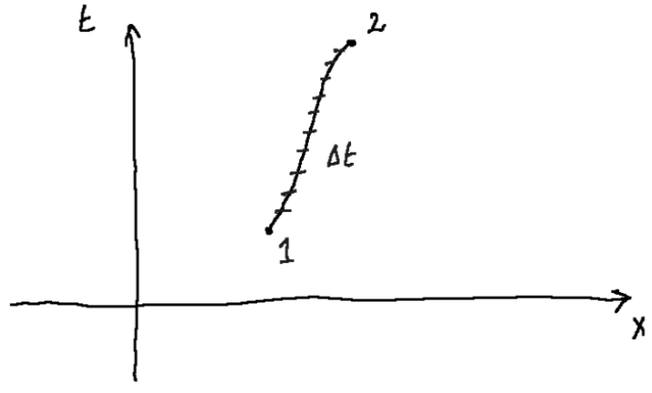
$$A = -m \int_1^2 \sqrt{-g_{\mu\nu}(X) \, \frac{dX^\mu}{dt} \, \frac{dX^\nu}{dt}} \, dt \tag{13}$$

At *each* time t, the quantity $\sqrt{-g_{\mu\nu}(X) \, \frac{dX^\mu}{dt} \, \frac{dX^\nu}{dt}}$, which is integrated over time, has a *definite* value along the trajectory.

It is a *certain function* of the velocity and the position X.

So we have *transformed* our expression (11) for the action into a conventional integral over time along the trajectory in figure 7.

The *integrand* in equation (13) is the Lagrangian.



It is the *quantity* which, in the calculation of a geodesic, plays *exactly* the same role as the lagrangian when we apply the principle of least action to calculate the trajectory of a particle in classical non-relativistic physics.

Action by *definition* is equal to the integral of the lagrangian, which is itself a function of velocities and positions.

$$A = \int \mathcal{L}(\dot{X},\ X)\ dt \tag{14}$$

In summary, with our problem of making stationary the action given by equation (13), we return to a problem that we already discussed in classical mechanics.

How to *calculate* an equation of motion from an action ?

In order to do that we *solve* the Euler-Lagrange equation (or equations) that the lagrangian must satisfy.

In our present problem, the lagrangian is

$$\mathcal{L} = -m\sqrt{-g_{\mu\nu}(X)\ \frac{dX^{\mu}}{dt}\ \frac{dX^{\nu}}{dt}} \tag{15}$$

A good question: is the *quantity* inside the square root *positive*?

Can we *take* its square root, even though there is a minus sign in front of $g_{\mu\nu}$?

Answer : *yes*, it is positive.

The quantity $-g_{\mu\nu}(X)\, dX^{\mu}\, dX^{\nu}$ is the *square* of the proper time over a small element dX along the trajectory, see equation (1).

It is *always* positive for a *time-like* trajectory.

And particles *always move* on time-like trajectories.

Real particles *do not move faster than the speed of light*.

In case you have forgotten, a space-like trajectory is one where the point moves faster than the speed of light.

Let's recall what the Euler-Lagrange equations, that the lagrangian must satisfy, are.

First of all we are going to partially differentiate $\mathcal{L}$ with respect to each of the variables $\dot{X}^{\mu}$.

But the first of these variables, which is the derivative of time with respect to time, is just one.

So there is no correponding equation.

We are only concerned with the three partial derivatives of $\mathcal{L}$ with respect to the components of ordinary velocity.

On the left hand side, for each of these partial derivatives, we take the derivative with respect to time, and equate it to the partial derivative of $\mathcal{L}$ with respect to the corresponding component of position.

Therefore the Euler-Lagrange equations are the three following equations

$$\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{X}^m} = \frac{\partial \mathcal{L}}{\partial X^m} \tag{16}$$

where *m* runs from 1 to 3.

We *learned* them earlier.

The *point* is : if you know the metric $g_{\mu\nu}(X)$, you *can work out* from equation (16) an equation of motion for the particle.

And the particle's motion *will be a geodesic* in the sense of the trajectory with shortest proper time integrated along the curve.

How does this relate to the definition of a geodesic given by equation (6)?

Answer : if you work out exactly equation (16) with the given metric $g_{\mu\nu}(X)$, you will discover that you *end* up exactly with equation (6).

I will not do the calculation here but I will reiterate that it can be shown that given the metric $g_{\mu\nu}(X)$, one can show that the Euler-Lagrange equations (16) which minimizes the proper time along a trajectory in space-time,

$$\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{X}^m} = \frac{\partial \mathcal{L}}{\partial X^m}$$

where $\mathcal{L}$ is given by equation (15), are *equivalent* to the definition of a geodesic given by equations (6), which say that the tangent vector to the trajectory in space-time stays *constant*,

$$\frac{d^2X^m}{d\tau^2} = -\Gamma^{\mu}_{\sigma\rho}\ \frac{dX^{\sigma}}{d\tau}\ \frac{dX^{\rho}}{d\tau}$$

It is generally true that to work with the action defined by equation (13), and with the Euler-Lagrange equations (16) that come from minimizing this action, is much *easier* than to work with a geodesic defined by equations (6).

We are *going* to do it in this lesson.

We are *going* to work out some equations of motion for particles in a particular metric.

The metric will be everybody's favorite metric : the *Schwarzschild* metric.

Schwarzschild metric

We now come to the problem of studying the metric and the motion of a particle in a *real* gravitational field, the gravitational field of the *Sun*, the *Earth* or a *black hole* or whatever you have, in other words a *massive spherically symmetric object*.

Figure 8 : Massive object and its gravitational field.

We are *outside* the mass of the object, *far away* from it, figure 8.

And we are *interested* in the metric of space in there.

So first of all we are going to *write* a formula for the metric and then we will *check* and see that this formula really does make sense.

By that we mean that, using the equations of motion (16), it would *give rise* to something that looks very familiar, namely Newton's equations for a particle moving in a gravitational field – at least when we are *far away* from the gravitating object, where the gravitation is fairly weak.

Here is the metric we are going to use.

First of all if we were in a *flat space*, it would have the form

$$\tau^2 = dt^2 - \frac{1}{c^2}\, dX^2 \tag{17}$$

where henceforth dX^2 *stands* for $dx^2 + dy^2 + dz^2$.

We *expect* indeed, as we go far away from the gravitating object, that the space-time there look *flat*.

But we know that the gravitating object does *something* to space-time.

So equation (17) *should not be quite right* for the metric of the space-time created by the object.

It should be right *only in the limit* when we are far away.

Therefore we *add a coefficient* in front of dt^2, and we leave the part with dX^2 as it is.

$$\tau^2 = \left(1 + \frac{2U(X)}{c^2}\right) dt^2 - \frac{1}{c^2}\, dX^2 \tag{18}$$

$U(X)$ is the *gravitational potential* due to the object in figure 8.

Let's *check* that, as long as our particle is moving slowly, its geodesic equation, in the lagrangian form (16), *just becomes* Newton's equation for a particle moving in a gravitational potential $U(X)$.

The general form of the lagrangian is given by equation (15), where inside the square root is just $d\tau^2$.

Now that we know the metric, from equation (18), we can write the lagrangian or the action more *explicitly*

$$A = -mc^2 \int \sqrt{\left(1 + \frac{2U(X)}{c^2}\right) dt^2 - \frac{1}{c^2}\, \frac{dX^2}{dt^2}\, dt^2} \tag{19}$$

Noting that $\frac{dX^2}{dt^2}$ is $\dot{X}^2$, and doing a bit of cleaning, this is the same as

$$A = -mc^2 \int \sqrt{\left(1 + \frac{2U(X)}{c^2}\right) - \frac{\dot{X}^2}{c^2}} \; dt \tag{20}$$

When we work with c explicitly present in the formulas, the expression for the action must carry a c^2 next to the mass.

We are familiar with that.

It can be derived from reasoning on dimensions, or we can remember that the action has units of energy multiplied by time.

The most basic action in classical mechanics is $A = \int \frac{1}{2} m v^2 dt$.

We could work with the lagrangian, in equation (20), as it is, but what we are interested in is what happens when we move with slow speed relative to the speed of light.

So the next step is to approximate the lagrangian, and look at the Euler-Lagrange equations with the approximate lagrangian.

As usual, we rely on the mathematical fact that, in this case, it is okay to approximate the lagrangian first and then to solve the Euler-Lagrange equations rather than solve properly the Euler-Lagrange equations first and then look at the limit case when $\dot{X}$ is small.

In other words, we can invert two big operations.

We are interested in slow motion because we want to show that equation (20) really does give rise to Newton's equations in the non-relativistic limit.

The non-relativistic limit is the one where c is taken to be very large.

We could be tempted to just erase anything with a $1/c^2$ beneath the integral sign, but – boy ! – this would kill just about everything in equation (20).

Inside the square root we can reorganize terms as one plus a small quantity :

$$\sqrt{1 + \frac{1}{c^2}\left(2U - \dot{X}^2\right)} \tag{21}$$

Next we use the binomial theorem :

$$\sqrt{1+\epsilon} \approx 1 + \frac{\epsilon}{2}$$

So expression (21) can be approximated by

$$1 + \frac{1}{2c^2}\left(2U - \dot{X}^2\right) \tag{22}$$

Going back to the expression of the action, in equation (20), we obtain

$$A = \int \left(-mc^2 - mU + \frac{m}{2}\dot{X}^2\right) dt \tag{23}$$

Inside the integral, we have the lagrangian when the speed of the particle is small, that is, when we can make a non- relativistic approximation.

If we use this lagrangian in the Euler-Lagrange equations, the constant term $-mc^2$ has no effect.

The only thing we do with a lagrangian is differentiate it.

When we differentiate a constant we get zero.

So we can disregard that term.

The other two terms in the lagrangian are a conventional kinetic energy, $\frac{m}{2}\dot{X}^2$, minus a potential energy, $mU(X)$, which actually depends on X.

For the potential energy we get to choose the function we like.

Incidentally in a gravitational problem the potential energy of a particle is always proportional to its mass.

Finally, when we use this lagrangian, the Euler-Lagrange equations will of course simply produce Newton's equation for a particle in a gravitational field $U(X)$, just as they did with exactly the same example calculations in classical mechanics.

The equation will be

$$m\ddot{X} = -m\frac{\partial U}{\partial X} \tag{24}$$

The term on the right hand side is a force.

On the left hand side is an acceleration multiplied by the mass.

And the mass cancels.

The main point here was that the action given by equation (13), which is equivalent to the geodesic given by equation (6), is easily worked out just by thinking of the Euler-Lagrange equations.

And it is even *easier* to work out if you are in the *non-relativistic* limit, where you just say c is very large, $1/c^2$ is very small, and just *expand* the square root in the expression of the action.

But the *important* point that we *learned* in doing the above calculation is that in some first approximation – or perhaps no approximation – we can write

$$-g_{00} = \left(1 + \frac{2U(X)}{c^2}\right) \tag{25}$$

Of course in the equation (18) we wrote for τ^2, there *could be even smaller terms* than $1/c^2$, terms in $1/c^4$ or $1/c^6$, etc.

But they would *not be important* for in non-relativistic limit.

So we *cannot* say with complete confidence that $-g_{00}$ is one plus twice the potential energy of a particle divided by c^2.

But we *can* say that it must be true to the *first order* in small quantities – small quantities meaning quantities with one over c^2.

Now for the gravitational potential energy created in space by a body of mass M, see figure 8, we shall use

$$U(X) = -\frac{MG}{r} \tag{26}$$

where G is Newton's constant, and r is the distance away for the center of the body.

Now we can write down our *first guess* at the metric of space-time surrounding a gravitational mass in figure 8.

Here it is :

$$d\tau^2 = \left(1 - \frac{2MG}{c^2\ r}\right) dt^2 - \frac{1}{c^2}(dx^2 + dy^2 + dz^2) + \ldots \tag{27}$$

where $r = \sqrt{dx^2 + dy^2 + dz^2}$

And the ... after the spatial terms stand for *smaller* things, that is things one or more orders of magnitude smaller than $1/c^2$.

Is that beginning to look *familiar*?

You have some familiarity with cosmology and thus may *recognize* the Schwarzschild metric, or the metric of a black hole – but not quite.

Next, let's *consider* the term

$$dx^2 + dy^2 + dz^2 \tag{28}$$

It is the *ordinary* metric of three-dimensional space.

In equation (27) defining the metric of space-time created by a massive body, *only* the time-time component of the metric has been *fiddled* with.

The space-space components of the metric and everything else have *not yet been fiddled with*.

They will in a moment but *not too much*.

The space-space components in equation (27) *are* the metric of the ordinary flat space.

Let's take flat space in *three dimensional polar coordinates*.

These coordinates are *characterized* by a radius, namely the distance from the center of the Sun – if we think of the body in figure 8 as the Sun –, and a pair of angles.

They *can* be a polar angle and an azimuthal angle.

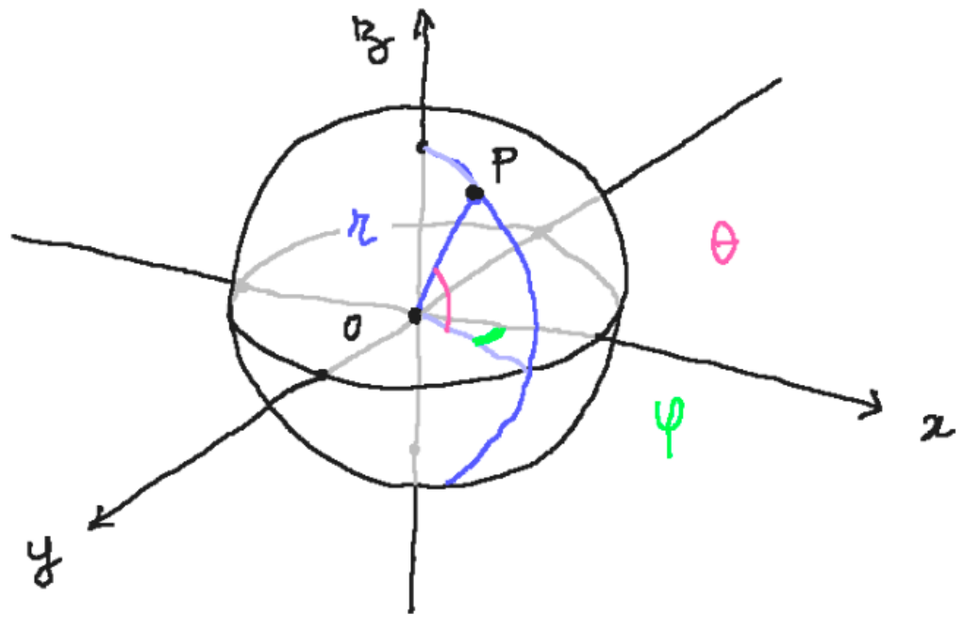


Figure 9 : Spherical polar coordinates.

r is the distance from the Sun O, to P.

The angle θ can be taken from the pole or from the equator.

We *choose* to measure it from the equator.

It is also called the *latitude*.

And ϕ, the azimuthal angle, is also called the *longitude*.

How is the ordinary 3D Euclidean metric $dx^2 + dy^2 + dz^2$ expressed in spherical polar coordinates?

The *formula*, that you may have learned in trigonometry at the end of high school, is

$$dx^2 + dy^2 + dz^2 = dr^2 + r^2\ (d\theta^2 + cos^2\theta\ d\phi^2) \tag{29}$$

Despite its complicated look, it is *simply* the square of the length element on the surface of the sphere, $r^2\ (d\theta^2 + \cos^2\theta\ d\phi^2)$ that we have already met, to which is added dr^2 to *complete* Pythagoras theorem in three dimensions.

Why are we considering polar coordinates?

Answer : simply because polar coordinates are the *good* coordinates for studying the central force problem.

We don't want to use x, y and z to study the motion of a particle in a gravitational field.

We *want* to use the polar coordinates.

So we are going to write the metric given by equation (27) in terms of *polar coordinates*.

We are going to *make* one change : we will give $d\theta^2 + \cos^2\theta\ d\phi^2$ a name, so we don't have to write it over and over.

We *call* it $d\Omega^2$.

There is no other reason than to avoid writing it all the time in its full expression with θ and ϕ.

Equation (27) can be rewritten

$$d\tau^2 = \left(1 - \frac{2MG}{c^2\ r}\right) dt^2 - \frac{1}{c^2}\ dr^2 - \frac{1}{c^2}\ r^2\ d\Omega^2 + ... \tag{30}$$

The variable r is just the *radial* distance.

And $d\Omega^2$ is the *metric on the surface of a sphere*.

If you keep r fixed, the *term* $r^2\ d\Omega^2$ is the *contribution* to the space-time metric when you move some angles θ and ϕ on the sphere.

Now if we *look* at this metric, given by equations (30) or (27), there is something *terribly wrong*.

It is *fine* when we are *far away*.

But there is *something deeply wrong* when we get *close to the center*.

What *happens* when r becomes *very small*?

At some point $2MG/c^2r$ *becomes bigger than one*, and the *coefficient*

$$(1 - 2MG/c^2r)$$

becomes negative.

In other words, the time-time part of the square of the proper time $d\tau$, the coefficient $-g_{00}$, becomes *negative*.

Since the other terms – the space-space parts – in the definition of $d\tau^2$ given by equations (30) or (27) are *also negative*, we run into a *problem*.

The metric tensor or matrix $g_{\mu\nu}(X)$ then *only has positive terms* on its diagonal (and the terms off diagonal are *all zero*).

Therefore it has four positive eigenvalues *instead* of one negative eigenvalue and three positive eigenvalues.

Where does the problem arise?

It *arises* when $1-2MG/c^2r$ *changes* sign, that is when

$$r \geq \frac{2MG}{c^2} \tag{31}$$

The quantity $2MG/c^2$ is *just* some positive number, *specific* of the body we are looking at in figure 8.

And the metric *no longe*r has three positive eigenvalues and one negative.

If you have studied the Earth's gravitational field you know if we enter inside the Earth, we will only be submitted to the gravitational field created by the mass at radius shorter than where we are, the mass farther away than us, all around the Earth, *will play no role*.

Of course this *happens* if all the mass of the body is somehow point-like or almost point-like.

This *leads* us to the topic of black holes.

Black holes

A *black hole* is a massive body whose mass M is *extremely concentrated* in a very small sphere.

It is concentrated in a sphere of radius *smaller* than the value $2MG/c^2$, which we *discovered* to be causing a problem in the metric.

$2MG/c^2$ is a distance *characteristic* of the black hole.

It is called the *Schwarzschild radius* of the black hole.

When we get closer and closer to the black hole, at some point, when we *cross* the Schwarzschild radius, the coefficient $-g_{00}$ *changes sign and becomes negative*, or equivalently g_{00} becomes positive, and the whole metric of equation (30) now has *four positive eigenvalues*.

The *meaning* of that is we somehow passed into a region where there are *four space directions and no time direction*.

That is bad!

That is something we *don't want*.

It is *not a good thing*.

What really happens?

When we study the *field equation*s of general relativity – *which we are going to do* – we will discover *really what happens*.

For the time being let me just say that *at the point where* $1 - 2MG/c^2r$ *changes sign*, the rest of equation (30) *will also change*.

We will put an *extra coefficient* in front of the term dr^2/c^2 in equation (30) to the *effect* that when the time-time component of the tensor changes sign, the *r-r* component will *also change sign*.

So what happens?

The term $1 - 2MG/c^2r\ dt^2$ *turns into a space direction*, but the term with dr^2 *turns into a time direction*.

The *signature* of the metric is *maintained* : one time-like direction and three space-like directions.

What can we put, in equation (30), in front of the dr^2/c^2 *to make it flip sign*, when the time-time component flips sign?

Well, we could *also* put in $1-2MG/c^2r$ in front of the dr^2/c^2.

So if we cross the Schwarzschild radius the first term changes sign and the dr^2 term changes signs too.

But that *turns out not to be quite the right thing*.

Einstein's field equations give us a *different* answer.

Instead of using the same coefficient as in front of dt^2, use its *inverse*.

From now on *disregarding* the smaller terms in $1/c^4$ or more, the metric *becomes*

$$d\tau^2 = \left(1 - \frac{2MG}{c^2\ r}\right) dt^2 - \left(\frac{1}{1 - \frac{2MG}{c^2\ r}}\right) \frac{1}{c^2}\ dr^2 - \frac{1}{c^2}\ r^2\ d\Omega^2 \tag{32}$$

When we cross the Schwarzschild radius we do have *both* coefficients changing sign at the same time.

Therefore the signature $-+++$ of the metric is *preserved*.

But something very odd happens : the thing that we are calling *t becomes* a space-like direction, and the thing that were calling *r becomes* a time-like direction.

They flip.

This is not easy to visualize but *I will give you the tools to visualize it* as we go along.

Something *happens* as we cross a certain *threshold* – the *surface of a sphere of radius r equals* $2MG/c^2$.

The radius *r becomes* a time variable, and the time *t becomes* a space variable.

This is completely *mysterious* for the moment, but *we will explain in some geometric detail* what is happening at the Schwarzschild radius.

We *will discover* that, *when* we are moving radially toward the black hole, to cross the Schwarzschild radius *takes an infinite amount of coordinate time t but a finite amount of proper time* τ.

In other words, *somebody falling, with a wristwatch, will say it takes him or her a finite amount of time to cross that threshold*.

But somebody, *watching from the outside that person fall through, will say it takes an infinite amount of time*.

That is a *characteristic* of the metric given by equation (32).

And that is one of the things *we want to work out* in this lesson.

Event horizon of a black hole

Let's *think about* where it is happening that this peculiar phenomenon of time and space switch.

It is *happening* at coordinate $r = 2MG/c^2$.

The speed of light c is a *huge* number, and its *square* is in the denominator.

So it is a *very small distance*, but *only if M* itself is not huge because the Schwarzschild radius depends on the mass of the gravitating object.

For the Earth, if its mass were concentrated in a smaller distance than that, this distance would be about a *centimeter*.

When we cross the Schwarzschild radius, you may *wonder* what happens to the wristwatch.

Does it *begin* to record spatial distance?

Answer : *no*.

It *keeps ticking* in a time direction.

Clocks *don't care* about what coordinates we use.

We are going to discover that the phenomenon of time and space switch is an *artifact* of our coordinate system.

There is *nothing special* going on at the Schwarzschild radius.

We will see that (t, r, θ, ϕ) are *just awkward coordinates* that make it look like something funny is happening at $r = 2GM/c^2$.

In truth there is *nothing funny happening* at the Schwarzschild radius.

We are going to *work out and see exactly why*.

But for the moment it *does look like* there is something going on there.

Looking at the metric tensor *expressed* in the coordinates (t, r, θ, ϕ), when the g_{tt} coefficient *vanishes*, the g_{rr} coefficient becomes *infinite*.

It looks like some *terrible things happen* to the geometry.

But in fact the *geometry is completely smooth* over the threshold $r = 2GM/c$.

There is *nothing special happening*.

The *light cones are perfectly healthy all over the region around the Schwarzschild radius*.

There is *nothing special* going on.

But we have to work harder to see it.

Here is some *historical note*.

Einstein *submitted* his field equations to the Prussian Academy of Science on November, 25 1915.

Then Schwarzschild *studied* Einstein's paper and, in December 1915, while fighting on the Russian front during WWI, *wrote* down equation (32) for the metric of a gravitating body (outside the body).

He unfortunately *died* shortly afterwards.

What did Einstein already know of it?

He *knew about* the first term in front of dt^2 and he knew *part of the second term* in front of dr^2 in equation (32).

To *understand* what we mean, let's *examine* this second term when r is *not too small*, that is when $2MG/c^2r$ *is small*.

Using our *favorite tool* to make approximations – the *binomial theorem* –, the *coefficient* in front of dr^2, in equation (32), can be *expanded* as follows

$$\frac{1}{1-\frac{2MG}{c^2\ r}} = 1 + \frac{2MG}{c^2\ r} + \left(\frac{2MG}{c^2\ r}\right)^2 + \ldots \tag{33}$$

On the *right hand side* the first term after 1 has $1/c^2$, the second term $1/c^4$, etc.

In other words, *what we are doing*, in equation (32), is *correcting* the dr^2/c^2 of equation (30) with corrections *that are very small* – of the order of magnitude of $1/c^4$ or smaller.

They *contribute* to the motion of the particles, *but not* in the non-relativistic limit.

Let's *close* this historical note, and *come back* to the metric of equation (32).

Before we analyze it and see what is going on at a point where we cross the horizon of the black hole – another name for the surrounding sphere of Schwarzschild radius –, *let's use it* and see if we can figure out *how particles move in the presence of this metric*, not in the non-relativistic limit but without taking any approximation.

It is *not that hard*.

By the way, if we are not going to do any approximation, we might as well set c equal to *one*.

It is just a *choice* of units.

Then the metric *looks* like this

$$d\tau^2 = \left(1 - \frac{2MG}{r}\right) dt^2 - \left(\frac{1}{1 - \frac{2MG}{r}}\right) dr^2 - r^2 d\Omega^2 \tag{34}$$

Two kinds of trajectories are easy to study: *circular* orbits – they reserve some surprise if you do them –, and *radial* trajectories.

In this lesson, let's study a *radial* trajectory, a spacecraft falling along a *straight* line and a priori going to *crash* on the black hole.

Figure 10 : Radial trajectory of a spacecraft falling toward the black hole.

The *question* is : *how long* does it take for the spacecraft to reach the horizon?

We are going to *measure* this duration *from two points of view*: that of an *external observer*, using *coordinate time t*, and that of an *astronaut* in the spacecraft, using the *proper time* τ of the spacecraft and himself.

The results are very different.

The *main point* of these calculations is to show you how to work with the metric of equation (34).

When you have to do any such calculation, you always *start* with the lagrangian for the particle.

Then you *work out* the Euler-Lagrange equations or whatever you need to do with the lagrangian.

And you *solve* the equations.

There are *tricks* you can do.

I'm going to show you some of the tricks that *allow you* to do the problem.

So we *go back* to the lagrangian for the in-falling particle.

Remember that particles follow geodesics in space-time, so, *from equation* (11), the lagrangian is just $-m\sqrt{d\tau^2}$.

We *get* $d\tau^2$ from equation (34).

Assuming that the spacecraft goes straight and that Ω *doesn't change*, we can *omit* the term $r^2 d\Omega^2$.

Therefore the *action*, that must be stationary, is

$$A = -m \int \sqrt{\left(1 - \frac{2MG}{r}\right) dt^2 - \left(\frac{1}{1 - \frac{2MG}{r}}\right) dr^2}$$

As usual, to *make sense* out of this integral, below the integral sign we divide inside the square root by dt^2, and multiply outside the square root by dt.

$$A = -m \int \sqrt{\left(1 - \frac{2MG}{r}\right) - \left(\frac{1}{1 - \frac{2MG}{r}}\right) \frac{dr^2}{dt^2}} \; dt$$

The ratio dr^2/dt^2 is just the *radial velocity* squared, that is $\dot{r}^2$.

So we got our *lagrangian*

$$\mathcal{L} = -m \sqrt{\left(1 - \frac{2MG}{r}\right) - \left(\frac{1}{1 - \frac{2MG}{r}}\right) \dot{r}^2} \tag{35}$$

The *coordinate* in this case is r.

And *this* is $\mathcal{L}(\dot{r}\ , r)$.

Fortunately *applying* the Euler-Lagrange equations to compute the trajectory is easy.

Or *at least* it is easy to see what is happening at the Schwarzschild radius.

Let's *not worry* about the exact solutions.

Exactly solving the problem is a little bit *hard*.

But seeing what happens is *easy*.

There is *always* a conserved quantity.

What is the conserved quantity that is always there?

Energy.

What is *energy* in terms of the lagrangian?

The *energy* is the *hamiltonian*.

And the hamiltonian is *expressed* in terms of the lagrangian.

To *compute* the hamiltonian, the first thing we do is *calculate* the generalized conjugate momentum to r.

It is

$$p = \frac{\partial \mathcal{L}}{\partial \dot{r}}$$

In the *general* case, when instead of one coordinate r we have a collection of coordinates q_{i} and their conjugates p_{i}, the general formula for the hamiltonian is:

$$H = \sum_i p_i \dot{q}_i - \mathcal{L}$$

In *our* case it becomes simply

$$H = p\dot{r} - \mathcal{L}$$

We will *not* do the detailed calculations.

The *result* is

$$H = \frac{m\ (1 - 2MG/r)}{\sqrt{(1 - 2MG/r) - \dot{r}/(1 - 2MG/r)}} \tag{36}$$

It is a *complicated* ugly expression.

But it is a *definite* thing.

What does it depend on?

It *depends* on the mass m of the particle or spacecraft, the distance r to the black hole center, and the velocity $\dot{r}$.

It is the *energy*.

We know that this energy is *conserved* over time.

We can call it E instead of H, and it *does not* change with time.

Then equation (36), giving the energy, *enables* us to express $\dot{r}$ as a function of that energy E.

With a little bit of *algebra* we get

$$\dot{r}^2 = (1 - 2MG/r)^2 - \frac{(1 - 2MG/r)^3}{E^2} \tag{37}$$

This expression again looks *unwieldy*.

But it *doesn't* really matter.

What is *important* is that we can *easily* see what *happens* when r gets *near* $2MG$, that is, when the spacecraft gets *close to the horizon* of the black hole.

Equation (37) *tells us* that as r gets near $2MG$, in other words as the spacecraft *approaches* the horizon of the black hole, its *velocity slows down to zero*.

Contrary to intuition, when the spacecraft *falls towards* the Schwarzschild radius, *instead of accelerating*, its velocity gets *smaller and smaller*.

We can show that from equation (36) for the energy, and equation (37) for $\dot{r}^2$, it *follows* that

$$\dot{r} \approx \sqrt{\frac{r - 2MG}{2MG}} \tag{38}$$

Equation (38) offers *another* way to see that as r approaches $2MG$ the spacecraft *decelerates*.

It is a surprise because we *might think* that the thing would accelerate like, in elementary physics, a particle would do *falling* toward a massive body.

But *instead* it *slows* down.

Notice that we *still* are in classical physics in the sense of *not quantum*.

The *only special things* from elementary physics are

1. particles move along geodesics in space-time
2. we look at a massive body whose mass is *extremely* concentrated
3. we look at what happens at the Schwarzschild radius, which is ordinarily very small (recall that it is $2MG/c^2$)

The speed goes *asymptotically to zero*, and the spacecraft or particle itself *never passes* the Schwarzschild radius.

You might say figuratively that *it takes forever to reach the horizon*.

Let's *go back* to the Schwarzschild metric given by equation (34).

Let's *write it again* below leaving aside the $d\Omega^2$ piece which does not play any role

$$d\tau^2 = \left(1 - \frac{2MG}{r}\right) dt^2 - \left(\frac{1}{1 - \frac{2MG}{r}}\right) dr^2 \tag{39}$$

How does a radial light-ray move?

Think of a light-ray beamed from some point either toward the center of the black hole or moving away, it doesn't matter for our analysis: *it moves radially*.

Figure 11 : Radial light-ray beamed from r in each direction.

The *trajectory* of the light-ray is a *solution* to equation (38) that says that it is *light-like*.

Light-like *means* that $d\tau^2 = 0$.

So a light-ray *satisfies*

$$\left(1 - \frac{2MG}{r}\right)^2 dt^2 = dr^2$$

or *equivalently*

$$\frac{dr}{dt} = \pm\left(1 - \frac{2MG}{r}\right) \tag{40}$$

That is a light-ray.

It is the *fastest* thing that can ever be.

When r is *big* the right hand side of equation (40) is *almost* 1, that is the *speed* of light in our units.

So there is *nothing troublesome* there.

But as r *decreases* to the Schwarzschild radius, we see that the *magnitude* of the radial velocity, which is $(r - 2MG)/r$, *goes to zero*.

In other words, *even* a light-ray somehow *gets stuck* as it *approaches* the black hole horizon !

Is it moving with the speed of light?

Answer : *Of course it is* moving with the speed of light.

What else can a light-ray do?

But the speed of light *has this property* that in the *particular set of coordinates* (t, r, θ, ϕ), the speed of light measured by dr/dt *goes to zero* as you get closer and closer to the Schwarzschild radius.

Therefore nothing even including a light-ray *can get past the surface* of the horizon... *or so it seems*.

The velocity goes to zero at the horizon, *whether* we are incoming or outgoing.

So wherever we are, *something odd* happens beyond the horizon.

And *we are going to work that out*.

Now let's *turn to someone moving with the spacecraf*t of figure 10, or with a photon of figure 11.

From the *point of view of a person falling in*, time is the *proper time* τ.

He or she *will just go sailing right through* the Schwarzschild radius.

The phenomenon of dr/dt going to zero is an *artifact* at $r = 2MG/c^2$ of the *peculiar* spherical polar coordinates used for the stationary frame.

There is *nothing really going* on that is strange at that distance.

And the person *will reach the black hole in a finite amount of proper time*.

We will see this phenomenon in *more details in lesson 6*.

In particular we will see the *relation between stationary time t and proper time* τ.

Equation (39) *shows* that when r approaches the Schwarzschild distance, the coefficient $(1 - 2MG/r)$ *goes* to zero.

When the coefficient is going to zero, a *given amount of dt corresponds* to a smaller and smaller amount of proper time $d\tau$.

As a consequence, *proper time in some sense "slows down"*.

That is *why* something can take an infinite amount of time in one frame (the frame where time is t), and a finite amount of time in the other frame (the frame where time is τ).

One last point that we should emphasize is that if we were really talking about the Sun or the Earth, these bodies are *not compact* enough to be black holes.

A black hole is a body whose mass is *contained within* its Schwarzschild radius.

The figure is about *3 kilometers* for the Sun, and about *9 millimeters* for the Earth.

However the *Earth radius* is 6000 kilometers.

Equation (34) for the metric, or equation (37) for $\dot{r}$ are *only valid outside the gravitating body itself*.

It is *only* if the Earth somehow would collapse within a radius of 9 millimeters, keeping its mass, that there *would be* a Schwarzschild radius to speak about, leading to *dr*/*dt* going to zero and so forth.

Same thing for the Sun: only if it *collapsed* within a sphere less than 3 kilometers in radius would it have a black hole horizon.

Inside the gravitating body, the metric is *different* from equation (34).

And *nothing bizarre happens* in the center of the Earth.

The center of the Earth *does not have* a black hole horizon.

To have the *complete metric in space-time* created by a gravitating body you have to *piece together the metric outside and the metric inside the body*.

Black holes do have peculiarities.

But they are *somewhat different* from what popular science says.

For example, in popular science, it is said that we can throw things – tissues, cars, planets, anything – into a black hole, *and make it grow*.

But I showed you that these things thrown at the black hole *will never get there*.

So there is a *problem*.

Next chapter will be *devoted* to deepening our knowledge of black holes.

Lesson 6 : Black holes

Schwarzschild metric

Last lesson, we began discussing the properties of a Schwarzschild black hole.

Let's *review* quickly what we did, before going on.

The S*chwarzschild solution is an idealized solution* in the *sense* that the Newtonian force law

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

is an idealization under the *assumption* that the mass creating the gravitational field is a *point* at the center of coordinates.

If the mass is spread out then of course we *don't really believe equation (1) to hold* in the interior of where there is mass.

And in the *interior* of the Earth equation (1) is *not correct*.

But it is *true in the exterior*, out beyond the surface of the Earth at least if we ignore the atmosphere and the other forms of gravitating mass.

The *point mass* in Newtonian mechanics is an idealization because you never really ever reach the volume of a point mass.

Real materials have a certain *stiffness*.

You can compress them but you can *never* squeeze them to infinitely small radius.

They just have a certain *resistance* to being squeezed.

And gravity is *only so strong*, even though the gravitational field appears to get very very large at short distances.

The nature of the material is always such that if you *squeezed* it down to arbitrarily small distance, it would *spring back*.

This is *not so in the general theory of relativity*.

We are going to see an *example* where, if matter gets too close to the singularity at $r = 0$, it gets sucked in a way that *doesn't happen* in Newtonian gravity.

For *example* in Newtonian physics, if we have a force center, and we shoot a particle with infinite precision along a radius toward the force center, it *will indeed hit the center*.

But if we have the smallest deviation away from perfection and we don't quite – even with a very small impulse – shoot in a radial direction, the particle *will not hit the center*.

It will *swing around* it as in figure 1.

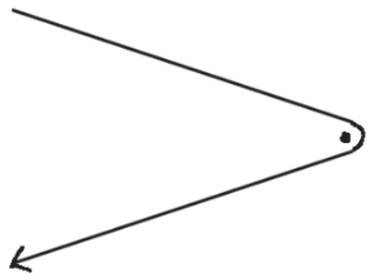


Figure 1 : Shooting a particle toward a force center without perfect precision, in Newtonian physics.

Effectively there is a kind of force repelling the particle from the center.

It is a *centrifugal* force.

If your aim is not perfect, the particle will have a little bit of *angular momentum* relative to the center, and the effective centrifugal force *will keep* it from going to the center.

So in Newtonian mechanics, even though the center pulls very hard when you get close to it, nevertheless it is infinitely unlikely for an in-falling point particle to actually hit the center.

The situation is *different* in the general theory of relativity.

Gravity is *even stronger.*

It is so strong that it *overwhelms* the centrifugal barrier, and *pulls* anything into the center.

Let's now *go back* to the Schwarzschild metric that we began to study in the preceding lesson.

Notice that we *did not derive* the metric.

We *just wrote it* with partial heuristic justification.

Later, when we establish Einstein's field equations in lesson 9, we *will show* that the Schwarzschild metric is a solution to them.

But *so far* for us it is a specific metric that is a given, a specific geometry, and we are going to analyze the *consequences* of the metric.

The metric is *given* by this equation

$$d\tau^2 = \left(1 - \frac{2MG}{r}\right) dt^2 - \left(\frac{1}{1 - \frac{2MG}{r}}\right) dr^2 - r^2 d\Omega^2 \tag{2}$$

Remember that $d\Omega^2$ is just the ordinary metric on the unit sphere expressed with two spherical angles, the angle θ from the equator and the azimuthal angle ϕ (see figure 9 of lesson 5).

$$d\Omega^2 = d\theta^2 + cos^2\theta \; d\phi^2 \tag{3}$$

Remember also that as *r* crosses the value $2MG$ the coefficient in front of dt^2, in equation (2), *changes sign*, but *so does* the coefficient in front of dr^2, therefore the signature of the metric dS^2 (which is the opposite of $d\tau^2$) *remains* one negative sign and three positive signs.

Regarding the two spherical angles, θ and ϕ, *sometimes* we may be interested in only one of them.

For *instance*, when we study an orbiting particle, we are going to locate it in space-time with t and ϕ.

Radius r will remain constant and $\theta = 0$. (We could also fix ϕ and consider θ as the variable angle of the orbiting particle.)

Schwarzschild radius or black hole event horizon

Let's draw a picture of the metric, with the time axis t vertical, and *two* spatial axes x and y forming a horizontal plane at $t = 0$, figure 2.

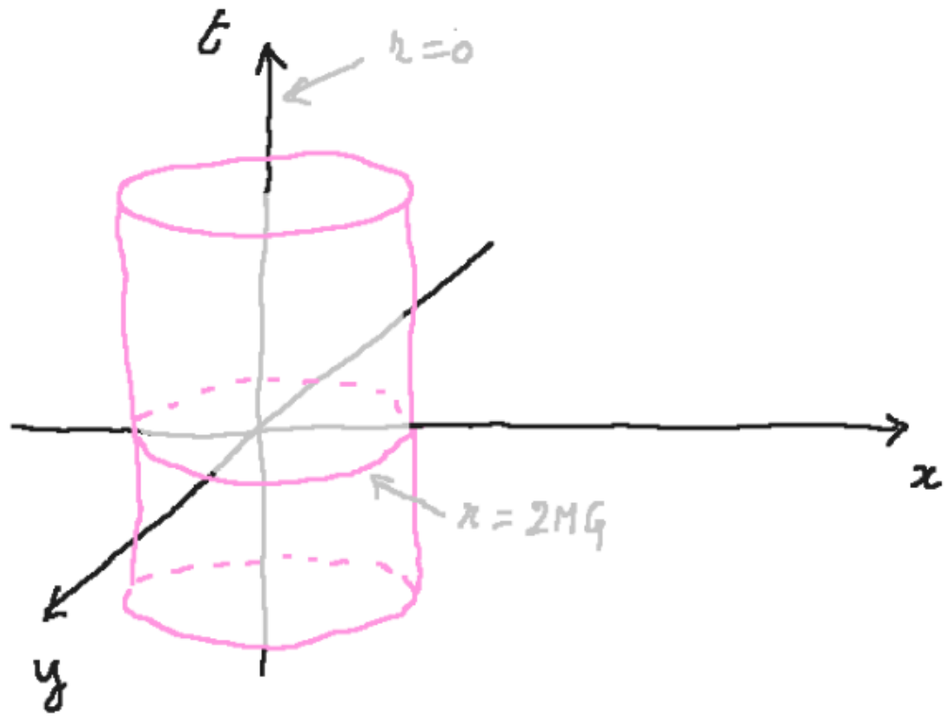


Figure 2 : Space-time in the vicinity of a black hole.

Of course there should be three spatial axes, x, y and z, and the plane at $t = 0$ should really be a 3D space, but it is *not possible* to draw such a thing, together with the time axis, on a page.

The radius $r = 2MG$, corresponding to a *cylinder* in figure 2, is called the *Schwarzschild* radius, or the *event horizon* of the black hole.

We will *understand why as we go along*.

Let's look at what is going on *very far away*.

When r is very *big*, the coefficient in front of dt^2 becomes *almost* 1, and the coefficient in front of dr^2 *almost* -1.

Then it is *just flat space-time* in polar coordinates.

But as we *move* into $r = 2MG$ something *happens* and the geometry is *characteristically different*.

Something looks *bad* at that *value* of r.

First of all the coefficient of dt^2 becomes *zero*.

But even worse the coefficient of dr^2 becomes *infinite*.

So something is *blowing up* in the mathematics.

However it is *only a consequence* of the mathematical coordinates (t, r, θ, ϕ) *chosen* to chart space-time.

We are going to see that *nothing* of any real physical significance is *happening* at that point.

On the other hand at $r = 0$ there is also a *problem* in the equation, and this time it corresponds to a singularity in the physics as well.

The coefficient in front of dt^2 becomes *infinite*, and that in front of dr^2 *zero*.

What is going on at $r = 0$ is that the *curvature* is becoming infinite.

At the Schwarzschild radius, $r = 2MG$, *no* large curvature or enormous deviation from flat space *is observed*.

But at $r = 0$ *all hell breaks loose*.

If we *calculated* the curvature tensor to find out how curved the geometry is at the center, we would find that *all of its components become infinite*.

There is *another way* to describe that phenomenon.

Curvature *means* tidal forces.

Anything falling in, when it would hit $r = 0$, would *experience* infinitely strong tidal forces and be *torn* to pieces.

It would be *stretched* radially and *squeezed* in the angular direction an infinite amount.

So $r = 0$ is the real place where *serious disaster occurs* – *not* the event horizon.

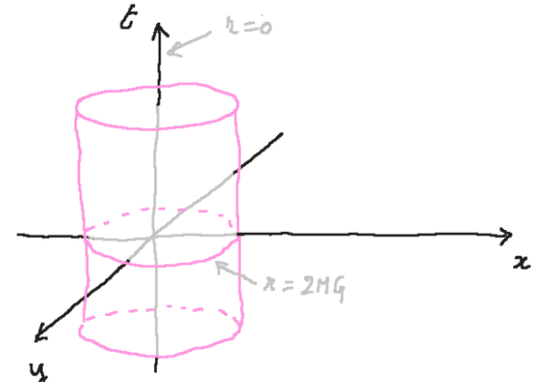
We *talked* a little bit about radially in-falling things.

Let's *go over* what the conclusion was.

We *won't* need to go into the mathematics again, because we are *going to see exactly* what is going on in terms of *pictures*.

We *will see* that it takes an infinite time for something falling in *to go past* $r = 2MG$.

In figure 2, the *surfaces* of constant time t are the *horizontal* planes at different height.



Something *falling in*, say *along* the x-axis, would have a *trajectory* shown in figure 3.

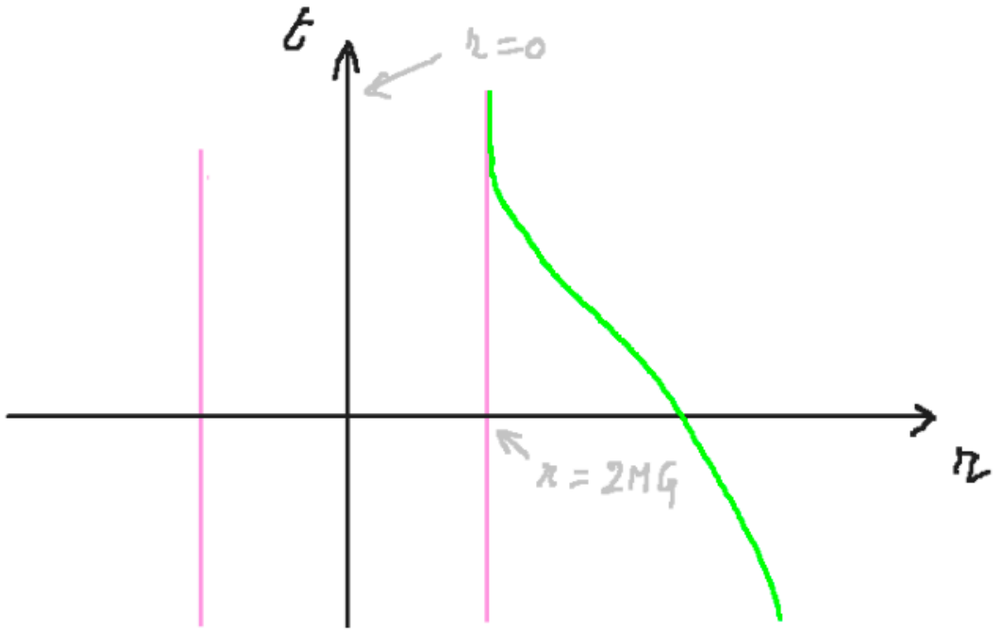


Figure 3 : Trajectory of a particle falling toward a black hole, in spherical polar coordinates (with $\theta = \phi = 0$).

The particle or object *accelerates* for a *while*.

But as it *gets closer and closer to the horizon*, its trajectory becomes *asymptotic* to the vertical axis $r = 2MG$, while *never quite getting there*.

If the object is a *meter stick* falling in the long way, its front end and back end would *follow* trajectories as shown in figure 4.

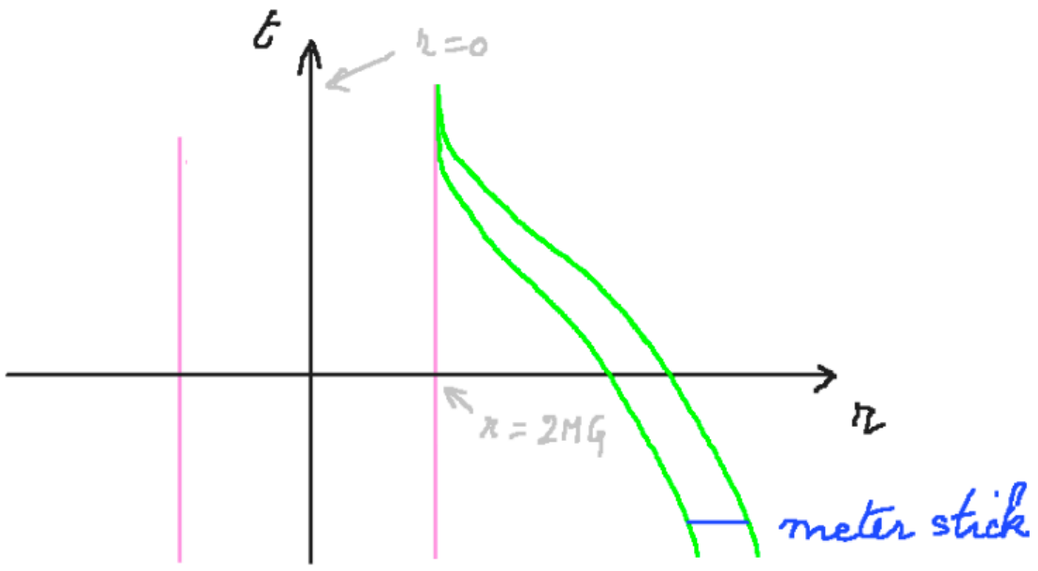


Figure 4 : Meter stick falling in radially.

The *only way* to make sense out of this phenomenon is to say, **at least in these coordinates**, that the gap in *r* between the front end and the back end of the meter stick *shrinks* to zero.

It is *actually* a form of Lorentz contraction.

As something falls in, even though it *appears* to be slowing down, its *momentum* is actually *increasing*, and it is *Lorentz contracting*.

Think also about a clock.

Let's *imagine* that the meter stick was a clock.

How do we build a clock that is also a meter stick?

Well, we can put *two mirrors* at either end of the meter stick and *allow* a light beam to go back and forth, and simply *count* the number of reflections.

But *clocks* as they move – whatever motion they have – *don't* tick off time *t* of equation (2).

They tick off *proper time*.

That is the *meaning* of the proper time: it is the *ticking of a clock moving together with the moving object under study*.

Time t of equation (2) is called *coordinate* time.

It is *just* the coordinate that we are using to *describe* the entire system.

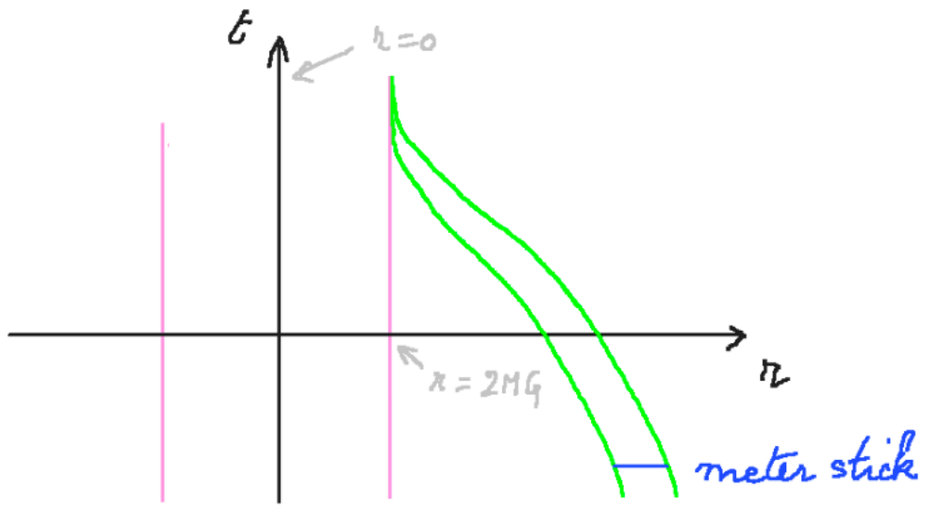
Notice that if we are *very far away*, that is if r is *very big*, and if we are *not moving*, then $d\tau$ is *equal* to dt, at least to a high approximation, because $1 - 2MG/r$ is essentially 1, and dr and $d\Omega$ are 0.

So *very far away*, proper time and ordinary time are the *same*.

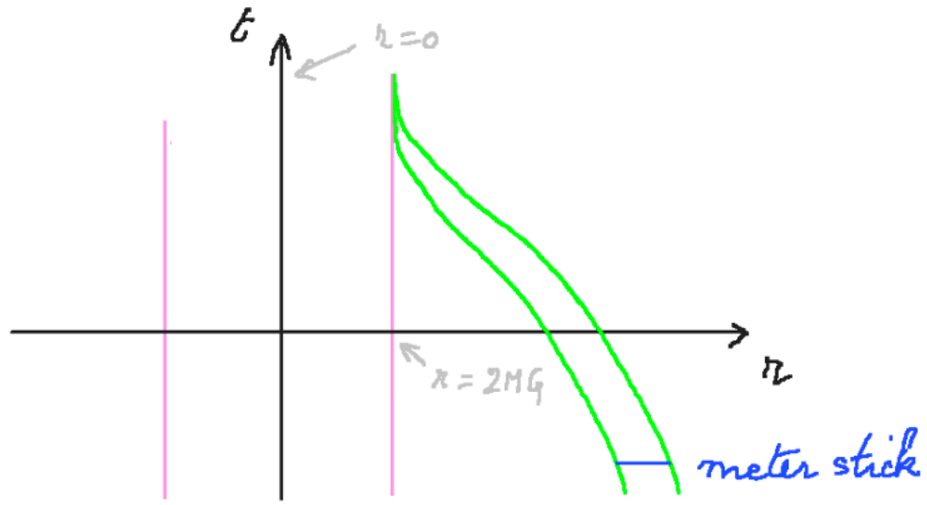
But as we *get closer and closer in*, for a *given* amount of dt, the incremental proper time $d\tau$ gets *smaller* and smaller.

For example, *consider* the coordinate time elapsing from t_1 to t_2.

Think of t_1 and t_2 as *horizontal* lines in figure 4.



How much proper time elapses along the trajectory of the object between these two horizontal lines?



When t_2 increases by dt, the corresponding proper time increment $d\tau$ gets smaller and smaller, as the value of t_2 gets bigger.

So, while it appears that the amount of t it takes to fall into the black hole is infinite, we could ask how much τ has elapsed by the time an in-falling object eventually arrives at the at the horizon.

And that proper time is *finite*.

Again we are going to see this in *pictures*, without doing any mathematics or calculations.

We will *see* it just from simple diagrams.

Consider somebody *falling* in with a clock.

Will that person *experience* an infinite amount of time before he or she gets to the horizon, or a finite amount?

By the way, the clock *doesn't* have to be an mechanical device.

By "experience" we mean the time the person *feels*.

That can be heart beats, or any other physiological process that is also a *time-keeper*.

The answer is : they *experience the proper time*.

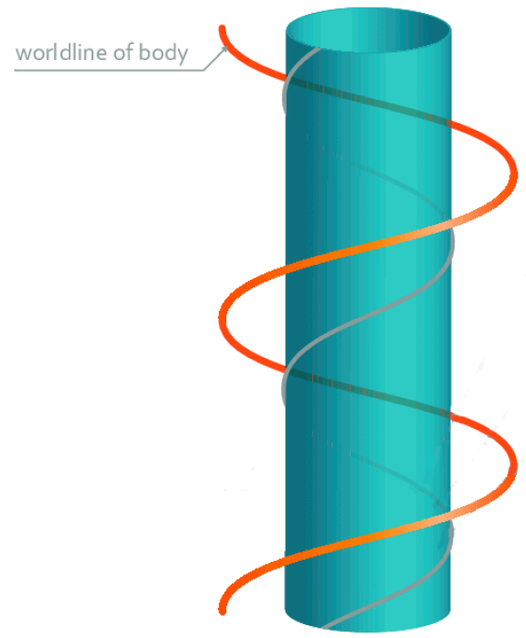
So the question is : *how much* proper time does it take to reach the horizon ?

To answer the question, we *could* solve equations, we *could* calculate, do integrals.

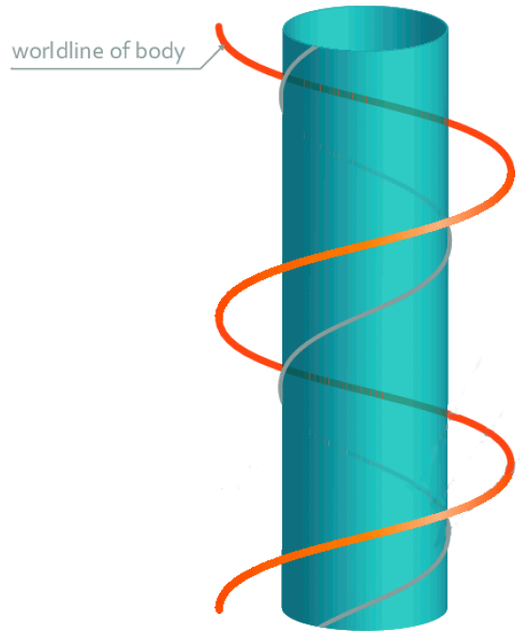
But fortunately we *don't* really have to.

Diagrams will take care of it.

Before we get to the diagrams, let's solve another problem which does not have to do with somebody falling into a black hole, but with another question : *orbiting the black hol*e.



Light-ray orbiting a black hole



Orbiting a black hole *means* going around it as time *t* goes forward.

If we think about a circular orbit, in the space- time representation of figure 2, the trajectory will have the shape of *corkscrew* (a usual "right" corkscrew, or a "left" corkscrew).

If we don't try to represent the coordinate time *t* axis, we can figure out the trajectory as in figure 5.

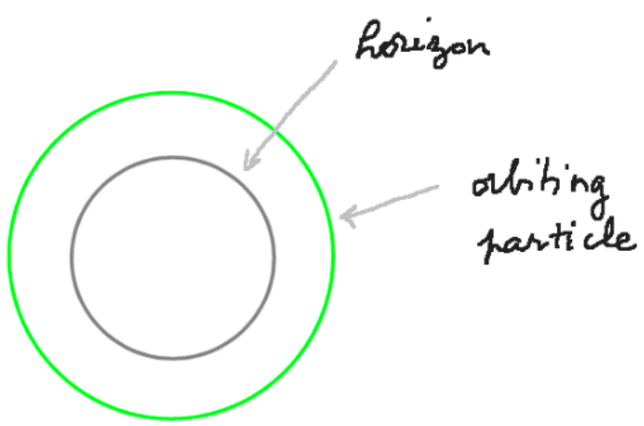


Figure 5 : Particle orbiting a black hole.

The center has coordinate $r = 0$, the horizon $r = 2MG$, and the particle r equals some *larger* value.

In fact we won't study the trajectory of any particle, we shall study the trajectory of a *light-ray*.

As we know from Einstein and from many other sources, a light-ray passing by a massive body is *bent*.

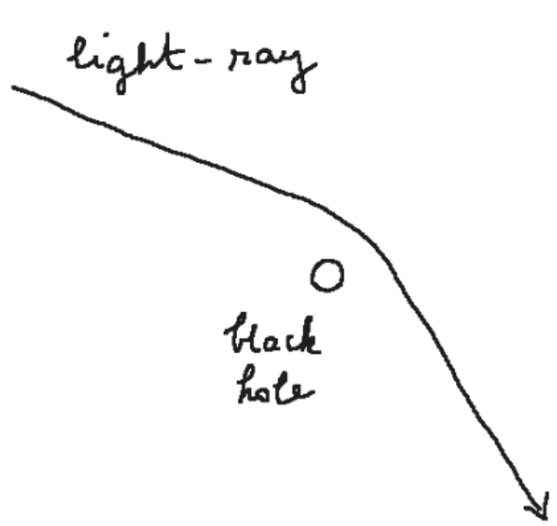


Figure 6 : Light-ray trajectory bent by a black hole.

In lesson 1 we *saw* that in Newtonian physics a uniformly accelerated elevator has an *effective* gravitational field inside it.

From that, we inferred the **equivalence principle** : accelerated frames and gravitation fields are *equivalent* (at least locally when the curvature of a real gravitational field can be neglected).

Secondly we saw that a light-ray crossing the elevator appears bent to someone in the elevator, so we deduce that real gravity should bend light-rays.

Then in lesson 4, we *saw* that in special relativity again, a uniformly accelerated frame of reference – which was something trickier to define – shows an *effective* gravitational field.

In figure 6 the *same* thing is happening : gravity *bends* trajectories.

And obviously as the light-ray passes *closer* to the black hole it will be bent *more*.

Why?

Because gravity becomes *stronger*.

As we are moving toward the Schwarzschild radius, the light-ray really *whips* around the black hole.

And at some point – it is *not* at the horizon, it is *further out* than the horizon – there exist trajectories which will just go in *circle* around the gravitating object.

By the way, this will *not* happen near the surface of the Earth.

You would have to *squeeze* the Earth to a small enough volume to transform it into a black hole.

We already said that this would be a ball – with the same mass – of radius less than *9* millimeters.

So we are *not* going to see light-rays circling around the Earth.

Nevertheless it is very *interesting* to try to understand orbiting light-rays around black holes.

The next step is a little bit calculation-*intensive*.

I'm going to show you the *sequence* of operations to *calculate* the circular orbits.

We are going to do it using the *rules* of classical mechanics.

The reason we will do it this way is *because* it is the only easy way, and even then it is not so easy.

An important point to note is that the things we learn in classical mechanics are *not* disconnected from the rest of physics and from what we are doing now.

The principle of least action is *universal* as we already discussed.

It applies in classical mechanics, quantum mechanics, special and general relativity, Maxwell electrodynamics, quantum field theory, etc.

The hamiltonian, derived from the lagrangian, and the energy are *always* the same thing, etc.

We are going to *use* some conservation laws, *plus* the basic principle of stability which says the following : equilibrium *happens* at stationary points of potential energy.

And we are going to *use* the mechanics of an object orbiting, but not an object in Newtonian mechanics, an object moving with the *Schwarzschild metric* given by equation (2).

We talked about it in the *last* lesson.

We shall *review* it quickly.

And *then* we will go into the calculations.

We start with the *action*.

Whenever you are working out equations of motion for anything, it is almost always *easiest* to start with the action principle.

From the action principle *derive* a lagrangian, and *from* the lagrangian *derive* the Euler-Lagrange equations of motion.

Or, from the action, *use some other trick of the same kind*.

We are going to use *conservation* laws.

So we start with the *metric*, which we repeat below

$$d\tau^2 = \left(1 - \frac{2MG}{r}\right) dt^2 - \left(\frac{1}{1 - \frac{2MG}{r}}\right) dr^2 - r^2 d\Omega^2$$

And the *action* of a moving particle (We *begin* the calculations with a particle of a certain mass.

Then we shall turn to photons) along a trajectory is (see lesson 5)

$$A = -m \int d\tau \tag{4}$$

Remember that an action *has* units of energy multiplied by time.

So in order for the units to be correct in equation (4), there has got to be a mass in front.

There is also an *implicit* $c^2 = 1$, because we *know* that mc^2 is an energy.

In this lesson we don't need to consider *explicitly* the speed of light, so we work with $c = 1$.

And the integral in equation (4) is taken *along* the trajectory of the particle.

The *proper time increment* $d\tau$ is the *square root* of the right hand side of the metric given above, which *defines* $d\tau^2$.

Let's write the action more *explicitly*.

But in order not write over and over the coefficient $(1 - 2MG/r)$, let's *call* it $\mathcal{F}$, and let's *call* its inverse $\mathcal{G}$.

With these notations, the *action* can be written

$$A = -m \int \sqrt{\mathcal{F} dt^2 - \mathcal{G} dr^2 - r^2 d\Omega^2} \tag{6}$$

For a *circular* orbit, $\dot{r} = 0$.

And the formula will be a little bit *simpler*, but we *don't* want to quite go there yet.

What do we *read* off as the lagrangian?

The *lagrangian* is the thing under the integral sign, when $-m$ has been brought in as well.

Let's *call* it $\mathcal{L}$.

$$\mathcal{L} = -m \sqrt{\mathcal{F}(r) - \mathcal{G}(r)\dot{r}^2 - r^2\dot{\theta}^2} \tag{7}$$

There are two conservations laws that we want to use for this problem.

1. *Conservation of energy* : when the lagrangian of a system doesn't explicitely depend on t, the energy is conserved.
2. *Conservation of angular momentum* : the angular momentum of an orbiting pa.ticle is the conjugate momentum to θ. When the lagrangian is invariant under rotation, it is also conserved

The angular momentum is the *analog* for rotating objects of the linear momentum – simply called momentum – for objects moving in a straight line.

And remember that, if $\mathcal{L}$ is *invariant* under translation, the momentum of the object, or system of particles, is *conserved*.

In lagrangian mechanics, we *define* a quantity called conjugate momentum: if we have a coordinate called q – also called a **degree of freedom** – the momentum associated with q is *equal* to the derivative of the lagrangian with respect to $\dot{q}$.

This quantity *arose* in my early discussion.

Here it is *again*.

Our *degree of freedom* is the angle θ.

And let's call L its conjugate momentum.

It is the *angular momentum*.

$$L = \frac{\partial \mathcal{L}}{\partial \dot{\theta}} \tag{8}$$

Using the expression for $\mathcal{L}$ given by equation (7), and the *chain* rule of differentiation, we first write the *derivative* of the square root of the intermediate variable

$$g = \mathcal{F}(r) - \mathcal{G}(r)\dot{r}^2 - r^2\dot{\theta}^2$$

then *multiply* it by the derivative of g with respect to $\dot{\theta}$, carrying along the factor $-m$.

Two factors 2 and two minus signs *cancel*.

We get

$$L = \frac{mr^2\dot{\theta}}{\sqrt{g}} \tag{9}$$

where $\sqrt{g}$ *stands* for

$$\sqrt{\mathcal{F}(r) - \mathcal{G}(r)\dot{r}^2 - r^2\dot{\theta}^2}$$

Since the lagrangian $\mathcal{L}$ *doesn't* depend on θ – only on $\dot{\theta}$ –, the angular momentum of the particle, given by equation (9), is *constant*.

It is also *proportional* to the mass m of the particle.

We are going to express L, using the Greek uppercase letter *lambda*, as follows

$$L = m\Lambda \tag{10}$$

In other words, Λ is *everything* on the right hand side of equation (9) with the *exception* of the mass.

Equivalently, we can write

$$\Lambda = \frac{L}{m} = \frac{r^2\dot{\theta}}{\sqrt{\mathcal{F}(r) - \mathcal{G}(r)\dot{r}^2 - r^2\dot{\theta}^2}} \tag{11}$$

This quantity Λ is sometimes called the **reduced angular momentum**.

It is the angular momentum *per* unit of mass.

It is a *conserved* quantity because L is conserved, and m is a constant.

And we see that Λ is a *function* of r, $\dot{r}$ and $\dot{\theta}$.

For a *circular* orbit it *simplifies* a little bit because $\dot{r}$ is *equal* to 0.

But let's *not* do the circular orbit yet.

L is the momentum *associated* with θ.

There is *another* momentum, P_r, *associated* with r – the *other* degree of freedom of the system.

To find P_r we also have to *differentiate* the lagrangian, this time with respect to $\dot{r}$:

$$P_r = \frac{\partial \mathcal{L}}{\partial \dot{r}} = \frac{m\mathcal{G}\dot{r}}{\sqrt{g}} \tag{12}$$

where $\sqrt{g}$ stands for $\sqrt{\mathcal{F}(r) - \mathcal{G}(r)\dot{r}^2 - r^2\dot{\theta}^2}$ as earlier.

Unlike the angular momentum L, the radial momentum P_r is *not* conserved.

The lagrangian *depends* on r.

As a *consequence* there is a radial potential.

Therefore there *are* radial forces, that we could *calculate* from the Euler-Lagrange equations.

There are *accelerations* along the radial axis.

If there are accelerations along the radial axis P_r *cannot* be conserved.

But what is conserved is the *energy*.

So we are going to *write* the general expression for the energy.

It is the good old *hamiltonian*.

Remember the expression of the hamiltonian.

If we have N generalized coordinates q_i, with i running from 1 to N, and the p_i's are the conjugate momenta, and $\mathcal{L}$ is the lagrangian, then the *hamiltonian* H is

$$H = \sum_i p_i \dot{q}_i - \mathcal{L} \tag{13}$$

The *point* is that there is always an energy.

It is always given by a certain *construction* – equation (13) – that defines it.

And the energy is *conserved* so long as the lagrangian $\mathcal{L}$ does not *explicitly* depend on time.

So *here* the energy is conserved.

In our case, there are only two coordinates, r and θ.

From equation (13), we can *calculate* H.

We get

$$H = \frac{\mathcal{F}(r)m}{\sqrt{g}} \tag{14}$$

We *won't* need P_r any further.

We *only* needed it to calculate the energy.

And H, in equation (14), is the *energy*.

Henceforth we shall *denote* it E.

$$E = H$$

And it is *conserved*.

It *does not* change with time.

Let's *reexpress* the energy with equation (14), but *now* we shall write explicitly the quantity that we denoted g, and we are going to use the fact that we are on a circular orbit.

On a circular orbit, the radius *doesn't* change.

So $\dot{r} = 0$.

And we can *write*

$$E = \frac{\mathcal{F}(r)m}{\sqrt{\mathcal{F}(r) - r^2\dot{\theta}^2}} \tag{15}$$

This is *partly* kinetic energy due to the motion of θ.

There is *no* kinetic energy due to the motion of r, because r is *not* changing on a circular orbit.

The energy E given by equation (15) is a *combination* of potential energy, because it depends on r, and kinetic energy because it also depends on the angular velocity $\dot{\theta}$.

And it is *conserved*.

What about the reduced angular momentum?

We got it in equation (11).

It has the *same* denominator as the energy, and now can be written.

$$\Lambda = \frac{r^2\dot{\theta}}{\sqrt{\mathcal{F}(r) - r^2\dot{\theta}^2}} \tag{16}$$

It is *also* conserved.

Now *what do we do* with these two conserved quantities, given by equations (15) and (16) ?

First of all, we *solve* equation (16) for $\dot{\theta}$.

This will *give* us the angular velocity as a function of the reduced angular momentum.

Then we *plug* it back into equation (15) for the energy.

The *result* will be an expression of the energy E as a function of r and the reduced angular momentum.

We will write in a second what we get.

Remark : physics always consists of this alternation between principles – having *fun* figuring out the principles, and what they tell us to do –, and then the *boring* work of doing a little calculations.

And then *back* to the principles again.

Let's *shortcut* the boring calculations part, here is what we get for the *energy*

$$E = m\,\frac{\sqrt{\mathcal{F}(r)}\,\sqrt{r^2\Lambda^2 + r^4}}{r^2} \tag{17}$$

This does *not* depend on $\dot{r}$.

It depends *only* on r and Λ.

What we are *interested* in is the value of *r* which corresponds to a position of equilibrium of the energy.

The equilibrium position *depends* on the angular momentum, or its reduced version Λ.

Now, we want to go to the *limit* of a photon.

We are interested in the *bizarre* question of whether a photon can orbit in the geometry.

Can a photon get close enough to the black hole that *even* the speed of light is not enough to *eject* it from a circular orbit around the black hole ?

So we are going to *look* for a circular orbit.

And the *rule* is to *minimize* the energy as a function of *r* for *fixed* angular momentum.

But now there is *something* going on: photons have *zero* mass.

So we got a bit of a *funny* problem here.

Since photons have zero mass, it *looks* from equation (17) that they have *zero* energy.

But they *don't* have zero energy!

A moving photon, for instance a high energy photon, *doesn't* have zero energy.

Photons *do* have energy.

So *something had better get big to cancel the fact that m goes to zero*.

Photons *also* have momentum and angular momentum.

A photon moving around in orbit, or even if it is just going past the star *does* have an angular momentum.

Remember, however, that the angular momentum *L satisfies* the equation

$$L = m\ \Lambda \tag{18}$$

What *happens* as m goes to zero?

The thing that you hold *fixed* is the angular momentum L.

Necessarily Λ *becomes* infinite.

In other words, the right way to take the limit of a photon, that is, to study the trajectory of a photon of fixed energy and fixed angular momentum, is to take the limit as the mass m goes to *zero*, and the reduced angular momentum Λ goes to *infinity*, in such a way that the product of m times Λ stays *fixed*.

What happens in equation (17) *when* m goes to zero and Λ *gets* very big, because $\Lambda = L/m$ and L is a *given*?

As Λ gets very big, the term $r^2\Lambda^2$ in the second square root eventually makes r^4 *negligible*.

The *square root* of $r^2\Lambda^2$ is just $r\Lambda$.

It is *multiplied* by m.

So, taking into account that $m\Lambda$ is L, equation (17) *becomes*

$$E = L\,\frac{\sqrt{\mathcal{F}(r)}}{r} \tag{19}$$

This *shows* that, for a given angular momentum L, which is just a *multiplicative* factor, the energy of the photon is a function *only* of r.

You can think of this E as a *kind* of potential energy.

The radial velocity $\dot{r}$ has been *eliminated* out, because on a circular orbit there is *no* radial velocity.

The angular velocity $\dot{\theta}$ has been eliminated by *expressing* it in terms of the angular momentum L, which is a given.

That is *why* E depends only on r.

Let's *study* the energy E of the photon as a function of r, in particular its point of equilibrium.

Photon sphere

Where is the equilibrium of E in terms of r?

To find the value of r, all we have to do is *differentiate* the left hand side of equation (19) with respect to r and *set* the derivative equal to zero.

At that value of r, the function $L\sqrt{\mathcal{F}(r)}/r$ will *either* have a maximum or minimum value.

Remember that the *Schwarzschild* factor, that we decided to denote $\mathcal{F}(r)$ to make the expressions lighter, *is*

$$\mathcal{F}(r) = 1 - \frac{2MG}{r}$$

Let's *plot* the function given by equation (19) to see what it looks like, and identify the point where it is stationary, figure 7.

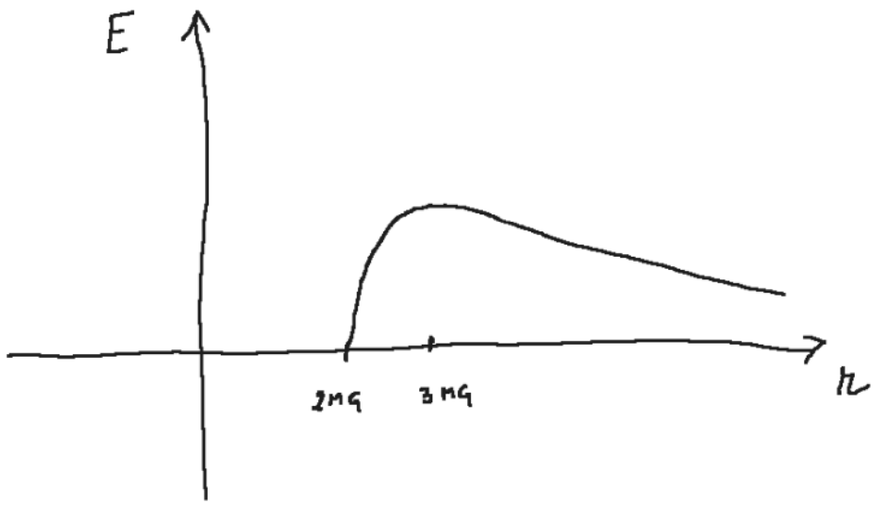


Figure 7 : E as a function of r.

When r becomes very *big*, E is *asymptotically* equal to L/r.

And when r gets *close* to the horizon, $2MG$, E gets *close* to zero.

In *between* E is positive, and there is a *point where* its derivative is *zero*.

Can show that the point where E is *stationary* is at

$$r = 3MG \tag{20}$$

It *does not depend* on the angular momentum.

The *interesting* fact is that where E is stationary, it is at a *maximum*.

If you have a potential energy that has a *maximum*, does that correspond to an *equilibrium* position?
Answer : yes, but it is an *unstable* equilibrium.

It is like *placing* a marble ball on top of a smooth metal sphere : if you place it absolutely *precisely*, the marble will *stay* there.

But the *slightest* misplacement off the top, or the *slightest* tap on the marble, and the marble will *roll* off.

The *smaller* the tap you give it, however, or the *smaller* the imprecision in putting it right up at the top, the *longer* it will last at or near the top before eventually rolling down.

Here we are talking *not* about a marble and sphere, but about a *circular* orbit of a massless particle *around* a black hole.

What we *found* is that *right* at $r = 3MG$ a photon *can orbit* a black hole.

It is *not* at the Schwarzschild radius.

It is at *one and a half* times the Schwarzschild radius.

The *sphere* at that radius is called the *photosphere*, or the *photon sphere*.

The *obvious* reason is that it is a sphere on which a photon can orbit.

You might *expect* that if you had a black hole, and you looked at that distance you would find all kinds of photons orbiting around it.

After all any photon that gets *started* in that orbit, will *stay* in that orbit.

But of course if the photon starts slightly away from the photon sphere, it will whip around and come out – not necessarily in the opposite direction.

The *closer* you get to the photon sphere the *more* angle the photon will sweep before it eventually goes away.

A photon coming in *very close* can circle around many times before escaping, figure 8.

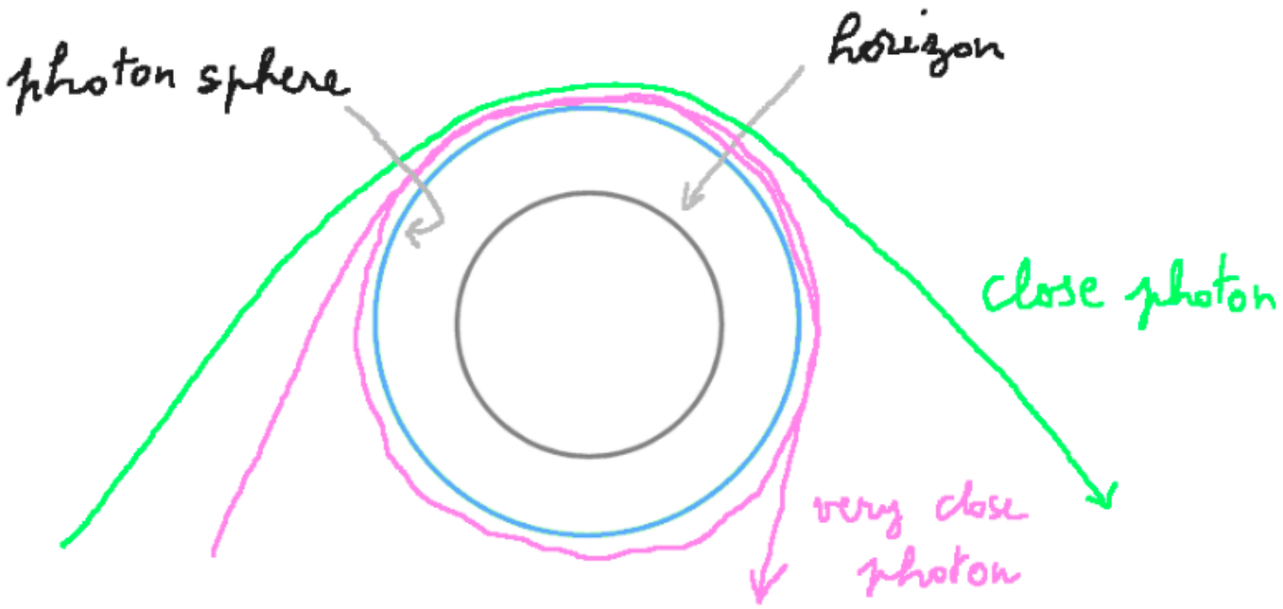
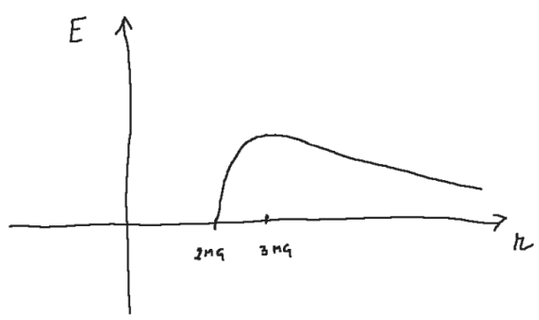


Figure 8 : Photons passing slightly outside the photon sph

Photons whipping around, or temporarily orbiting the black hole before escaping, *correspond* to a closest distance to the black hole slightly off to the right of the point $3MG$ in figure 7, and then *slowly* going outward.

What happens, on the other hand, if a photon is on the *left* of $3MG$ in figure 7?



Answer : it will get *pulled* in.

It *can't* withstand the gravitation if it is just *slightly* inward the photon sphere.

It *might* orbit a few times, as above, but it will *spiral* into the horizon and then of course *eventually* spiral into the singularity.

This was an *example* of calculations with the Schwarzschild metric.

And we showed the *surprising* fact that massless particles can have an *unstable* equilibrium orbit at radius $r = 3MG$.

If one *compares* the calculations we did with the corresponding Newtonian calculations, then one will see all the pieces are the *same*.

But the *outcome* is quite *different*.

In Newtonian physics, light-rays *don't* orbit anything.

In inertial frames of reference, they move in *straight* lines *whatever* the field they go through.

In inertial frames of reference, massive particles, plunged in a central gravitational field, move along *conics*.

And for certain *configurations* of E and L massive particles can follow a circular orbit around a force center at any distance, *including* at distance $r = 3MG$ or smaller.

Whereas, in space-time with the Schwarzschild metric, particles below the photosphere, whether massive or massless, will *inescapably* get sucked in.

The *phenomenon* of photons temporarily orbiting in the photon sphere before escaping again has interesting *consequences*.

Let's *look* at a black hole from a *distance*.

The black hole itself *emits* no light, but *light-rays coming from outer space behind the black hole*, relative to you, *display* strange patterns.

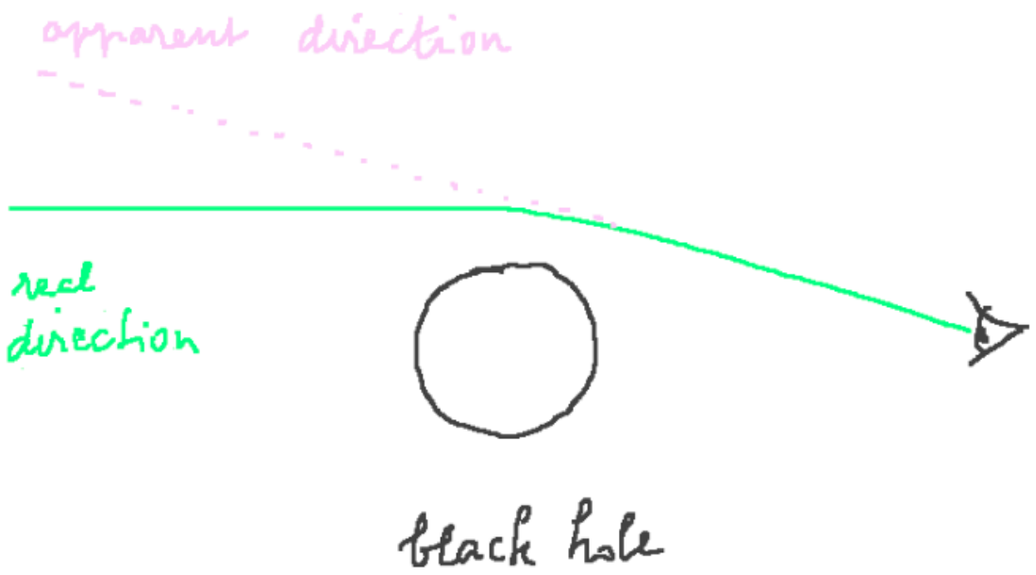


Figure 9 : Distant light emitter apparently displaced.

For *instance* a light-ray arrives from a light emitter, say, a distant star, in outer space behind the black hole. It *bends* a little bit, and then hits your eye, figure 9.

So *instead* of seeing the light-ray coming from its real direction in the backdrop, you see it *apparently* coming from another direction.

It is *worse* than that.

A light-ray *coming* from the *same* distant source, may fly by the black hole *near* another point, for instance not above but *below* the black hole in figure 9.

Then it will be *bent* differently and *appear* to be coming from another direction than the first one.

Therefore the light emitter behind the black hole will *appear* to your eye as a ring around the black hole.

A black hole is a *bad* place for your enemy to hide behind from you.

What you *see* is even worse.

Other points in the backdrop may emit light-rays that will come *near* the photosphere, orbit for a while and happen to leave in a direction that will *hit* your eye, figure 10.

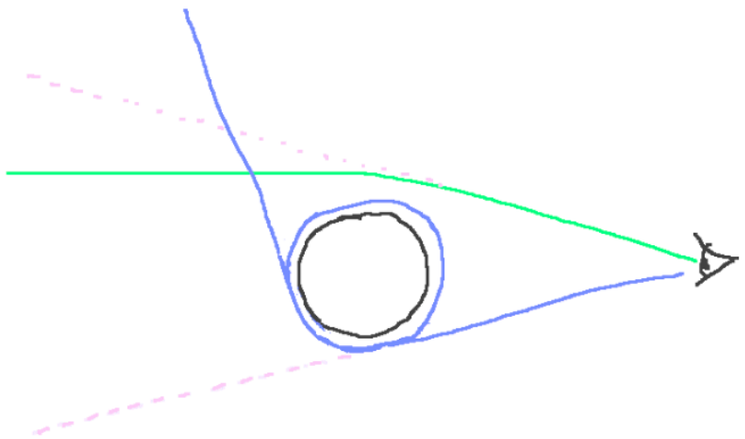


Figure 10 : Another light emitter apparently displaced.

Light emitters *don't* have to be far away to display this phenomenon.

Light emitting points *near* the horizon of the black hole, *anywhere* on the sphere, can emit light-rays that will *come out* of the photosphere.

Some can *end* up in your eye.

In short, *what* you will see from the backdrop *as well as* from the vicinity of the black hole will form a *complicated* pattern in your eye due to light-rays trajectories modified in various ways by the black hole.

That makes looking at a black hole a *dizzyfying* experience.

Notice that the ring created by a source behind the black hole, called *Einstein's* ring, is a physically *observed* phenomenon.

The first complete one was observed by the Hubble space telescope in *1998*.

Black holes are *convenient* massive bodies to observe light *deviation*, because they *don't* emit light on their *own*.

Therefore, *unlike* with the Sun for instance, we don't have to *wait* for an eclipse to study light-rays flying in their vicinity.

Hyperbolic coordinates revisited

Now we want to *understand* better what is happening at the *event horizon* of the black hole, figure 11.

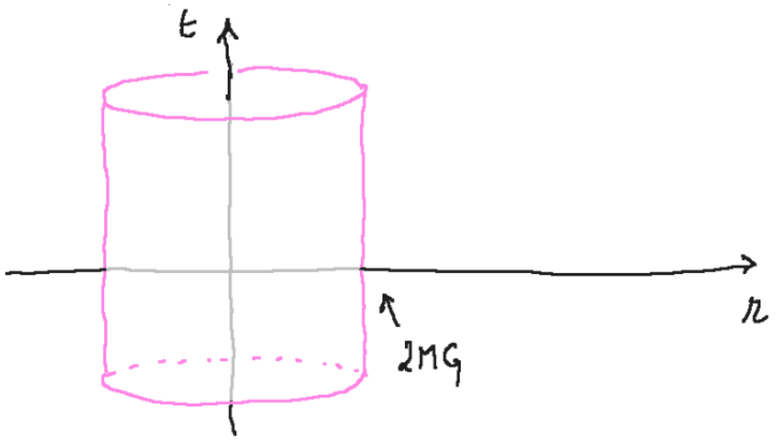


Figure 11 : Black hole horizon.

Among other things, something *curious* happens in the *equation* of the metric.

$$d\tau^2 = \left(1 - \frac{2MG}{r}\right) dt^2 - \left(\frac{1}{1 - \frac{2MG}{r}}\right) dr^2 - r^2 d\Omega^2$$

The *sign* of the coefficient in front of dt^2 *changes*.

At the *same* time the *sign* of the coefficient in front of dr^2 *changes* too – thus *preserving* the signature of the metric.

We *know* that clocks "slow down", in the sense that when r gets *close* to $2MG$, for a given dt the *corresponding* $d\tau$ is smaller and smaller.

We saw that a stick falling in vertically *contracts*, etc.

But I also told you that in some other sense, *nothing* is happening : *someone* falling in will not *experience* anything special while crossing the horizon.

Many *things* that happen in gravity can be understood by first *understanding* them for accelerated coordinate frames.

That is the *equivalence* principle.

Gravity is *not exactly the same* as a uniformly accelerated coordinate system of course, but it is often a good idea, when *studying* a question, to first look at how it *manifests* in a uniformly accelerated coordinate system.

So let's *go back* to that problem.

Remember that studying a uniformly accelerated reference frame in relativity is a little more complicated than it is in Newtonian physics.

It is *best studied* in a coordinate system which is the *analog* of polar coordinates – but *hyperbolic* polar coordinates.

Let's do a *quick* review.

We are in a *flat* space-time now.

It is *not* a black hole, it is *not* the gravitational field of a massive body, it is *plain* old flat space-time.

Any point P *has* standard Minkowski coordinates, which we *denote* (T, X), and hyperbolic polar coordinates, which we *denote* (ω, ϱ).

There *may be* more spatial coordinates Y and Z, but we are *not interested* in them, because they *won't* play a role in the subsequent analysis.

We have *called* coordinate ω the *hyperbolic angle*, because it is the *analog* of the ordinary angle in polar coordinates.

However, *instead* of varying from 0 to 2π, it *varies* from $-\infty$ to $+\infty$.

It is a *kind* of time.

Coordinate ϱ is sometimes *called* the hyperbolic radius.

A point in figure 12 can be *located* with its T and X coordinates, and also with its ω and ϱ coordinates.

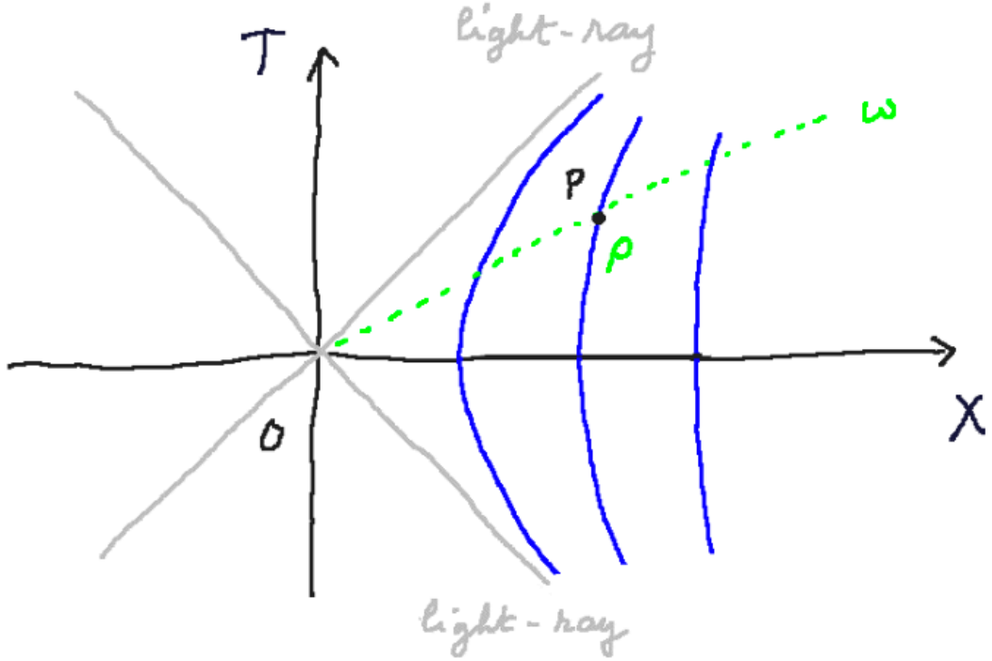


Figure 12 : Flat space-time with hyperbolic polar coordinates.

Remember the following facts in figure 12 :

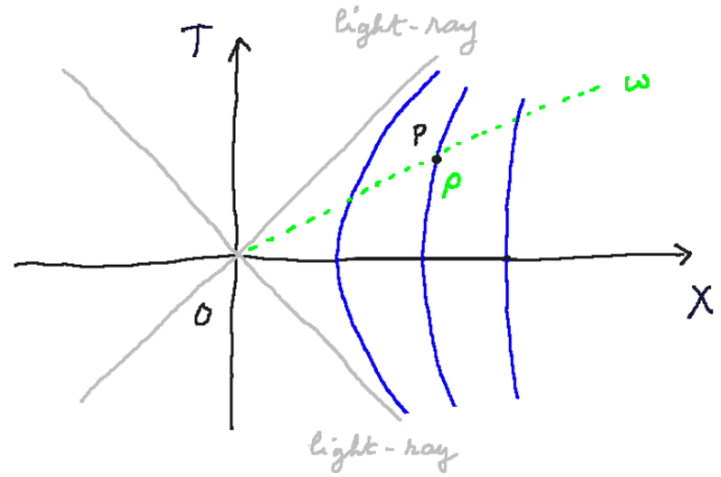
- points on a horizontal line have the same T
- points on a vertical line have the same X
- points on the same dotted line have the same ω
- points on the same hyperbola have the same ρ

Finally the *algebraic correspondence* between the Minkowski coordinates and the hyperbolic polar coordinates of P is

$$\begin{aligned} X &= \rho \cosh \omega \\ T &= \rho \sinh \omega \end{aligned} \tag{21}$$

What about $X^2 - T^2$?

It is $\varrho^2(\cosh^2\omega - \sinh^2\omega)$ which is *equal* to ϱ^2.



So on each hyperbola in figure 12, $X^2 - T^2$ is *constant* when ω *varies*.

In lesson 4 we *studied* such a transformation.

It *defines* a **uniformly accelerated reference frame** in special relativity.

Let's *recall* briefly the reasoning :

1. In Newtonian physics, the concept of a uniformly accelerated collection of points, fixed relative to each other, is intuitive and clear.
2. In Newtonian physics, points rotating around a circle, all fixed relative to each other, are all submitted to the same acceleration toward the center.
3. Finally, in special relativity, points with fixed ρ's and moving with the same ω, are said to be uniformly accelerated. In this latter case, it is a definition.

The *metric* of ordinary flat space-time, *expressed* in the hyperbolic polar coordinates, becomes

$$d\tau^2 = \rho^2 d\omega^2 - d\rho^2 \tag{22}$$

That is the *analog* of $r^2 d\theta^2 + dr^2$ in ordinary polar coordinates on the Euclidean plane, that is in *ordinary* geometry.

And equation (22) is the *same* in Minkowski space-time.

Now, in order to study *conveniently* what is going on in the vicinity of the event horizon of a black hole, we are going to *change* coordinates.

We *don't* change coordinates for the pleasure, but to make life *simpler*.

The *new* coordinates will be denoted ξ and t, and defined as follows

$$\xi = \frac{\rho^2}{4}$$
$$t = 2\omega \tag{23}$$

And we *rewrite* the metric.

Elementary *algebraic* manipulations on equation (22) yield

$$d\tau^2 = \xi dt^2 - \frac{d\xi^2}{\xi} \tag{24}$$

We notice a *similarity* with the Schwarzschild metric.

In the Schwarzschild metric, equation (2), which we reproduced at the beginning of this section, we *have* a coefficient in front of dt^2, and we have the *inverse* of the same coefficient in front of dr^2.

If we think of ξ as r, equation (24) bears some *family resemblance*.

Notice another thing : $(1 - 2MG/r)$ and its inverse *change* sign when r gets in too close to the singularity.

Somehow analogously, ξ *changes* sign simply when it becomes negative, and then *so* does $1/\xi$.

Does ξ becoming negative *mean* something?

Yes.

We will see that ξ becoming negative *does* mean something.

So there is a *similarity*.

But the *basic similarity* is a coefficient and another coefficient, which are inverses of each other, and which, we will see, meaningfully change sign as ξ become negative.

In what follows, we are going to *use* the flat metric of equation (24).

But let's *first* turn again to the Schwarzschild metric of the black hole.

On the black hole, we want to *zoom* in on the place where $r = 2MG$.

This is *another* way of saying that we shall study the Schwarzschild metric of equation (2) in the *region* where r is almost equal to $2MG$.

We use a *microscope* to look at points near the horizon.

The *first* coefficient in the Schwarzschild metric is $(1 - 2MG/r)$, which can be rewritten

$$\frac{r - 2MG}{r}$$

Since we won't let r change very much near $2MG$, to the first order approximation (in $\varepsilon = r - 2MG$) it is *equal* to

$$\frac{r - 2MG}{2MG}$$

This is *like* in Newton's equation : if you are *near* the surface of the Earth, for many purposes you can set r *equal* to the radius of the Earth.

Omitting for the moment the term $r^2 d\Omega^2$ which only concerns changes in the spherical angles, the Schwarzschild metric becomes

$$d\tau^2 = \left(\frac{r - 2MG}{2MG}\right) dt^2 - \left(\frac{2MG}{r - 2MG}\right) dr^2 \tag{25}$$

In order to *simplify* the expression, we shall set the mass of the black hole to a specific number, so that $2MG$ equals 1.

We can always do that without *loss* of generality.

It is just a *choice* of units.

What are the *units* of $2MG$?

Units of *length*, because $2MG$ is the Schwarzschild radius.

For a particular black hole we can choose *units* of length where the Schwarzschild radius is just equal to one.

We can't do it for all black holes simultaneously but if we are interested in a *particular* black hole we can do that.

That makes formula (25) much *simpler*.

It is now

$$d\tau^2 = (r-1)dt^2 - \left(\frac{1}{r-1}\right)dr^2 \tag{26}$$

The next step is to *redefine* $r-1$ and call it ξ.

It is just a *change* of variables.

Since with our preceding change of units the horizon had coordinate $r = 1$, we are now *translating* coordinates so that the *horizon* has coordinate $\xi = 0$.

And ξ *measures* the deviation from the horizon.

Equation (26) becomes

$$d\tau^2 = \xi dt^2 - \frac{1}{\xi}dr^2 \tag{27}$$

Notice that dr and $d\xi$ are the *same* thing since $\xi = r-1$.

So after all these various *changes* (local analysis at the horizon, change of units of mass of the black hole, and change of variable near the horizon) the Schwarzschild metric is *now*

$$d\tau^2 = \xi dt^2 - \frac{d\xi^2}{\xi} \tag{28}$$

This equation (28) is exactly the *same* as equation (24)!

This is *telling* us is that the *vicinity* of the horizon is in some *mysterious* way just *flat* space-time, or *approximately* flat space-time.

Nothing *dramatic* is happening *there*.

No *large* curvature, *only* more or less flat space-time.

This may be *surprising* because of what we saw earlier – the *slowing* down of clocks, the *contraction* of lengths, etc. – but these were *just coordinate-related phenomena*.

With an *appropriate* change of coordinate, the space-time of the black hole *displays no* special features at the horizon.

It is *locally* more or less *flat* like everywhere else (*except* at $r = 0$).

Of course, the space-time *near* the horizon is flat in a *uniformly* accelerated coordinate system – just *like* on the *surface* of the Earth inside a *free-falling elevator* there is no gravity.

Uniform gravity or uniform acceleration *don't* make the space *non*-flat.

The essential point in the above analysis is that nothing special happens at the black hole horizon.

Why would we be interested in a *coordinate* system, for a *gravitating* object, which is *mimicking* uniform acceleration?

Well, standing where I am, I am *experiencing* an acceleration of 10 meters per second per second.

I feel exactly the same things that I would if I were in free space and the floor under me was being accelerated upward with an acceleration of 10 meters per second per second.

Similarly, in studying the gravitating black hole from the perspective of somebody standing still, being *supported* let's say *above* the horizon of the black hole or wherever he or she is, you are *effectively* doing physics in a *uniformly* accelerated reference frame in space-time.

What does physics feel like there?

What does it do there?

It does *whatever* the uniformly accelerated reference frame does.

And that is exactly what we *experience* near the horizon of a black hole: we experience whatever is *effectively* created by a uniformly accelerated coordinate system – just like you, in Newtonian physics, sitting in your chair.

But the space-time itself is *flat*.

Interchange of space and time directions at the horizon

Now let's talk about the *interchange* of *space to time and time to space*, that we have already *alluded* to, when we *cross* the black hole horizon.

We shall now make *full use* of the *fact* we just established, *that* in the *vicinity* of the black hole horizon the space-time gravitational field is *equivalent* to that of a uniformly accelerated reference frame.

We saw that we *defined* ξ as

$$\xi = r - 1$$

or, if we *put back* M and G, ξ would be *proportional* to $r - 2MG$.

ξ can *change* sign.

How?

It changes sign by going from *outside* the horizon t*o inside* the horizon.

It makes sense – at least **in the black hole context** – to say that ξ can change sign.

Notice that, even though in the Schwarzschild metric, space-time is *everywhere* locally flat – except at $r = 0$ –, there is still a *special* role played by $r = 2MG$, coming from the *definition* of the metric in equation (2).

This *requires* some clarifications *because* in a uniformly accelerated reference frame in space-time, from equations (23), we *also* have

$$\xi = \frac{\rho^2}{4} \tag{29}$$

How can ϱ^2 change sign?

ϱ is a *real* number and ϱ^2 is *always* positive.

But that is *not* the right way to look at ϱ^2.

The *right* way to look at ϱ^2 is to *look* at the right hand side of

$$X^2 - T^2 = \rho^2 \tag{30}$$

For a *fixed* right hand side, this is the equation of a *hyperbola*.

If ϱ^2 is *positive*, it is a hyperbola in the *first* quadrant and *third* quadrant.

If ϱ^2 is *negative*, it is a hyperbola in the *second* and *fourth* quadrant, see figure 13.

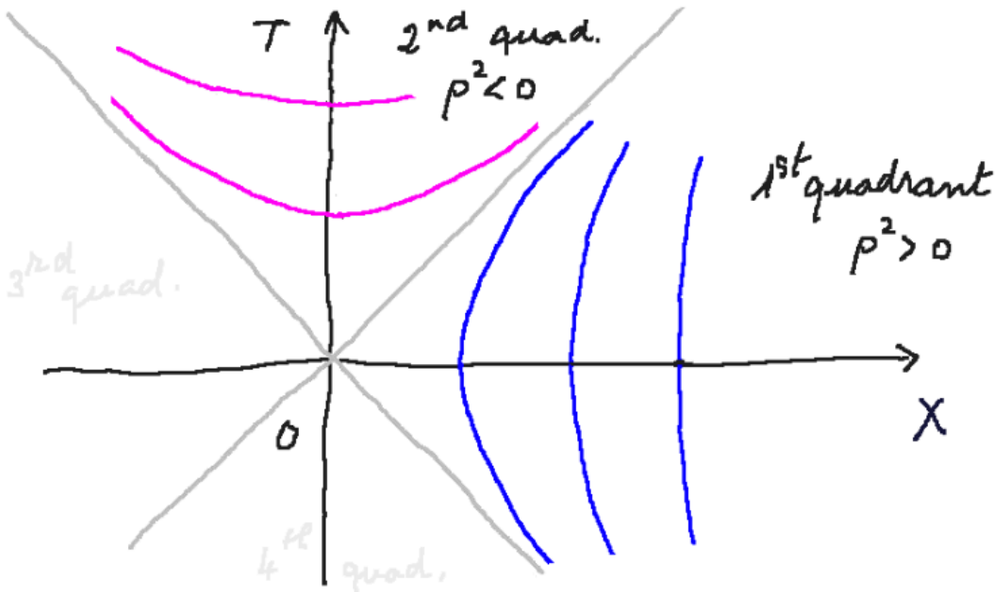
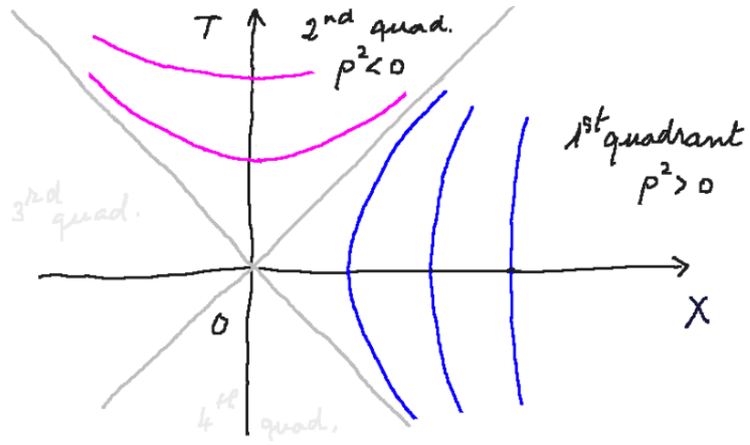


Figure 13 : Hyperbolas with positive ϱ^2 and with negative ϱ^2. The point O is **not** the center of the black hole, but a point **on** its **horizon**.

Let's *forget* about the *left* and *lower* quadrants, and *concentrate* on the *right* and *upper* quadrants.

ϱ^2 being *negative* simply *corresponds* to points in the second or upper quadrant.



So in *going* from ξ positive to ξ negative, we are simply *passing* from the *right* quadrant to the *upper* quadrant.

Imagine that we *follow* the space coordinate ξ till we *get* to $\xi = 0$.

Then *where* would we go from there?

In other words, we are *somewhere* on the *horizontal* axis.

The *hyperbolic* angle ω is, for instance, *equal* to zero.

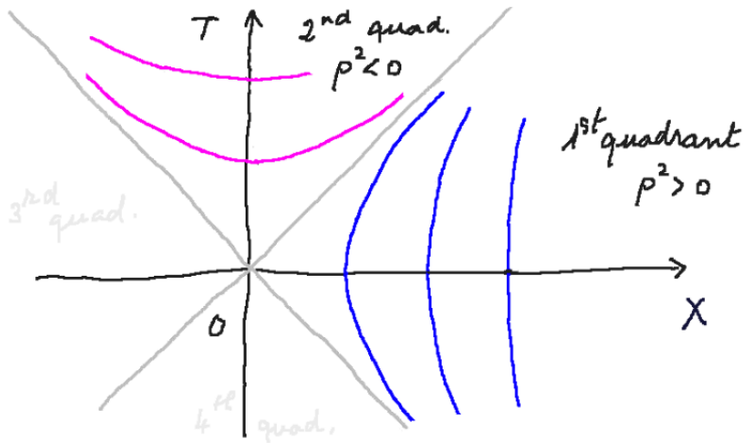
And we are *following* our way with ξ *decreasing*, while $\omega = 0$, *toward* the origin.

At the origin O – which, remember, is *not* the center of the black hole, but the point where we *cross* the horizon (with θ and ϕ *fixed*) –, ξ *changes* sign.

Then *where* do we go, if we *keep* moving with ξ now varying in the *negative* numbers, and ω *still* zero?

We *don't* go left into the *third* quadrant, we now go *upward* into the *second* quadrant.

On the OX *axis*, in figure 13, $\omega = 0$ and $\xi > 0$.



And on the OT axis, $\omega = 0$ and $\xi < 0$.

What happens is that a space-like coordinate, namely ξ, *jumps to become* a time-like coordinate.

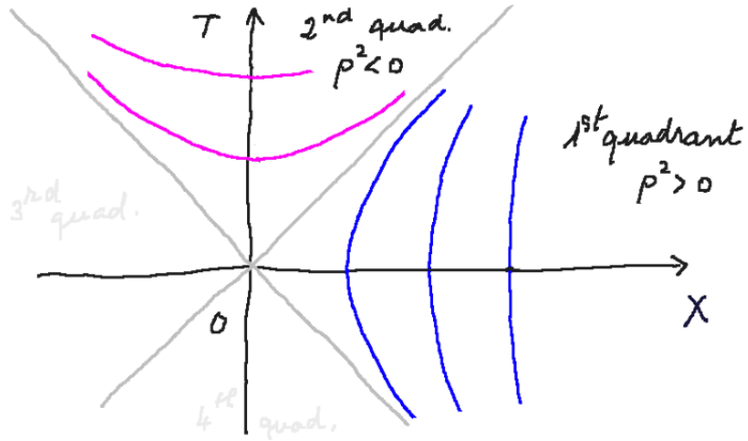
Indeed, when ξ is *positive*, a variation in ξ entails an interval displacement *of the space-like type*.

Then, when ξ is *negative*, a variation in ξ entails an interval displacement *of the time-like type*.

Let's *now turn* to the coordinate ω.

In the *first* quadrant, ω – also called the hyperbolic angle – is a *time-like* coordinate.

When it varies *alone*, we move on a *time-like* interval.



What *happens* to it after we *passed* the origin O?

Now when we are in the *second* quadrant, and let ω *vary*, with ξ *fixed*, we are *moving along* a hyperbola of the *second* quadrant, see figure 13, that is *along a space-like interval*.

In *summary*, when we *pass* O, in other words when we *cross* the horizon, the space and time directions in the coordinate system (ξ, ω) are *interchanged*.

Let's *not* be mistaken : ξ *remains* ξ, and ω *remains* ω.

But the *former*, which was space-like, becomes time-like.

And the *latter*, which was time-like, becomes space-like.

All these considerations concern coordinates of events in space-time.

Nothing physically meaningful *happens* at the events or, more *accurately*, to the particles going from event to event, following trajectories in space-time – with the *exception* of the singularity at the center of the black hole, which we shall study *next*.

It is *important* to *understand* that time is *just* one of the coordinates to locate events in space-time.

We have *freedom* in selecting what is our time coordinate, *provided* that when it varies alone we *move* on a time-like interval.

We have *used* in turn T in the standard Minkowski diagram, ω in equations (21), t in equations (23), etc.

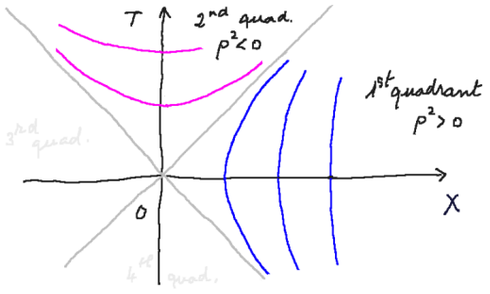
And a *trajectory* is by definition a collection of events *indexed* by a time coordinate.

ξ and ω are just *funny* coordinates where ξ *goes from* being a space coordinate to being a time coordinate, and *conversely* ω *goes from* being a time coordinate to a space one.

However, *remember* that *all* this is just a flat space.

There is *nothing going on* in the space-time region around the horizon.

Somebody who *jumps* from the right quadrant to the upper quadrant sees *nothing* special.



There is *nothing* happening.

It is just *peculiar coordinates* that have some *funny behavior* as you go from one quadrant to the other

The *same* is true of the *Schwarzschild coordinates*. ξ is $r-1$ or $(r-2MG)/r$.

When $r-2MG$ changes sign, you are doing *exactly* the *same* thing as you would going from the right quadrant to the upper quadrant in figure 13.

Black hole singularity

Let's go on and *think* about the black hole itself now.

The coordinate ξ is *simply* $r - 1$.

Decreasing r until we got to 1 *consisted in reaching and then crossing the event horizon*.

This *also consisted* in reaching and crossing the origin in figure 13.

Then *continuing* to let r decrease toward 0, we were *now moving up* along the vertical axis.

That is the *nature of these coordinates*, and is *not* the most important point from a physical point of view.

Eventually we will reach $r = 0$.

And that is what we want to *study* now.

At $r = 0$ something *nasty* happens.

Where is $r = 0$?

It is in the *upper quadrant* on the hyperbola shown in figure 14.

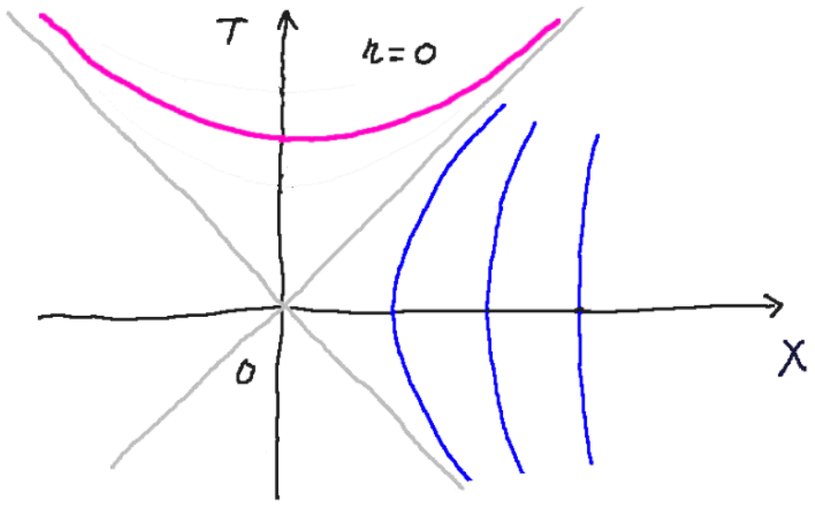
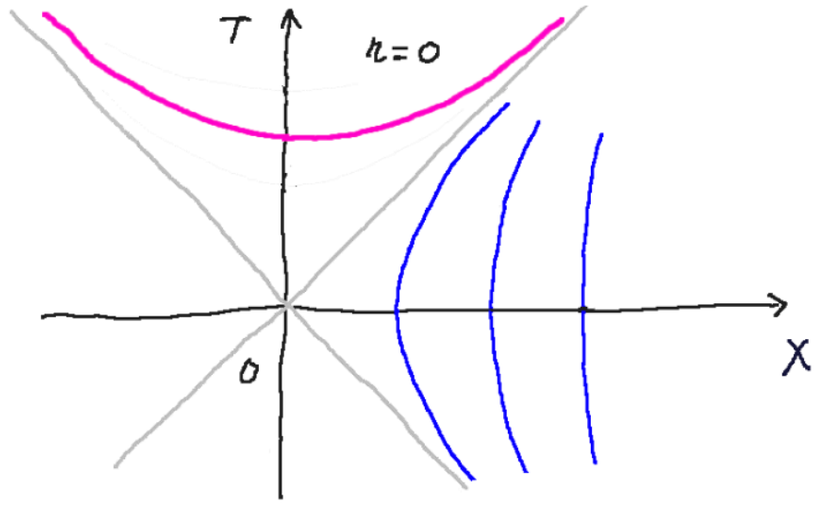


Figure 14 : r = 0. It is **not** a vertical straight line, **but** a time-like curve.

At $r = 0$ we are *behind* the horizon, *inside* the sphere of radius $r = 2MG$.

The *coefficient* $(1-2MG/r)$ has *changed* sign and *went* to $-\infty$.



And lo and behold : the *singularity* at $r = 0$ is *not* a place, it is a *time*.

More *accurately*, it is *not* a single place but *many* different places all *on* a time-like curve.

The surface $r = 0$ is *not what we would normally think of*.

Normally a place corresponds, in the *Minkowski* diagram, to a *vertical* line.

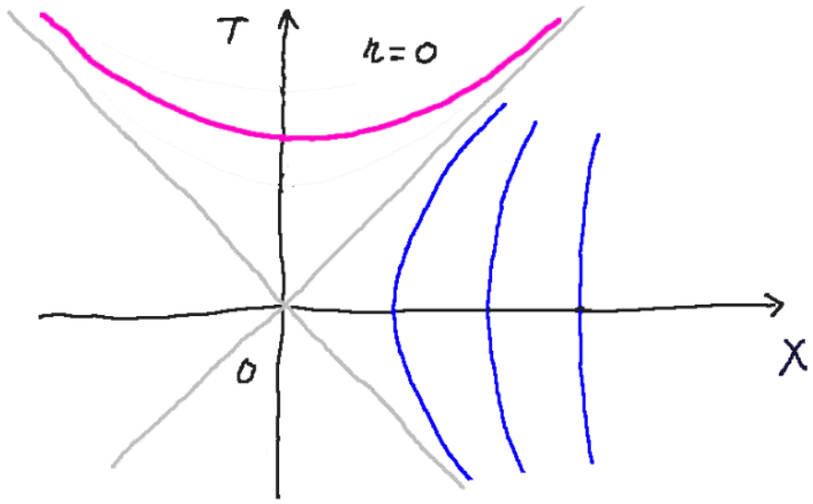
But this is *not* the case of the *singularity* at $r = 0$.

In this case it is the *hyperbola* in the *upper* quadrant in fig 14.

And it is more time-like than space-like.

This is *difficult* to figure out and at *first confusing* of course.

But this is the way the black hole really is.



You will get used to it, if you think about it and play with it.

We *cannot* avoid the singularity *when* we got past the horizon.

Why? Because we *cannot avoid the future*.

There is *no way to escape* from the future the way we can *avoid a place*.

Think of a place in front of you : you can *avoid* it, you can go around it.

What you *cannot* do is avoid the future, any future time.

Once we are past the origin in fig 14, once we are *inside* the horizon, that is in the *upper* quadrant, no matter what we do *we will run* into the hyperbola r = 0.

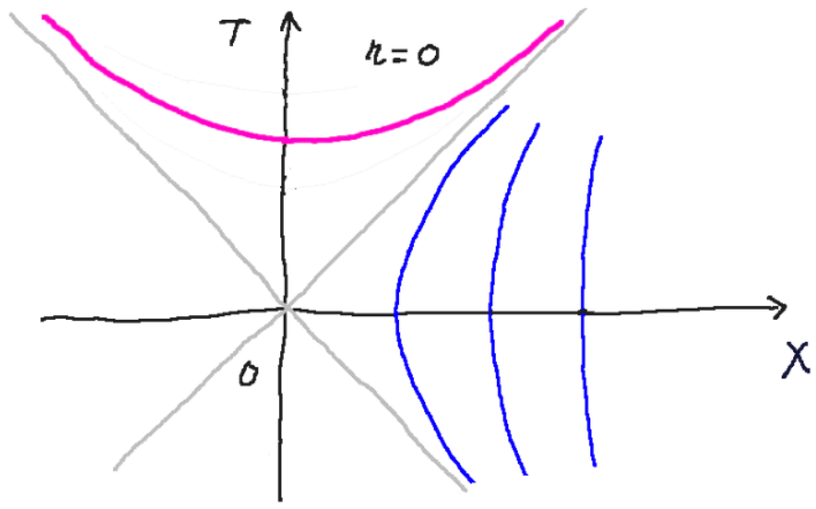
The singularity is *not* avoidable the way it would be in *Newtonian physics*.

In Newtonian physics the center of the coordinates is a *place*.

We *can* go around a place, but we *cannot* go around a time.

So that is the nature of the black hole singularity.

What about the *horizon*?



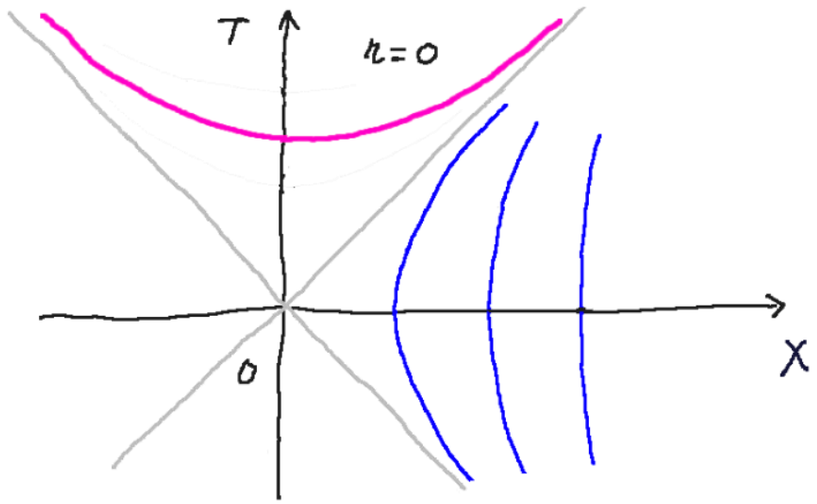
The horizon is the *point O* in figure 14.

But we are going to think about it a *little bit differently* eventually.

We are going to think about the *whole straight line* at 45° as being the horizon.

Figure 14 *explains* something to us.

It *allows* us to understand much better the *idea* that it *takes* an *infinite* amount of time for a thing to *fall* through the horizon – at *least* an infinite amount of *coordinate* time.



But still, it falls through in its own time keeping.

And with that time it falls through in a finite time.

Let's *redraw* the figure and see if we can *understand* that.

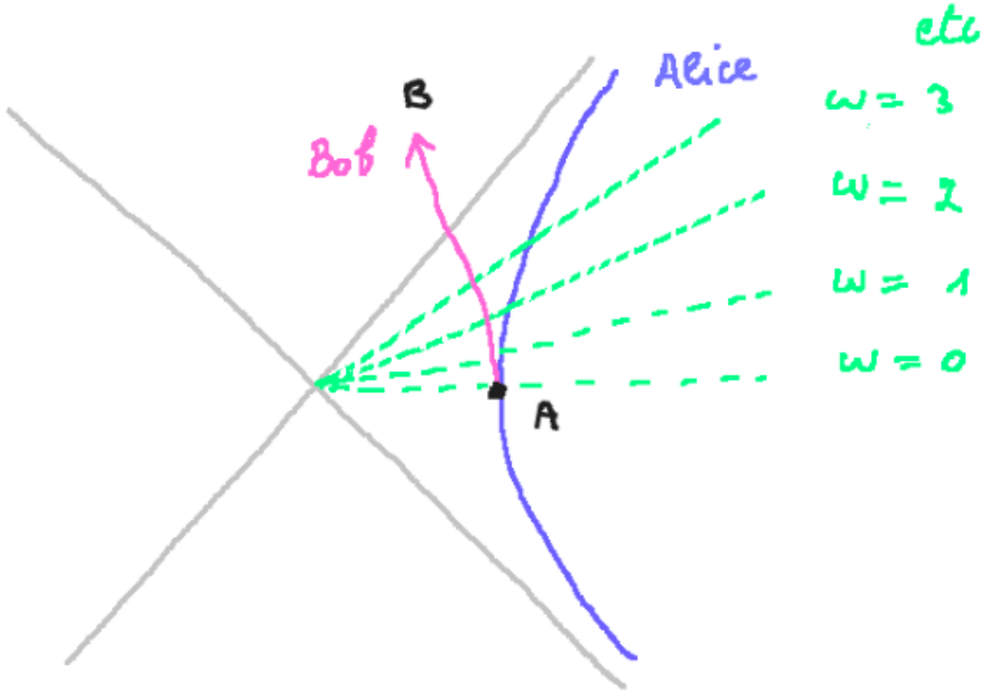


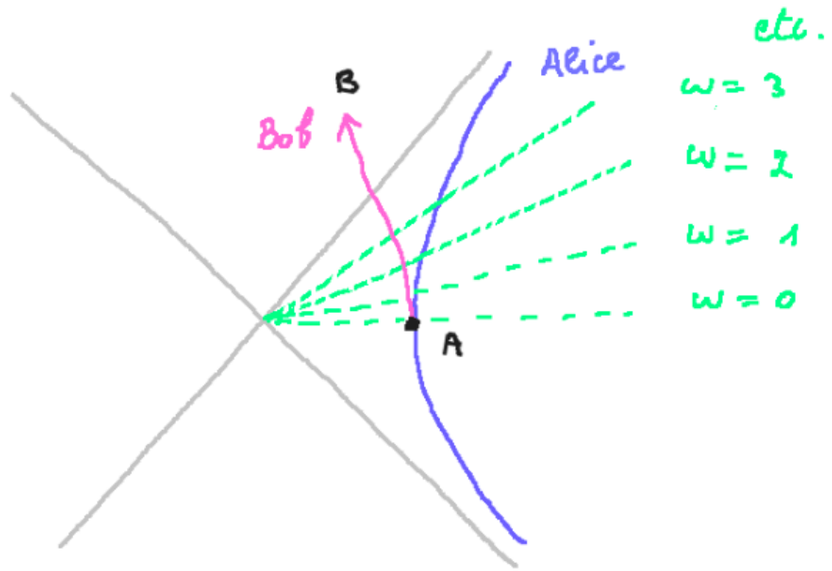
Figure 15 : Alice stays **outside** the black hole horizon and Bob **falls in**.

Consider the two famous protagonists : *Alice and Bob*. Alice, one way or another, stays *outside* the horizon of the black hole at a *fixed distance* ϱ from the horizon.

And she *pushes* Bob or at least *observes* him *fall* into the black hole.

The *origin*, in figure 15, is the point on the *horizon* below her, as in figure 14.

The *straight* lines from the origin are *time slices of constant* ω.



Remember that ω is *like* time.

The *relationships* between *standard* Minkowski coordinates and *hyperbolic* coordinates are given by these equations

$$X = \rho \cosh \omega$$
$$T = \rho \sinh \omega$$

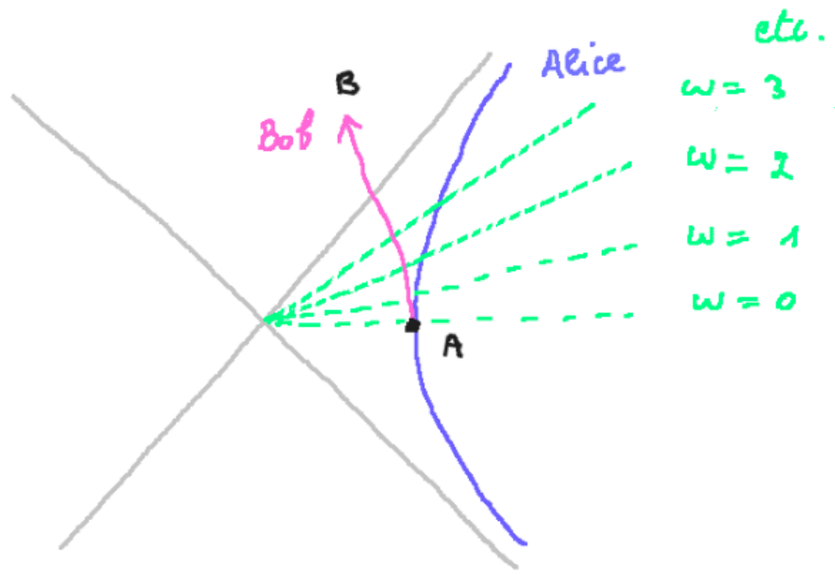
Here, we have *only* translated the horizontal axis so that the *origin* of the diagram is *not* at the center of the black hole, *but* on its horizon.

So Alice's *trajectory* is the *hyperbola* in fig 15.

Even though *she doesn't move*, it is *not* a vertical straight line *because* of the Schwarzschild metric which is a *gravitational* field.

And we saw that, at any point except the center of the black hole, it is *everywhere* locally equivalent to a uniformly accelerated frame, even in the vicinity of the horizon.

So a motionless point above the horizon follows a *hyperbola*.



Alice's ϱ doesn't change, and the time, measured by ω, ticks as she passes each radial line in the picture.

Remember that ω is proportional to her proper time, see equation (22).

ω asymptotically goes to infinity when the radial lines becomes closer and closer to the 45° line.

Alice says that the surface with 45° slope corresponds to infinite time.

Notice that the coordinate time T – the time of the physicist observer of the whole thing if you like – becomes infinite if and only if ω becomes infinite.

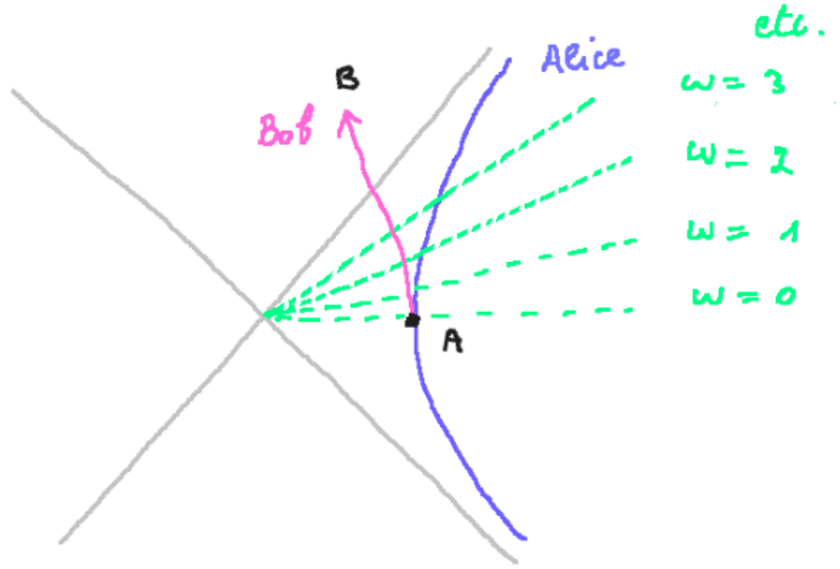
If we consider now Bob unfortunate trajectory, it is not until an infinite amount of Alice's proper time has elapsed, that Bob has gone through the horizon.

In other words she never sees Bob go through the horizon.

According to her reckoning, it takes an infinite amount of ω before Bob falls through the horizon.

That translates to an infinite amount of T.

On the other hand, just looking at the diagram in fig 15, we can see immediately that the amount of proper time it takes Bob to go from point A to point B is finite.



In other words Bob goes through the horizon in a finite time.

The surface $X = T$ is the horizon.

We will come back to it.

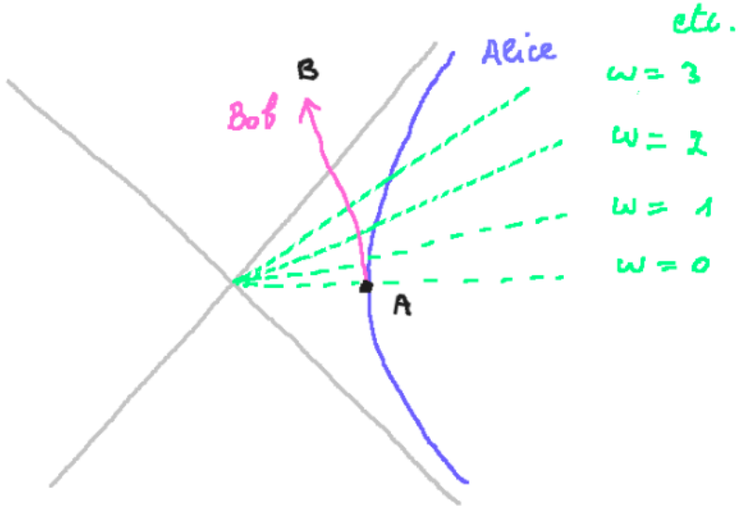
It is only a funny coordinate property which leads to Alice's reckoning that Bob takes an infinite amount of time to fall through the horizon, while Bob says : no, I watched it on my clock, it only takes a finite amount of time.

The reason is the discrepancy between proper time and coordinate time.

And Alice uses her proper time, which is linked to coordinate time by $T = \varrho \sinh \omega$, while Bob uses his proper time.

We can figure out how Alice "sees" Bob.

It requires somehow for Alice to look back in time, because when we speak of looking we refer to light arriving to our eyes, and light takes time to travel.



But let’s talk of "looking" in a different way : when we say Alice looks at Bob at time ω, we mean she figures out what does Bob do at this simultaneous time ω.

In other words, where is Bob on the radial line of value ω. Well, we know from the picture where he is.

So we see that Alice sees Bob slow down. Each heartbeat of Bob takes longer in Alice’s time.

Now how does Bob see Alice?

We apply the same reasoning : by what Bob "sees" of Alice, we mean what happens to Alice at the same Bob's time, when he "looks back", that is going into the past coordinate time, but simultaneous for him.

Those lines of simultaneous times for Bob are –45° lines which link points on Bob's trajectory and Alice's trajectory, figure 16.

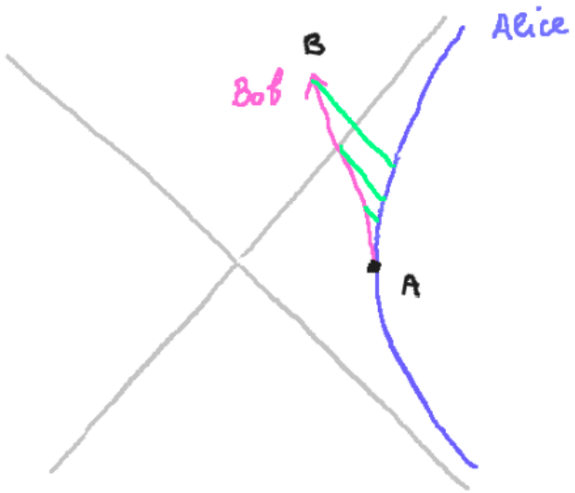


Figure 16 : How Bob sees Alice.

So we see that Bob doesn't see anything special.

He sees Alice before he plunges through the horizon.

He sees Alice while he is at the horizon.

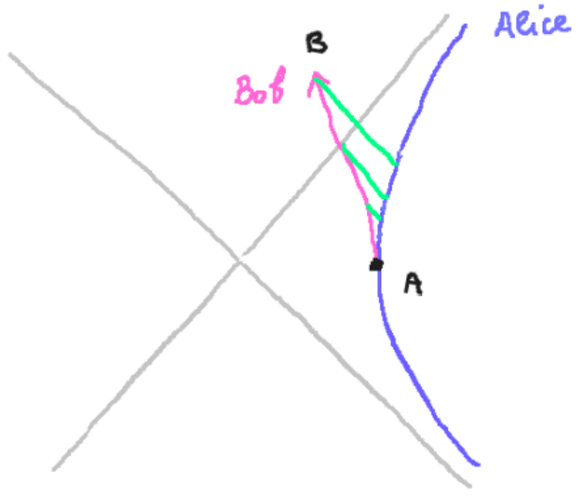
And he even sees her afterwards without problem.

While on his trajectory, on which he does not notice anything particular happening when crossing $r = 2MG$, Bob doesn't see Alice shrink.

He does see Alice accelerate away from him, that's true.

But that is all.

At any point, he sees Alice perfectly normally, except that she is accelerating away from him.



Notice that when Bob is at B, he can see Alice.

Whereas Alice can never see Bob at point B.

So it is a very asymmetric situation.

What happens when Bob hits the singularity?

By that point the tidal forces will become so large that Bob will no longer be with us.

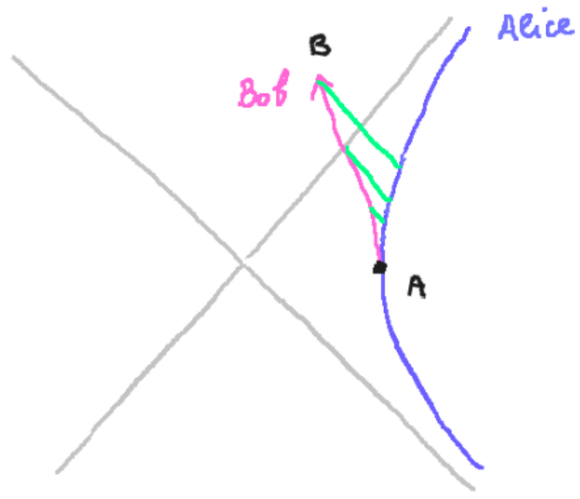
Bob will have been destroyed by the time he gets to the center of the black hole field.

That is why we don't think about trying to answer the question : what does Bob see when he gets to, say, a negative r.

By the time he gets to negative r, he has experienced infinitely strong tidal forces and there is no more Bob.

Notice that in the upper quadrant, Bob will probably use r for time or something like r.

Inside the horizon, where r varies from $r = 2MG$ to $r = 0$, Bob's clocks would have more to do with r than with t of equation (2).



And t would be more like position.

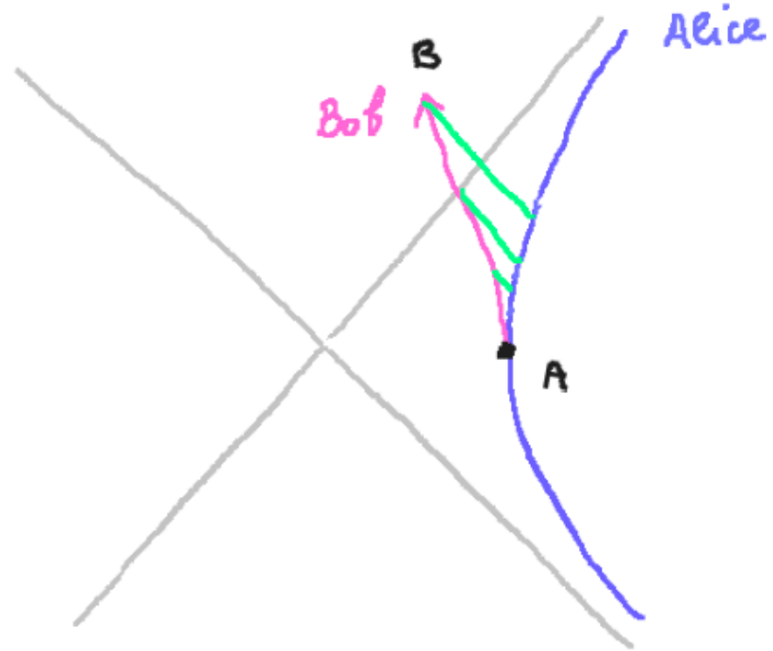
But, as I said, Bob doesn't feel any funny thing happening with space or time.

His clock doesn't become a meter stick, and his meter stick doesn't becomes a clock.

Nothing like that happens. He just sails through.

Yet the line at 45° in figure 16 is rather special.

Let's see what can and cannot happen once something – a person, an object, a particle or even a photon – has crossed the horizon.



No escaping from a black hole

Anything that crossed the horizon can no longer escape.

Remember, in the coordinates that we are using light moves with a 45 degree angle.

Therefore light cannot escape from the upper quadrant in figure 17.

All it can do is eventually hit the singularity.

And anything that is moving slower than the speed of light has a slope closer to the vertical, and will also hit the singularity.

So anybody who finds himself or herself in the upper quadrant is doomed.

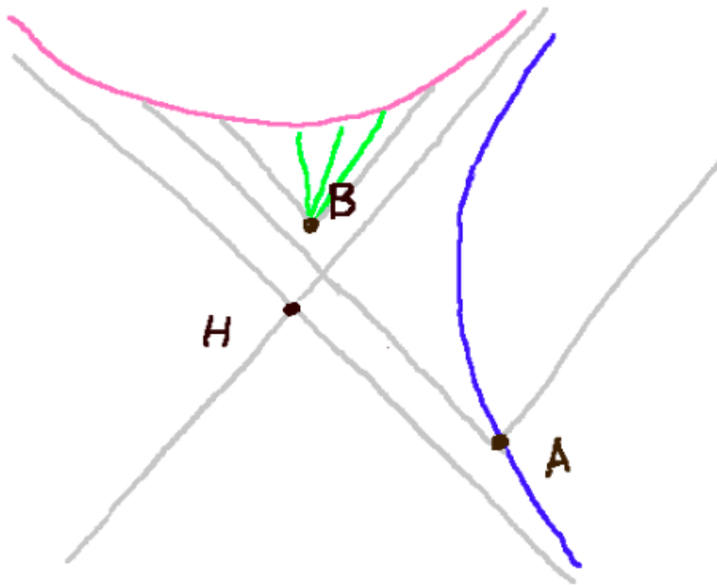
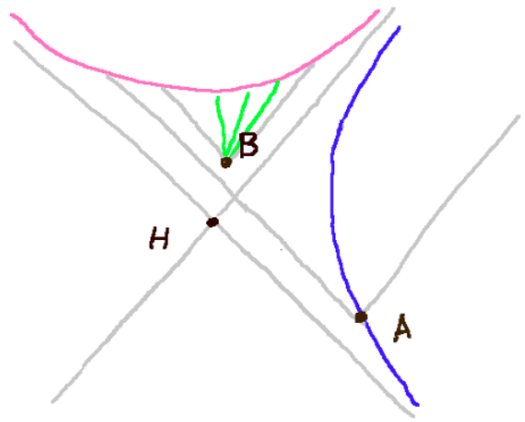


Figure 17 : Trajectories starting inside or outside the horizon.

Somebody in the right quadrant has the possibility of escaping.

Let’s talk again about light-ray.



If it is moving radially outward, it will escape.

It will simply just keep going.

If it is pointing radially inward it is of course also doomed.

It will hit the singularity.

And there is everything in between.

If a light-ray is pointing out of the page, depending on whether it is inside or beyond the photon radius, it will fall in or not.

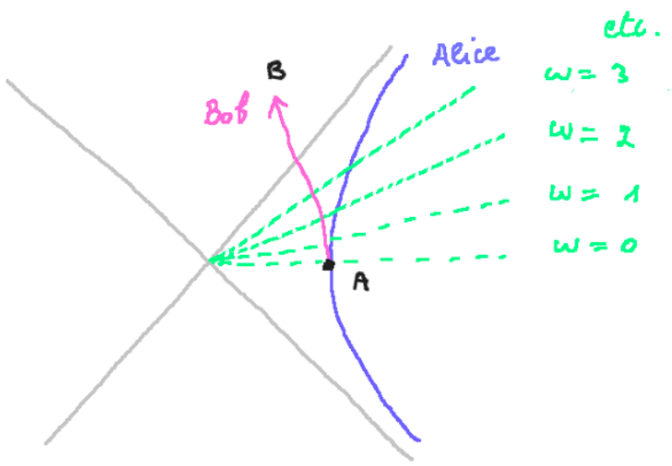
Figure 17, and its variants fig 15 and 16, are the pictures you want to get used to.

If you want to understand and be able to resolve funny paradoxes about who sees what in the black hole, you should go back to this picture.

Alice is outside the black hole at a fixed distance ϱ, subjected to gravitation locally equivalent to uniform acceleration.

Alice's time corresponds the dotted lines shown in fig 15.

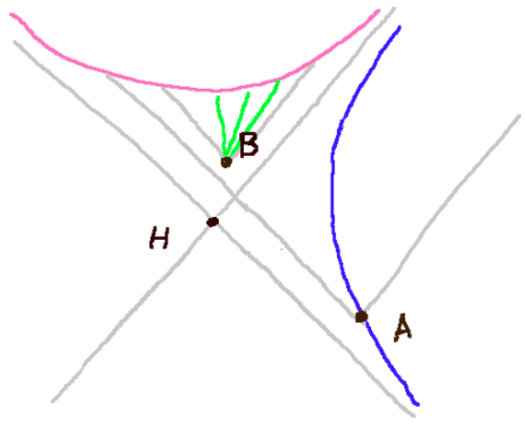
Up to the fixed factor ϱ, her clocks record the hyperbolic angle ω, because Alice's proper time $d\tau$ is, with hyperbolic coordinates, $\varrho d\omega$.



See equation (22) which says

$$d\tau^2 = \rho^2 d\omega^2 - d\rho^2$$

Bob starts outside the black hole, and passes inside below $r = 2MG$.



The trajectory he follows can be a free fall or any other trajectory.

In fig 17, outside the black hole corresponds to the right quadrant, and inside the black hole corresponds to the upper quadrant.

That is due to the peculiar Schwarzschild metric given by equation (2), whose first and second coefficients interchange signs at the horizon.

Bob's clocks record Bob's proper time.

Like for anyone else, it is $d\tau$ given by equation (2).

It is the strong mismatch between Alice's proper time and Bob's proper time which produces all the peculiarities showing up around a black hole.

A good understanding of black holes is a prerequisite to understanding general relativity.

The reason is that black holes are the simplest, ideal form of massive bodies: their mass is concentrated at a point.

It is the equivalent in relativity of point masses in Newtonian physics.

Point masses in Newtonian physics give rise to Newtonian gravitation, trajectories solved with Newton's equation, etc.

Their physics has a singularity at the central point itself, but nowhere else.

In relativity the analog of point masses is black holes, and their metric.

It is Schwarzschild metric.

Black holes display of course a singularity at their center, but also strange phenomena at $r = 2MG$.

However the peculiar phenomena happening at $r = 2MG$ only come from the coordinates, not from real physical strangeness.

The next two lessons will be devoted to deepening our knowledge of black holes.

Lesson 7 : Falling into a black hole

Introduction

The lesson is devoted to reviewing, getting acquainted with, and deepening what we began to learn in the previous lesson, namely what happens around a black hole, outside the horizon, at the horizon, and at the singularity.

We will go again over the change of coordinates and the corresponding diagram which facilitate understanding the physics at the horizon and in its vicinity.

We will study people and objects falling in, and possible and impossible communications between inside and outside.

Schwarzschild metric, event horizon, singularity

Let’s get back to the Schwarzschild metric, named after Karl Schwarzschild (1873 - 1916).

In the early years of last century he was a professor at Göttingen, with among his colleagues Hilbert and Minkowski.

I don’t know much more about him because that was before my time.

But we shall know his metric very well.

We shall simplify its equation a little bit by rescaling the coordinates.

This won't change anything about the physics, but will make the equations simpler to work with.

Let's write it first as we learned it.

But we put primes to the time and radius variables, so that after the change of variables, the new variables will be as usual (t, r, θ, ϕ).

$$d\tau^2 = \left(1 - \frac{2MG}{r'}\right) dt'^2 - \left(\frac{1}{1 - \frac{2MG}{r'}}\right) dr'^2 - r'^2 d\Omega^2 \tag{1}$$

r′ is the radius and t′ is the time used in the last lesson.

This is exactly the same equation as in the previous lessons, except with just a slight change of notation.

Let's give a new name to the Schwarzschild radius $2MG$:

$$R_S = 2MG \tag{2}$$

R_S reads "R Schwarzschild".

Next, we rescale the radius.

We define r as follows

$$r = \frac{r'}{R_S} \tag{3}$$

Normally radii have units of length.

But when we divide a length by a length, we get something which is dimensionless.

So our new r, which is not the same r that we used in the last lesson, is dimensionless.

We do the same thing with time.

What are the units of time?

Time.

But since in our equations we carry an implicit $c = 1$, which has units length over time, for us units of time and units of length are the same.

Our new t is

$$t = \frac{t'}{R_S} \tag{4}$$

You can think of R_S as both a distance and a time.

It is the distance from the center of the black hole to the horizon.

And it is also the time light takes to cover that distance.

Finally, the equation of the metric takes the following simpler form

$$d\tau^2 = \left[\left(1 - \frac{1}{r}\right) dt^2 - \left(\frac{1}{1 - \frac{1}{r}}\right) dr^2 - r^2 d\Omega^2 \right] R_S^2 \tag{5}$$

Inside the brackets is the metric that we would have written down in the first place if $2MG$ were equal to 1.

What we find out from this exercise is that basically the metric of all Schwarzschild black holes is the same, except for an overall factor which is proportional to the square of the radius of the black hole.

Moreover, everything inside the brackets is dimensionless.

All of the dimensions are in R^2_S.

For most purposes in studying the geometry of the black hole, we can really just set R_S equal to one.

It is like studying the metric of a sphere.

The metric of all spheres is the same, with just a different radius.

There are bigger spheres and smaller spheres, but, up to a factor which is their radius, they all have the same geometry.

Analogously, all black holes are the same, up to their radius.

Now let's turn to a question we haven't asked ourselves for a while.

What is the curvature of space-time near a black hole?

Remember the curvature tensor curvature tensor that we studied in lesson 3.

It is the tool which tells us whether the space-time is flat or curved.

It is analogous to tidal forces.

It is the tendency for the geometry to distort objects.

And in fact tidal forces are curvature.

The formula is

$$\mathcal{R}^{\ t}_{srn} = \partial_r \Gamma^t_{sn} - \partial_s \Gamma^t_{rn} + \Gamma^p_{sn}\Gamma^t_{pr} - \Gamma^p_{rn}\Gamma^t_{ps} \tag{6}$$

Our question is : how big are the components of the curvature tensor, let's say at the horizon?

The horizon corresponds to $r = 1$.

A preliminary question is : what are the units of curvature?

Remember that a Christoffel coefficient can be calculated.

It is expressed in terms of the metric components as follows

$$\Gamma^t_{mn} = \frac{1}{2}\, g^{rt}\ [\, \partial_n g_{rm} + \partial_m g_{rn} - \partial_r g_{mn}\,] \tag{7}$$

We also remember that

$$dS^2 = g_{\mu\nu} dX^\mu dX^\nu \tag{8}$$

From equation (8), it follows that $g_{\mu\nu}$ is dimensionless.

And so is its inverse $g^{\mu\nu}$.

Then, since the Christoffel symbols, in equation (7), contain spatial derivatives of g, they have units one over length.

And we conclude, from equation (6), that the curvature tensor components have units one over length squared.

Using the standard notation with brackets for dimension, this is expressed by writing

$$[\mathcal{R}] = \frac{1}{m^2} \tag{9}$$

Let's come back to the question : what is the curvature at r=1?

And let's use a nice argument resting only on dimensions to find the answer.

Looking again at the metric, the only quantity on the right hand side of equation (5) which has units of length is R_S.

Everything else is dimensionless.

Therefore when we calculate the curvature – doing all sorts of algebraic manipulations on the dimensionless elements in equation (5) –, the only possibility for it to have units one over length squared is if the curvature is inversely proportional to R^2_S itself, that is, to the Schwarzschild radius squared.

We write it

$$\mathcal{R}_{Horizon} \sim \frac{1}{R_S^2} \tag{10}$$

where ~ stands for "proportional to", with a fixed proportionality ratio, the same for all black holes.

To convince yourself of this argument, think of what would happen if you multiplied R_S by any dimensionless constant k and changed the units before doing all the calculations.

For a given black hole, if at $r = 1$ the curvature has a certain value, at $r = 2$ the curvature will be four times smaller.

Conversely at radius $R_S/2$, the curvature will be four times bigger.

As we approach the singularity, the curvature becomes infinite.

For different black holes, the bigger the black hole – i.e., the more massive it is – the smaller the curvature at its horizon.

So the tidal forces associated with a large black hole are less severe than the tidal forces at a small black hole.

A small black hole is a much nastier thing near its horizon than a very big black hole.

Now let’s concentrate on the geometry created by the metric of equation (5).

And let’s throw away the R^2_S , in other words let’s replace it by 1.

We consider, if you like, a "unit black hole", that is, a black hole of radius 1.

Let’s go back to something we already talked about in the last lesson.

We have only more or less outlined it.

Let’s do it again.

We want to use new coordinates.

You may think that we keep changing coordinates but that will make things simpler.

And it is what relativity is all about : it is changing coordinates.

We will introduce a new coordinate to replace r.

Figure 1 below shows the r axis.

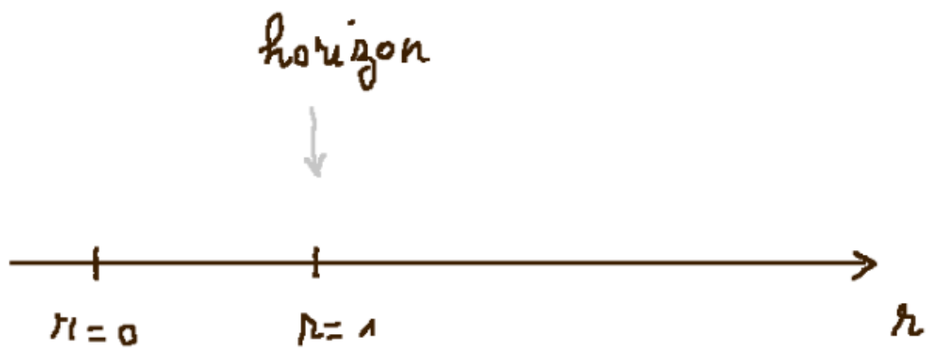


Figure 1 : Coordinate r.

We want to invent a new coordinate which is just proper distance from the horizon.

How do we calculate the distance from the horizon?

We go back to the expression of proper time given in equation (5), which we rewrite below, with $R_S^2 = 1$,

$$d\tau^2 = \left(1 - \frac{1}{r}\right) dt^2 - \left(\frac{1}{1 - \frac{1}{r}}\right) dr^2 - r^2 d\Omega^2 \tag{11}$$

Remember that $d\tau^2$ is proper time squared, and dS^2 is proper distance squared.

These are just minus each other.

So proper distance is given by almost the same equation as equation (11)

$$dS^2 = -\left(1 - \frac{1}{r}\right) dt^2 + \left(\frac{1}{1 - \frac{1}{r}}\right) dr^2 + r^2 d\Omega^2 \tag{12}$$

Now let's move away from the horizon radially outward.

Consider two points A and B at distance r and $r + dr$.

What is the proper distance between these two points?

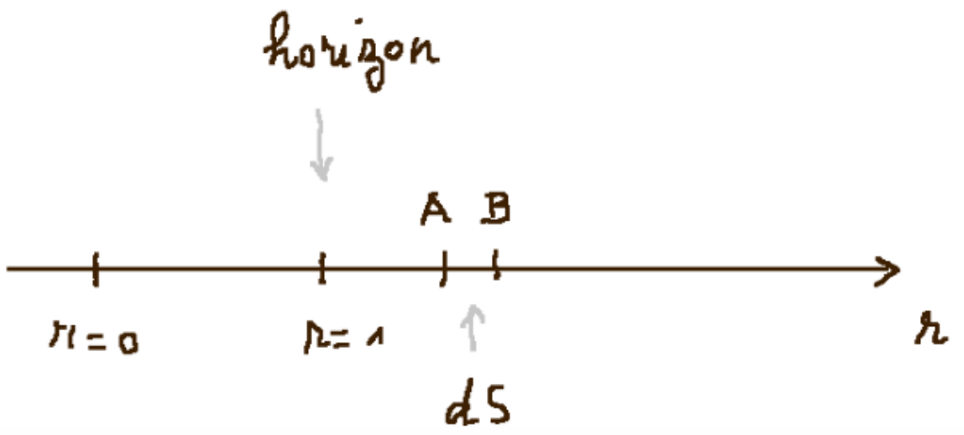


Figure 2 : Proper distance between A and B.

To find out we use equation (12).

First of all we compare two points at the same time, so $dt = 0$.

Secondly we move radially, so Ω doesn't change either.

Therefore between A and B, the proper distance squared is

$$dS^2 = \frac{dr^2}{1 - \frac{1}{r}}$$

or

$$dS = \sqrt{\frac{r}{r-1}}\, dr \tag{13}$$

Let's give a name to the proper distance from the horizon.

We call it ϱ.

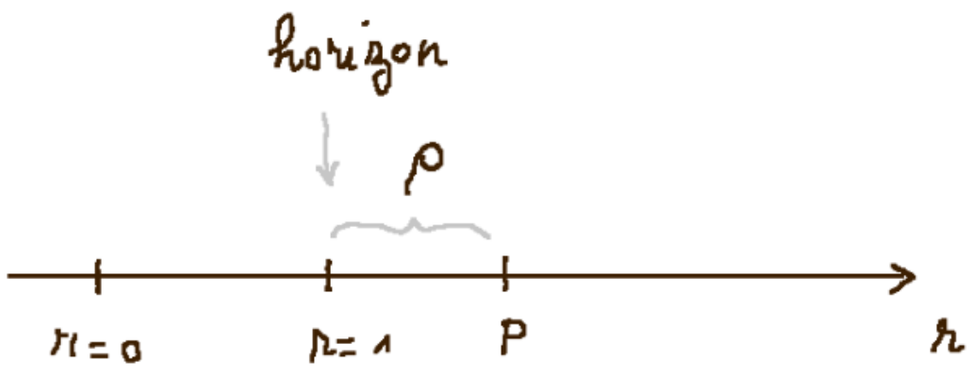


Figure 3 : ϱ = proper distance from the horizon.

If $(0, r, 0, 0)$ is the coordinate of P in the coordinate system (t, r, θ, ϕ), notice that the proper distance between P and the horizon is not $r - 1$.

The quantity $r - 1$ is a certain kind of distance but it is not the proper distance from the horizon.

How do we find the proper distance between P and the horizon?

Well, we know its differential when we move a distance dr along the r axis, figure 2 and equation (13).

Therefore ϱ, the actual proper distance from the horizon to P, is the integral

$$\rho = \int_{u=1}^{u=r} \sqrt{\frac{u}{u-1}}\, du \tag{14}$$

This is some function of r.

The integral is a doable integral, but we don't need to calculate it.

ϱ is just some function of r, which we denote $\varrho(r)$.

Moreover, the metric given by equation (11) can be expressed in terms of ϱ instead of r, because r is some function of ϱ.

From equation (14), if we know r we know ϱ, and if we know ϱ we know r.

To make equations nice, let's also modify a bit the time.

We define ω as follows

$$\omega = \frac{t}{2} \tag{15}$$

Now let's reexpress $d\tau^2$ of equation (11) with the variables ϱ and ω instead of r and t.

The second term $r/(r-1)\, dr^2$ is simply $d\varrho^2$.

We see this from equation (13), remembering that we decided to call the proper distance ϱ.

Concerning the first term $(r-1)/r\, dt^2$, it can be expressed as a function of ϱ times $d\omega^2$.

For reasons that will become clear in a moment let's write this function $F(\varrho)\varrho^2$.

With this change of variables from (t, r) to (ω, ϱ), the metric giving $d\tau^2$ becomes

$$d\tau^2 = F(\rho)\rho^2 d\omega^2 - d\rho^2 - r(\rho)^2 d\Omega^2 \tag{16}$$

It is possible to calculate explicitly the last coefficient $r(\varrho)^2$ but we don't need to.

This form of the metric is sufficiently interesting as it is in equation (16), plus a little bit of knowledge about what $F(\varrho)$ and $r(\varrho)$ are like.

We already saw some of this in the last lesson, in the section entitled "Hyperbolic coordinates revisited" of lesson 6.

But we are now going into more details.

And we use now explicitly the proper distance which we denote ϱ, whereas in lesson 6 we worked differently.

We stayed near the horizon, at $r \approx 1$, and worked directly with the quantity $r - 1$, which we denoted ξ, see equations (26) to (28) of lesson 6.

Where is the horizon in terms of ϱ?

It is at $\varrho = 0$, because ϱ is the (proper) distance from the horizon.

In equation (16), we can say a number of things.

First of all, when ϱ becomes very big, far away, it is easy to show that

$$\lim_{\rho \to +\infty} F(\rho)\rho^2 = 4 \tag{17}$$

That just reflects the metric far from the black hole.

That is one thing.

Next, let's look at what happens when ϱ goes to zero.

That is getting right up against the horizon.

Again we easily establish that

$$\lim_{\rho \to 0} F(\rho) = 1 \tag{18}$$

The first term in the expression of $d\tau^2$, in equation (16), becomes $\varrho^2 d\omega^2$.

You may find that familiar.

That was the reason, by the way, for defining the term function of ϱ in front of $d\omega^2$ as $F(\varrho)\varrho^2$.

Indeed, that way, close to the horizon $F(\varrho)\varrho^2$ is just ϱ^2, or equivalently $F(\varrho)$ is just equal to one.

Finally, let's look at the limit of $r(\varrho)$ when ϱ goes to zero.

That we can work out.

$\varrho = 0$ means at the horizon.

There $r = 1$, but we can do a little bit better.

We can show that

$$\lim_{\rho \to 0} r(\rho) = 1 + \frac{\rho^2}{4} \tag{19}$$

This non standard way to write a limit means, if you prefer, that $r(\varrho)/(1 + \varrho^2/4)$ tends to one.

Equations (17), (18) and (19) are basically everything that we need to know to study the metric.

Now the most important thing is to look at equation (16) when ϱ is small.

It is approximately

$$d\tau^2 = \rho^2 d\omega^2 - d\rho^2 - d\Omega^2 \tag{20}$$

That is exactly flat space in hyperbolic polar coordinates.

So let's review for a minute hyperbolic polar coordinates in the vicinity of a point H on the horizon, figure 4.

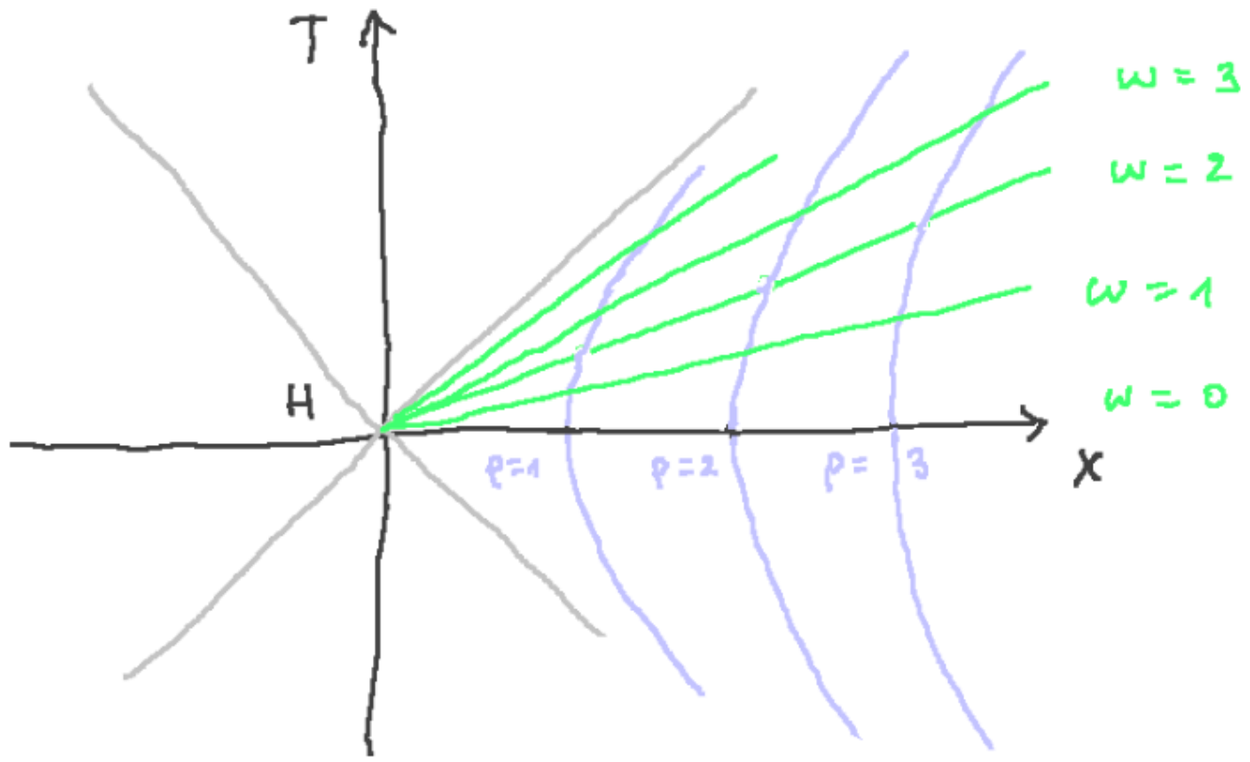


Figure 4 : Hyperbolic polar coordinates in the vicinity of the horizon.

There are two important quadrants in this picture, the right quadrant of events located farther than the horizon, and the upper quadrant of events located inside the horizon.

Let's look first at the right quadrant.

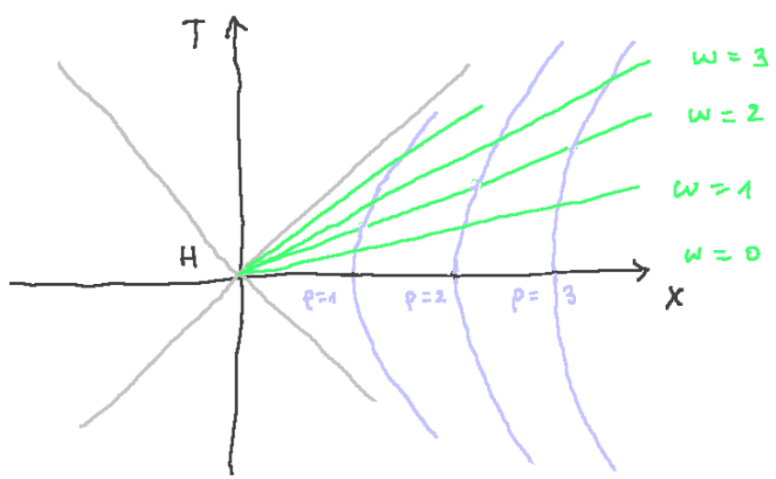
The first hyperbola corresponding to $\varrho = 1$ is the trajectory in time of a particle fixed at one unit of ϱ away from the horizon.

The second hyperbola is the trajectory of a particle fixed at $\varrho = 2$, etc.

That is just looking at the metric of equation (16) and noticing that, apart from the coefficient $F(\varrho)$, near the horizon it just has the form of flat space in hyperbolic polar coordinates that we studied in lesson 4.

The lines of constant ω are the straight lines fanning out of the origin in figure 4, that is event H.

We drew $\omega = 0$, that is the horizontal axis, $\omega = 1$, $\omega = 2$, $\omega = 3$, etc.



Where is $\omega = +\infty$?

It is right along the light-cone, which on the diagram is the line at 45°

Remember what is ω.

It is just $t/2$.

And we had defined t and r as follows

$$X = r\cosh t$$
$$T = r\sinh t \tag{21}$$

So we have discovered some coordinates, (ω, ϱ), in which time is like a hyperbolic angle.

And time infinity is just the light cone at 45°.

What about the upper quadrant?

We already talked about it in the last lesson.

It is the region where $r < 1$.

When $r < 1$ the sign of the first two terms in equation (11), or equivalently in equation (16), interchange.

In the upper quadrant ϱ^2 is negative.

We can see this from equation (19).

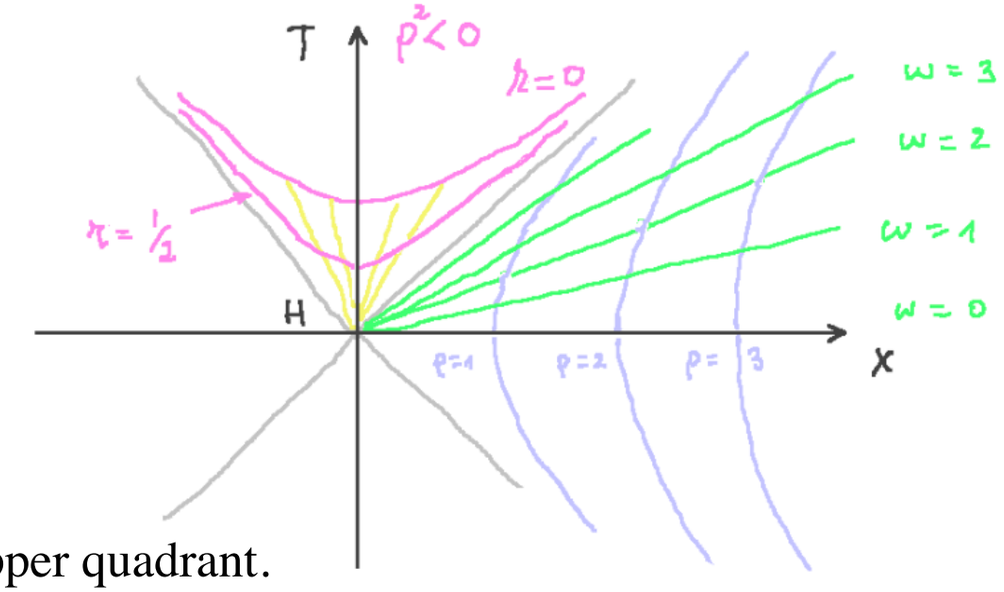


Figure 5 : Constant ϱ's and constant ω's in the upper quadrant.

In the upper quadrant the curves of constant ϱ, or equivalently constant r, are the hyperbolas shown in figure 5.

We drew $r = 1/2$ and $r = 0$.

Nothing crazy physically is going on in the upper quadrant.

Time has not become space, space has not become time.

We have just introduced coordinates which have the funny property that when you go from the right quadrant to the upper quadrant, what was time-like becomes space-like and conversely.

It is a complete coordinate artifact.

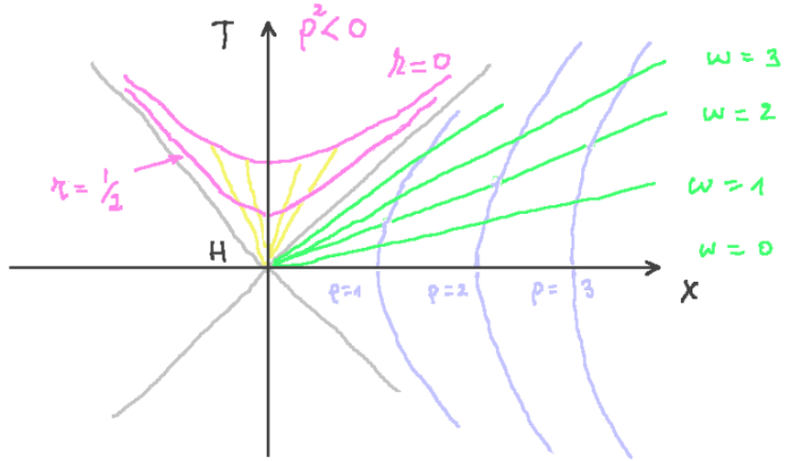
Yet there is something interesting about it.

If we go far away from H, we have to remember that $F(\varrho)$ is not just one.

And $r(\varrho)$ is not just one either.

That means the metric along the ω direction, $F(\varrho)\varrho^2$, does differ from flat space.

But the difference shows up in the way $F(\varrho)$ varies as we move away from the black hole.



Same comment on $r(\varrho)$ in front of $d\Omega^2$.

The origin H in figure 5 is $\varrho = 0$ but $r = 1$.

What happens as we move up into the upper quadrant and r decreases?

Eventually we hit the nasty point where $r = 0$.

Then that is no longer just a coordinate glitch, we are at a real physical singularity.

If we calculated the curvature near the hyperboloid $r = 0$, it is not too hard to figure out what we would find.

As r goes to zero, the sphere is getting smaller and smaller and the curvature is getting larger and larger.

It becomes infinite on the hyperboloid $r = 0$.

This is a true singularity.

It is a place of infinite tidal forces, and it is basically a place where you wouldn't want to be.

The problem with the singularity is that once we are in the upper quadrant, once we are below the horizon of the black hole, we can't avoid it.

It is not really a place.

As we discussed in the last lesson, it is really, in a sense, a time.

And, if we can avoid running into an obstacle in space – we just go around it –, we cannot avoid "running into the future".

If you are orbiting in a space station far away from the horizon of the black hole, as long as you don't do anything stupid, like jump off and allow yourself to get sucked in by gravity, you are safe in your capsule.

Nothing bad will happen to you.

But if you are foolish enough to think "I want to explore what is in there" and decide to go and see, as soon as you pass the horizon of the black hole, you are doomed.

Not only there is no way you can get out, but there is even no way you can avoid flying into the singularity and being destroyed by tidal forces.

In order to get out you would have to exceed the speed of light.

That corresponds to the 45° line in figure 5, or to the light cone of 45° lines if we add the other spatial dimensions.

Once you are in the upper quadrant the best you can do to try to escape – if we assume that nothing can exceed the speed of light – is follow a 45° trajectory.

And they all run into the hyperboloid $r = 0$ in a finite amount of your proper time.

Where is the point of no return?

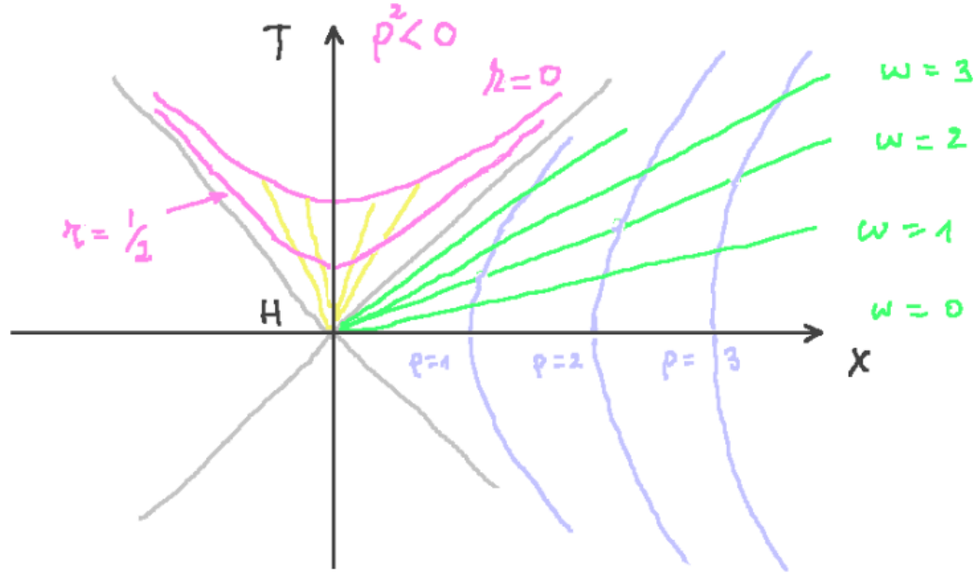
It is actually the entire line at 45°.

To understand why, let's ask ourselves : what is that light-cone?

In the right quadrant each hyperboloid is at a different value of ϱ.

And the entire hyperboloid corresponds to its value of ϱ.

In figure 5 we see the hyperboloids $\varrho = 3, \varrho = 2, \varrho = 1,$ etc.



At the point $H, \varrho = 0$.

But in fact $\varrho = 0$ all along the limit hyperboloid which is formed by the two lines at 45° and –45°.

There is something very different about this kind of geometry than ordinary geometry.

Suppose we drew ordinary geometry, as in figure 6, and asked : where is the point $r = 0$?

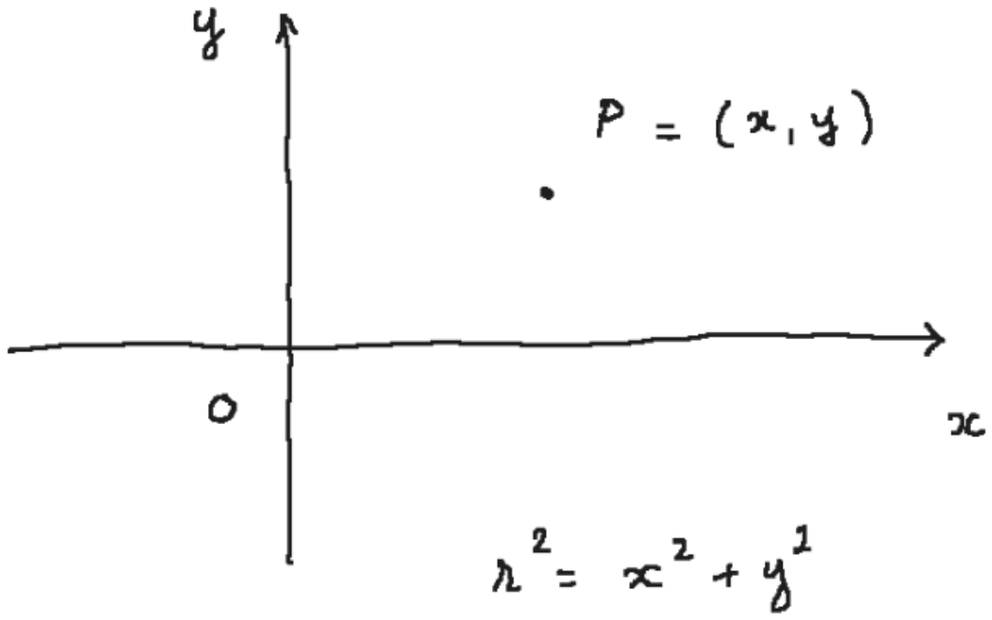


Figure 6 : Flat plane in ordinary Euclidean geometry.

Well we know that there is only one point where $r = 0$.

It is the origin.

Where is the point where $r = \varepsilon$?

It is not a point, it is a tiny circle around O.

As ε goes to zero, the circle gets closer and closer to being just point O.

Where is $\varrho = \varepsilon$?

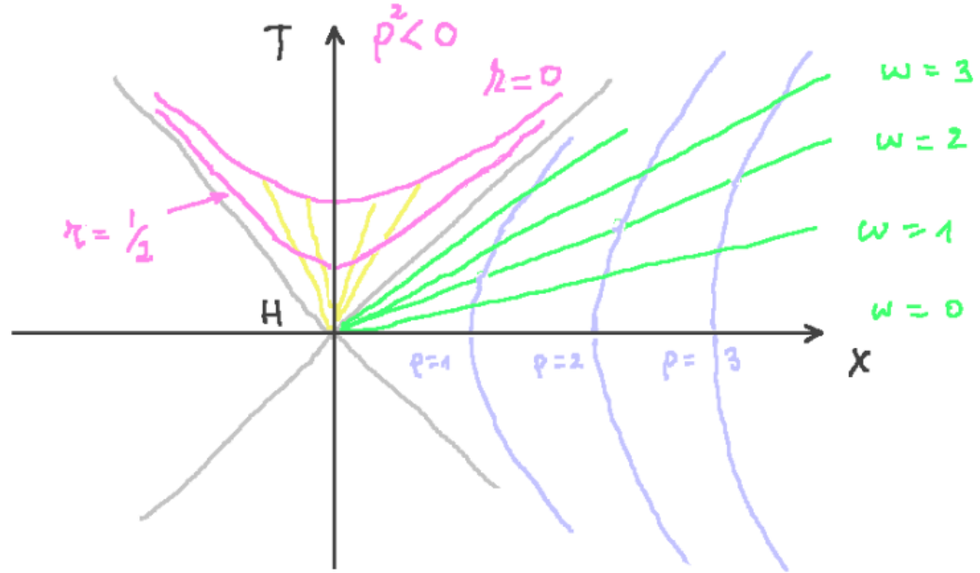
It is an entire hyperboloid close to the light-cone.

As ϱ gets smaller and smaller it tends to the light-cone, not just to the point H.

And $\varrho = 0$ is the horizon.

So the horizon is not just the point H at the center.

The horizon is all along the lines at 45° and –45°.



Anybody who finds himself or herself beyond a point on these lines is inside the black hole, inside the horizon.

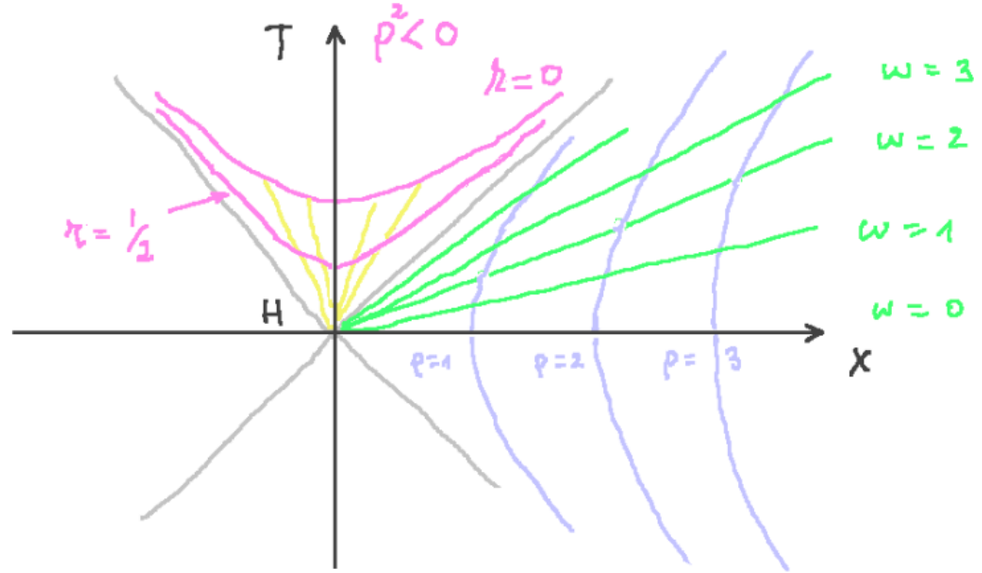
And as we can see just looking at figure 5 such a person is doomed.

To summarize, the point of no return is indeed $r = 1$, or equivalently $\varrho = 0$.

But it is not only the point H.

It is the whole line at 45°.

The whole line in figure 5 is the horizon of the black hole.



Fundamental diagram

It is fundamental to understand the diagram of figure 5 very well in order to understand gravitational fields created by massive bodies, and eventually the theory general relativity.

We reproduce it below.

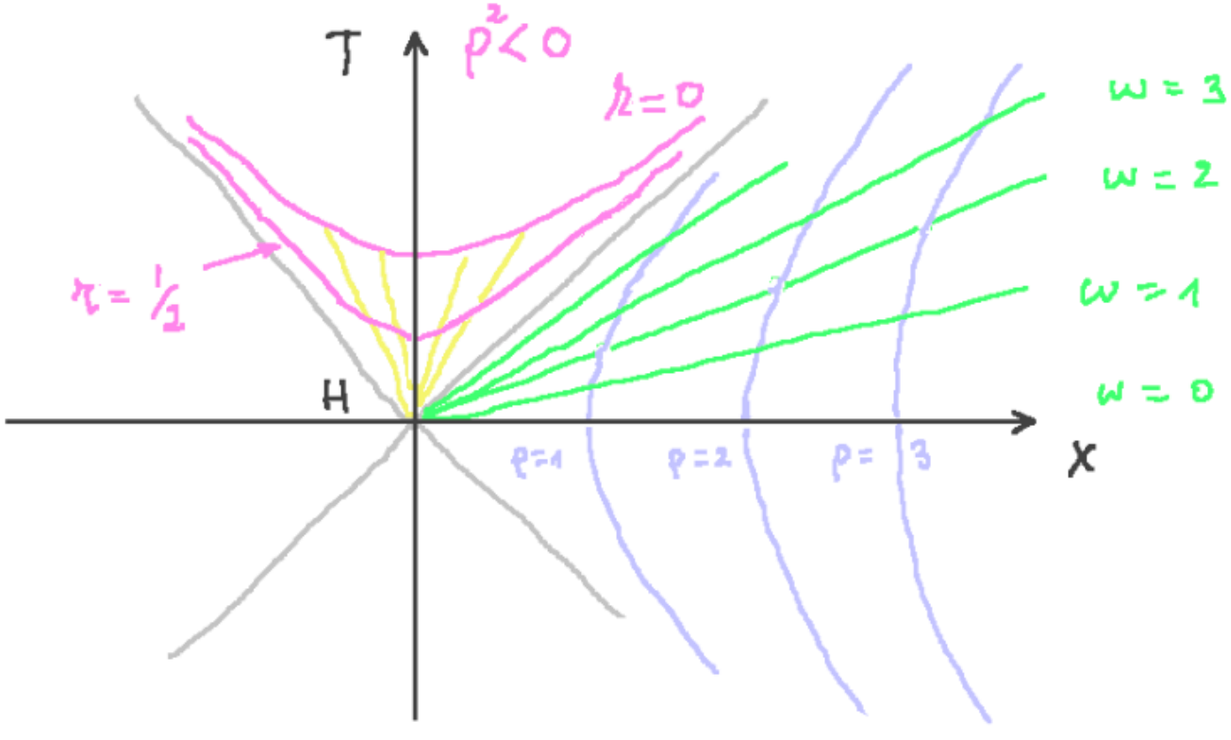


Figure 5b : Fundamental diagram to understand the gravitational field created by a point mass in general relativity.

The point mass is not at the point H, but at the curve $r = 0$.

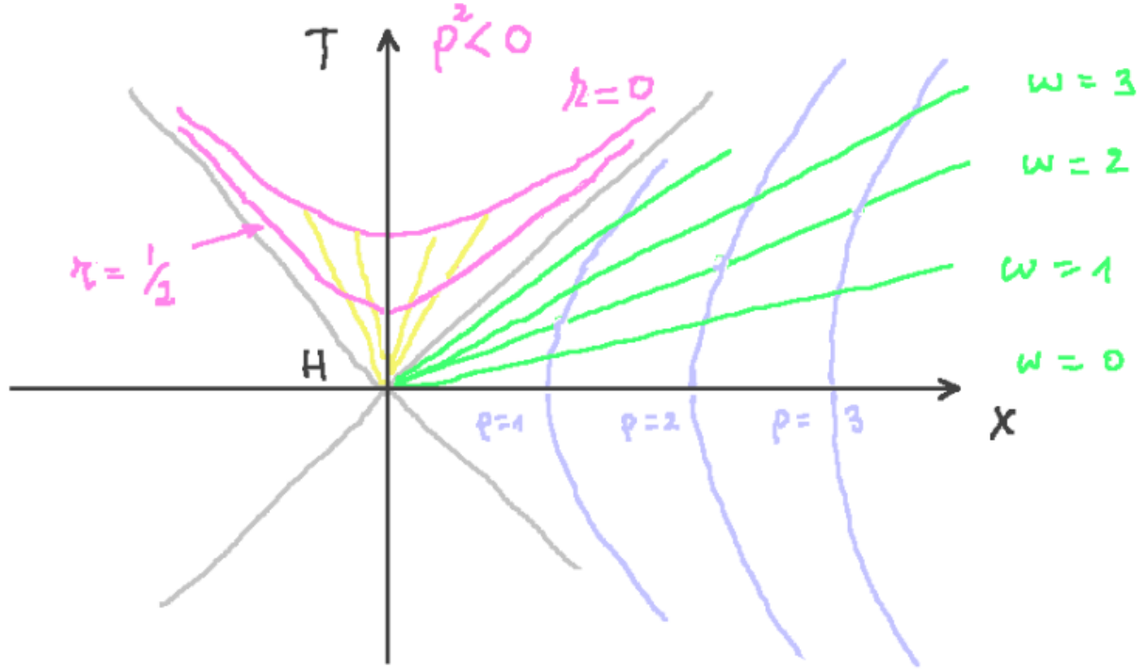
Point H and the two lines at 45° are the horizon of the point mass, called a collapsed star or a black hole.

Let's make miscellaneous comments on this diagram.

1. In the upper quadrant, the hyperbolas of constant r, or equivalently constant ρ, correspond to $\rho^2 < 0$. Therefore the ρ's are imaginary.

2. In the same upper quadrant, the straight lines coming out of H are still lines of constant t, or equivalently contant ω. But they are now space-like. And the hyperbola of a given r is time-like.

3. **What are the coordinates** $(X,\ T)$**?** They are the coordinates of an observer who is near the horizon of the black hole, and is falling freely in the gravitational field of the black hole.

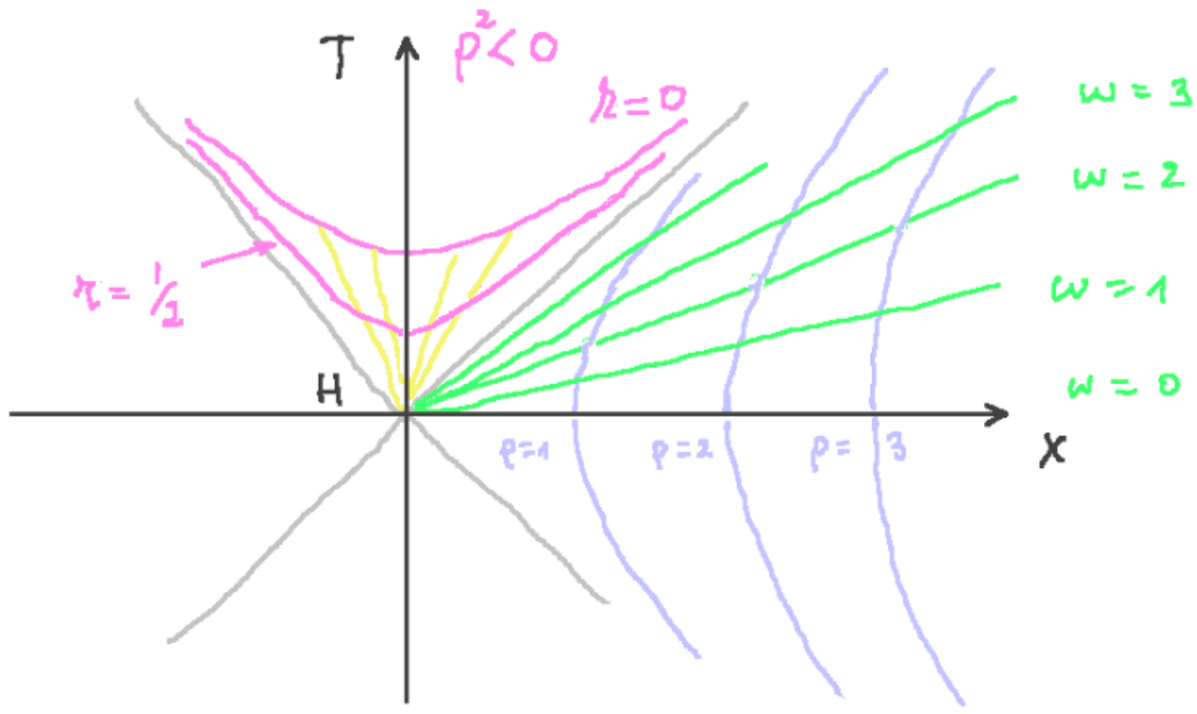


4. Someone falling freely through the horizon might indeed find it convenient to use the rectangular coordinates $(X,\ T)$ to chart what is going on around him or her. T is by definition the proper time of the observer. It is sometimes called the *coordinate time*. And X of course measures distances with the stick of the observer.

5. A particle which, unlike the observer, is at a fixed distance from the black hole, for instance at $\rho = 2$ from its horizon, follows in space-time a trajectory which, in the frame of reference of the observer, is the hyperbola in the right quadrant indexed with $\rho = 2$.

6. Time T and time t are different. Time T is the time of the observer. Time t is the time of any particle fixed relative to the black hole. Such a particle is living on one of the hyperbolas in figure 5b. More accurately, for each particle, its proper time is proportional to t, or equivalently proportional to ω which is $t/2$.

7. It is important to understand that, just like in Newtonian physics we need a reference frame to represent space and time – time being a neatly separate and universal coordinate –, in relativity we also need a reference frame to represent space-time. Since, due to massive bodies in the universe, there are no genuine uniform fields, no frame is clearly superior to any other. However the frame of reference of free-falling observer is particularly convenient.



The coordinates (ω, ϱ) have a name.

They are called the Kruskal coordinates.

History of black holes

When Schwarzschild wrote in late 1915 the Schwarzschild metric, he didn't know that the horizon was a horizon.

As far as he, as well as Einstein and all the people who studied these things for a while, could tell, $r = 1$ was a nasty place where two coefficients in the metric given by equation (5) changed sign and also one of them – the one in front of dr^2 – went through infinity.

They said "oh dear this coefficient is divergent at $r = 1$, something bad is happening there".

And so the conclusion was that the horizon was some sort of singularity.

In other words, the first people did not realize that the horizon was a smooth nonsingular ordinary place.

Sometime in the 1950's Finkelstein realized that the horizon of a black hole was the point of no return, but that at the horizon itself nothing nasty happened to someone falling through.

Only some finite proper time later would that person be annihilated at the singularity.

He rediscovered coordinates that had already been written down by Eddington.

So they were called the Eddington-Finkelstein coordinates.

They are not exactly the Kruskal coordinates, but are of the same type.

Martin Kruskal was not a relativity expert.

He was a plasma physicist but very good at equations and very good at changing coordinates.

He loved to change coordinates.

Somebody showed him the Schwarzschild metric.

He tried many changes of coordinates and eventually, in 1960, found these coordinates (ω, ϱ) in which the metric has the nice form of equation (16), that is,

$$d\tau^2 = F(\rho)\rho^2 d\omega^2 - d\rho^2 - r(\rho)^2 d\Omega^2$$

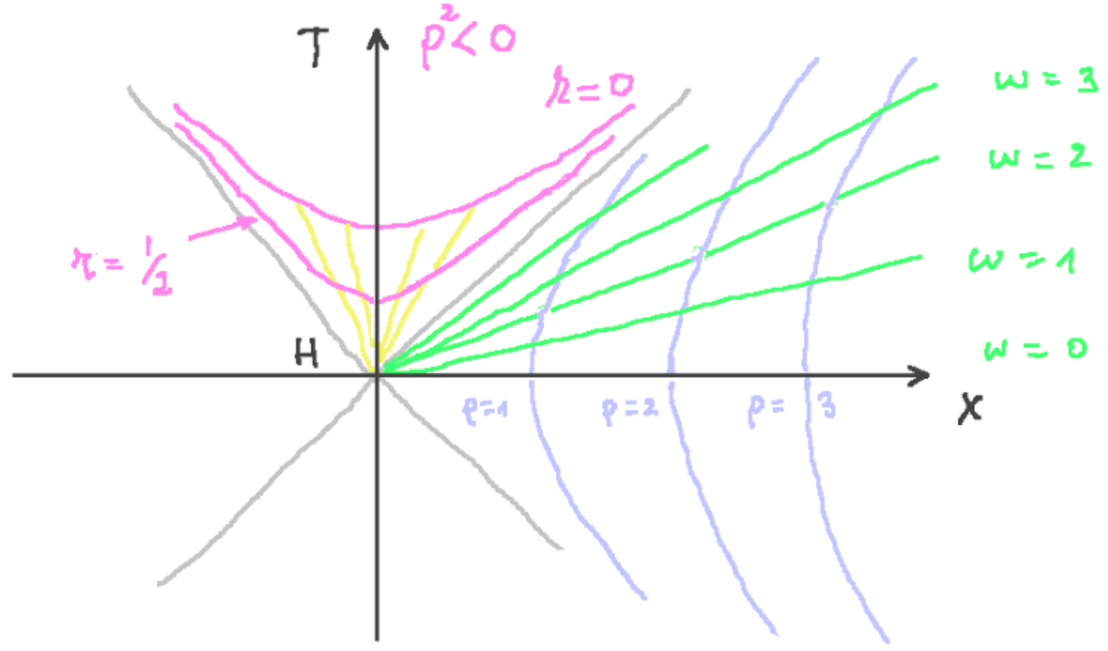
And he proposed the now usual diagram that goes with them, figure 5b.

Before being called a black hole, this kind of object was called a collapsed star or massive collapsed star.

The term black hole was coined by John Wheeler.

John Wheeler was a very nice and very sweet man.

I met him once when he was visiting Hans Bethe.



That caused a stir.

The Physical Review did not want to publish it.

It was before my time as an active scientist actually, or just about, but I knew there had been a problem.

I didn’t know what it was.

I learned it was not just the Physical Review being its usual conservative self.

No, they were being hyper conservative : they thought that the term black hole was obscene.

So they refused to publish the paper at first.

And he fought and fought and fought with them, and eventually won.

Then just to get back at them he entitled his next paper "black holes have no hair".

Incidentally, what does it means for black holes not to have any hair?

It means that if you take a non-rotating black hole, gravity is so strong that it will always pull it together into a sphere.

Even if it starts very asymmetric, like two rocks coming together, after a very small amount of time the horizon will pull itself into a sphere, and become indistinguishable from a perfect sphere.

And John Wheeler called that characteristic of not having any visible structure nor any visible defects on it, the property of not having hair.

So the statement that the black holes have no hair is just the statement that gravity is so strong that it will pull the horizon of any black hole into a perfect sphere.

If it is rotating then the sphere can get deformed.

It can get into an oblate spheroid.

But the nature of the oblate spheroid depends only on the angular momentum.

So black holes have no hair.

Now, let’s talk about things or people falling into a black hole.

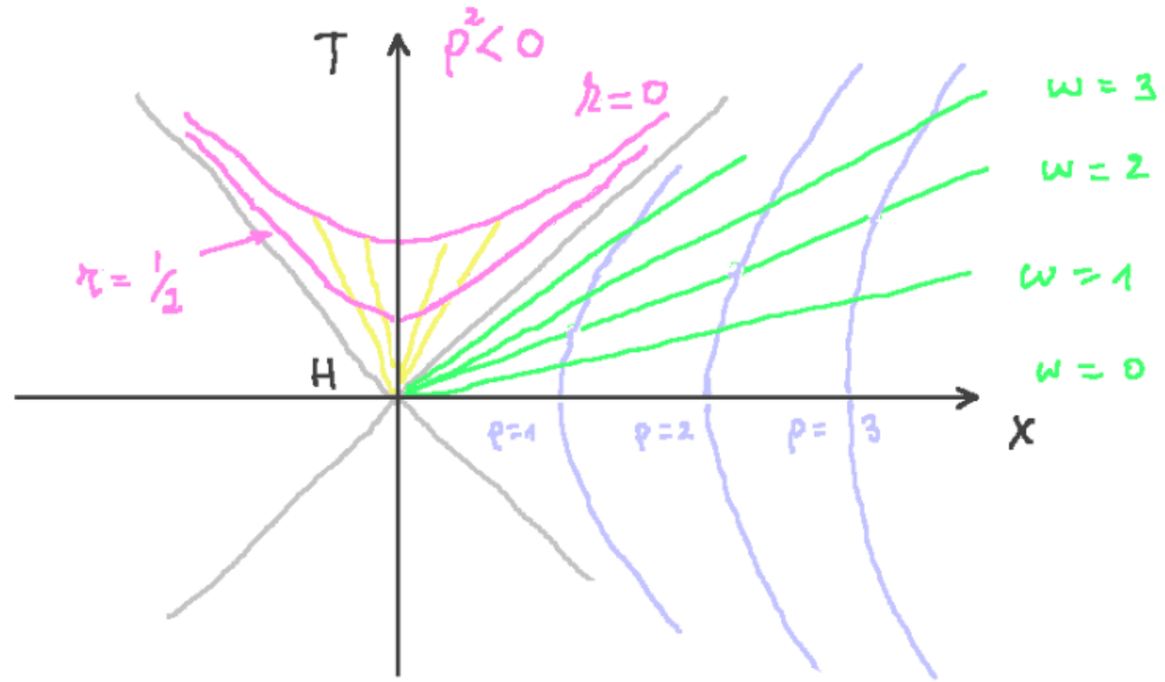
Falling into a black hole

Back at the diagram in figures 5 or 5b, you might ask what do the left and bottom quadrants represent.

We are going to see that this other half of the diagram, the half below the –45° line, has no real significance.

We will come to it.

But for the moment we are mainly interested in the upper right hand part.



The exterior of the black hole is the right quadrant, and the interior of the black hole is the upper quadrant.

Let's redraw the picture, figure 5.

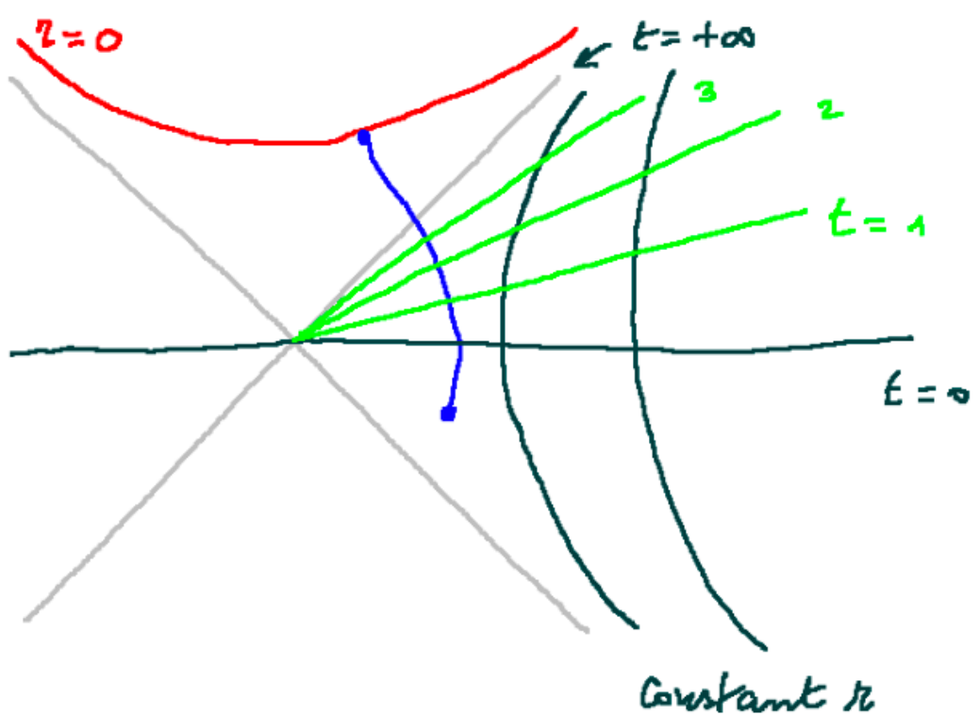


Figure 6 : Alice falling into the black hole

It is Alice's turn to fall in.

For simplicity of notations, we use coordinates (t, r), which are equivalent to (ω, ϱ), the close correspondence being given by equations (14) and (15).

Bob stands somewhere outside the black hole at a fixed position.

Therefore his trajectory is a hyperbola of constant r as his proper time, proportional to t, ticks.

Looking at the diagram, we can see that while Alice is falling in, she doesn't pass the horizon until t equals infinity.

It is in that sense that an in-falling object never passes the horizon.

It is a statement in **Bob's time**.

And indeed Bob cannot see an object nor a person fall through the horizon.

But of course Alice does pass the horizon.

We can see her in the diagram sailing into the singularity.

Nothing crazy happens to her as she passes the horizon.

It is just the strange set of coordinates that says that $t = +\infty$ when Alice crosses the horizon of the black hole.

So you might think: well, it doesn't really mean anything to anybody the fact that Alice never gets in.

But on the other hand Bob is out there, staying at a fixed position, that is, on a hyperbola of constant r.

And Bob is watching Alice.

Let's examine more precisely what he sees of Alice, what does that mean to see Alice.

How does he see Alice ?

To "see Alice" at time t means for Bob to receive a light-ray emitted by Alice sometime earlier , and arriving at Bob at time t.

The lines of simultaneous time for Bob are the straight lines fanning out from the origin of the diagram

In other words, Bob "looks into the past".

For instance, at time t_1, when Bob is at point P_B, he sees Alice at point P_A.

At time t_2, when Bob is at point Q_B, he sees Alice at point Q_A, figure 7.

Remember that light travels at 45° angle.

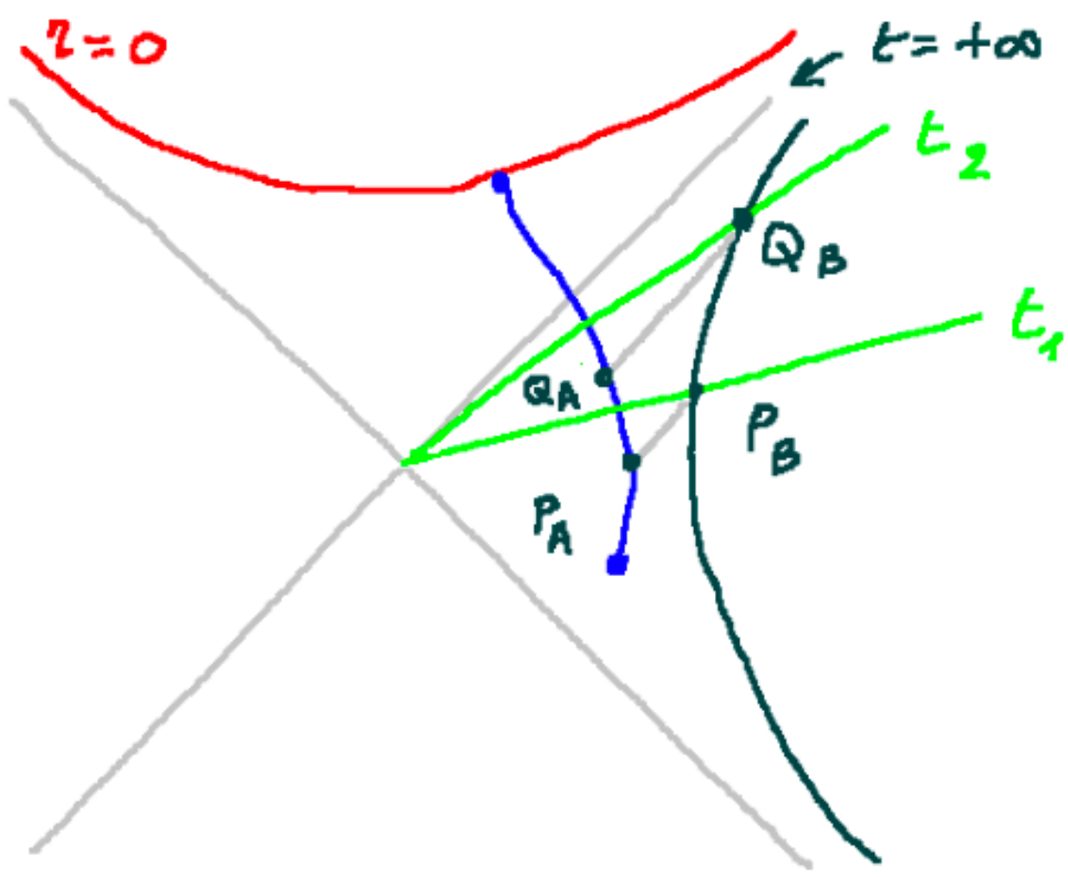


Figure 7 : Bob sees Alice.

Figure 7 makes it clear that Bob will never see Alice cross the horizon, that is, the line $t = +\infty$.

As long as Bob stays outside the black hole and looks back, what he will see is Alice getting closer and closer to the horizon, but never passing it.

So in fact, you may say that it is not just a coordinate artifact that Alice "never crosses the horizon".

It is physically what Bob (doesn’t) see.

As far as Bob is concerned Alice never passes the horizon.

If Alice in her own frame of reference passes the horizon, she cannot send a signal to Bob warning him "Hey Bob, I just passed the horizon".

Why?

Because such a signal would travel on a 45° line in figure 7 and never reach Bob.

Everything that Bob can observe and measure, his entire physical observations, his entire universe, all involve Alice outside the horizon.

As far as he knows, her heart slows down as she approaches the horizon.

It somehow stops beating, in the sense that each beat takes longer and longer.

So for him, she all but dies at the horizon – in an infinite amount of time.

He can have no idea about what happens to her past the horizon, because for him it is beyond the end of times.

And she is indeed doomed then...

One important point ought to be mentioned : we are considering a very big black hole.

Indeed, we assume that the tidal forces at the horizon are negligible.

The apparent squashing of Alice at the horizon, viewed by Bob, has nothing to do with tidal forces.

It is only a variety of Lorentz-FitzGerald contraction already studied in the last lesson.

No tidal force related deformation happens to Alice at the horizon.

For a collapsed star with the mass of the Sun, the Schwarzschild radius is approximately 3 kilometers, and the tidal forces at the horizon would already be huge.

The Sun when it collapses will make a small black hole, a black hole of the nasty type as we saw.

On the other hand at the center of our galaxy, there is a humongous black hole, which is rather mild even close to its horizon – but of course you would be ill-advised to go in and explore.

When Bob watches Alice, he sees her slow down.

Even her heartbeats slow down – rather intrepid for somehow about to fall into a black hole!

Does this mean that Alice sees something special happening to Bob when she watches him ?

Let's study what Alice sees when she looks at Bob.

Again all the answers are given by the diagrams.

See now figure 8.

As Alice sails along her trajectory she sees light-rays emitted by Bob sometimes in her past with no problem.

When she is at R_A, outside the horizon, she sees Bob at R_B.

When she crosses the horizon at S_A, she sees Bob at S_B.

Until she hits the singularity, where she will be annihilated, she continues for a while to see Bob without difficulty.

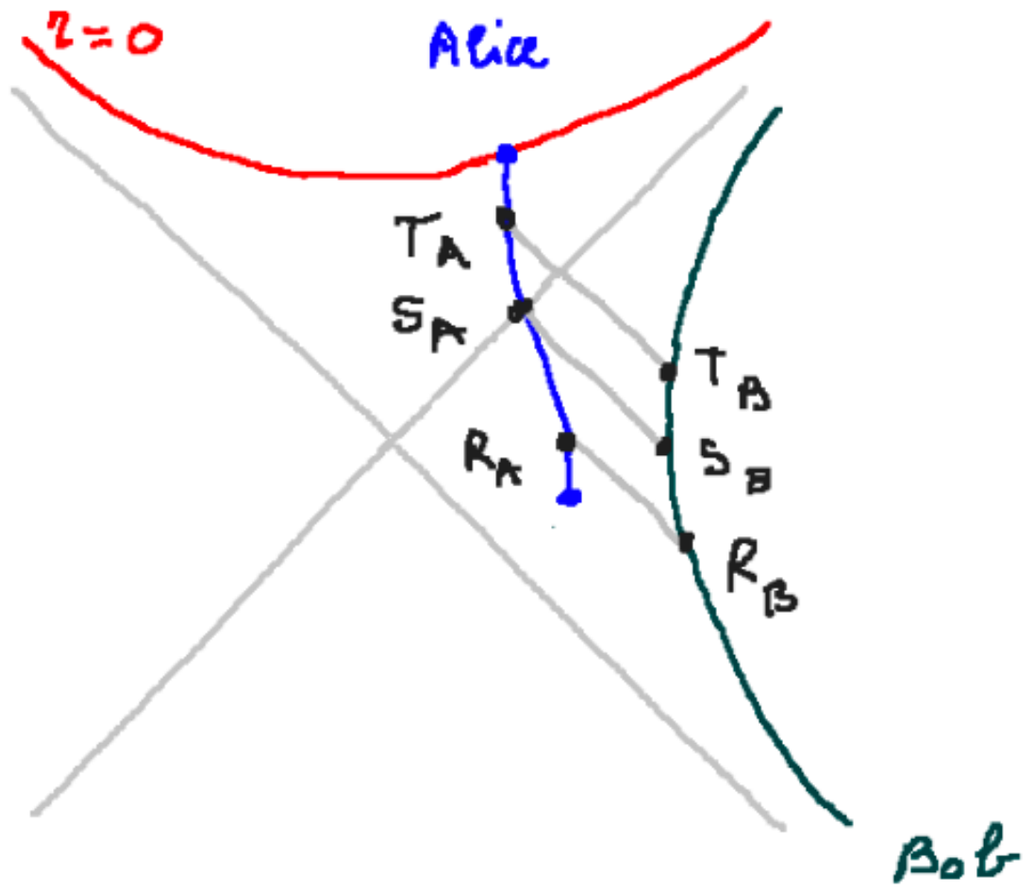


Figure 8 : Alice sees Bob.

Moreover there is nothing indicating that she would see Bob speed up or slow down.

Alice can still see Bob, when Bob can no longer see her.

And Alice doesn't see anything particular happening to Bob.

To her, everything is normal as usual.

Of course Alice cannot see Bob past a point a little after T_{B}.

But that is only because, by then, the tidal forces will have torn her apart and terminated her.

In conclusion, there is a total asymmetry between Bob and Alice.

One can describe what a third person, say Charlie, following Alice sometime later would see of Alice and of Bob at different times.

So long as Charlie or Alice is still outside the black hole,

Bob can receive a message from him or her.

And, for Bob, the "last" messages will be infinitely late in time.

So he does not see either of them ever cross the horizon.

One can also analyze how Alice and Charlie are able communicate with each other.

In the case that Alice and Charlie are both in free fall, their trajectories will be parallel and look almost like straight lines.

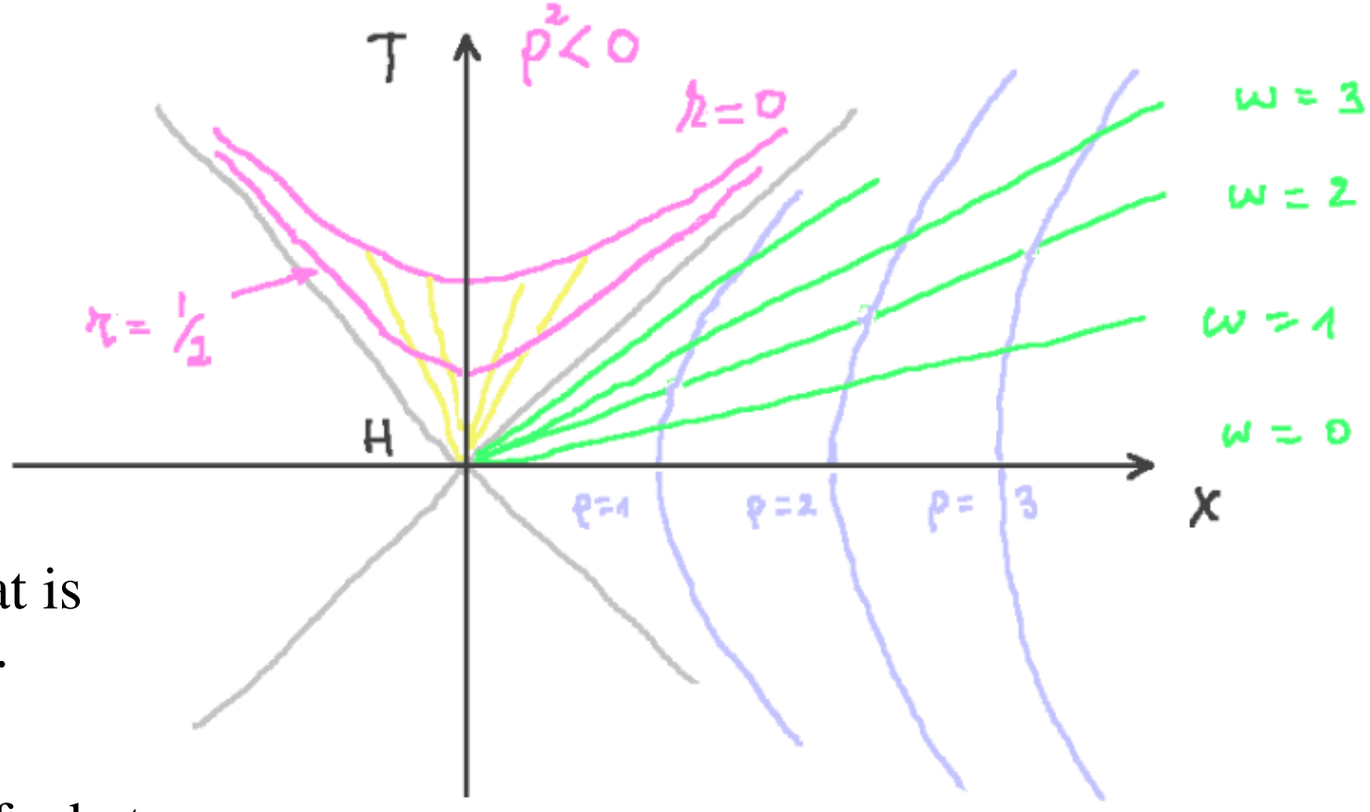
However they fall, you can think of Bob being eventually unable to see them as a consequence of the fact that he has to accelerate to stay outside the black hole.

He has to accelerate toward the exterior viewed in their frame or frames of reference.

Remember that the equivalence principle says that being at rest and feeling gravitation in frame 1 is equivalent, viewed from a frame 2 with no gravitation, that is, in free fall in the uniform gravitational field of frame 1, to being accelerated upward

In all cases, what one of the protagonists can see of the others is simply and strictly related to the pair of points, in their respective trajectories, that 45° light-rays can link.

In short, the diagram in figure 5 says it all.



Miscellaneous Thoughts

There are some movies that purport to show what is happening when someone falls into a black hole.

Andrew Hamilton has made some simulations of what it looks and feels like to fall into a black hole.

We said that nothing special happens at the horizon but of course the light that is coming at you is coming in very peculiar and funny ways.

The movies are not very illuminating.

You won't get much out of them but it might be fun to watch one.

They really are disorienting.

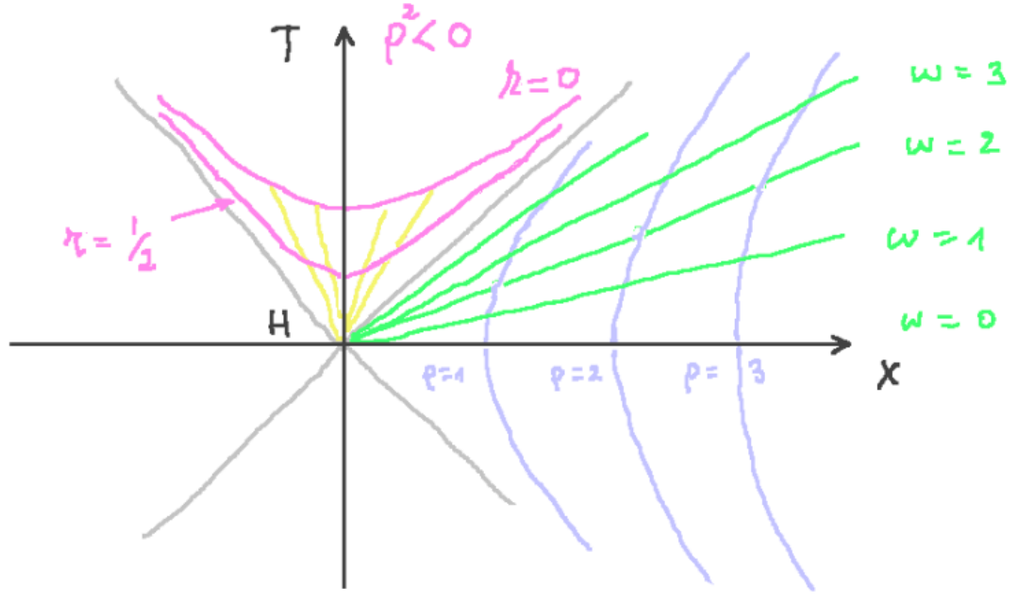
If you are in an auditorium and one of them is projected on a big screen in front of you, they can make you nauseous and seasick.

You might wonder what happens at the origin in the diagram of figure 5, that is at the point H on the horizon.

Answer : nothing much because in fact for a real black hole, that is formed for example by stellar collapse or something like that, the center part of the diagram isn't even on the figure.

Only a portion of the diagram in the upper and right parts is relevant.

Figure 5 is that of an idealized black hole.



For a real black hole which forms by collapse only a portion of the diagram, which does not include H, means anything.

As a black hole gulps a mass coming from the outside it own mass gets larger.

Therefore the Schwarzschild radius gets larger too.

As a black hole gulps a mass coming from the outside it own mass gets larger.

Therefore the Schwarzschild radius gets larger too.

The pictures we drew are appropriate only when the objects or persons falling in are much lighter than the black hole.

So the black hole doesn't react strongly to them.

If a big mass, of a size not negligible compared to that of the black hole, falls into it, when the mass approaches the black hole, the horizon of the black hole will bulge to merge with the coming mass, see figure 9.

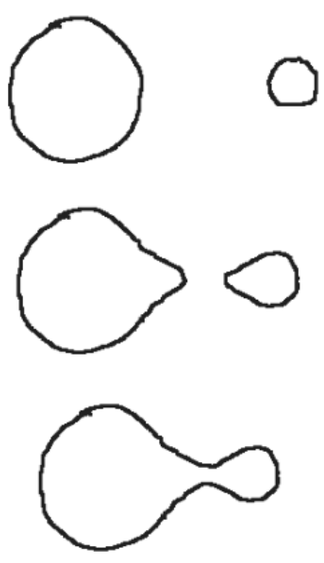


Figure 9 : Black hole (left) gulping another sizable mass (right).

Then because black holes have no hair, it will quickly pull itself together back into a sphere.

But the sphere will be slightly larger.

The process in figure 9 can be calculated with Einstein field equations, which we will study in lesson 9.

How can we see the black hole at the center of the Milky Way ?

Answer : we don’t "see" the black hole itself.

We identify phenomena that are due to the presence of the black hole.

The light that we see from a black hole is not coming from the horizon.

It is coming from all sorts of stuff, hot stuff, that is circulating around the horizon and that is heated up by energy and collisions.

Remember for example that at a certain distance around the black hole we can have photons of light moving around in circles.

All sorts of very complicated collisions can take place around the black hole, which sends out particles and light some of which ends up in our eyes or measuring apparatus.

In figure 5, we were talking about the idealized situation of a black hole in empty space.

But in reality there are all kinds of stuff around it.

There is an accretion disk.

There are plenty of things falling in.

And what you see is not coming precisely from the horizon.

It is coming from all this activity at some distance from the horizon.

The biggest black hole that is known has a mass something like 10^9 solar masses.

We haven't talked about what happens inside the black hole when a mass is absorbed.

Next lesson, we are going to talk exactly about how a black hole is formed, how the horizon forms, and if you throw in more material how the horizon responds.

We will examine a simple example, and see that things are different then expected.

They are surprising and yet logical.

Does anything change if Bob is orbiting the black hole?

No, not much.

It becomes a more complicated problem to be quantitative about exactly what he sees.

But there is no fundamental change.

A black hole is the analog in relativity of a point mass in Newtonian physics.

All the mass is at the center.

There is no such thing as getting deeper and deeper into the mass of the black hole after we pass the horizon, like what happens when we enter below the surface of the Earth and the gravity caused by the outer layers for us gets out of the equations.

The vast majority of things or collections of things in the universe have some angular momentum.

As a consequence the vast majority of black holes rotate fast.

Why?

Because in the process of collapsing and forming a black hole, even if at first the material does not show much rotation, it has some angular momentum.

Then, as its dimension shrinks, like an ice skater pulling his or her arms along the body, it will spin faster and faster.

Stars form when materials agglomerate and start to radiate.

The interior of the star burns hydrogen or helium or whatever.

When the star eventually runs out of fuel, the forces, that is, the radiation pressure that is preventing it from collapsing disappears and the star begins to contract.

Depending on how heavy it is, it might contract into a white dwarf, that is a more less ordinary thing, pretty dense but still made out of nuclei of atoms and so forth.

If it is heavier it might collapse into a neutron star.

That is a very compact object, but it can still support itself because the material is strong enough to support itself against gravity.

If it is still heavier then it may contract past its own Schwarzschild radius.

Once it falls past its own Schwarzschild radius, it is doomed : it becomes a black hole.

The Sun, when it eventually "turns off", will not form a black hole.

If nothing else happens to it, like merging with something else, it will form a white dwarf.

In theory black holes can "evaporate" and disappear. But the process is very long.

For instance a black hole with the mass of Mount Everest would take longer than the age of the universe to evaporate.

Aside from gravity collapse that we just described, another process of formation of black hole is via violent collisions.

Velocity can replace gravity and slam stuff together.

If there is enough violence a black hole can form – even though there was not enough material for gravity collapse.

The smallest imaginable black hole is a black hole of Planck mass.

$$m_P = \sqrt{\frac{\hbar c}{G}} \approx 1.22 \text{ x } 10^{19} \text{ GeV/c}^2 \approx 2.18 \text{ x } 10^{-8} \text{ kg}$$

By slamming particles hard enough you could theoretically make small black holes.

Similar collisions happen naturally, with higher energies than in man-made accelerators, when cosmic rays hit the Earth upper atmosphere.

Therefore possible artificial black holes, created for instance at the LHC in Geneva, would be no more dangerous.

The study in more detail of the formation of black holes is the subject of next lesson.

Lesson 8 : Formation of a black hole

In the last lesson, we explained the Kruskal coordinates – also called Kruskal-Szekeres coordinates – for a black hole.

We shall go over them again, for practice.

Then we shall use a different type of diagram to represent the geometry of space-time near a black hole, and see what other phenomena it reveals.

Finally we shall examine the formation of real black holes.

Kruskal–Szekeres coordinates

Remember the fundamental diagram :

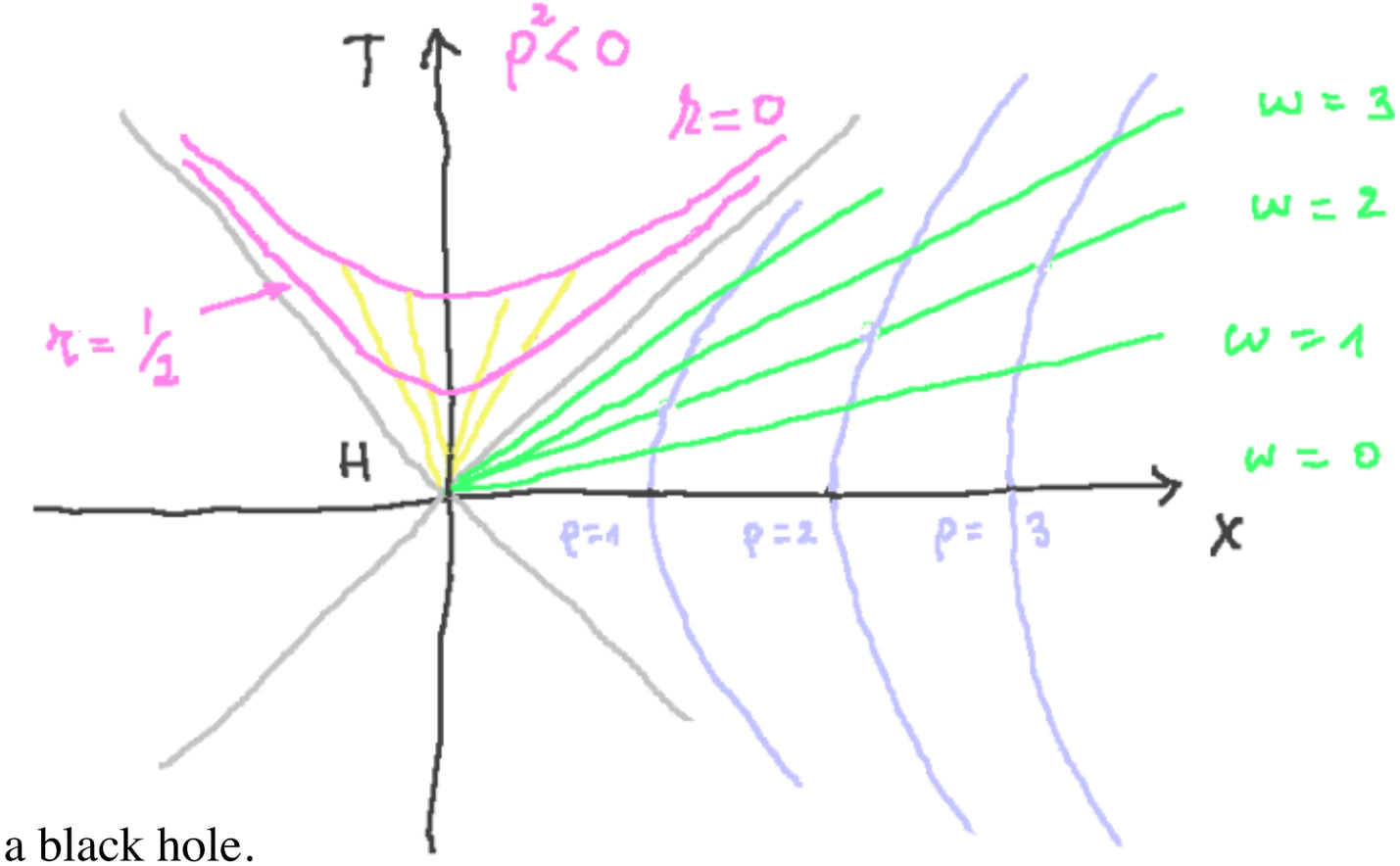


Figure 1 : Kruskal diagram near the horizon of a black hole.

In the last lesson we saw that the geometry of a Schwarzschild black hole near the horizon looks like flat space .

In fact, it looks locally like flat space anywhere in space-time except at the singularity, just like any smooth surface in Euclidean space looks locally like a plane.

We built near the horizon hyperbolic polar coordinates which illustrated that points at fixed distances outside the horizon behave as if they were fixed in a uniformly accelerated coordinate frame.

We remember that somebody standing in the right quadrant is outside the horizon of the black hole.

Somebody in the upper quadrant is inside the horizon of the black hole.

The 45° lines represent the motion of light-rays either going to the right or going to the left.

We saw that not only point H but the entire lines at 45 ° form the horizon.

We are going to see that the left quadrant and the bottom quadrant don't have any physical significance for a real black hole.

It is only the right quadrant and the upper quadrant which have real significance.

Outside the black hole, somebody at rest, that is, at a fixed distance r, or equivalently at a fixed distance ϱ from the horizon – with a trajectory on a hyperbola –, is of course experiencing acceleration.

ϱ is the proper distance.

When $r > 1$, ϱ is related to r by a simple increasing function which we studied in lesson 7.

This is the same kind of acceleration which you and me experience even when we are standing still.

Our weight is due the gravitational field of the Earth, and is equivalent to being constantly accelerated.

On the surface of the Earth, we represent the geometry of space and time with a Euclidean space plus a universal time coordinate.

We reason on only one dimensional spatial axis, which is radial from the center of the Earth.

A uniformly accelerated frame is simply that of an observer inside an accelerated elevator.

Various points fixed above each other in the elevator, viewed in the reference frame of the stationary observer resting on Earth, have trajectories which are parallel lines when time varies.

Lines of fixed time don't converge to a center point but are parallel too, orthogonal to the trajectories.

And the time coordinate is the same in any reference frame.

In relativity the geometry of space-time is different.

Space and time are intimately related. Simultaneity, for instance, is frame dependent.

However, a uniformly accelerated reference frame, far away from the origin of the Minkowski diagram, is very much like an ordinary accelerated reference frame near the surface of the Earth : trajectories are almost parallel vertical lines (if the time axis is vertical), and time and proper time are almost the same.

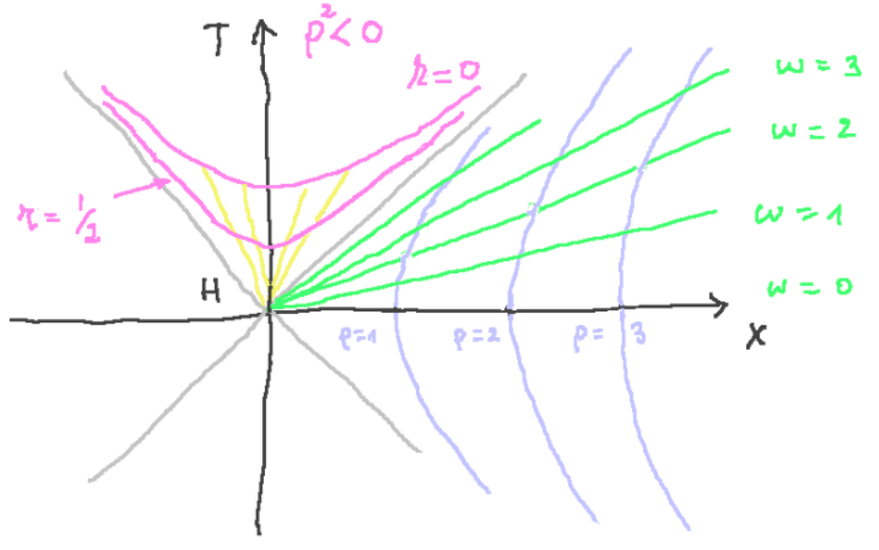
We are so accustomed to the acceleration which we permanently feel at the surface of the Earth that we tend to forget it.

But somebody who has been in free fall in outer space for too long, lost the sense of what a gravitational field is like.

If you place him or her on the surface of the Earth that person will be very much aware of a new feeling.

The person will be experiencing a sense of upward acceleration.

Spread on a radial axis jutting out of the horizon of the black hole (angles θ and ϕ play no role, they are constant) we then have a collection of people, at proper distances respectively $\varrho = 1$, $\varrho = 2$, $\varrho = 3$, etc.



The closer to the horizon a person is, the stronger the acceleration he or she feels.

In figure 1, the coordinates T and X are those of a free falling observer, who is anywhere in space-time except at the singularity $r = 0$, and happens to look, in his or her coordinate frame, what is going on near the horizon of the black hole.

The observer sees the collection of persons at the various ϱ's as being at fixed positions in a uniformly accelerated reference frame.

In lesson 4, we defined what is a uniformly accelerated reference frame in relativity.

Pay attention to the fact that in such a uniformly accelerated reference frame, the actual acceleration – the physical push felt by each person – is not the same at different ϱ's.

The smaller ϱ is, the stronger is the acceleration.

This is shown by the growing acuteness of the hyperbolas near H.

On the other hand, the acceleration stays the same along a hyperbola of constant ϱ.

In the right quadrant, the straight lines fanning out of H are the lines of constant proper time t or ω for the people in the accelerated frame.

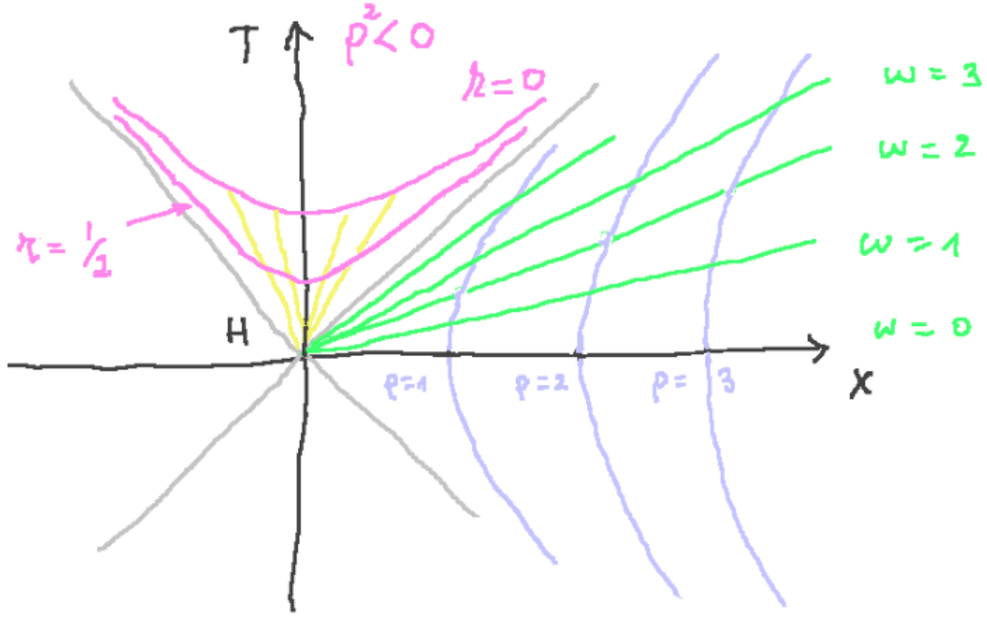
Remember that ω is simply $t/2$.

As ω tends to $+\infty$, the lines become closer and closer to the 45° line of a light-ray emitted from H.

In the discussion so far we have described the outside of the black hole.

The inside of the black hole is represented in the upper quadrant.

In this upper quadrant, there is a point where the geometry of space-time is no longer equivalent to locally flat.



There is something nasty happening.

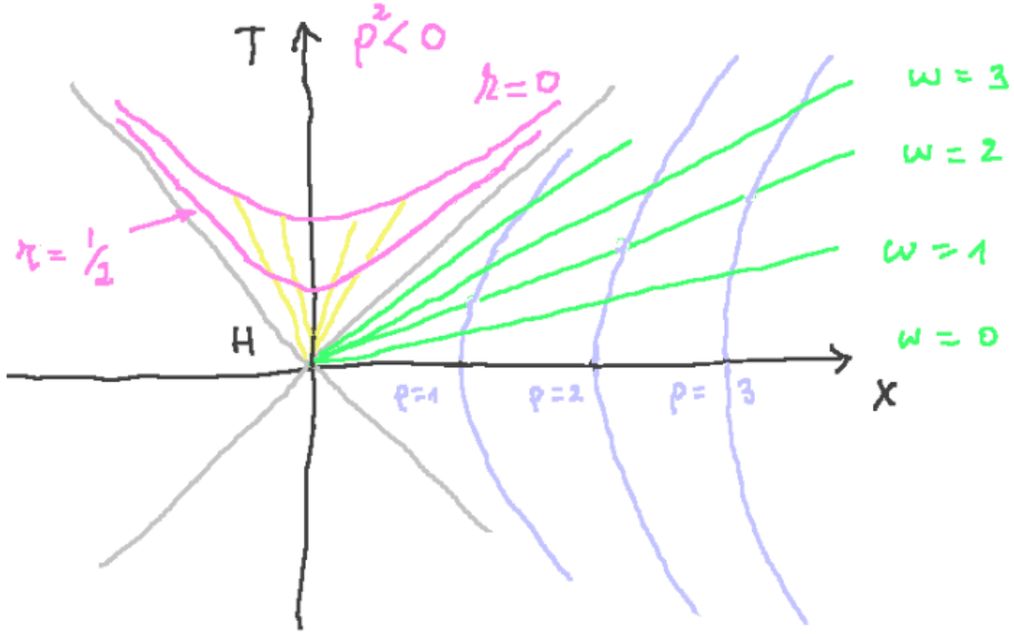
It is at the singularity.

Notice that we would have the same kind of nasty singularity for the Earth, even in Newtonian physics, if we considered that all its mass was concentrated at its center.

Strangely enough, the singularity is represented by a whole curve in figure 1.

It is the hyperbola marked with $r = 0$.

The tidal forces are very strong near the singularity.



There is no longer any analog in Newtonian physics.

This is due to the metric of space-time having three positive signs and one negative sign.

Furthermore the Schwarzschild metric is such that the radial space and the time coefficients exchange signs at the horizon.

Therefore the trajectories for $r < 1$ are space-like and no longer time-like as usual.

The problem for an adventurous person exploring space-time is that once you find yourself inside the horizon of a black hole, that is, in the upper quadrant of figure 1, there is no way out – unless you could exceed the speed of light.

And you will inescapably hit the singularity in a finite amount of your own time.

So you are doomed.

In summary, outside black holes you or objects or light-rays can get into the black hole, or they can stay out.

But inside the black hole nothing can get out, and everything ends up at the singularity in a finite amount of time.

That is all there is to know about the geometry of black holes.

It is just not equivalent to plain flat space.

The geometry is warped because of the singularity.

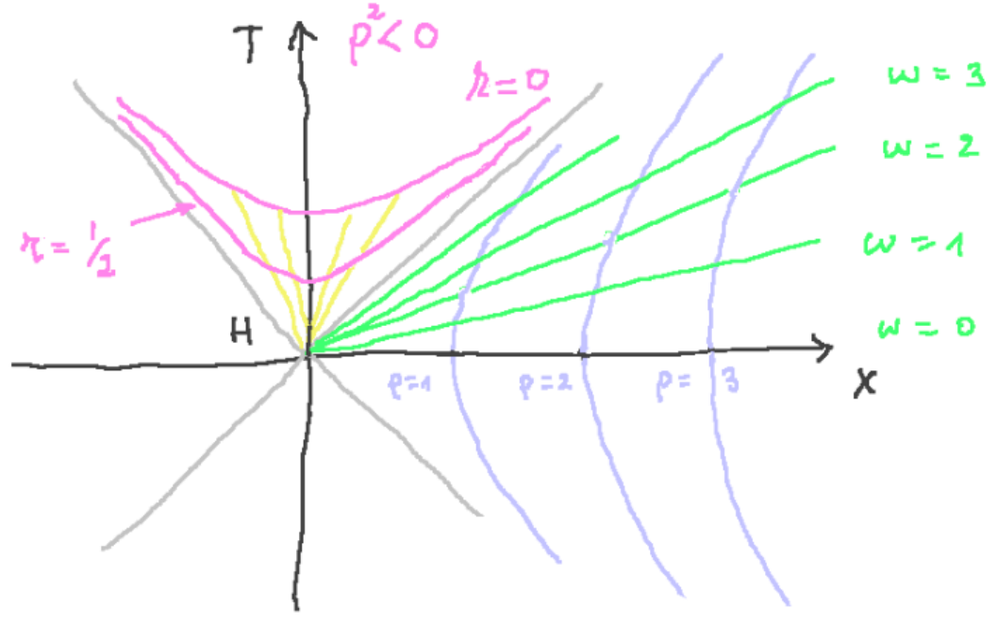
Penrose diagrams

It is very convenient to redraw the diagram in figure 1 differently.

We are going to end up with Penrose coordinates and a Penrose diagram.

But these are built in several steps.

So let's take the steps one by one.



In figure 1, space is infinite.

Time may come to an end at the singularity, but the light-like directions off at 45° are also infinite.

Therefore we cannot draw the entire space-time in the limited dimensions of the page.

For many purposes it is useful, however, to be able to draw the entire space-time on a finite sketch.

It provides some visual tools and helps intuition.

Let's try doing that but let's start with good old flat space-time.

We will do a coordinate transformation which pulls the whole thing into some finite region of the page.

This is useful, incidentally, for geometries in space-time which display rotational symmetry.

Rotational symmetry, or invariance, means that what happens at any event in space-time doesn't depend on the direction of the event from some center.

Let's focus for a moment on ordinary 3D Euclidean space.

To describe a system which displays rotational symmetry, it is usually useful to use ordinary polar coordinates.

And when we have rotational symmetry, we often don't need to worry too much about the angular direction.

Aside from time, the other coordinate that matters is the radial distance.

Therefore, to start with – before shrinking the whole space-time on the page –, we think of ordinary flat space-time as having a time axis T and a space axis R, which is a radial direction, see figure 2.

The time goes from $-\infty$ to $+\infty$.

And the spatial coordinate R goes from 0 to $+\infty$.

We put some markers: $T = 0$, $T = 1$, $T = 2$, $T = -1$, $T = -2$, etc, and $R = 0$, $R = 1$, $R = 2$, $R = 3$, etc.

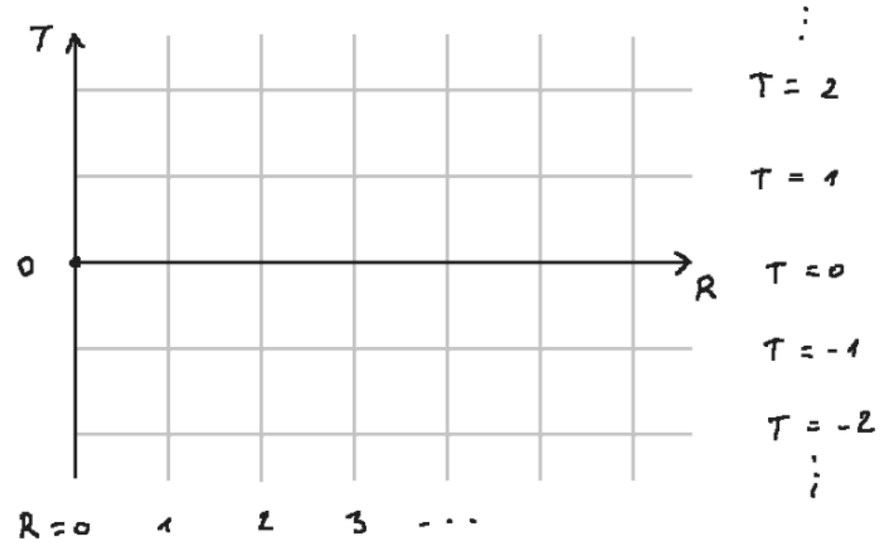


Figure 2 : Flat space-time.

The entire space-time is not yet shrunk on the page.

So far T can go up to heaven, and down to hell.

Similarly, the radial axis R doesn't have any limit on the right.

In all cases, we will imagine using units so that the speed of light is $c = 1$.

For instance, if we use seconds for time, then the space units will be light-seconds.

Then some light-rays follow 45° lines.

But not all of them.

Light-rays coming from the past aimed at the spatial origin and then going on into the future, form a pair 45° half lines as shown in figure 3.

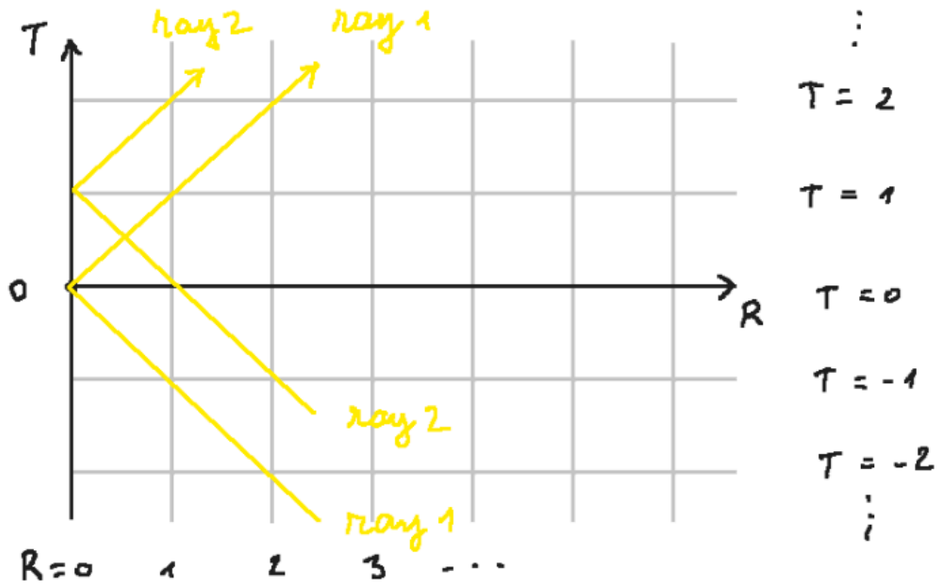


Figure 3 : Two light-rays aimed at the spatial origin.

Another picture will help us understand why some light-rays follow paths like in figure 3, and some don't.

Figure 4 shows space in polar coordinates centered on the same point as above.

The time axis is not represented.

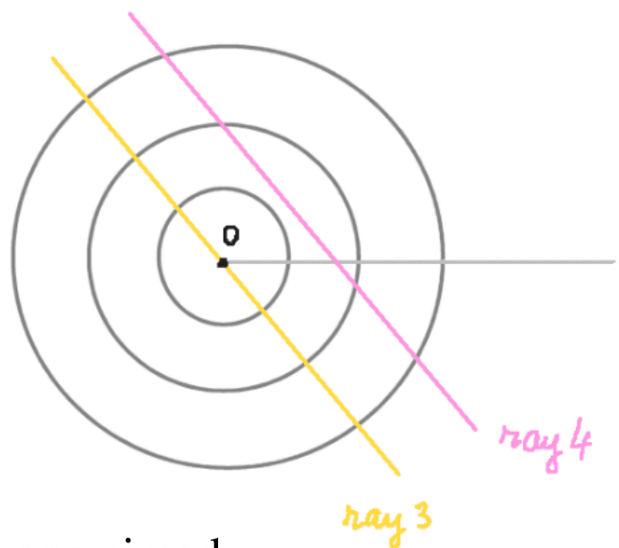
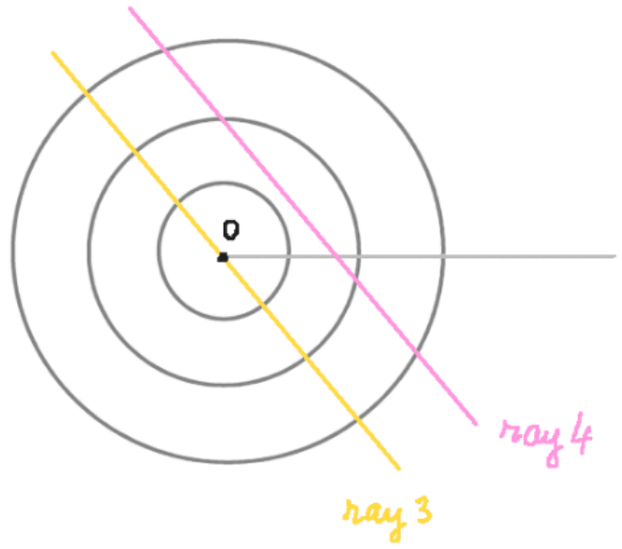


Figure 4 : Two light-rays, one aimed at the center, and not the other.

In figure 4, ray 3 is aimed at the center.



That ray, represented in figure 3, would appear to bounce against the vertical line $R = 0$, like ray 1 and ray 2.

But ray 4, if we represented it in figure 3, would not bounce against $R = 0$.

It would be coming from far away in the past, from what we call light-like infinity.

As long as it is far away, unless we have very good instruments, we could not say whether it would hit the center.

But near the vertical line $R = 0$ it would swerve before hitting it, see figure 5.

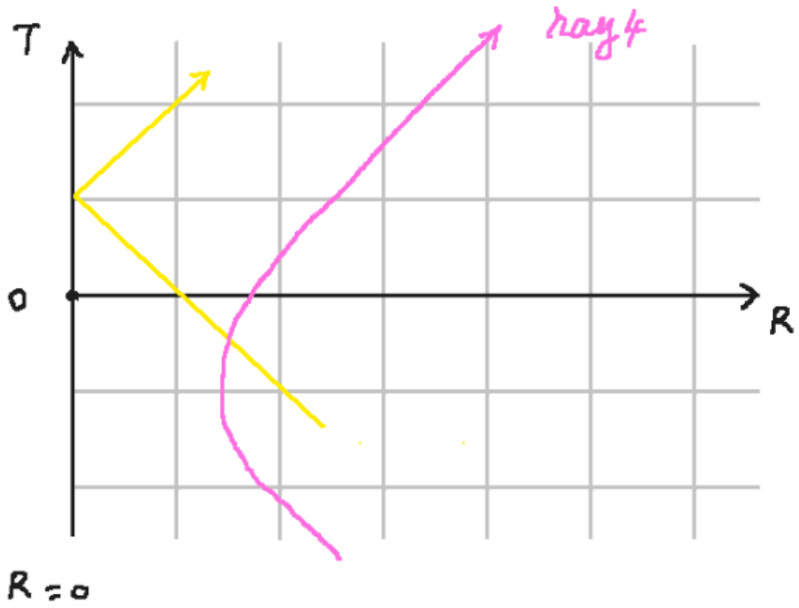
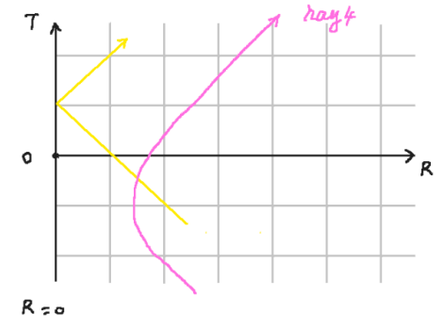


Figure 5 : Light-ray 4 passing by the origin, but not hitting it.

Light-ray 4 doesn't hit the origin because it was not aimed at the origin.



Far away from the small radiuses it follows almost 45° lines in figure 5.

In fact the trajectory of ray 4 is a hyperbola (but that has nothing to do with the hyperbolas of the Kruskal diagram).

In figure 5, light-ray 4 looks like it is repelled from the origin.

Of course it is not repelled from anything, it just passes by near the origin before flying off again, all in a straight line as shown in figure 4.

But the phenomenon appearing in figure 5 really is what we call *centrifugal* force.

It is the centrifugal force that keeps the light-ray from hitting the origin

We went over these things in figures 2 to 5 to get some landmarks and prepare for the next step.

The next step is to squeeze the diagram in figures 2, 3 and 5, to fit the entire space-time on the page.

In particular what we dubbed the light-like infinity will now appear as a point somewhere on the graph.

Of course the diagram will be deformed.

It is not going to look the same.

But we are going to keep one feature fixed : light-rays will keep following 45° lines or asymptotes.

That is a useful thing to do because then we can see how light rays move, and we can see what is going slower or faster than the speed of light.

The shrinking of the whole space-time into a limited diagram is done mathematically in two steps as follows.

We introduce a first set of new coordinates.

$$T^+ = T + R$$
$$T^- = T - R \qquad (1)$$

In this set of coordinates, lines of constant T^+, for instance, are diagonal lines at –45° angle, see figure 6.

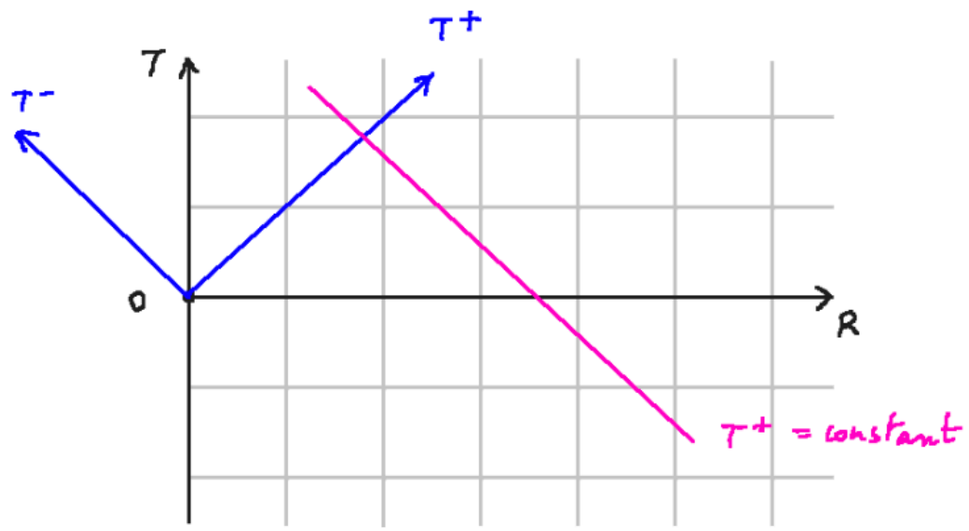


Figure 6 : Coordinates $T^+ = T + R$ and $T^- = T - R$.

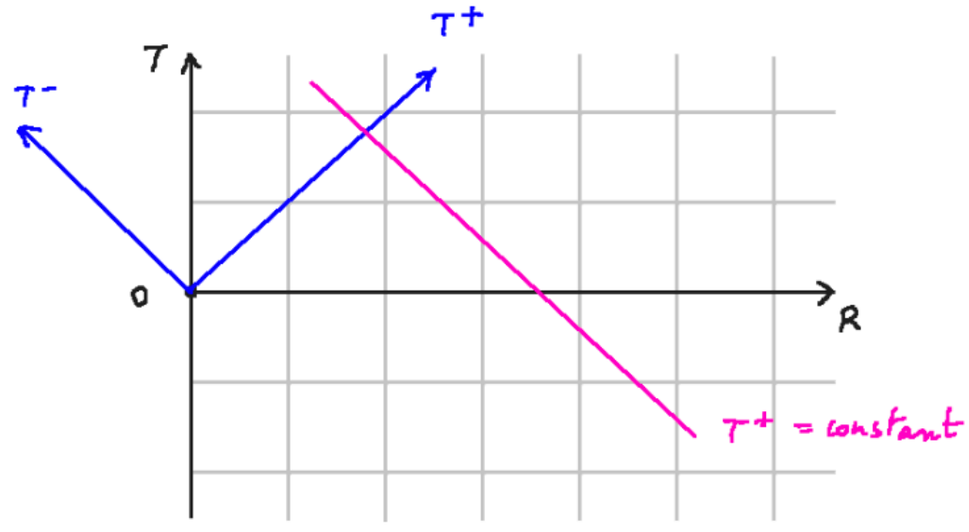
Similarly, lines of constant T^- are diagonal lines at +45°.

So now we have two sets of coordinates to describe our flat space-time.

We have (T, R) and we have (T^+, T^-).

Let's look at the vertical line going through O.

It is $R = 0$ in the old coordinates (T, R).



And it is easy to see that in the new coordinates (T^+, T^-), it has the equation

$$T^+ = T^- \tag{2}$$

One way to see it is to notice that it is the usual equation for the line splitting the angle formed by the two axes T^+ and T^- in half.

Or we can write $T^+ = T^-$ as $T + R = T - R$, which becomes $2R = 0$, or equivalently $R = 0$.

For obvious reasons, the coordinates (T^+, T^-) are called light-like coordinates.

Now we introduce a second set of new coordinates, so as to shrink the whole plane (T^+, T^-) into a bounded area of the page.

The second set is U^+ and U^-.

U^+ will be an increasing function of T^+ such that when T^+ goes from $-\infty$ to $+\infty$, the coordinate U^+ goes from -1 to $+1$, see figure 7.

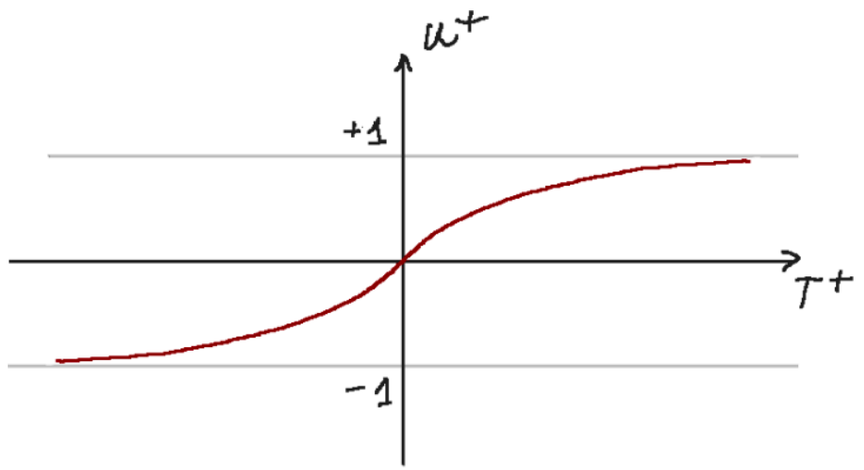


Figure 7 : U^+ as a function of T^+.

There are plenty of such functions.

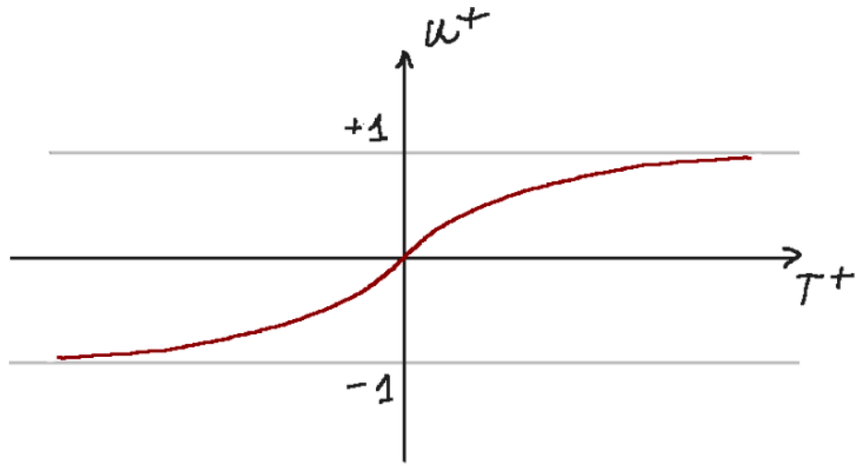
The one that is usually used is hyperbolic tangent.

And we will apply the same transformation to T^- as well.

So we define

$$U^+ = \tanh T^+$$
$$U^- = \tanh T^- \qquad (3)$$

You don't need to know much about hyperbolic tangent, except that is has a graph as in figure 7.

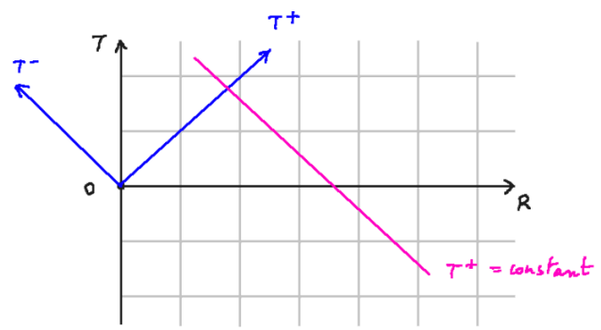


Now we represent the flat space-time in the coordinates (U^+, U^-).

When T^+ goes to infinity, U^+ never gets bigger than one.

Same for T^- and U^-.

Figure 6 becomes the following figure 8.



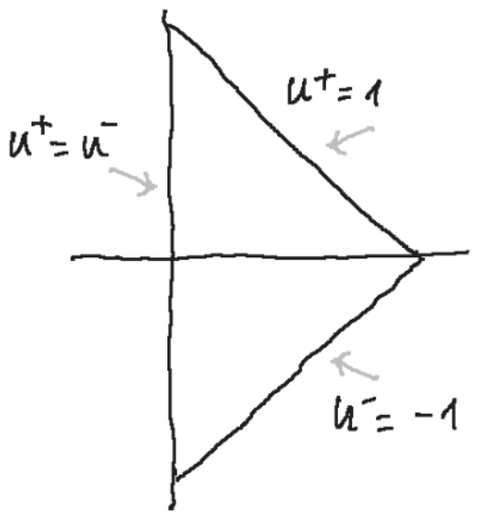
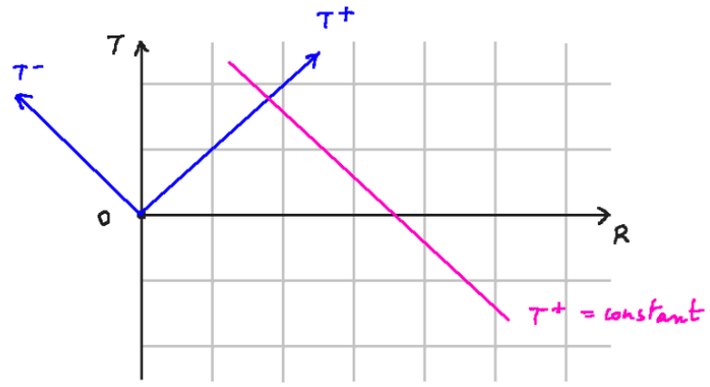


Figure 8 : Space-time in coordinates (U^+, U^-).

All we've done is squish the geometry of figure 6, vertically and horizontally, onto a finite triangle.

This will become clearer when we look at light-rays.

Let's look at an in-going light-ray aimed at the origin.



It corresponds to

$$T^+ = constant$$

After hitting the origin, it becomes an outgoing light-ray from the origin.

It then corresponds to

$$T^- = constant$$

The coordinate transformation has the property that such light-rays are still straight lines as shown in figure 9.

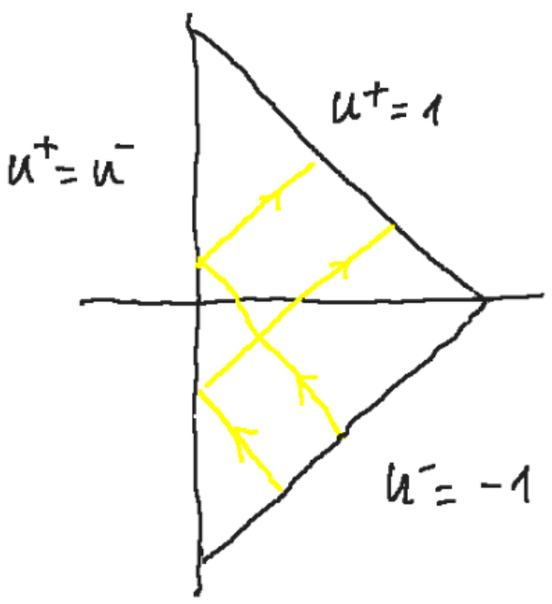


Figure 9 : Light-rays going through the origin, in (U^+, U^-) coordinates.

Another enlightening graph is that of constant times, see figure 10.

We have squished "spatial infinity" into the point $R+\infty$.

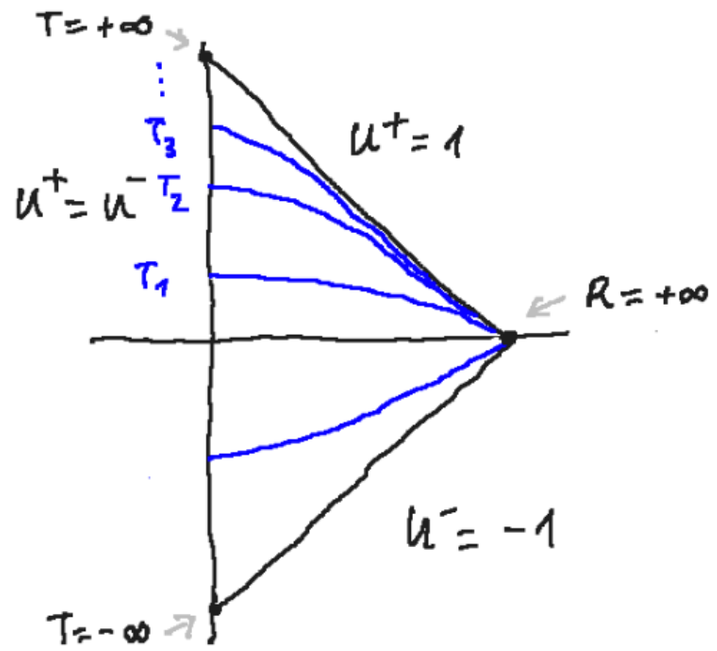


Figure 10 : Lines of constant times.

It is called space-like infinity.

And as we look at bigger and bigger fixed times, the corresponding curves will be closer and closer to the line $U^+ = 1$.

Now let's look at fixed spatial positions.

They will all go to $T+\infty$, coming from $T-\infty$.

They are shown in figure 11.

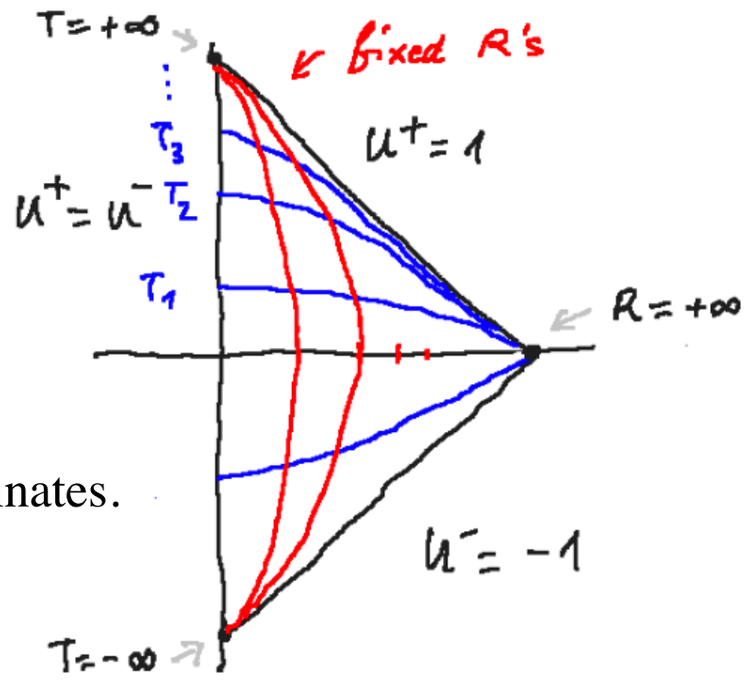
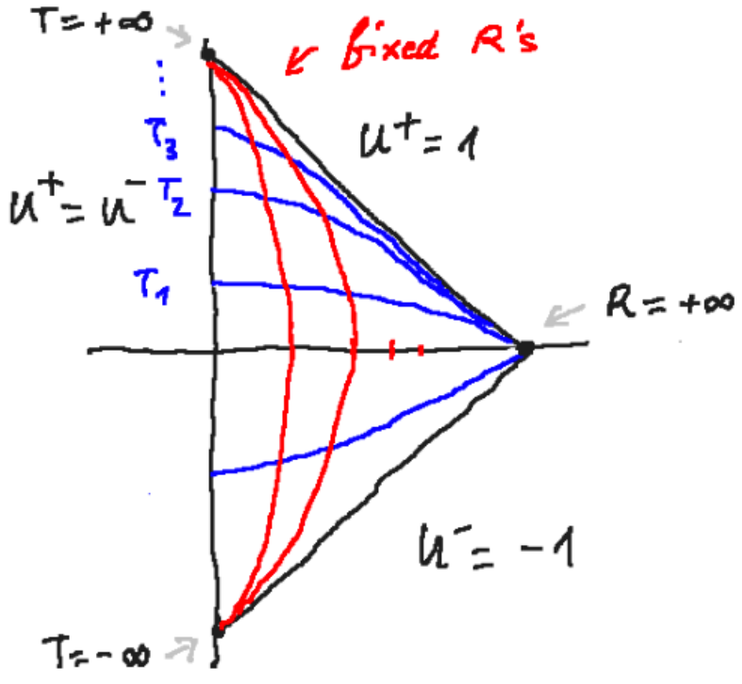
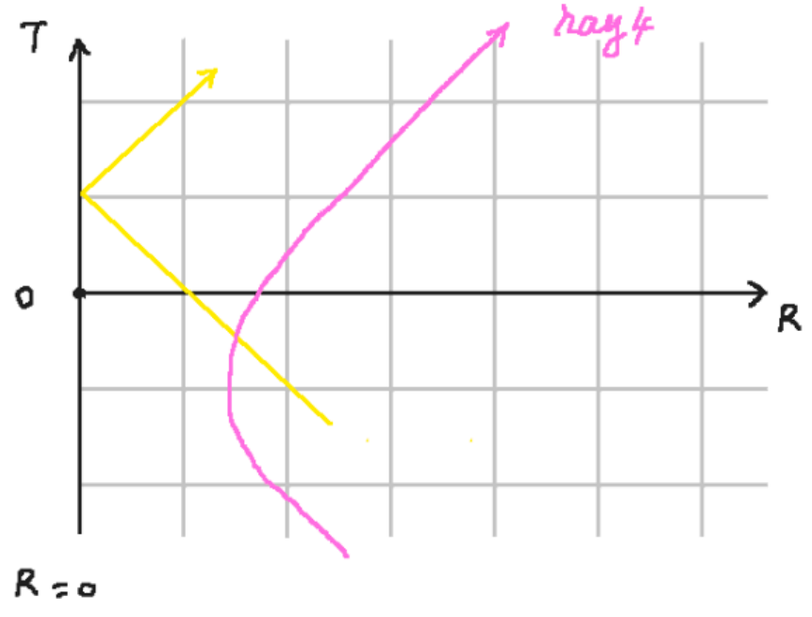


Figure 11 : Fixed position R, in (U^+, U^-) coordinates.

The diagram of figure 11 is "the same" as that of figure 5.





But we have operated a change of coordinates so that every point which had some coordinates (T, R) gets mapped somewhere onto a finite triangle.

The process of bringing everything from figure 5 into a finite diagram (fig 11), without changing the way radial light-rays trajectories appear (fig 9), is called *compactification*.

And figure 11 is called a *Penrose diagram* or a *Carter-Penrose diagram* because actually the Australian physicist Brandon Carter invented it first.

That closes our section on the Penrose diagram of a flat space-time.

Now we turn to black hole geometry.

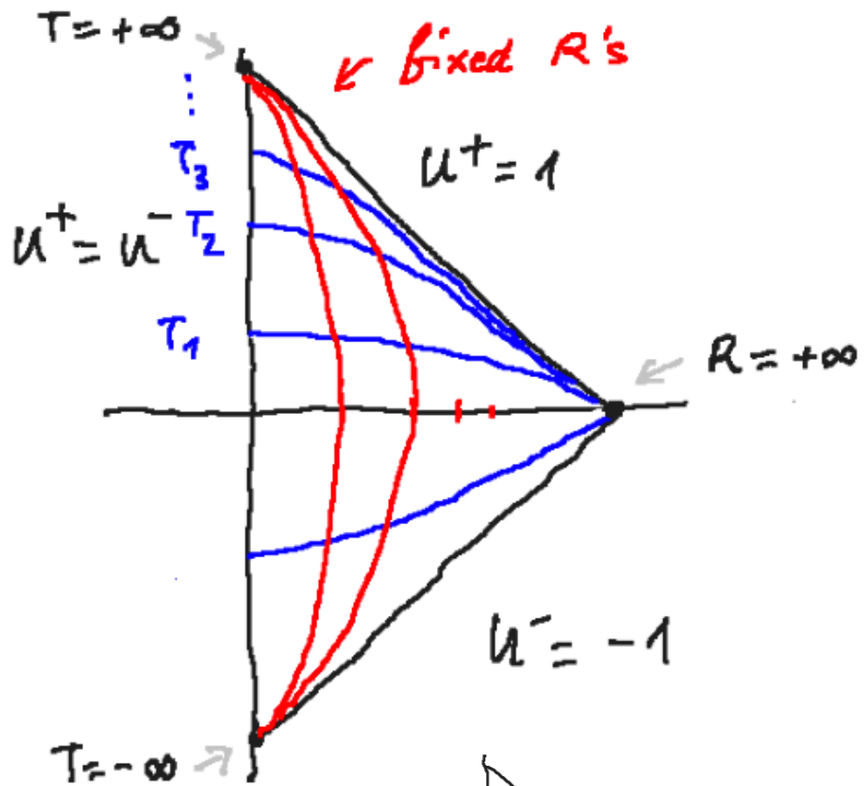
And this will lead us to a discussion of wormholes.

Wormholes

We succeeded in representing the entire flat space-time on the page.

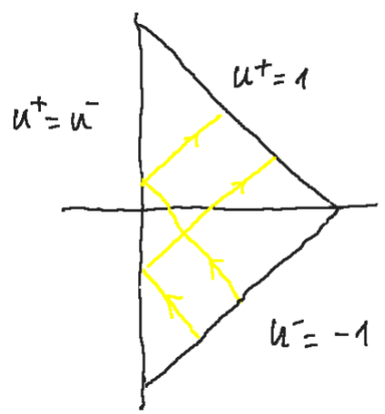
Now the question is : can we apply the technique to black hole geometry ?

Some terminology to start with : the right tip of the triangle in figure 11, where $R = +\infty$, is called spatial infinity, or space-like infinity.



The tip of the triangle on top, along the right side, where $T = +\infty$, is called future time-like infinity.

And the tip at the bottom is called past time-like infinity.



The 45° line between $T_{-\infty}$ and $R_{+\infty}$ is where all light-rays come from (fig 9).

And the line between $T_{+\infty}$ and $R_{+\infty}$ is where they all go, after hitting the origin or swerving past the origin.

The standard notation for the segment $T_{-\infty}$ to $R_{+\infty}$ is I^{-}, read "script i minus", or sometimes "scry minus".

And the segment $T_{+\infty}$ to $R_{+\infty}$ is I^{+}, read "script i plus", or "scry plus".

They are called respectively past light-like infinity and future light-like infinity.

They are the places where light-rays begin and where light-rays end.

This is all summarized in figure 12.

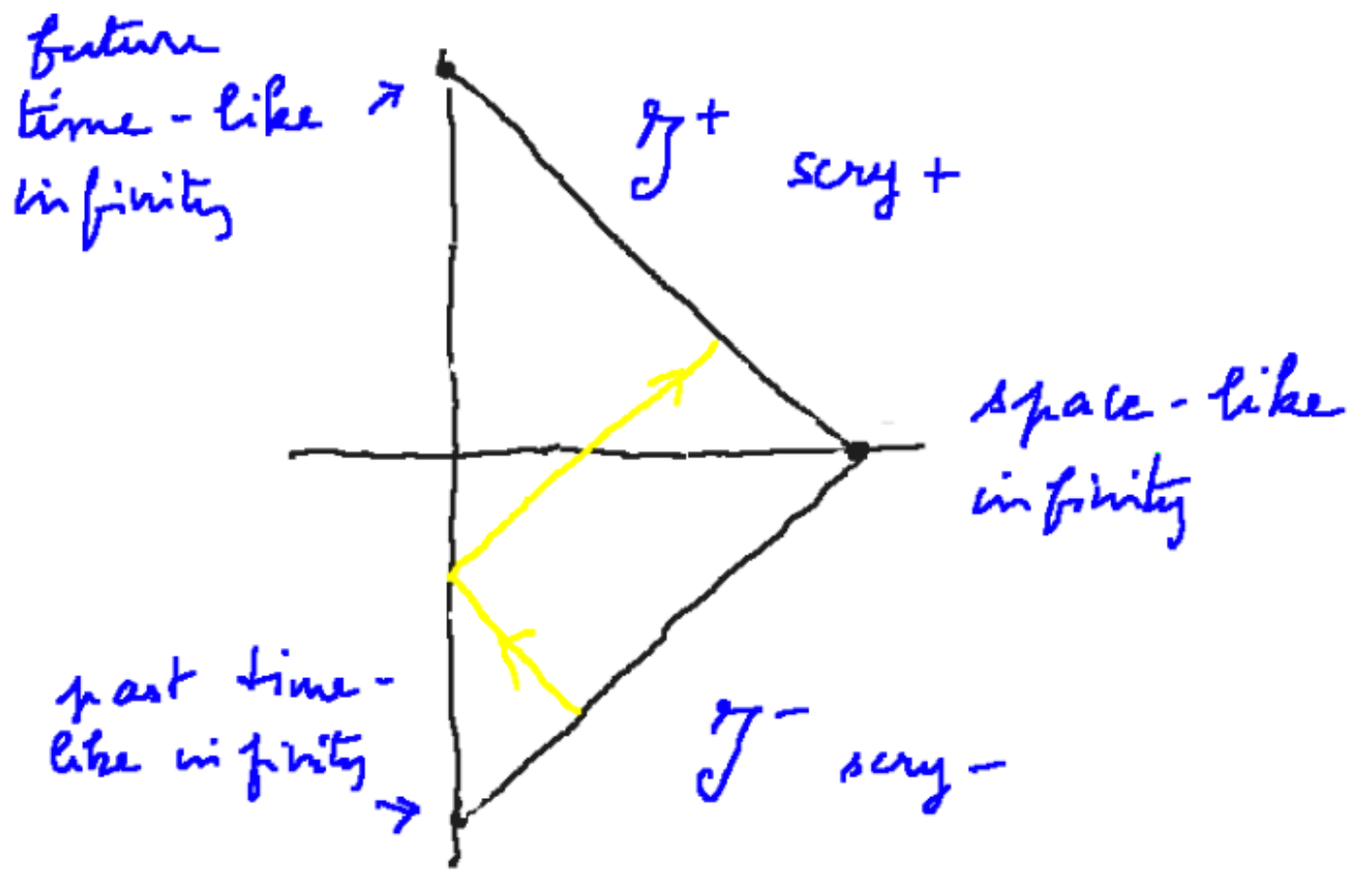


Figure 12 : Flat space-time compactified.

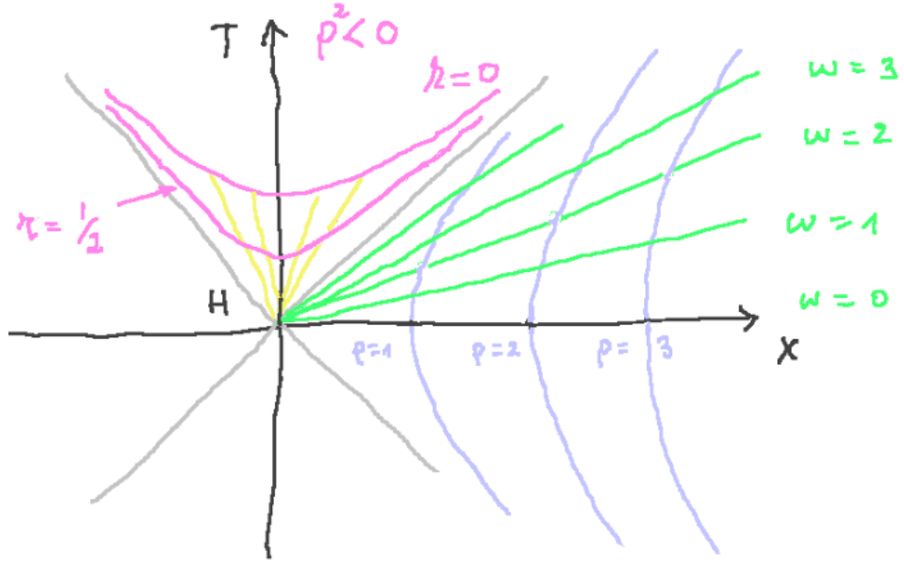
Figure 12 is what flat space-time looks like when compactified.

Now what about a black hole.

What does that look like ?

We can do exactly the same thing.

We can take the entire diagram in figure 1 (where we can rename the X-axis the R- axis), that is, the original Kruskal coordinates and diagram.



We can again introduce light-like coordinates, $T^+ = T + R$ and $T^- = T - R$, shrink them into functions varying between -1 and $+1$.

We do exactly the same operations.

What will we get ?

What it will look like is the diagram in figure 13.

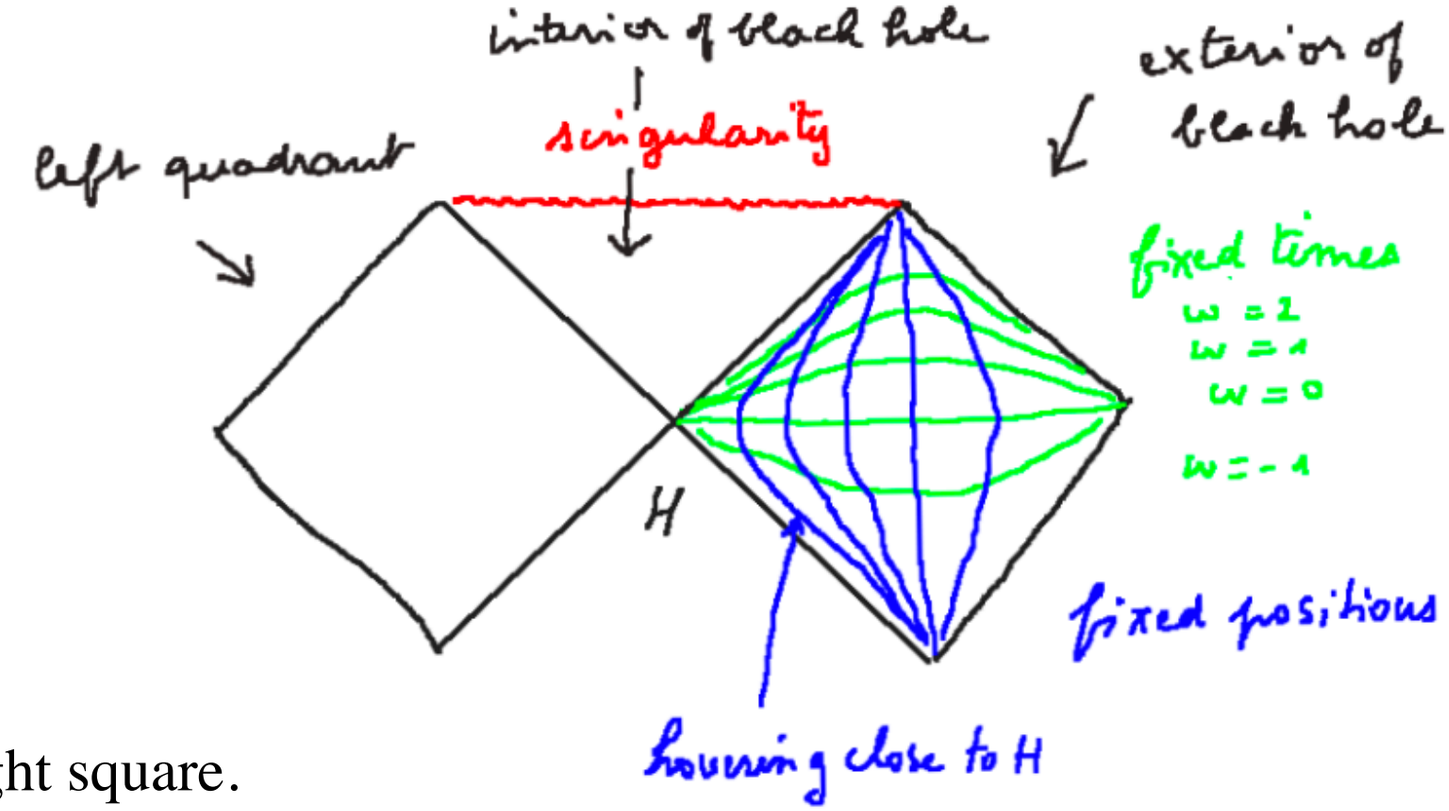


Figure 13 : Black hole space-time compactified.

The exterior of the black hole is the right square.

The hyperbolas of fixed positions hovering above the horizon H become the blue curves.

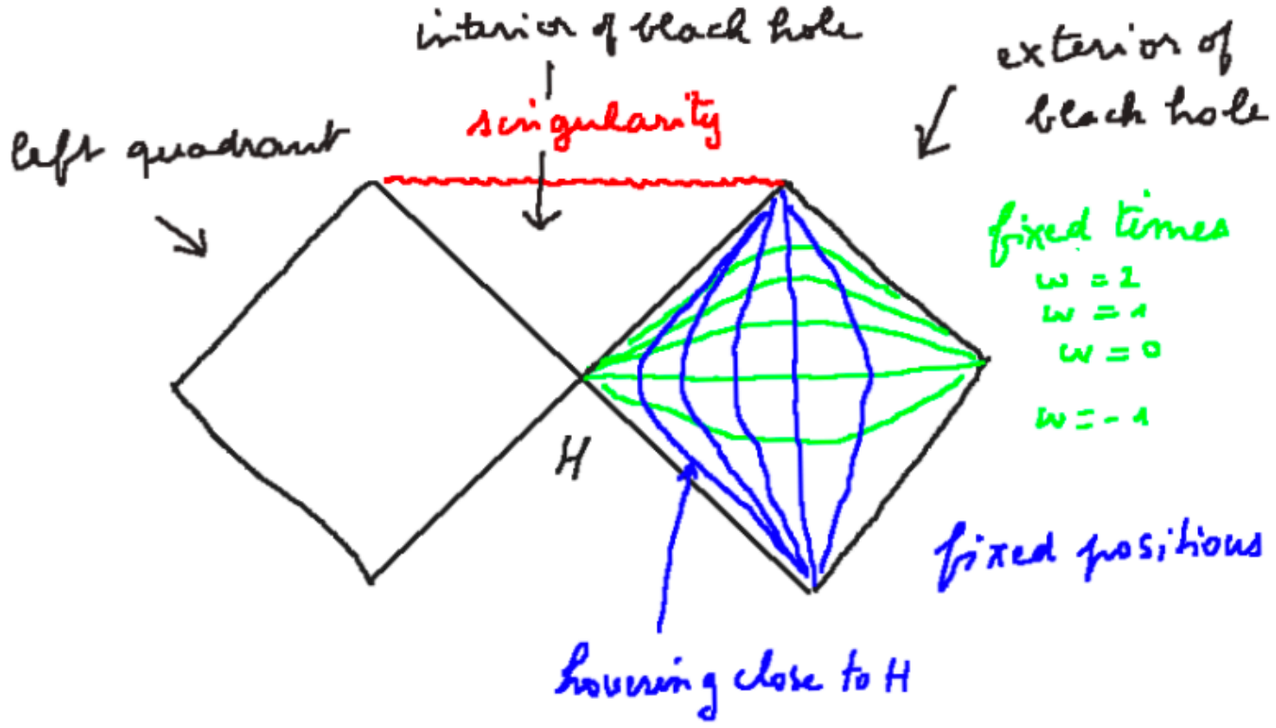
The original straight lines of constant time ω become the green curves.

The interior of the black hole is the upper part.

It is the place where we are doomed.

It is the half square bounded by the singularity.

The singularity is not off at infinity.



Remember that once we are inside the horizon, we cannot stay at a fixed position on the graph because we cannot avoid the future, and it takes a finite amount of time to travel to the singularity and be annihilated by infinite tidal forces.

Figure 13 shows the entire space-time of a black hole in a Penrose diagram.

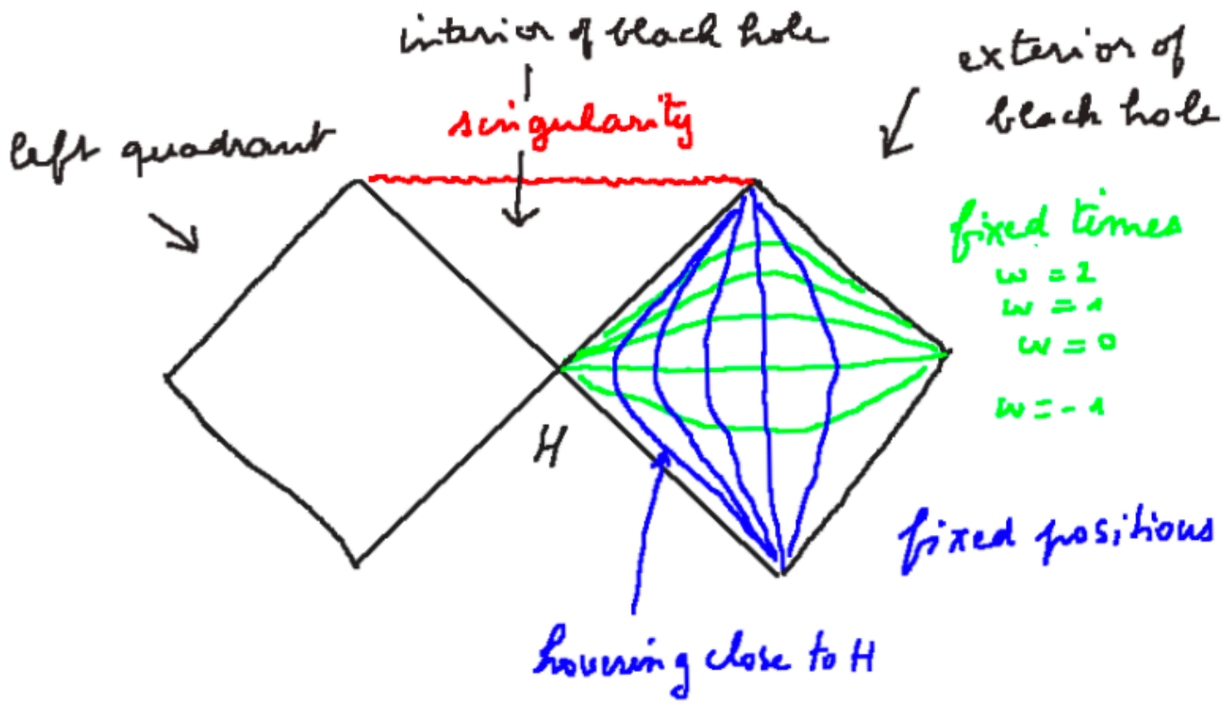
If we are in the right part we can stay where we are, but that takes acceleration in the sense that we have to oppose gravitation pulling us in.

We can even escape to farther away – but, remember, all outer space is now represented in the right square.

We can also fall beyond the horizon.

Once we are inside the horizon, we cannot go back to the right part.

We are doomed.



The picture in figure 13 asks an obvious question : what could possibly correspond to the left square ?

Same for the lower part.

These are the compactified representations of the left quadrant and the lower quadrant of figure 1.

The left quadrant and the lower quadrant have no real meaning for a real physical black hole.

We will see that when we work out a real physical black hole, how it forms, etc.

Nevertheless, we can ask ourselves : what kind of geometry is described by the full extended Kruskal-Penrose diagram of figure 13.

It seems to have two exterior regions – the right and left squares – connected together at the horizon H. Remember that, at the horizon, $R = 2MG$.

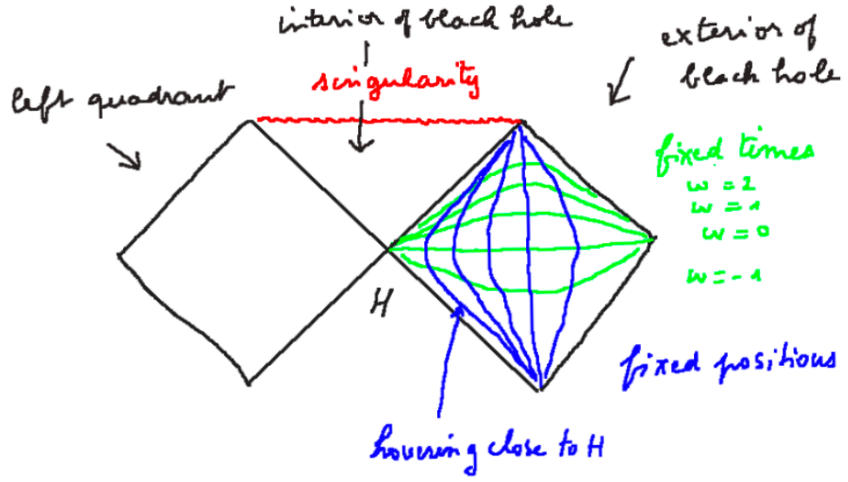
This value of R is also denoted R_S, and called the Schwarzschild R.

It is the place where $(1 - 2MG/R)$ changes sign.

We arrive from above, that is from the exterior region, the coefficient changes sign, and we continue in the interior region, represented in the upper quadrant or upper half square.

Now suppose, at a fixed time, say $T = 0$, we look at a slice of space, or at the entire space.

What does it look like?



In figure 13, it is just the complete horizontal segment from the right the left of the figure.

But in all our diagrams so far, for the sake of drawing, space was only one-dimensional.

Think of space as three dimensional.

Then for each R it is the surface of a sphere – also called a 2-sphere – of radius R.

If, for drawing purposes, we think of space as only two-dimensional, for each R the 2-sphere becomes a "1-disk", that is, the boundary of a disk, in other words a circle.

Flipping the diagram from horizontal to vertical with H down below, we can represent these circles in figure 14.

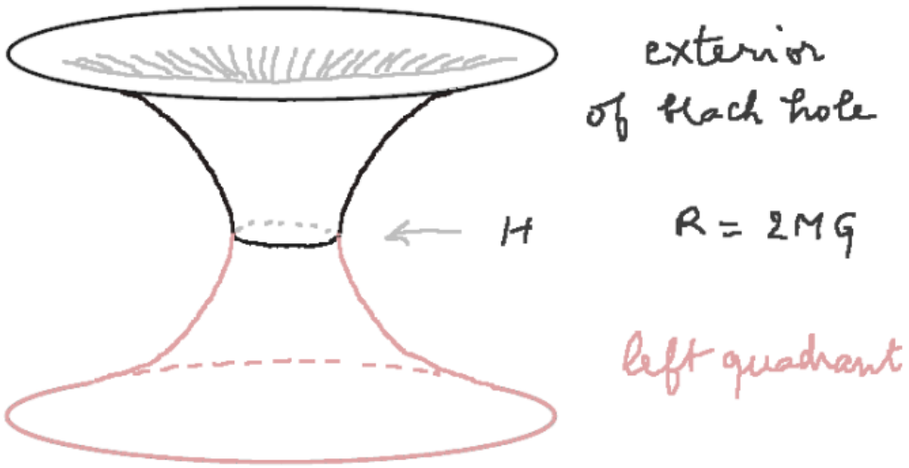


Figure 14 : Wormhole.

On the diagram, at the horizon, instead of turning into the interior region (upper quadrant or upper half square), we just "kept going".

That is the lower part of figure 14.

It represents the left quadrant, or left square, of the previous figures.

Once we have passed the horizon in this way, the space of each R starts expanding again.

It looks like we can pass through from one side to the other by going through what people call a *wormhole*.

It is also called an *Einstein-Rosen bridge*.

I have appended to this part a short mathematical description of the embedding process that produces the famous wormhole diagrams.

It connects two seemingly external regions of the black hole, which get bigger and bigger as we move away from the bottleneck.

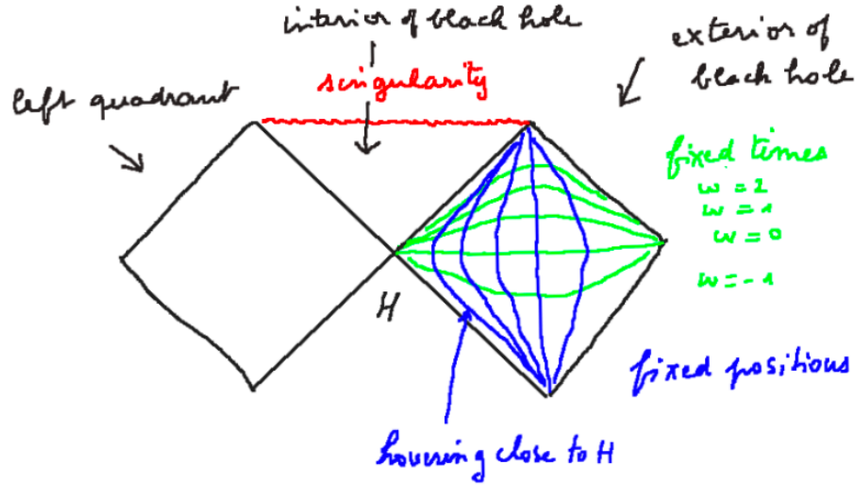
In other words, it looks like the black hole is connecting two universes, or two asymptotic regions.

You might think : well we could pass through the bottleneck at the horizon going from right to left.

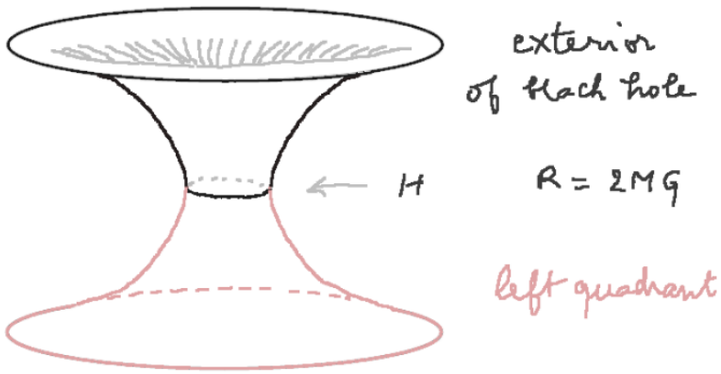
But we can't.

Let's think of somebody who wants to make such a trip.

If that person starts anywhere on the right and is to end up anywhere on the left, his or her trajectory at some point must have a slope smaller than 45°, that is, he or she will have to exceed the speed of light.



In fact the diagram of figure 14 is somewhat misleading.



It shows everything at an instant of time.

But there simply isn't really time to go from the upper part to the lower part.

We cannot go from the real exterior of the black hole to the left quadrant (shown in the lower part of fig 14).

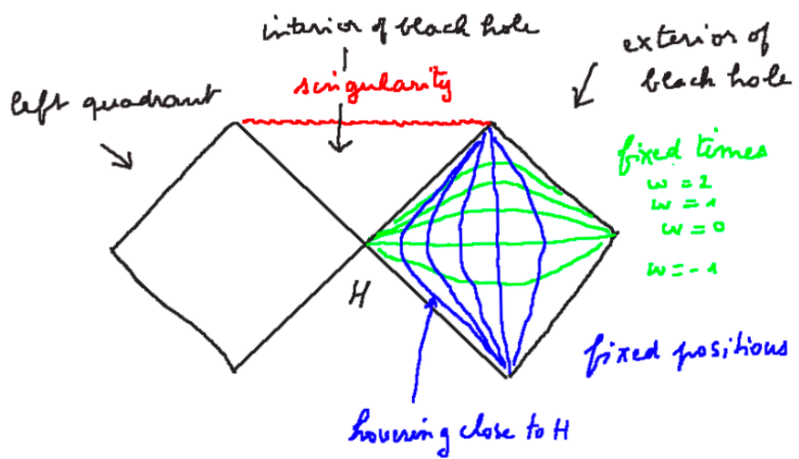
Therefore the Einstein-Rosen bridge isn't really a bridge.

One way to think about it is that the bottleneck opens up and closes again before anything can pass through it.

But the best way is just to look at figure 13 and say : yeah if we could exceed the speed of light and move horizontally yes we would move from right to left going through the neck at the center.

However we are not allowed to do that.

We can only move along 45° lines or steeper.



This travelling through wormholes has been the source of all sorts of science fiction, passing from one universe to another through the horizon of a black hole (not into its inside but into its left region).

However as you can observe it cannot happen.

These wormholes don’t allow you to access other universes.

They are of a kind called "non traversable wormholes", meaning to say you can’t traverse them.

This may come as a disappointment that we won’t be able to go on a field trip to other universes at the end of the lectures.

But in fact, as we shall see in the next section, there is no real meaning to the left hand side of figure 13 anyway.

It is not a real place.

Let’s talk now about creating a black hole, not in the laboratory because that is too hard, but in an infinite space by having some in-falling matter.

Formation of a black hole and Newton's shell theorem

In order to study the formation of a black hole, we are going to take a very special kind of in-falling matter. .

We begin with a point in space-time, where is no black hole nor anything, and with a very distant shell, figure 15.

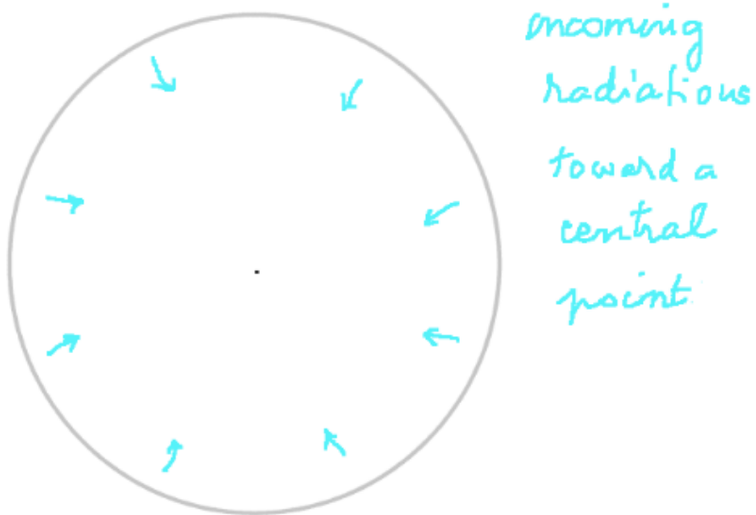


Figure 15 : Shell of incoming radiations.

The shell is not made of iron or other matter like that.

It is a thin shell of incoming radiations.

Radiation carries energy.

Radiation carries momentum.

For one reason or another, it has been created far away.
And it is incoming, with spherical symmetry, at the speed of light, toward a central point.

At some point in time for the observer, as the shell gets close enough together, there will be so much energy in a small region that a black hole will form.

This is the simplest version of black hole formation.

A star isn't really made up out of things which fall in with the speed of light.

What we are looking at is the problem of a black hole being formed by stuff coming in with the speed of light, the stuff being on a thin spherical shell.

It is the simplest of all of relativity problems.

There are only two important things we need to know to understand this formation of a black hole :

1. The shell moves in with the speed of light.
2. We need to know Newton's theorem in classical mechanics, and a version of it in relativity.

Here is Newton's theorem in classical mechanics.

If we have a shell of matter forming a 2-sphere, that is, uniformly spread on the surface of a sphere, then

a) The gravitational field *inside* the sphere is null.

b) The gravitational field *outside* the sphere is identical to that of a point mass at the center with the same amount the mass.

That is Newton's theorem or a special case of Newton's theorem : in the interior of a distribution of mass forming a spherical shell you see no gravitational field, and outside you see the gravitational field as it would be if all of the mass was concentrated at the central point.

That is even true if the shell is moving, for example collapsing toward the center.

Of course, as time goes, there will be less inside space, with no gravitational field, and more outside space with a gravitational field.

But it is still true.

This theorem is also true in general relativity.

It now says the following : if you have a shell of in-falling mass or energy of any kind, then the interior region is just flat space-time.

It is like space-time where there is no point source, or no mass.

So for the interior, it is similar to classical mechanics.

But for the exterior region, it doesn't look like a Newtonian point source with an ordinary gravitational metric around it – because there is no such thing in general relativity.

What does it look like then?

It looks like the Schwarzschild metric, that is, the solution to Einstein field equations that we shall study in lesson 9.

So in general relativity, the inside of a shell of mass or energy (that is the same thing) is flat space-time.

And the outside of it looks like a Schwarzschild black hole.

If you have a static, non-moving shell, what you would do to construct the actual solution, that is, the actual metric, is sort of paste together the metric in the interior, which is flat, with the metric in the exterior, which is the Schwarzschild metric.

Notice that we do the same thing in Newtonian physics : we paste together no gravitational field on the inside, with, on the outside, the standard gravitational field due to a central point source.

And that is the way we solve problems of Newtonian physics involving a 2-sphere of mass.

You do exactly the same thing in general relativity.

Also, notice a similar fact in electromagnetism : consider a conducting body with electric charges in a stable, that is static, configuration.

Then – whatever the shape of the body – the charges are on the outer surface, and there is no electric field inside the body.

Let's redraw the Penrose diagram of a black hole to illustrate what we are going to find, figure 16.

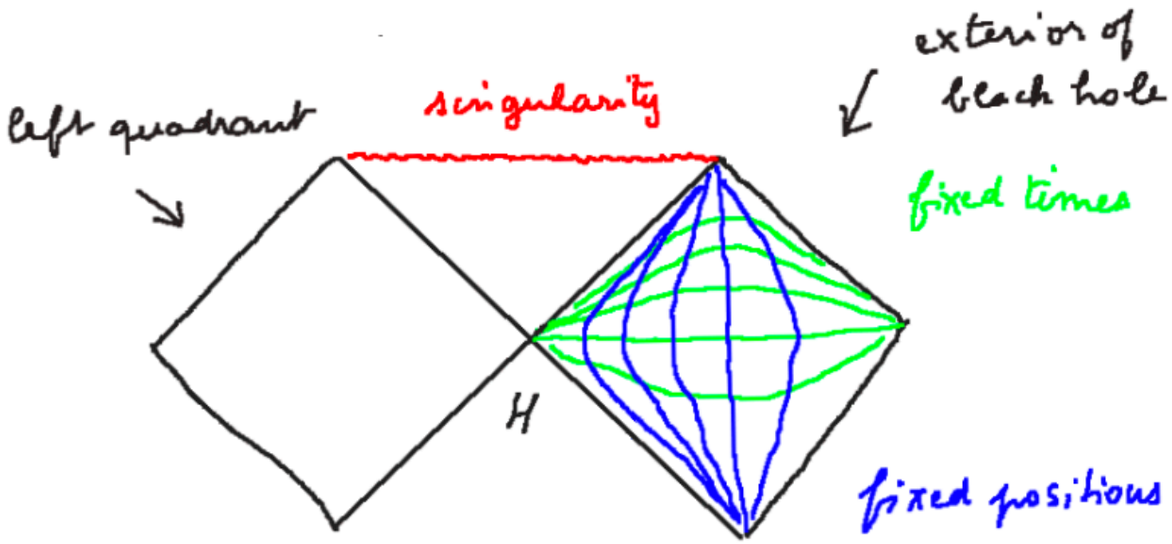


Figure 16 : Penrose diagram of a black hole.

Let's also redraw the Penrose diagram of flat space-time.

But we now add an incoming shell of radiations, figure 17.

We can think of the incoming shell of radiations as a sort of pulse of incoming photons, a pulse distributed nicely on a sphere.

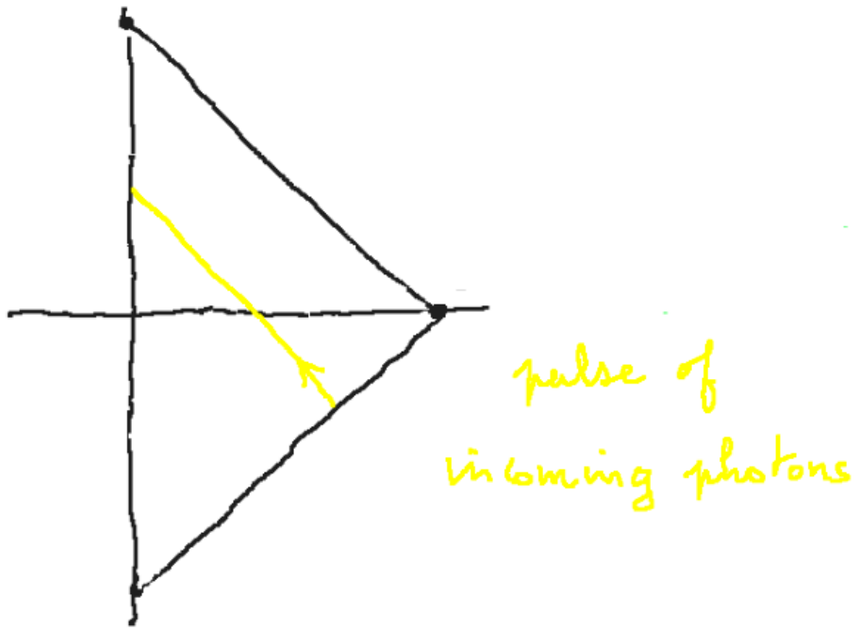


Figure 17 : Pulse of incoming photons in flat space.

The pulse of incoming photons comes from past light-like infinity (what we also call scry-).

In 3D they form a continuous collection of shells focussing on the origin, but in figure 17 where we can draw only one spatial dimension they are represented simply as a straight line.

Nonetheless, let's think of it as a shell.

At any given instant of time, where is the interior and where is the exterior of the shell?

The diagram makes it easy to answer.

The triangle represents space-time with R and T.

But both shrunk so that the entire half plane of figure 2 fits into the triangle.

A point in time and space (being represented with only one dimension) is a point in the triangle, figure 18.

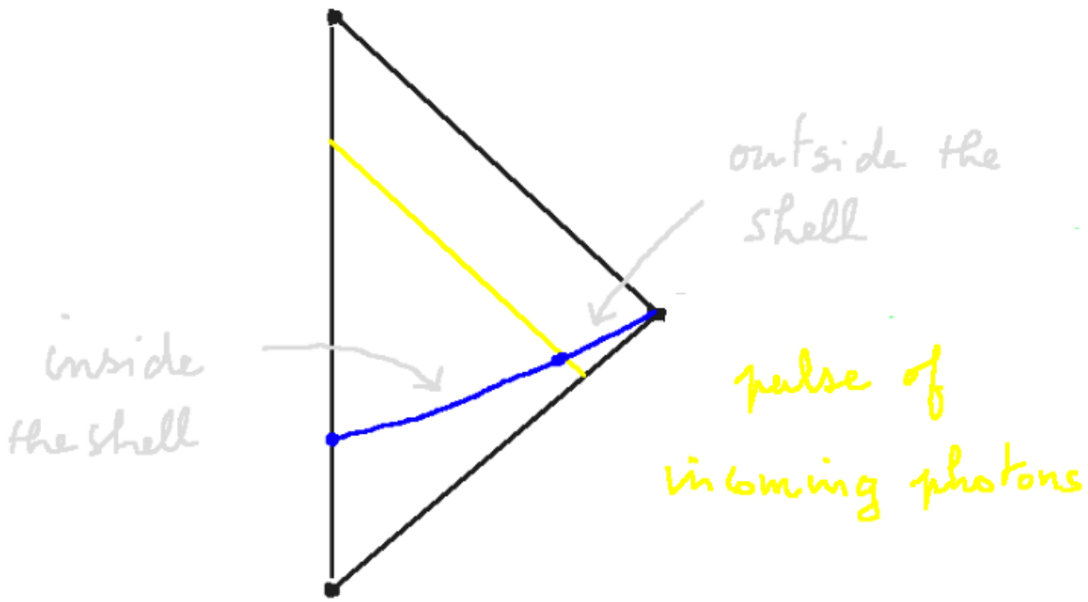


Figure 18 : Interior and and exterior of the shell at any given instant of time.

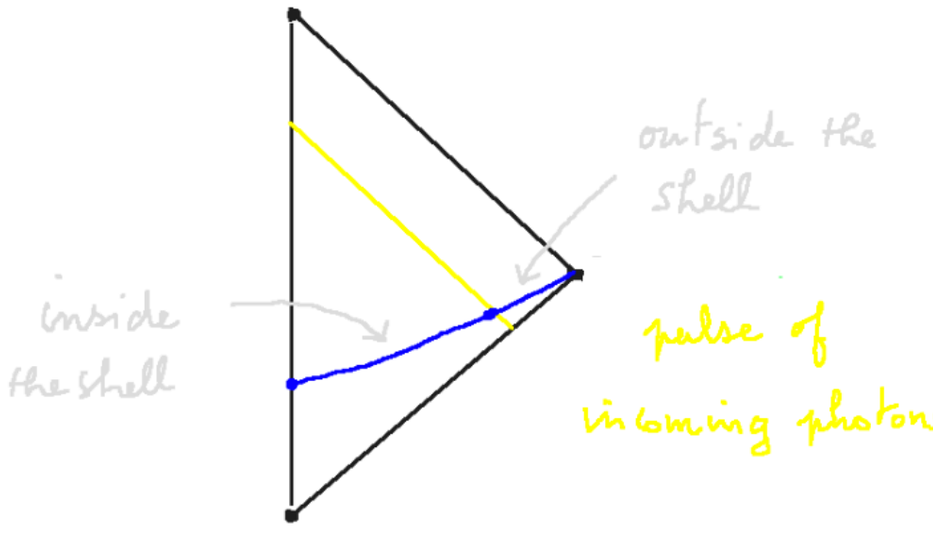
The trajectory of the shell is the yellow line.

All the events at a given time is the blue curve.

At a given time the shell is at the intersection.

The interior of the shell is the part of the blue curve to the left.

The exterior goes from the shell to space-like infinity.



What Newton's theorem (or its general relativity version) says, is that on the interior of a shell everything is flat space-time.

On the diagram of figure 18, we have a dynamic view of the interior of the shell has it moves.

In other words, it is correctly represented by the space-time drawn in figure 19, which was just a representation of flat space-time.

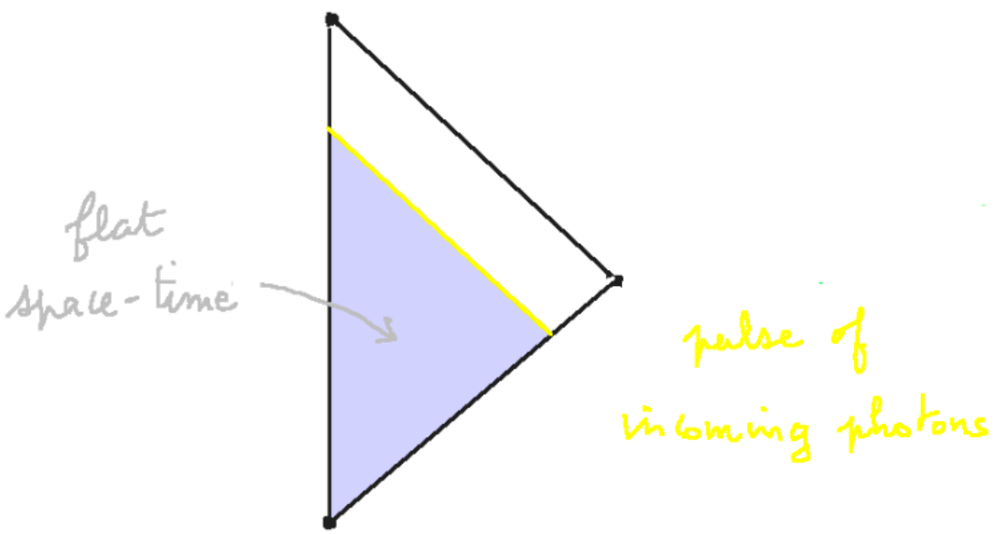
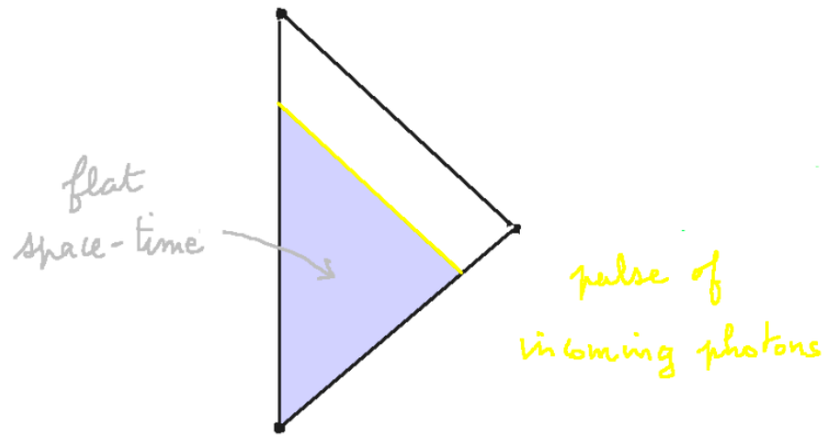


Figure 19 : Flat space-time inside the shell, as the shell moves in time.

So, first of all, for the interior of the shell the Penrose diagram of flat space-time is the correct representation of everything that is going on.

Now what about the exterior?

On the exterior we are told that there is a gravitational field.



It does not look like flat space-time.

So the upper white region of the triangle in figure 19 is not the correct representation of the physics or of the geometry of the in-falling shell.

Out beyond the shell, what is the correct representation?

It is the representation of the Schwarzschild black hole of figure 16.

Let's redraw it without the unnecessary details of fixed times and fixed positions, figure 20.

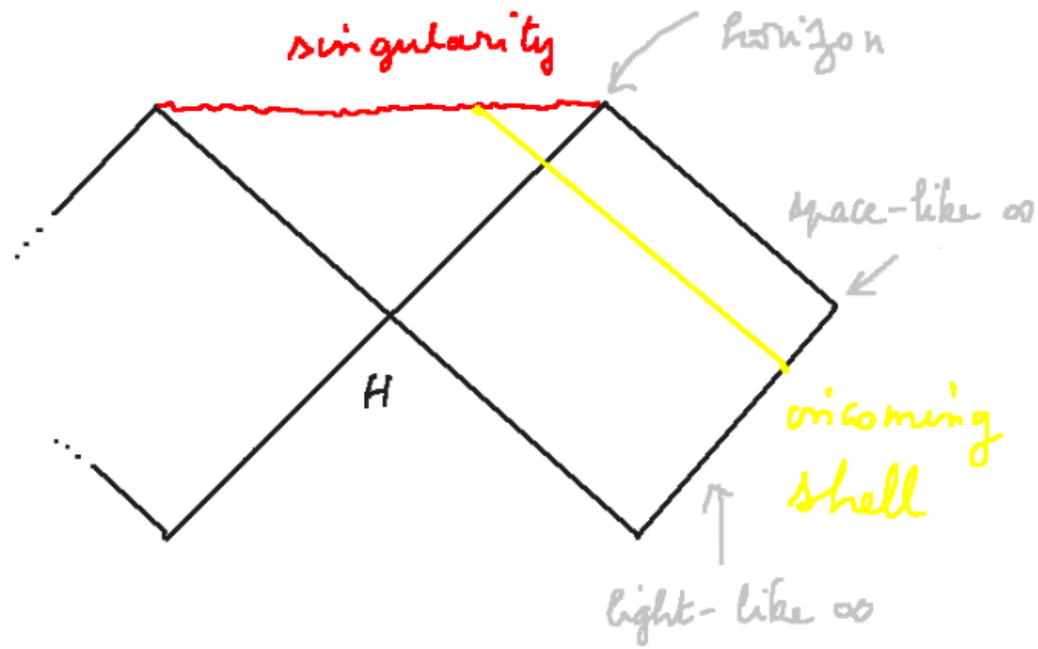
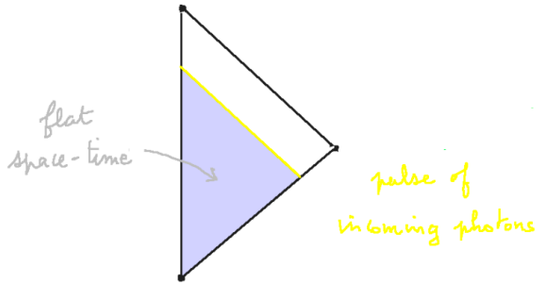


Figure 20 : Shell in the black hole diagram, as it moves in time.

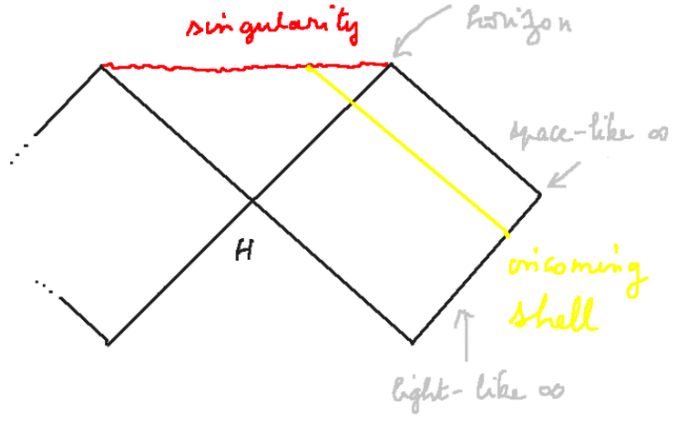
Somebody on the outside (the right square) throws in the shell.

The in-falling shell, represented with one spatial dimension, is a radially incoming light-ray.

It must move along a 45° straight line.



It comes in from far away.



It experiences nothing special when it crosses the line called horizon.

It keeps going and eventually hits the singularity.

But now which part of the diagram, in figure 20, is correctly representing the physics that we are doing?

Notice that the interior of the shell is not correctly represented by the black hole diagram of figure 20, because its correct representation is the flat space-time diagram of figure 19.

The exterior of the shell in the black hole diagram is shown in figure 21.

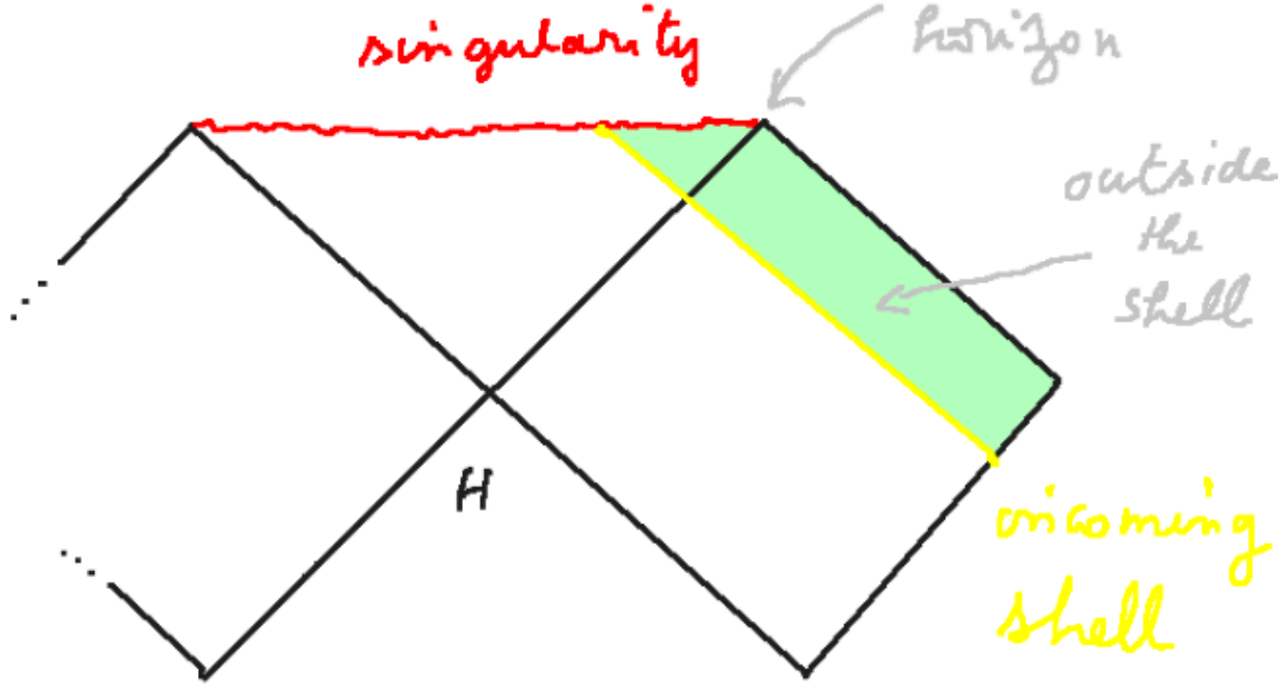


Figure 21 : Outside the shell, over time, in the black hole diagram.

So we have two diagrams, one representing correctly the inside and incorrectly the outside, and the other representing correctly the outside but incorrectly the inside.

How do we put the two together in order to make a single geometry?

It is pretty easy.

In each diagram we throw away the wrong part, then we paste together the correct parts, figure 22.

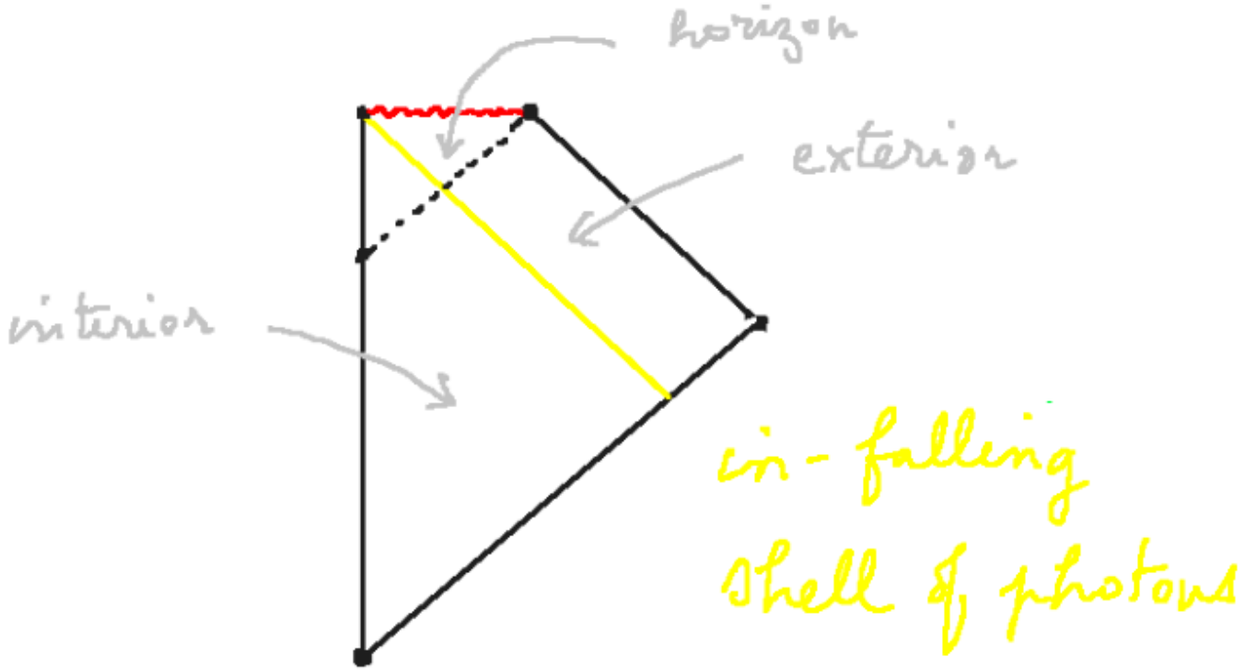


Figure 22 : Interior and exterior of the shell pasted together.

The in-falling shell is the thing that the two parts of figure 22 have in common.

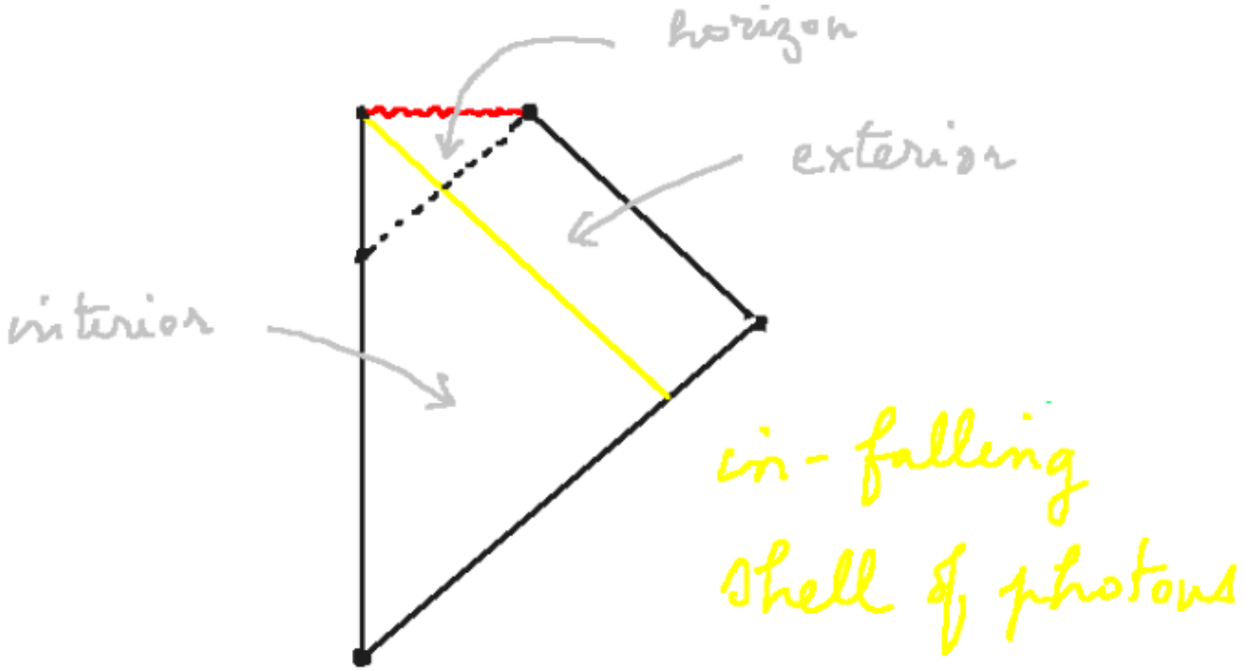
On one side of the shell we put flat-space geometry.

On the other side we put black hole geometry.

The complete diagram represents the geometry of a black hole that is formed by an in-falling shell.

In figure 22, we also plotted the horizon.

It is the 45° dotted line ending up at the intersection of the singularity and scry+.



Nothing special is experienced by someone crossing the horizon.

But once you are inside the horizon, that is, above the dotted line in the diagram, there is no way you can avoid eventually hitting the singularity.

Therefore once inside you are doomed.

On the other hand if you are outside the shell and outside the horizon, you can get away.

You can escape the scry+ for example.

Notice something very curious.

Even in the little triangle shown in figure 23, where it seems that life should be trouble-free, you are doomed.

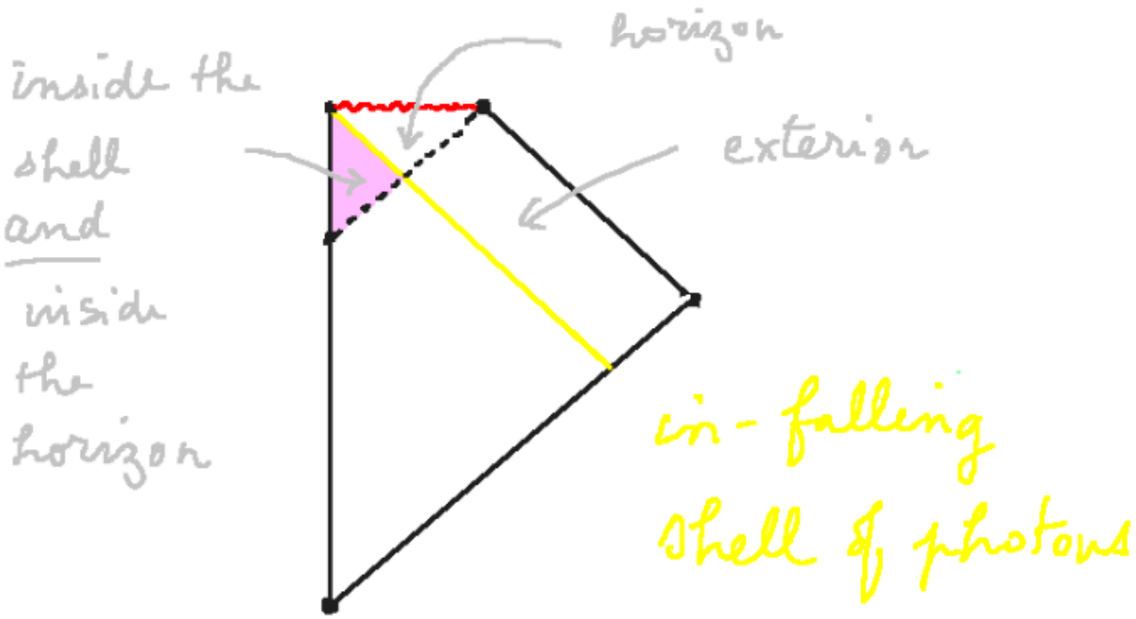


Figure 23 : Inside the shell and inside the horizon.

In the little triangle you are still in flat space-time.

The shell hasn't even gotten in yet.

You cannot see the shell. If you looked backward on your light cone, you would not see it, because for you looking backward means looking either at light-rays coming from the past parallel to the shell trajectory or on the other 45° line of rays coming from the other direction.

In both cases, they don't meet in the past.

So you don't know the shell is coming.

You are standing in the flat, apparently trouble-free, space-time region.

Yet you are also standing behind the horizon defined by the shell.

In other words, even though the shell is still far away, and the horizon still very small, you are already in the doomed region.

To understand this paradoxical situation, let's see what could happen.

If you choose not to move in space, the shell will come in and pass you.

Then there is no question that you will be inside the Schwarzschild radius.

If you choose to move, trying to get out, you can only move on trajectories steeper than 45°.

So you will hit the singularity anyway, and in a finite amount of your proper time at that.

Whether before the shell caught on you or after, you will hit the singularity.

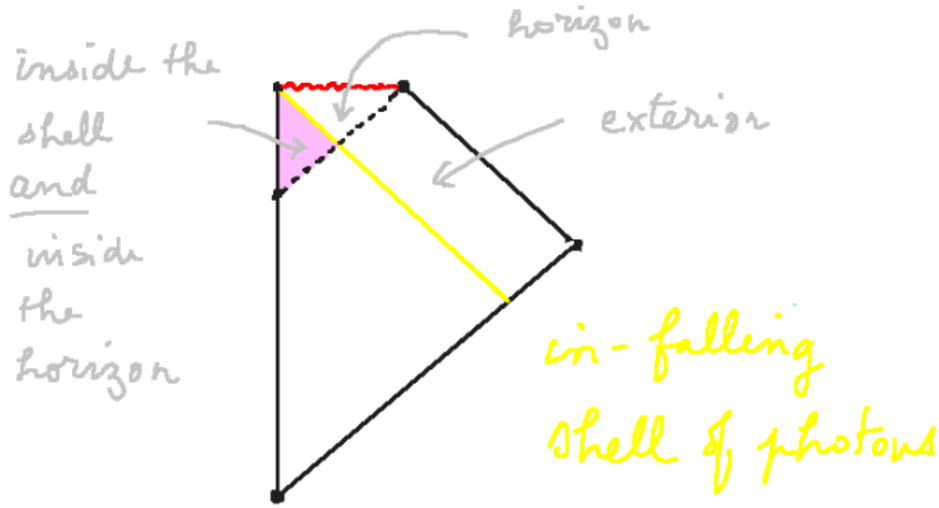
The horizon itself starts very small at some time in space-time (see next discussion for a detailed discussion).

Then, as the shell approaches, the horizon grows following the dotted line.

At some point in time, the shell crosses its horizon.

Then the horizon does not grow anymore.

In figure 23, surprisingly enough, the dotted line, after the intersection with the yellow line, corresponds to a fixed $R = 2MG$.



The usefulness of the Penrose diagram, combining the geometries inside and outside the shell, is to show us that – provided we cannot exceed the speed of light – there is no path out from beyond the horizon, no matter where the shell is.

Of course, if we are inside the shell but outside the horizon, we are not doomed.

As long as the shell has not passed us, we can stay where we are effortlessly.

Once the shell has passed us, to stay still we will have to fight gravitation with acceleration (like you and me do when we are sitting on a chair – even though we tend to forget it).

If we don’t fight acceleration, we will fall beyond the horizon and be annihilated when we hit the singularity.

But if we accelerate we can not only stay fixed, but we can even fly off to scry+ if we move fast enough.

Let’s look at the black hole formation with the more conventional concrete diagram, figure 24.

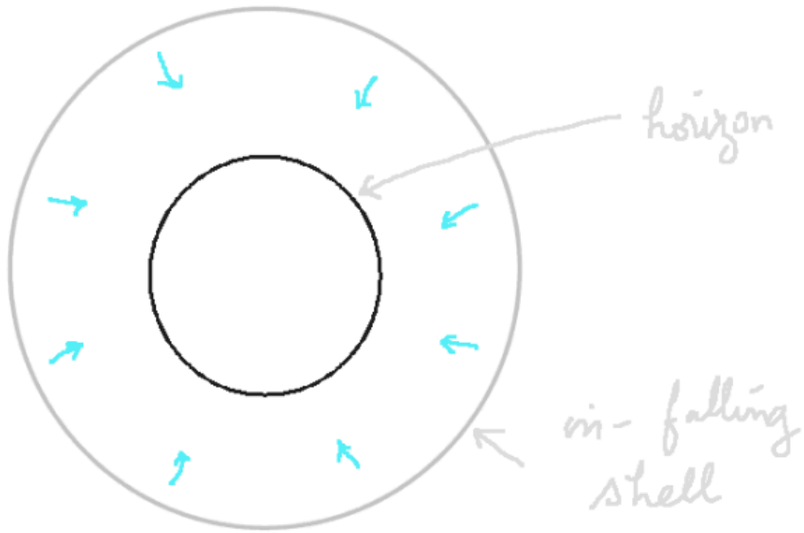


Figure 24 : Shell, black hole and its horizon.

Inside the shell, space-time is flat.

Outside the shell, space-time has the same metric as that of a black hole of the same mass at the center of the picture.

The horizon is growing toward the Schwarzschild radius, then it doesn’t grow anymore.

Whatever is inside the horizon is doomed, *even if the shell is not there yet*.

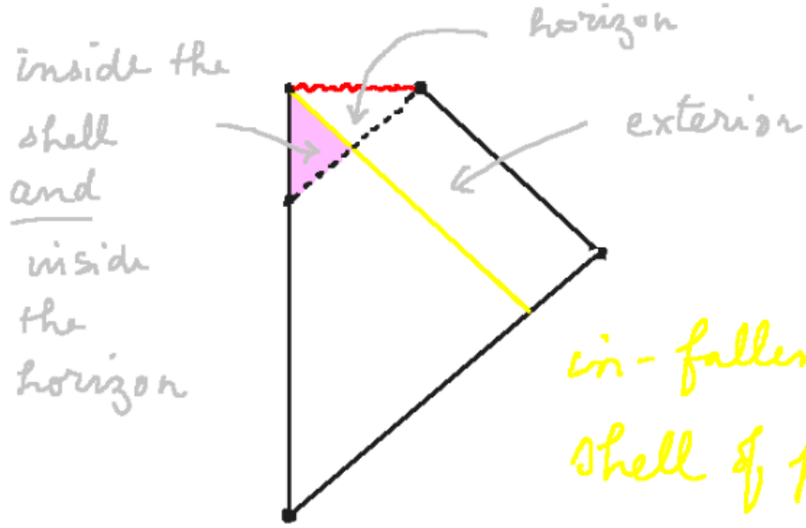
So long as the shell has not crossed its horizon, it could presumably change its mind and accelerate back out.

Well we can suppose that shells of photons don't have minds.

But if they had one the photons could turn around as long as they are farther away than 2MG.

When the shell crosses the horizon nothing special happens, except that now the photons are trapped for good.

They can no longer turn around.



Or, well, they could try to turn around, but would not get very far.

They are now doomed to hit the singularity.

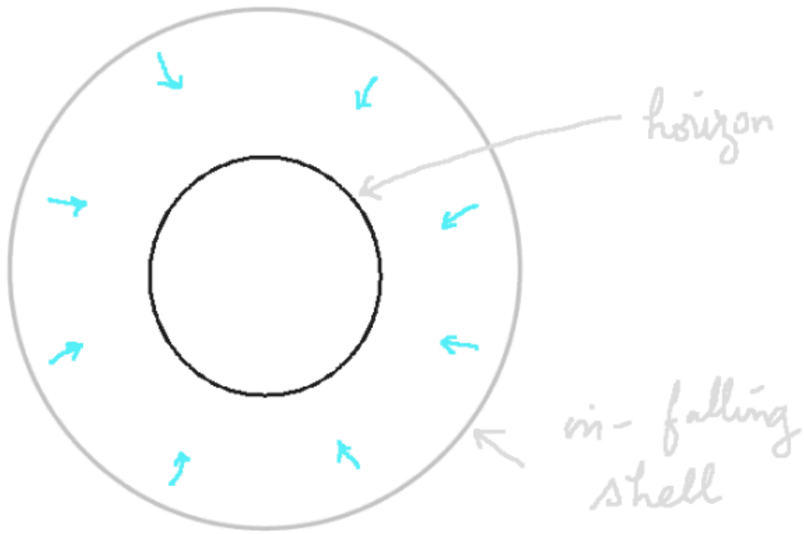
In figure 23, we can visualize the point where the shell (yellow trajectory) crosses its own horizon (dotted line).

It happens at $R = 2MG$.

Afterwards, we can say that the black hole has formed for good.

Discussion of the paradoxical aspect of the horizon before the shell has crossed it, figure 24 : it seems that as long as the shell is outside its horizon, as in figure 24, someone inside the horizon could possibly move outside of it, while still inside the shell, because in that region space-time is tranquil.

Therefore the person would escape from his or her doomed destiny.



But that is a mistaken view.

There would not be time.

Or, if you prefer, the person would have to exceed the speed of light.

That is the superiority of the diagram of figure 23 over the more conventional diagram of figure 24.

The Penrose diagram is more abstract, but it shows explicitly the possible and impossible space-time trajectories.

So the horizon is a funny thing.

In a ghostly manner, it begin to exists, at first very small, at some point in time after the shell has begun to fall in and is still far away.

It grows up to the Schwarzschild radius and then does not grow anymore.

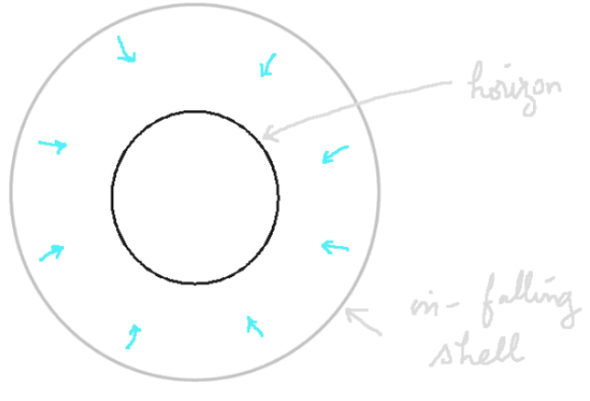
The horizon is the point of no return.

It is its definition.

Discussion of the time variable

Let's look more closely at the relationship between figure 23 and figure 24, and at the significance of the time variable.

Figure 24 is the conventional picture of the in-falling radiations.



It is a snapshot at an instant of time.

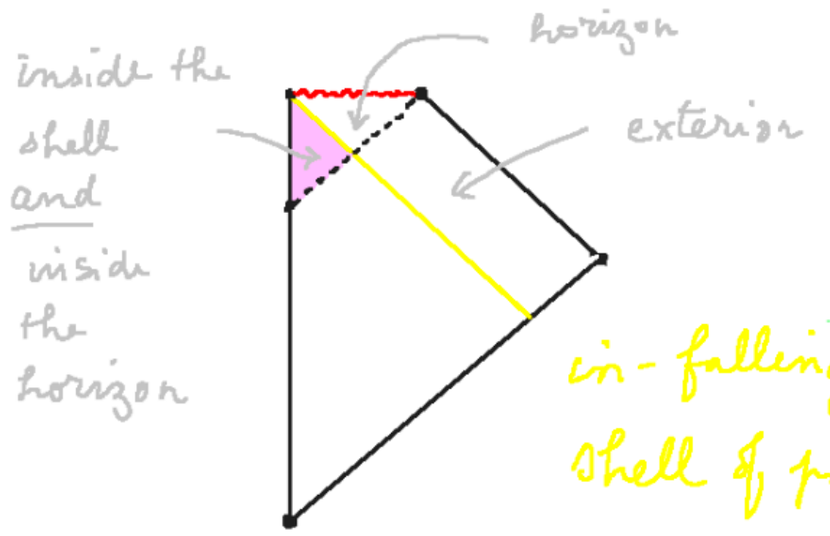
Figure 23 is the Penrose diagram for the formation of a black hole.

It shows in the same picture space and time, that is, the entire space-time.

In order to speak of an instant of time, we have to specify which time do we use ?

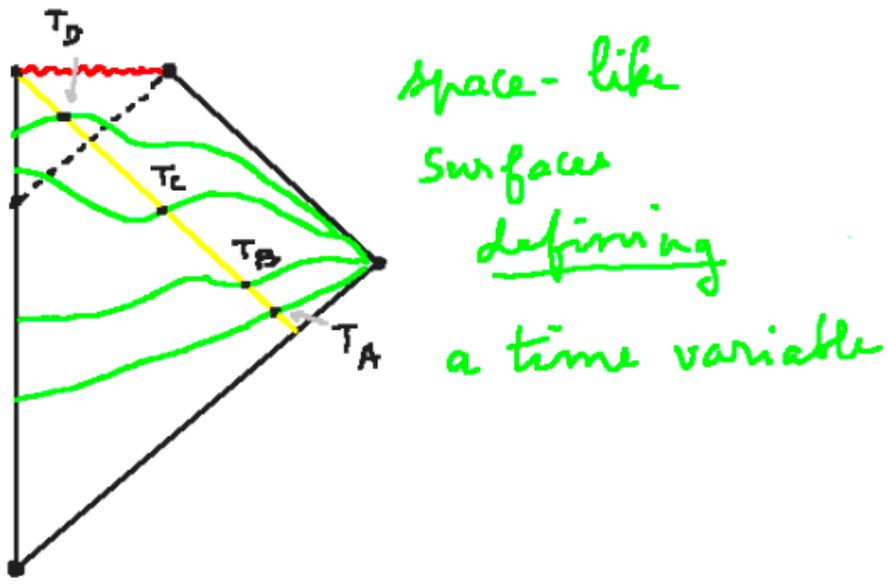
Time is just a coordinate.

We can make coordinate transformations.



To talk about an instant of time really means picking a surface, from a collection of surfaces which are everywhere space-like, see figure 25.

Figure 25 : Defining a time variable.



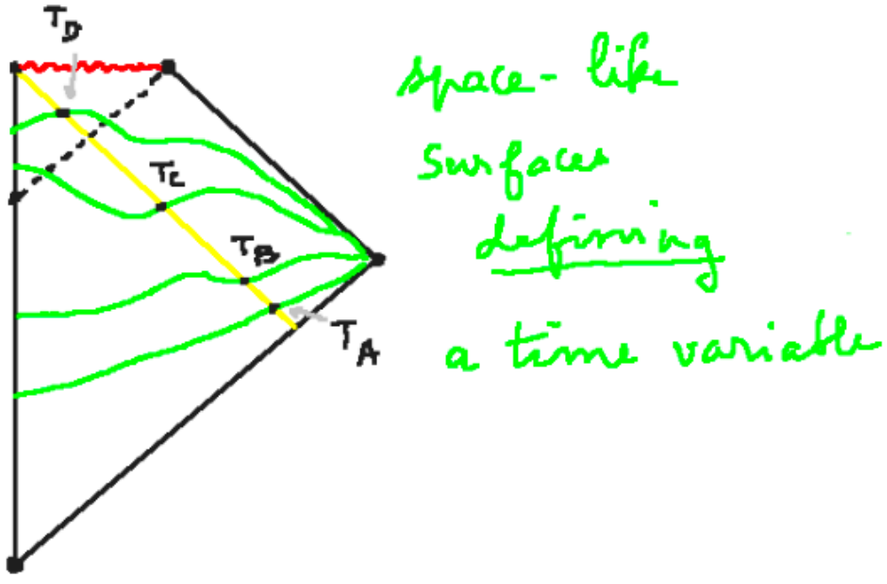


Figure 25 : Defining a time variable.

From this collection of space-like surfaces indexed by some number, time can be taken to be the index of the surface we are on.

If we follow this time variable, at time T_A there is a shell far away.

There is no horizon.

At time T_B the shell is closer.

The horizon has not begun to form yet.

At time T_C , the horizon exists, but the radiations are still outside.

Sometime between T_C and T_D, the shell crosses its horizon.

Then we can say that the black hole exists.

To summarize : as the shell falls in, at some point in time the horizon appears, at first as a tiny point.

It grows until the shell meets it, at $R = 2MG$.

Then it stays at the Schwarzschild radius.

As long as the horizon is below the shell, it is in the flat space-time region.

Someone crossing the horizon doesn't feel anything.

On either side the person is still in gravitation-free space.

However beyond the horizon, even in the flat region, the person is doomed.

The boundary is the dotted line of figure 25.

It is not when we cross this boundary, that is, the horizon, that we begin to feel gravity.

It is when we cross the shell.

The importance of the diagrams cannot be overemphasized.

If you try to do calculations to find out where for example the horizon forms, you will find that can't do it.

You will wind up drawing the picture, drawing the dotted line from the intersection of scry+ and the singularity, at 45°.

If you want to get a good picture of what is going on, the way to do it is to become familiar with all the diagrams in the lesson, particularly with the Penrose diagrams – for flat space-time, for black hole, and their combination.

If you try to reason the problems out, without a good picture in front of you that accurately represents the relationships between the different parts of space-time, you will run into difficulties.

By the way, the important relationship is not the relative size of things in the pictures.

A small circle at the intersection of scry+ and the singularity represents a huge space-time region.

The same small circle where the horizon is born, along the vertical side in fig 25, represents a tiny region.

The important relationships shown by the Penrose diagrams are what light-rays – which are 45° lines – and other trajectories can link or not.

The diagrams illustrate causal relationships, what can cause what, what can signals propagate, who can send a signal to whom.

Can a signal from Alice here get to Bob there ?

In technical language we would say that the diagrams reflect the causal structure, cause and effect.

It means : what can influence what.

The rule is that an event E – that is, a point in the space-time diagram – can only influence things in front of it, in other words things in the region where light emitted from E or slower travelling objects sent from E can get to.

E cannot influence events outside its light-cone.

The Penrose diagrams were built first of all to be able to draw and examine everything on one page, and second of all to reflect faithfully the causal relations, what can have an effect on what.

For that reason these diagrams are very valuable.

It is very difficult to think without them.

With the Penrose diagrams we see things much more clearly than without them.

For instance, we saw how the Penrose diagram of fig 23 is much more talkative than the more conventional diagram of fig 24.

If you play with the Penrose diagrams, it won't take you long to become familiar with them and to be able to use them efficiently.

You may say that the Penrose diagrams are abstract pictures.

Yet general relativity is still classical physics.

It is much less abstract than for instance quantum mechanics.

General relativity can be represented.

It is more or less easy of course, but in the end usual intuition works.

In the next chapter we shall finally arrive at Einstein's field equations, which are the keystone of general relativity.

And Lev Landau said of general relativity that it is the most beautiful physical theory ever built.

I have left the actual derivation of the field equations until this late because the detailed derivation is not really important to understanding the physics, which was what I was most interested in presenting.

We will now work out the field equations and solve them to obtain all the equations we have been using to understand the physics (*actually Mathematica will do most of the work*)

In addition we will use the field equation to derive the equations we need to discuss Cosmology.

APPENDIX

Embedding Diagrams - Mathematical but fun and result is worth it!!!!

A helpful way to visualize warped space is to use an embedding diagram.

As I said, this derivation is for mathematically inclined reader - others just need final result.

Full Schwarzschild solution to Einstein's equations for point mass

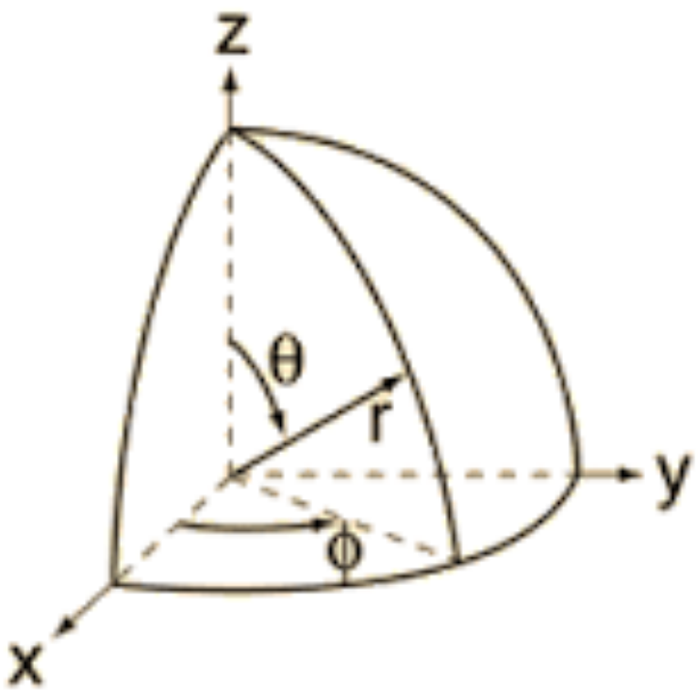
is represented by **interval** relation

$$ds^2 = -\left(1 - \frac{r^*}{r}\right) c^2 dt^2 + \frac{1}{\left(1 - \frac{r^*}{r}\right)} dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2$$

where

$$r^* = \frac{2GM}{c^2}$$

Variables (r,θ,ϕ) are standard spherical-polar 3D coordinates.



Since it is difficult to work with a full 3-dimensional curved space,

concentrate on 2-dimensional slice where we fix θ = π/2 (—> motion on a plane)

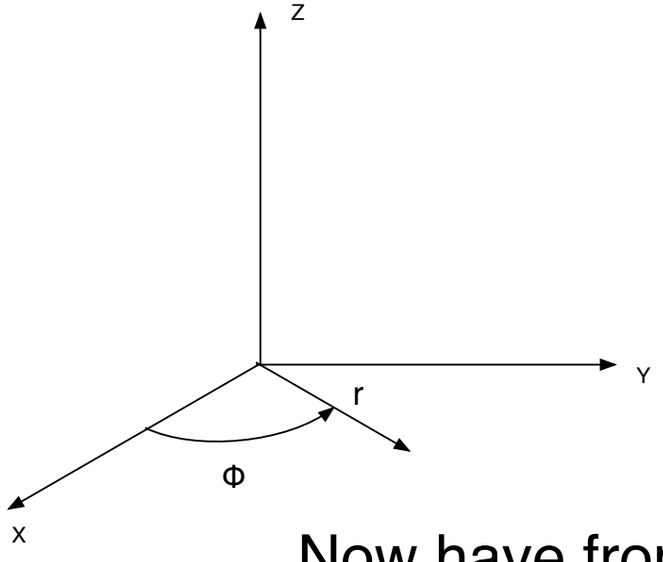
and consider only one instant of time(i.e., we are looking at only one instant of time).

In absence of gravity then have(flat space)

$$ds^2 = -c^2dt^2 + dr^2 + r^2d\theta^2 + r^2\sin^2\theta d\phi^2 \rightarrow dr^2 + r^2d\phi^2$$

Embedding procedure goes as follows:

1. We imagine this 2-dimensional surface as part of 3-dimensional Euclidean space with cylindrical coordinate system (r, ϕ, z) as shown with x = rcosϕ and y = rsinϕ.



2. Relationship z(r) between new coordinate z introduced to create fake Euclidean space and normal coordinate r is determined by assuming flat space interval expression

$$ds^2 = dz^2 + dr^2 + r^2d\phi^2$$

Now have from calculus

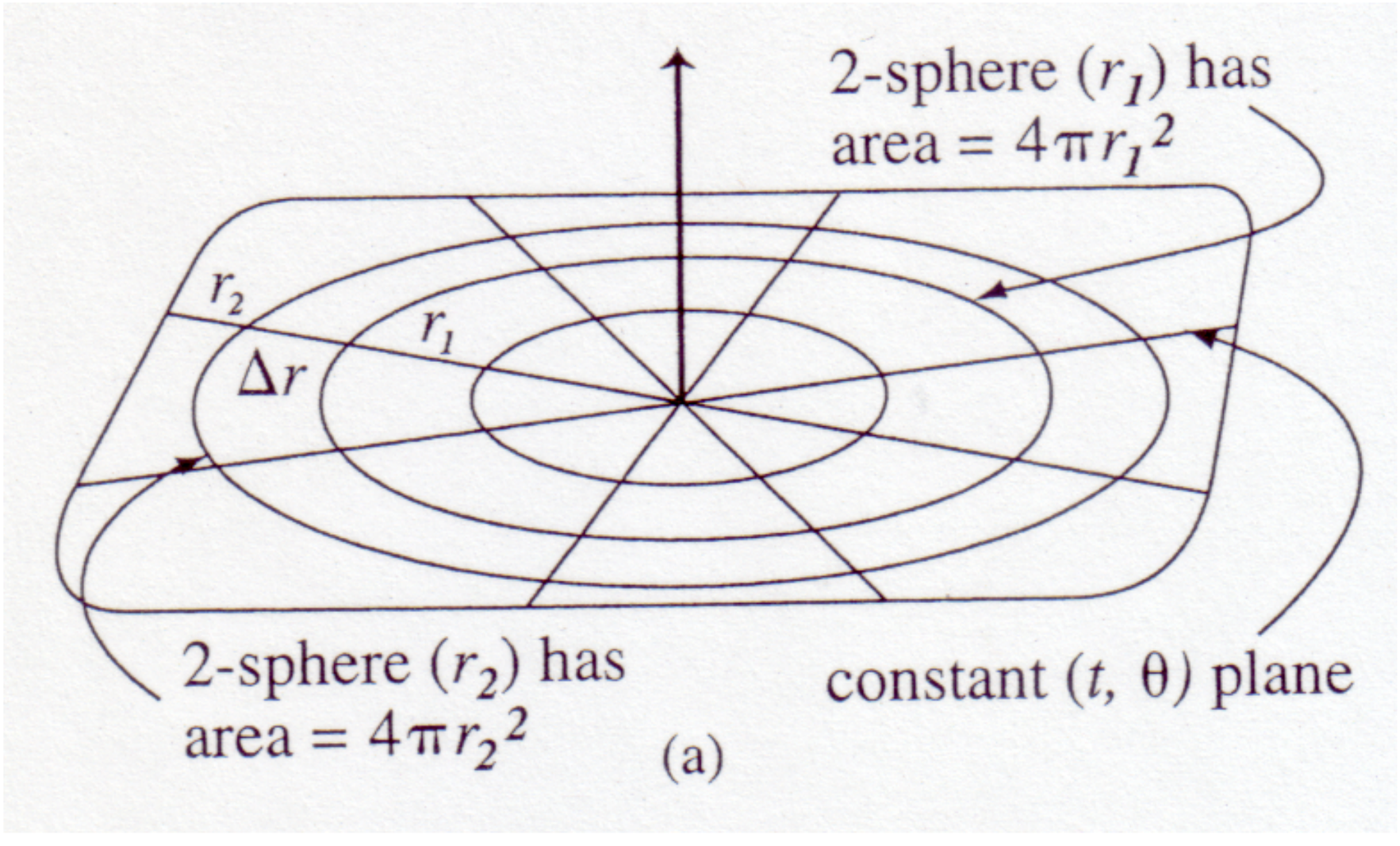
$$dz = \frac{dz}{dr}dr$$ so that obtain $$ds^2 = \left(\left(\frac{dz}{dr}\right)^2 + 1\right)dr^2 + r^2d\phi^2$$

3. Comparison with true interval expression in our special case says

$$ds^2 = \left(\left(\frac{dz}{dr} \right)^2 + 1 \right) dr^2 + r^2 d\phi^2 = dr^2 + r^2 d\phi^2$$

or that $$\frac{dz}{dr} = 0 \rightarrow z = \text{ constant}$$

This is just flat plane as shown below, i.e., 4-dimensional gravity-free spacetime is plane when embedded in 3 dimensions(fake).



If instead we carry out same procedure with Schwarzschild solution it goes as follows.

$$ds^2 = \left(\left(\frac{dz}{dr} \right)^2 + 1 \right) dr^2 + r^2 d\phi^2 = \frac{1}{\left(1 - \frac{r^*}{r}\right)} dr^2 + r^2 d\phi^2$$

or

$$\frac{dz}{dr} = \pm \frac{r^*}{(r - r^*)^{1/2}}$$

Using calculus then have

$$z = \pm 2r^* (r - r^*)^{1/2} \rightarrow z^2 = 4r^* (r - r^*)$$

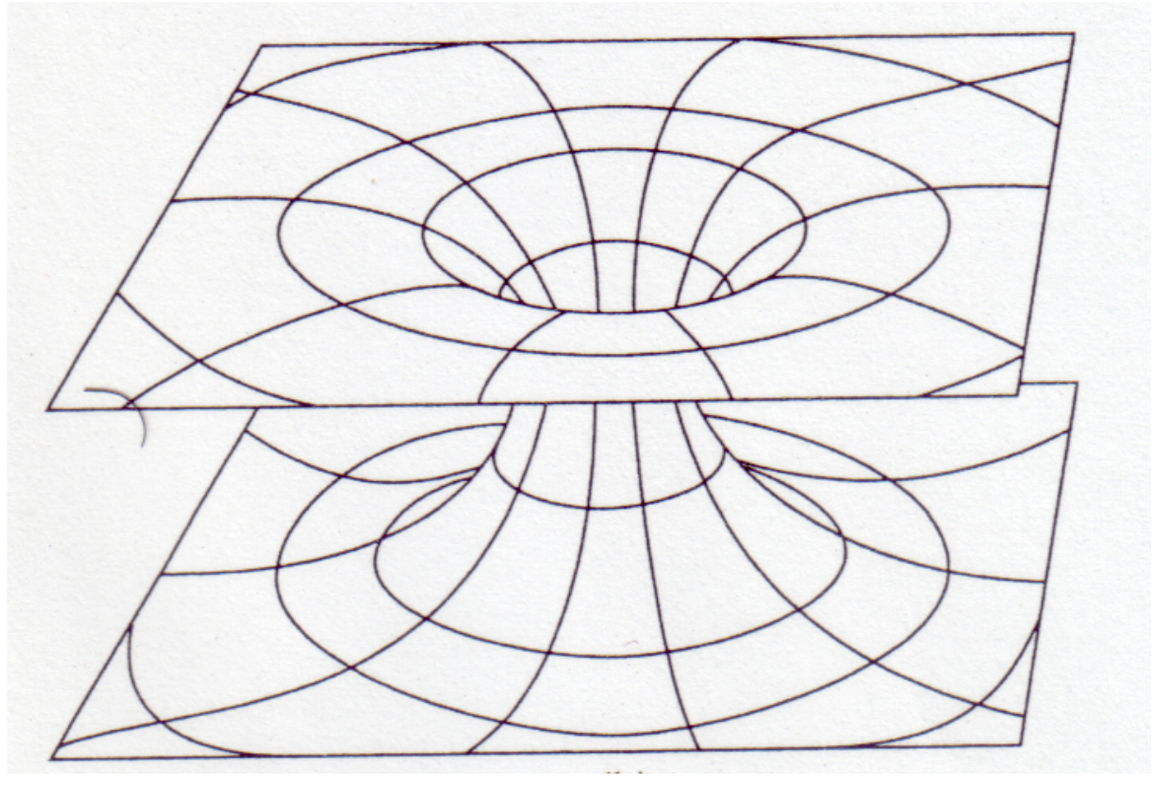
This is a sideways parabola in (r,z) plane

for given value of ϕ.

It is same for any ϕ value

so 3-dimensional representations

is as shown right:



This is called an **Einstein-Rosen bridge**.

Lesson 9 : Einstein field equations

Introduction

In this lesson, we finally arrive at Einstein field equations.

We shall introduce the continuity equation which says that certain things cannot disappear from here to reappear there without passing in between, and the energy-momentum tensor which extends the idea of distribution of masses in space and their motions.

Then we will be able to derive Einstein equations which are the analog in general relativity to Newton equations of motion in classical non-relativistic physics, in their version using the concept of field.

We are not going to get deep into solving the Einstein field equations at this time - I will solve them in deail later during our Cosmology discussions..

They are mathematically rather complicated.

Even writing them down explicitly is intricate.

That is a feature of general relativity we have already mentioned : the principles are pretty simple, but the equations are computationally nasty.

Almost everything we try to calculate gets complicated fast.

There are lots of independent Christoffel symbols, elements of the curvature tensor, derivatives, etc.

Each Christoffel symbol has a bunch of derivatives.

The curvature tensor has more derivatives.

The equations become hard to write on a single piece of paper.

The best way to solve them, or even write them, is just to feed them into your computer.

And Mathematica will spit out answers whenever it can.

As said, the basic principles are simple, but going anywhere past the basics principles tends to be computationally intensive.

The next 10 slides (pages) show a sample Mathematica code and output

In[1]:=
```
(*This is an example on using the GR
 package: the calculation of the Schwarzschild space-time with a perfect fluid stress-energy tensor*)
```

In[2]:=
```
(*Author: Giovanni Formighieri (giovform@gmail.com). Please report any bugs to this e-mail.*)
```

In[3]:=
```
(*Header package: removes all global symbols, resets assumptions (also in all parallel kernels),
sets the working directory to this notebook directory and loads the Utilities.wl package*)
Get[FileNameJoin[{NotebookDirectory[], "Header.wl"}]];
```

In[4]:=
```
(*Get the GR.wl package (the calculations are done using parallel kernels if available)*)
GetNotebookOrPackage["GR.wl"];
```

In[5]:=
```
(*Set the simplification function used in calculations: Refine, Simplify, FullSimplify or SuperFullSimplify (slow)*)
SetSimplificationFunction[FullSimplify];
```

In[6]:=
```
(*Defining symbols for shorter names and including dependencies for ρ and p on r (this is essential)*)
speedOfLightInVacuum = c;
gravitationalConstant = G;
massEnergyDensity = ρ[r];
hydrostaticPressure = p[r];
```

In[10]:=
```
(*Hide functions arguments for less output clutter on the printing functions*)
AddHiddenArgsOutputFunction[F[_]];
AddHiddenArgsOutputFunction[H[_]];
AddHiddenArgsOutputFunction[ρ[_]];
AddHiddenArgsOutputFunction[p[_]];
```

In[14]:=
```
(*Defining some assumptions on the symbols*)
AddAssumption[t ∈ Reals];
AddAssumption[r > 0];
AddAssumption[0 < θ < Pi];
AddAssumption[φ ∈ Reals];
```

In[18]:=
```
(*Defining the metric signature*)
signature = {-1, 1, 1, 1};
```

In[19]:=
```
(*Defining coordinates (notice that c and t have a space between them because they are two different symbols)*)
coordinates = {c t, r, θ, φ};
```

In[20]:= (*Defining the covariant metric tensor*)

$$\text{covMetricTensor} = \begin{pmatrix} -e^{2F[r]} & 0 & 0 & 0 \\ 0 & e^{2H[r]} & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2 \sin[\theta]^2 \end{pmatrix};$$

In[21]:= (*Initializing*)

```
Setup[coordinates, signature, covMetricTensor];
```

In[22]:= (*Calculate (this function can be called without a stress-energy tensor)*)

```
Calculate[CalculateCovariantPerfectFluidStressEnergyTensor[]];
```

In[23]:= (*Some of the symbols generated after calculation (pay attention to the pattern used in the symbols names)*)

In[24]:=
```
covMetricTensor
PrintCovariantMetricTensor[]
```

Out[24]= $\{\{-e^{2F[r]}, 0, 0, 0\}, \{0, e^{2H[r]}, 0, 0\}, \{0, 0, r^2, 0\}, \{0, 0, 0, r^2 \text{Sin}[\theta]^2\}\}$

Out[25]//TraditionalForm=

$$g_{\mu\nu} = \begin{pmatrix} -e^{2F} & 0 & 0 & 0 \\ 0 & e^{2H} & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2 \sin^2(\theta) \end{pmatrix}$$

In[26]:=
```
conMetricTensor
PrintContravariantMetricTensor[]
```

Out[26]= $\left\{\{-e^{-2F[r]}, 0, 0, 0\}, \{0, e^{-2H[r]}, 0, 0\}, \left\{0, 0, \frac{1}{r^2}, 0\right\}, \left\{0, 0, 0, \frac{\text{Csc}[\theta]^2}{r^2}\right\}\right\}$

Out[27]//TraditionalForm=

$$g^{\mu\nu} = \begin{pmatrix} -e^{-2F} & 0 & 0 & 0 \\ 0 & e^{-2H} & 0 & 0 \\ 0 & 0 & \frac{1}{r^2} & 0 \\ 0 & 0 & 0 & \frac{\csc^2(\theta)}{r^2} \end{pmatrix}$$

In[28]:=
```
covMetricTensorDeterminant
PrintCovariantMetricTensorDeterminant[]
```

Out[28]=
```
-ℯ^(2 (F[r]+H[r])) r^4 Sin[θ]^2
```

Out[29]//TraditionalForm=

$$\left|g_{\mu\nu}\right| = r^4 \left(-e^{2\,(F+H)}\right) \sin^2(\theta)$$

In[30]:=
```
covPerfectFluidStressEnergyTensor
PrintCovariantPerfectFluidStressEnergyTensor[]
```

Out[30]=
```
{{c^2 ℯ^(2 F[r]) ρ[r], 0, 0, 0}, {0, ℯ^(2 H[r]) p[r], 0, 0}, {0, 0, r^2 p[r], 0}, {0, 0, 0, r^2 p[r] Sin[θ]^2}}
```

Out[31]//TraditionalForm=

$$\mathrm{L}_{\mu\nu} = \begin{pmatrix} c^2\, e^{2F}\, \rho & 0 & 0 & 0 \\ 0 & e^{2H}\, p & 0 & 0 \\ 0 & 0 & p\, r^2 & 0 \\ 0 & 0 & 0 & p\, r^2 \sin^2(\theta) \end{pmatrix}$$

In[32]:=
```
covStressEnergyTensor
PrintCovariantMetricTensor[]
```

Out[32]=
```
{{c^2 ℯ^(2 F[r]) ρ[r], 0, 0, 0}, {0, ℯ^(2 H[r]) p[r], 0, 0}, {0, 0, r^2 p[r], 0}, {0, 0, 0, r^2 p[r] Sin[θ]^2}}
```

Out[33]//TraditionalForm=

$$\mathrm{g}_{\mu\nu} = \begin{pmatrix} -e^{2F} & 0 & 0 & 0 \\ 0 & e^{2H} & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2 \sin^2(\theta) \end{pmatrix}$$

In[34]:=
```
mixStressEnergyTensor
PrintMixedStressEnergyTensor[]
```

Out[34]=
```
{{-c^2 ρ[r], 0, 0, 0}, {0, p[r], 0, 0}, {0, 0, p[r], 0}, {0, 0, 0, p[r]}}
```

Out[35]//TraditionalForm=

$$\mathrm{T}_{\mu}{}^{\nu} = \begin{pmatrix} -c^2\, \rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}$$

In[36]:=
```
(*Some printing functions allow to set the number of display columns*)
christoffelSymbols
PrintChristoffelSymbols[3]
```

Out[36]=
$$\left\{\{\{0, F'[r], 0, 0\}, \{F'[r], 0, 0, 0\}, \{0, 0, 0, 0\}, \{0, 0, 0, 0\}\},\right.$$
$$\{\{e^{2F[r]-2H[r]} F'[r], 0, 0, 0\}, \{0, H'[r], 0, 0\}, \{0, 0, -e^{-2H[r]} r, 0\}, \{0, 0, 0, -e^{-2H[r]} r \text{Sin}[\theta]^2\}\},$$
$$\left.\{\{0, 0, 0, 0\}, \left\{0, 0, \frac{1}{r}, 0\right\}, \left\{0, \frac{1}{r}, 0, 0\right\}, \{0, 0, 0, -\text{Cos}[\theta] \text{Sin}[\theta]\}\right\},$$
$$\left.\left\{\{0, 0, 0, 0\}, \left\{0, 0, 0, \frac{1}{r}\right\}, \{0, 0, 0, \text{Cot}[\theta]\}, \left\{0, \frac{1}{r}, \text{Cot}[\theta], 0\right\}\right\}\right\}$$

Out[37]//TraditionalForm=

$$\Gamma^{t}{}_{tr} = \frac{\partial F}{\partial r} \qquad \Gamma^{t}{}_{rt} = \frac{\partial F}{\partial r} \qquad \Gamma^{r}{}_{tt} = e^{2F-2H} \frac{\partial F}{\partial r}$$

$$\Gamma^{r}{}_{rr} = \frac{\partial H}{\partial r} \qquad \Gamma^{r}{}_{\theta\theta} = -e^{-2H} r \qquad \Gamma^{r}{}_{\varphi\varphi} = -e^{-2H} r \sin^2(\theta)$$

$$\Gamma^{\theta}{}_{r\theta} = \frac{1}{r} \qquad \Gamma^{\theta}{}_{\theta r} = \frac{1}{r} \qquad \Gamma^{\theta}{}_{\varphi\varphi} = \sin(\theta)\,(-\cos(\theta))$$

$$\Gamma^{\varphi}{}_{r\varphi} = \frac{1}{r} \qquad \Gamma^{\varphi}{}_{\theta\varphi} = \cot(\theta) \qquad \Gamma^{\varphi}{}_{\varphi r} = \frac{1}{r}$$

$$\Gamma^{\varphi}{}_{\varphi\theta} = \cot(\theta)$$

In[38]:=
```
mixRiemannTensor;
PrintMixedRiemannTensor[3]
```

Out[39]//TraditionalForm=

$$\mathrm{R}^{\mathrm{t}}{}_{\mathrm{rtr}} = \frac{\partial F}{\partial r}\left(\frac{\partial H}{\partial r} - \frac{\partial F}{\partial r}\right) - \frac{\partial^2 F}{\partial r^2} \qquad \mathrm{R}^{\mathrm{t}}{}_{\mathrm{rrt}} = -\frac{\partial F}{\partial r}\frac{\partial H}{\partial r} + \frac{\partial^2 F}{\partial r^2} + \left(\frac{\partial F}{\partial r}\right)^2 \qquad \mathrm{R}^{\mathrm{t}}{}_{\theta\mathrm{t}\theta} = -e^{-2H} r \frac{\partial F}{\partial r}$$

$$\mathrm{R}^{\mathrm{t}}{}_{\theta\theta\mathrm{t}} = e^{-2H} r \frac{\partial F}{\partial r} \qquad \mathrm{R}^{\mathrm{t}}{}_{\varphi\mathrm{t}\varphi} = -e^{-2H} r \sin^2(\theta) \frac{\partial F}{\partial r} \qquad \mathrm{R}^{\mathrm{t}}{}_{\varphi\varphi\mathrm{t}} = e^{-2H} r \sin^2(\theta) \frac{\partial F}{\partial r}$$

$$\mathrm{R}^{\mathrm{r}}{}_{\mathrm{ttr}} = -e^{2F-2H}\left(-\frac{\partial F}{\partial r}\frac{\partial H}{\partial r} + \frac{\partial^2 F}{\partial r^2} + \left(\frac{\partial F}{\partial r}\right)^2\right) \qquad \mathrm{R}^{\mathrm{r}}{}_{\mathrm{trt}} = e^{2F-2H}\left(-\frac{\partial F}{\partial r}\frac{\partial H}{\partial r} + \frac{\partial^2 F}{\partial r^2} + \left(\frac{\partial F}{\partial r}\right)^2\right) \qquad \mathrm{R}^{\mathrm{r}}{}_{\theta\mathrm{r}\theta} = e^{-2H} r \frac{\partial H}{\partial r}$$

$$\mathrm{R}^{\mathrm{r}}{}_{\theta\theta\mathrm{r}} = -e^{-2H} r \frac{\partial H}{\partial r} \qquad \mathrm{R}^{\mathrm{r}}{}_{\varphi\mathrm{r}\varphi} = e^{-2H} r \sin^2(\theta) \frac{\partial H}{\partial r} \qquad \mathrm{R}^{\mathrm{r}}{}_{\varphi\varphi\mathrm{r}} = -e^{-2H} r \sin^2(\theta) \frac{\partial H}{\partial r}$$

$$\mathrm{R}^{\theta}{}_{\mathrm{tt}\theta} = -\frac{e^{2F-2H}\frac{\partial F}{\partial r}}{r} \qquad \mathrm{R}^{\theta}{}_{\mathrm{t}\theta\mathrm{t}} = \frac{e^{2F-2H}\frac{\partial F}{\partial r}}{r} \qquad \mathrm{R}^{\theta}{}_{\mathrm{rr}\theta} = -\frac{\frac{\partial H}{\partial r}}{r}$$

$$\mathrm{R}^{\theta}{}_{\mathrm{r}\theta\mathrm{r}} = \frac{\frac{\partial H}{\partial r}}{r} \qquad \mathrm{R}^{\theta}{}_{\varphi\theta\varphi} = \left(1 - e^{-2H}\right)\sin^2(\theta) \qquad \mathrm{R}^{\theta}{}_{\varphi\varphi\theta} = \left(e^{-2H} - 1\right)\sin^2(\theta)$$

$$\mathrm{R}^{\varphi}{}_{\mathrm{tt}\varphi} = -\frac{e^{2F-2H}\frac{\partial F}{\partial r}}{r} \qquad \mathrm{R}^{\varphi}{}_{\mathrm{t}\varphi\mathrm{t}} = \frac{e^{2F-2H}\frac{\partial F}{\partial r}}{r} \qquad \mathrm{R}^{\varphi}{}_{\mathrm{rr}\varphi} = -\frac{\frac{\partial H}{\partial r}}{r}$$

$$\mathrm{R}^{\varphi}{}_{\mathrm{r}\varphi\mathrm{r}} = \frac{\frac{\partial H}{\partial r}}{r} \qquad \mathrm{R}^{\varphi}{}_{\theta\theta\varphi} = e^{-2H} - 1 \qquad \mathrm{R}^{\varphi}{}_{\theta\varphi\theta} = 1 - e^{-2H}$$

In[40]:=
```
covEinsteinTensor
PrintCovariantEinsteinTensor[]
```

Out[40]=
$$\left\{\left\{\frac{e^{2\,F[r]-2\,H[r]}\left(-1+e^{2\,H[r]}+2\,r\,H'[r]\right)}{r^2},\,0,\,0,\,0\right\},\left\{0,\,\frac{1-e^{2\,H[r]}+2\,r\,F'[r]}{r^2},\,0,\,0\right\},\right.$$
$$\left.\left\{0,\,0,\,e^{-2\,H[r]}\,r\,((1+r\,F'[r])\,(F'[r]-H'[r])+r\,F''[r]),\,0\right\},\left\{0,\,0,\,0,\,e^{-2\,H[r]}\,r\,\text{Sin}[\theta]^2\,((1+r\,F'[r])\,(F'[r]-H'[r])+r\,F''[r])\right\}\right\}$$

Out[41]//TraditionalForm=

$$G_{tt}=\frac{e^{2F-2H}\left(2r\frac{\partial H}{\partial r}+e^{2H}-1\right)}{r^2}$$

$$G_{rr}=\frac{2r\frac{\partial F}{\partial r}-e^{2H}+1}{r^2}$$

$$G_{\theta\theta}=e^{-2H}r\left(\left(r\frac{\partial F}{\partial r}+1\right)\left(\frac{\partial F}{\partial r}-\frac{\partial H}{\partial r}\right)+r\frac{\partial^2 F}{\partial r^2}\right)$$

$$G_{\varphi\varphi}=e^{-2H}r\sin^2(\theta)\left(\left(r\frac{\partial F}{\partial r}+1\right)\left(\frac{\partial F}{\partial r}-\frac{\partial H}{\partial r}\right)+r\frac{\partial^2 F}{\partial r^2}\right)$$

In[42]:=
```
mixEinsteinTensor
PrintMixedEinsteinTensor[]
```

Out[42]=
$$\left\{\left\{-\frac{e^{-2H[r]}\left(-1+e^{2H[r]}+2\,r\,H'[r]\right)}{r^2},\,0,\,0,\,0\right\},\left\{0,\,\frac{-1+e^{-2H[r]}\,(1+2\,r\,F'[r])}{r^2},\,0,\,0\right\},\right.$$
$$\left.\left\{0,\,0,\,\frac{e^{-2H[r]}\,((1+r\,F'[r])\,(F'[r]-H'[r])+r\,F''[r])}{r},\,0\right\},\left\{0,\,0,\,0,\,\frac{e^{-2H[r]}\,((1+r\,F'[r])\,(F'[r]-H'[r])+r\,F''[r])}{r}\right\}\right\}$$

Out[43]//TraditionalForm=

$$G_t{}^t=-\frac{e^{-2H}\left(2r\frac{\partial H}{\partial r}+e^{2H}-1\right)}{r^2}$$

$$G_r{}^r=\frac{e^{-2H}\left(2r\frac{\partial F}{\partial r}+1\right)-1}{r^2}$$

$$G_\theta{}^\theta=\frac{e^{-2H}\left(\left(r\frac{\partial F}{\partial r}+1\right)\left(\frac{\partial F}{\partial r}-\frac{\partial H}{\partial r}\right)+r\frac{\partial^2 F}{\partial r^2}\right)}{r}$$

$$G_\varphi{}^\varphi=\frac{e^{-2H}\left(\left(r\frac{\partial F}{\partial r}+1\right)\left(\frac{\partial F}{\partial r}-\frac{\partial H}{\partial r}\right)+r\frac{\partial^2 F}{\partial r^2}\right)}{r}$$

In[44]:=
```
covFieldEquations
PrintCovariantFieldEquations[]
```

Out[44]=
$$\left\{\left\{\frac{e^{2F[r]-2H[r]}\left(-1+e^{2H[r]}+2\,r\,H'[r]\right)}{r^2}==\frac{8\,e^{2F[r]}\,G\,\pi\,\rho[r]}{c^2},\text{True},\text{True},\text{True}\right\},\right.$$
$$\left\{\text{True},\frac{1-e^{2H[r]}+2\,r\,F'[r]}{r^2}==\frac{8\,e^{2H[r]}\,G\,\pi\,p[r]}{c^4},\text{True},\text{True}\right\},$$
$$\left\{\text{True},\text{True},e^{-2H[r]}\,r\,((1+r\,F'[r])\,(F'[r]-H'[r])+r\,F''[r])==\frac{8\,G\,\pi\,r^2\,p[r]}{c^4},\text{True}\right\},$$
$$\left.\left\{\text{True},\text{True},\text{True},e^{-2H[r]}\,r\,\text{Sin}[\theta]^2\,((1+r\,F'[r])\,(F'[r]-H'[r])+r\,F''[r])==\frac{8\,G\,\pi\,r^2\,p[r]\,\text{Sin}[\theta]^2}{c^4}\right\}\right\}$$

Out[45]//TraditionalForm=

$$(\text{tt})\Rightarrow\frac{e^{2F-2H}\left(2r\frac{\partial H}{\partial r}+e^{2H}-1\right)}{r^2}=\frac{8\pi e^{2F}G\rho}{c^2}$$

$$(\text{rr})\Rightarrow\frac{2r\frac{\partial F}{\partial r}-e^{2H}+1}{r^2}=\frac{8\pi G e^{2H}p}{c^4}$$

$$(\theta\theta)\,(\varphi\varphi)\Rightarrow e^{-2H}r\left(\left(r\frac{\partial F}{\partial r}+1\right)\left(\frac{\partial F}{\partial r}-\frac{\partial H}{\partial r}\right)+r\frac{\partial^2 F}{\partial r^2}\right)=\frac{8\pi G p r^2}{c^4}$$

In[46]:=
```
CovariantMetricTensorToLineElement[covMetricTensor]
```

Out[46]=
$$-c^2\,dt^2\,e^{2F[r]}+dr^2\,e^{2H[r]}+d\theta^2\,r^2+d\varphi^2\,r^2\,\text{Sin}[\theta]^2$$

In[47]:=
```
PrintLineElement[]
```

Out[47]//TraditionalForm=

$$ds^2=-c^2\,dt^2\,e^{2F}+d\theta^2\,r^2+dr^2\,e^{2H}+d\varphi^2\,r^2\sin^2(\theta)$$

In[48]:=
```
CovariantToMixed[covPerfectFluidStressEnergyTensor]
PrintExpression[]
PrintMixedStressEnergyTensor[]
```

Out[48]= $\{\{-c^2\,\rho[r],\ 0,\ 0,\ 0\},\ \{0,\ p[r],\ 0,\ 0\},\ \{0,\ 0,\ p[r],\ 0\},\ \{0,\ 0,\ 0,\ p[r]\}\}$

Out[49]//TraditionalForm=

$$\begin{pmatrix} -c^2\,\rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}$$

Out[50]//TraditionalForm=

$$\mathrm{T}_{\mu}{}^{\nu} = \begin{pmatrix} -c^2\,\rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}$$

In[51]:=
```
(*ThetaTheta and PhiPhi field equations are the same in this case*)
PossibleSameEquationsQ[covFieldEquations[[3, 3]], covFieldEquations[[4, 4]]]
```

Out[51]= True

In[52]:=
```
(*Use print expression for easy reading output*)
covFieldEquations[[3, 3]]
PrintExpression[covFieldEquations[[3, 3]]]
```

Out[52]= $e^{-2\,H[r]}\,r\,((1 + r\,F'[r])\,(F'[r] - H'[r]) + r\,F''[r]) == \dfrac{8\,G\,\pi\,r^2\,p[r]}{c^4}$

Out[53]//TraditionalForm=

$$e^{-2H}\,r\left(\left(r\,\frac{\partial F}{\partial r} + 1\right)\left(\frac{\partial F}{\partial r} - \frac{\partial H}{\partial r}\right) + r\,\frac{\partial^2 F}{\partial r^2}\right) = \frac{8\,\pi\,G\,p\,r^2}{c^4}$$

In[54]:=
```
(*Then you can just go on using all these quantities*)
```

In[55]:=
```
eq1 = Solve[covFieldEquations[[2, 2]], p[r]]
```

Out[55]= $\left\{\left\{p[r] \to -\dfrac{c^4\,e^{-2\,H[r]}\,\left(-1 + e^{2\,H[r]} - 2\,r\,F'[r]\right)}{8\,G\,\pi\,r^2}\right\}\right\}$

```
In[56]:= (*S is the simplification function configured in the beginning of this notebook with all the assumptions
          automatically included*)
         S[covFieldEquations[[3, 3]] /. eq1][[1]]
         PrintExpression[]
```

Out[56]= $e^{2\,H[r]} + r^2\left(F'[r]^2 + F''[r]\right) == (1 + r\,F'[r])\,(1 + r\,H'[r])$

Out[57]//TraditionalForm=

$$r^2\left(\frac{\partial^2 F}{\partial r^2} + \left(\frac{\partial F}{\partial r}\right)^2\right) + e^{2H} = \left(r\frac{\partial F}{\partial r} + 1\right)\left(r\frac{\partial H}{\partial r} + 1\right)$$

So we won't do much computation.

We will concentrate on the meaning of the symbols.

And then we will see what happens when we try to solve them in various circumstances.

In the next and final lesson, we will do a little bit of solving when we talk about gravitational waves.

But in the present lesson, the topic is not gravitational waves, it is the fundamental equations of general relativity : Einstein field equations.

Newtonian gravitational field

Before we talk about Einstein field equations, we should talk first about the corresponding Newtonian concepts.

Newton didn't think in terms of fields.

He didn't have a concept of field equations.

Nevertheless, in classical non-relativistic physics, there are field equations which are equivalent to Newton's equations of motion.

Field equations of motion are always a sort of two-way street.

Masses affect the gravitational field and the gravitational field affects the way masses move.

So let's talk about the two way street in the context of Newton.

First of all, a field affects particles.

That is just a statement that a gravitational force F can be written as minus a mass m times the gradient of the gravitational potential.

The gravitational potential is usually denoted as ϕ which is a function of position x.

So we have

$$F = -m \nabla\phi(x) \tag{1}$$

This equation means that everywhere in space, due to whatever reason, there is a gravitational potential $\phi(x)$ which varies from place to place.

You take the gradient of $\phi(x)$.

You multiply it by m, the mass of the object whose motion you want to figure out, and put a minus sign.

And that tells you the force on the object.

Equation (1) can be one dimensional, in which case the gradient is just the ordinary derivative of the function ϕ with respect to the independent variable x.

And equation (1) equates two scalars.

Or equation (1) can be multi-dimensional.

It is the case when x, is a vector position, which we could then denote $X = (x, y, z)$.

The sign ∇, which reads "gradient" or "del", is then the vector of partial derivatives

$$\nabla\phi(X) = \left(\frac{\partial\phi}{\partial x}, \frac{\partial\phi}{\partial y}, \frac{\partial\phi}{\partial z}\right)$$

In that case, of course, the force F is also a vector.

For simplicity, however, we will keep with the notation x, be it one- or multi-dimensional.

Same for F and for the acceleration introduced below.

So equation (1) is one aspect of field $\phi(x)$.

It tells particles how to move.

In the case which occupies us, it does that by telling them what their acceleration should be.

This comes from Newton's equation linking the force F exerted on the particle to its acceleration a :

$$F = ma \tag{2}$$

Combining equations (1) and (2), the mass m cancels.

And we obtain

$$a = -\nabla\phi(x) \tag{3}$$

That is one direction of the two way street.

It is how a field tells a particle how to move.

On the other hand – there is the other direction – masses in space tell the gravitational field what to be.

The equation that tells the gravitational field what to be is the Poisson equation.

It says that the second derivative of ϕ with respect to space (or the sum of second partial derivatives) is related to the distribution of masses in space as follows

$$\nabla^2\phi = 4\pi G\ \rho(x,\ t) \tag{4}$$

∇^2, which is a notation for the formal dot product $\nabla \cdot \nabla$, is called the *laplacian*, and has its own notation : it is sometimes denoted $\triangle$.

Here is how to read equation (4) :

a) As said, if we are in several dimensions

$$\nabla^2\phi = \frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2}$$

b) The multiplicative factor 4π is essentially a convention. It originates from the geometry of the sphere, because we often deal with a field that is invariant under rotation, therefore spherical.

c) G is Newton gravitational constant. It is equal to

$$6.674 \text{ x } 10^{-11} Nm^2/kg^2$$

d) Finally, ρ is the *density of mass* at each position x, or $(x,\ y,\ z)$. And it can depend also on time because masses can move around.

That is how the distribution of masses in space determines the gravitational field.

So we have these two aspects : the field tells particles how to move, and masses, particles in other words, tell the field how to be – how to curve, we shall see.

We can easily solve equation (4) in the special case where the distribution of masses (or the density) ϱ is simply all the mass concentrated at one point.

The mass in any small volume around a point x being then the local density of mass $\varrho(x)$ multiplied by the small volume of space dx or $dx\,dy\,dz$.

In general, the density ϱ is by definition the amount of mass per unit volume.

In the case where $\varrho(x)$ is "concentrated", it is actually a Dirac distribution – we met these "extended" mathematical functions as “delta” functions..

We manipulate them essentially through their integral over some region.

At first, we may be considering a real star, or a planet or a bowling ball of mass M, see figure 1.

The only constraint is that it must be a spherically symmetric object of total mass M.

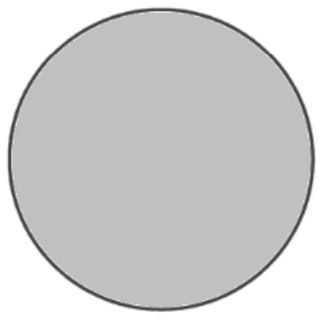


Figure 1 : Spherically symmetric object of mass M, equivalent to a point mass when we are outside its surface.

It does not matter whether the mass is uniformly distributed inside the volume.

It just has to be symmetric with respect to rotation.

There could be layers with more density near the center than near the outer surface.

The solution to equation (4) then is

$$\phi(r) = -\frac{MG}{r} \tag{5}$$

where we replaced the spatial coordinate x by a radius r.

If we consider that the body is a point mass, then equation (5) is the valid solution to equation (4) everywhere except at the central point (where ϕ is not defined because $-MG/r$ tends to infinity).

Let's pursue this example where ϕ is given by equation (5).

Taking its gradient just puts another r downstairs, while changing the sign.

Indeed, differentiating $1/r$ with respect to the radius, which is what ∇ does, produces $-1/r^2$.

We wind up with

$$F = \frac{mMG}{r^2} \tag{6}$$

This is a kind of *primitive* field theoretic way to think about gravitation.

Instead of action at a distance we have a gravitational field.

In truth it still implies action at a distance, because in Newton's theory if you move around a mass, ϕ *instantly* reacts to it and changes, from equation (4).

But it is a way of writing the theory, making it *look* like a field theory.

Equations (3) and (4), corresponding to "field → motion of mass" and "mass→ structure of field", is what we want to *replace* by something making sense in general relativity.

Just to get a little bit of a handle on what we are going to do in this chapter, let's remember something about the Schwarzschild geometry.

In lesson 5, we pulled the Schwarzschild geometry out of a hat.

Of course the point was that it is the solution to Einstein's equations in the case of a central point mass field.

But we did not have the tools to derive it.

Equations (3) and (4) are not Einstein's equations, they are Newton's equations.

And their solution, in the simple case of a central spherically symmetric field, is $\phi(r) = -MG/r$.

So we wrote down the Schwarzschild metric, see equation (32) of lesson 5.

We are not going to write the whole shebang again.

But let's examine what g_{00} was.

$$g_{00} = 1 - \frac{2MG}{r} \tag{7}$$

As usual, we have set everywhere the speed of light c equal to 1.

Our aim presently is to *guess* some correspondence between g_{00} given by equation (7), and anything in the Newton framework.

Using equation (5), the above equation (7) can be rewritten as $g_{00} = 1 + 2\phi$.

Therefore $\nabla^2 g_{00}$ is just twice $\nabla^2 \phi$.

This leads to

$$\nabla^2 g_{00} = 8\pi G\rho \tag{8}$$

This formula should be taken with a grain of salt.

It is just a mnemonic device to remember the relationship between some aspects of general relativity and matter.

Nevertheless, interestingly enough, it already suggests that matter or mass is affecting geometry.

When we make this correspondence between Newton ϕ and Einstein or Schwarzschild metric, we see roughly that matter is telling geometry how to curve so to speak.

We use the word roughly because we are going to be more precise.

Of course equation (8) is not really the part "mass → field" of Einstein's equations. In Einstein's equations, it is good deal more complicated than that.

And what about the other part "field → motion of mass"?

In Newton's framework we saw equations (1) or (2) which specify how the field dictates the motion of mass.

In general relativity, the equation $\ddot{x} = -\nabla\phi(x)$ is replaced by the statement that once we know the geometry, that is, once we know g_{00}, the rule is :

particles move on space-time geodesics.

And for the other direction, we saw that the Newtonian field equation, $\nabla^2\phi = 4\pi G\varrho$, is replaced by something which we naively wrote as $\nabla^2 g_{00} = 8\pi G\varrho$.

We are going to do better.

We have to figure out exactly how the mass distribution affects the field.

Well, a little more than a century ago, Einstein figured out exactly what goes there.

Before we write down the field equations, we need to understand more about the right hand side $4\pi G\varrho$ of Newton field equations, or $8\pi G\varrho$ of Einstein field equations.

We need to know more about the density of mass.

Mass really means energy.

Indeed, we saw in special relativity that $E = mc^2$.

If we forget about c, that is, if we set it equal to one, then energy and mass are the same thing.

Therefore what goes on the right hand side of the field equations is really energy density.

We need to understand better what kind of quantity energy density is in relativity.

It is part of a complex of things which includes more than just the energy density.

It is part of a complex.

In other words it is part of some kind of tensor whose other components have other meanings.

So let's go back and review quickly a little bit about the notion of conservation.

In the case that we are concerned with, it will be conservation of energy and momentum.

But we shall begin with a simpler case : conservation of charge.

Continuity equation

Conservation, densities, flows of things like charge and mass - let's review briefly these concepts.

Let's start with electric charge.

It is simpler than energy for reasons we will come to.

The total electric charge of a system is called Q.

It is the standard notation for the electric charge.

In many situations the electric charge density is called ϱ, but in order not to create confusion with the mass density or the energy density, which we have already called ϱ, we shall denote the charge density with the letter σ.

What is electric charge density?

Consider a small volume of space, a differential volume of space at a given point, take the electric charge in that volume, divide it by the volume.

That gives you the electric charge density at that point.

The density can be schematically written

$$\sigma = \frac{Q}{\text{Volume}} \tag{9}$$

It has units of charge divided by volume.

Let's look at an infinitesimal volume, figure 2.

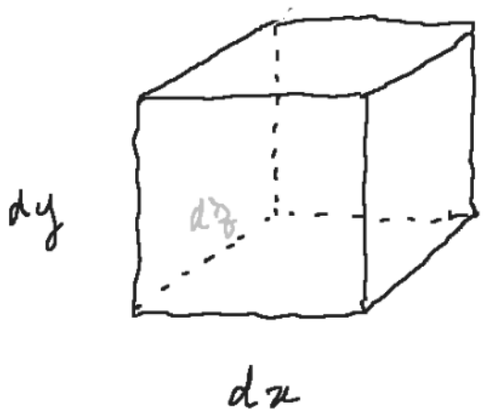


Figure 2 : Infinitesimal volume.

If we look at electric charges moving around in space, between t_1 and t_2, we see some entering the volume $dxdydz$, some leaving, some passing by, some passing through, etc., figure 3.

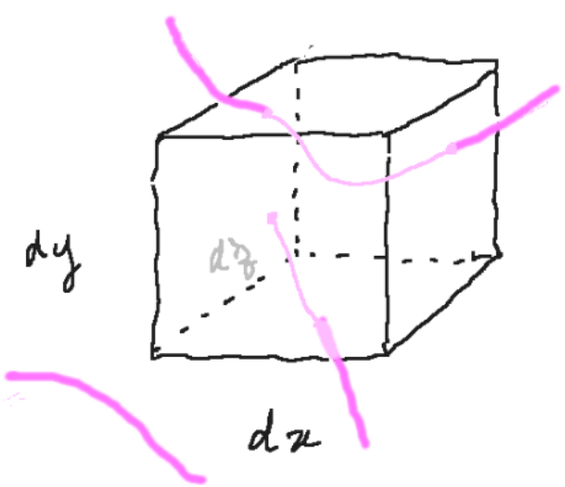


Figure 3 : Trajectory in space, between t_1 and t_2, of various moving electric charges.

The current in direction x is defined as the quantity of charge passing through the side *dydz* per unit time.

It is denoted J^{x}.

There are analogous definitions and notations for *y* and *z*.

Then, similarly to equation (9), notice that each current along one spatial dimension can be schematically written

$$J^m = \frac{Q}{\text{Area Time}} \tag{10}$$

But in relativity we have four dimensions more or less on an equal footing.

Time and space are nice and symmetric to each other.

So dividing by area times time is again dividing by three dimensions.

One happens to be time-like, two happen to be space-like.

Both equations (9) and (10) can be thought of as a charge divided by a three-dimensional volume.

In the case of equation (9) is it a pure spatial volume.

In the case of equation (10) it is a mixed space and time three-dimensional thing.

The combined density σ plus the three currents J^x, J^y, and J^z happen to form a 4-vector in the sense of relativity.

That is, in a change of coordinates they change according to the equations of a change of frame of reference

Following the standard notation, instead of (σ, J^x, J^y, J^z) we simply write

$$J^{\mu} \tag{11}$$

where the Greek index μ runs from 0 to 3.

J^0 is the electric charge density σ.

And the other three elements are simply the components of the current indexed with 1, 2 and 3 instead of x, y and z.

Now we arrive at an important physical fact :

the conservation of charge is a local property.

We haven't yet talked about the familiar relationship between the evolution of the charge Q inside the volume $dxdydz$ and the current going through its sides, because we want to stress that in theory, conservation could hold in a system without being local.

The electric charge carried by an object in a system, could at time t disappear from the object and *instantaneously* reappear elsewhere in another distant part of the system.

If the system we are considering is the entire universe, the charge could disappear from my pen and reappear in Alpha Centauri.

I always use Alpha Centauri as someplace which is so far away that it doesn't matter.

If that were possible, conservation would still hold.

You would say rightly : well charge is conserved.

But I would retort : who cares if charge is conserved, if it can just disappear to some arbitrarily distant place.

It is just as good as saying it wasn't conserved.

In the laboratory, however, charge doesn't disappear that way.

If it leaves the laboratory, it goes through the walls, or the windows, or whatever, or simply through the door.

In other words, the electric charge cannot change in a given volume without some current flowing through the boundary.

That idea is called *continuity*.

And there is an equation that goes with it.

The equation is the *continuity equation*.

If we look at a box of volume one in some units, the charge inside the box is σ, that is the charge density times one and the charge leaving the box over a unit time is $-\dot{\sigma}$.

In other words, it is minus the time derivative of σ.

Why minus?

Because it is leaving the box.

That has to be equal to the sum of the currents passing through the box.

A basic theorem of multivariate calculus, called the divergence theorem, also known as Gauss's theorem or Ostrogradsky's theorem, states that the diminution of charge inside the box is equal to the divergence of the current.

$$-\dot{\sigma} = \nabla . J \tag{12}$$

The formal dot product $\nabla \cdot J$ on the right hand side means

$$\frac{\partial J^x}{\partial x} + \frac{\partial J^y}{\partial y} + \frac{\partial J^z}{\partial z}$$

Remembering that t is one of the four components of space-time in relativity, this can be rewritten more nicely, first of all as

$$\frac{\partial \sigma}{\partial t} + \frac{\partial J^x}{\partial x} + \frac{\partial J^y}{\partial y} + \frac{\partial J^z}{\partial z} = 0 \tag{13}$$

Then, since J^0 is by definition σ, see expression (11), using the summation convention we can write this even more nicely

$$\frac{\partial J^\mu}{\partial X^\mu} = 0 \tag{14}$$

So the divergence theorem has lead us to a nice tensor-type equation satisfied by J^μ.

J^{μ} is 4-vector.

The X^{μ}'s are the four components of spacetime.

Therefore equation (14) has the valid look of a good equation involving the derivative of a tensor with respect to position, and contracting(making 2 indices identical) it.

Since equation (14) is true in any frame (because it is a tensor equation), it expresses a conservation law.

In general, an expression like equation (14) would be correct only in flat space.

In curved coordinates, we might have to replace the ordinary derivative by the covariant derivative of the tensor.

$$\frac{DJ^{\mu}}{DX^{\mu}} = 0 \tag{15}$$

Remember covariant derivative of tensors which we studied in lesson 3.

It turns out in the case of charge current it doesn't matter.

But in general it would matter.

When you go to curved coordinates, you should replace all derivatives by covariant derivatives, otherwise the equations are not good tensor equations.

Why do we want tensor equations?

We want tensor equations because we want them to be true in any set of coordinates.

That is the theory of electric charge flow of current and the continuity equation.

Equation (15) is called the continuity equation.

And the physics of it is that when charge either appears in or disappears from a volume, it is always traceable to current flowing into or out through the boundaries of that volume.

Now let's come to energy and momentum.

Energy-momentum tensor

Energy and momentum are also conserved quantities.

They can be described in terms of density of energy and density of momentum, that is, density of each component of momentum because it is a vector quantity.

We can ask how much energy there is in the form of particles or whatever it happens to be, including the mc^2 part of energy.

We can ask how much energy is in a volume.

We can ask how much momentum is in a volume.

Just look at all the particles within a volume and count their momentum.

Photons, or electromagnetic radiations, have both energy and momentum.

And that energy and momentum can be regarded as the integral of a density.

So, in that sense, each one of them, each energy and each component of the momentum, is like the electric charge.

They are conserved.

They can flow.

If an object is moving, the energy and momentum may be flowing.

The question is : how do we represent the same set of ideas for energy and momentum as we did for charge ?

Now there is a difference between charge and energy and momentum.

Electric charge is an invariant.

No matter how the charge is moving, the charge of an electron is always the same.

It does not depend on its state of motion.

Therefore charge itself is an invariant.

The density of charge and the current of charge, however, are not invariants.

For example, suppose we have a given charge, taking a certain volume of space, and I decide to look at it in a different frame of reference than you.

I walk by it with a certain velocity.

Because of Lorentz contraction I say that the volume of that charge is one thing.

You, sitting still relative to the charge, assign a different volume to it.

Since the charge itself does not change, we will not agree about the value of the charge density.

That's okay : charge density is a component of a 4-vector.

Remember 4-vectors.

Depending on the reference frame, their components change according to tensor equations of change of frame.

But if the 4-vector is zero in one frame, it is zero in every frame.

Similarly, I see charges moving, and you see them at rest.

You say there is no current.

I'm moving and I see a wind of charges passing me by.

I say there is current.

We are both right of course.

Charge density and charge current are not invariant.

They form together a 4-vector.

Energy and momentum are more complicated.

The total energy and momentum, not the density of them but the total energy and momentum, are not invariant.

I see a particle standing still – I'm talking about the whole particle not the density.

I say there is some energy of a certain magnitude.

You are walking past it and you see not just the $E = mc^2$ part of the energy, but you also see kinetic energy of motion.

You see more energy in the particle or object not because of any Lorentz contraction of the volume it is in, but because the same object when you look at it has more energy than when I look at it.

The same is true of the momentum, not the flow of it, not the density of it, but the momentum itself.

It is also frame dependent.

You see an object in motion, you say there is momentum there.

I see the same object at rest, I say there is no momentum.

So energy and momentum, unlike charge, are not invariant.

They together form the components of a 4-vector.

$$P^{\mu} = (E,\ P^{m}) \tag{16}$$

E is the energy.

It is also P^0.

And the P^m are the components of momentum, where the Roman index m labels the directions of space.

So zero is the index for energy, and the others for momentum.

Each of the four components of P^{μ} is like a charge Q.

In other words, it is a conserved quantity in the sense of invariance of 4-vectors : if two 4-vectors are equal in one frame, they are equal in every frame.

In equations (14) or (15), density was the zeroth component of J^{μ}.

So zero is also the index for density, and the others for flow.

Let's return to the energy in P^{μ}.

It is P^0.

And consider the density of energy.

In other words, how much energy is in a small volume.

We are going to denote this

$$T^{00} \tag{17}$$

The first index 0 is referring to energy as opposed to momentum, and the second 0 to the fact that we are considering a density.

It is a function of position, so we will sometimes write

$$T^{00}(X) \tag{18}$$

where X stands for (X^0, X^1, X^2, X^3).

Energy can also move.

Like charge it can disappear from a region, and like charge too it does so by crossing the boundary of the region.

In other words, the continuity equation applies to energy density and energy flow, as it applied to charge density and electric current

The amount of energy flowing through a unit surface, along the X^1 direction, per unit time – the "current of energy" you might call it – is denoted

$$T^{01}(X) \tag{19}$$

the first index 0 referring to an energy, and the second index 1 to a flow along the first spatial axis.

Likewise there is a T^{02} and a T^{03}.

In summary, the three components (T^{01}, T^{02}, T^{03}) form the flow of energy.

And T^{00} is the density of energy.

The continuity equation for energy is derived in exactly the same way as we did for charge.

It involves now the components $T^{0\nu}$.

And what does it say?

For the moment the first index 0 is just passive.

It signals that are talking about energy. .

It is the second index ν running over 0, 1, 2 and 3 which is like the index μ of J^{μ} in equation (15).

Since ν is a dummy index, we may as well use μ.

The *continuity equation for energy* then is

$$\frac{DT^{0\mu}}{DX^{\mu}} = 0 \tag{20}$$

Next step : everything that we said about energy we could now say about any one of the components of the momentum.

So let's go to the momentum components of the 4-vector P^{μ}.

It is the last three components, customarily indexed with the Roman letter *m* running from 1 to 3.

The component P^{m} also has a density, denoted T^{m0}.

The naught indicates that it is a density.

The *m* indicates that we are talking about the *m*-th component of momentum.

It is also a function of *X*.

Likewise we can consider the flow of the *m*-th component of momentum, along *n*-th direction of *X*.

It is naturally denoted T^{mn}.

Momentum is a conserved quantity, in the sense that, with the energy, the three momenta form a 4-vector.

Momentum can flow.

Each of its components can flow along some direction.

So we can go a little further.

We can say the same equation as equation (20) is true even if we replace energy by a component of momentum.

In other words we could replace 0 by n.

But then we have the same equation for all four possibilities of the first index.

So, switching μ and ν again for cosmetic reasons, we can finally write

$$\frac{DT^{\mu\nu}}{DX^{\nu}} = 0 \tag{21}$$

We can now step back and see where we arrived.

The densities and flows of energy and momentum form a tensor with two indices.

The first index tells us what were talking about, energy or momentum.

The other index tells us whether we are talking about density or flow.

That is called the energy-momentum tensor.

The energy-momentum tensor has an interesting property which we have not proved.

Take for example T^{m0}, that is, the density of the m-th component of momentum.

Compare it with T^{0m}, that is, the flow of energy in the m-th direction.

It is a general property of relativistic systems, which we are not going to prove now, that the matrix formed by the components of the energy-momentum tensor is symmetric.

In other words, $T^{m0} = T^{0m}$, and more generally $T^{\mu\nu} = T^{\nu\mu}$.

It takes a little bit of work to prove it.

Relativistic invariance allows you to connect T^{m0} with T^{0m}.

It is a theorem of all relativistic field theories, that the energy-momentum tensor is symmetric.

So let's take this into account and write the energy-momentum tensor in matrix form

$$T^{\mu\nu} = \begin{pmatrix} T^{00} & T^{01} & T^{02} & T^{03} \\ T^{01} & T^{11} & T^{12} & T^{13} \\ T^{02} & T^{12} & T^{22} & T^{23} \\ T^{03} & T^{13} & T^{23} & T^{33} \end{pmatrix} \tag{22}$$

We will come back to the meaning of these elements in a little while.

T^{00} is clear : it is energy density.

The components (T^{01}, T^{02}, T^{03}) are fairly clear : they are flow of energy.

T^{10} (which is the same as T^{01}) is momentum density.

Then we have flow of momentum.

So their meaning is pretty clear.

But we will find out that some of the elements of $T^{\mu\nu}$ have other meanings, connected with pressure, things of that nature.

The important idea is that the flow and density of energy and momentum are combined into an energy-momentum tensor.

And this energy-momentum tensor satisfies four continuity equations, one for each type of the stuff that we are talking about (energy, that is P^0, or momentum along one of the three spatial axes, that is P^m).

What we have learned it that the notion of energy density, ϱ, which we tried to relate to the metric in equation (8), reproduced below,

$$\nabla^2 g_{00} = 8\pi G\rho \tag{23}$$

is incomplete.

It is part of a complex of things.

It is part of a tensor of course.

This is a fundamental observation : the right hand side of equation (23) is part of a tensor, because ϱ is T^{00}.

So the left hand side must also be part of a tensor.

In physics an equation which says that some particular component of a tensor is equal to the same component of some other tensor is usually meaningless or wrong – unless its full version say that the two tensors are equal.

We will understand this with vectors, which are a simple type of tensor.

Suppose we have two 3D spatial vectors A and B.

Suppose we assert that there is a law of physics which says that A_3 is equal to B_3, the index 3 being the z direction.

Does this make sense as a law of physics ?

Well, it only makes sense as a law of physics if it is also true that $A_2 = B_2$ and $A_1 = B_1$.

Why is that?

Because if it is a law of physics, it means that it is true in every reference frame.

But we can always rotate the coordinates so that the first axis in the old frame becomes the third axis in the new frame, so A_1 must also be equal to B_1, etc.

This is an illustration of the general idea that vectors and tensors are geometric entities which exists independently of any consideration of coordinates or components.

"Go five steps over there", if I show you a direction while telling you that, is meaningful without referring to coordinates.

Of course, with appropriate coordinates, it may become "go three steps to the east, and four steps to the south".

But the coordinates are frame dependent.

My injunction may have first coordinate 3 in one frame and first coordinate 2 in another.

When we go to relativity, the same considerations are true for 4-vectors, including the time component, because it can be transformed into a space component through a Lorentz transformation.

And they are also true for any two tensors of the same type.

So tensor equations, to be good laws of physics, must equate all the components of two tensors, not just two components of them.

And if we have an equation, which we believe should reflect a law of physics, which equates two components of tensors, then it should be true for all the components of the two tensors.

Equation (23) is a formula which equates the zero-zero component of the energy-momentum tensor with something else.

If it is to be a law of physics, we should figure out which tensors it actually equates.

Let's not worry too much about whether the left hand side is correct or not.

We just made a guess of what the left hand side might look like, and found something related to the metric, that is, to the geometry.

But the right hand side is the energy density, and it is what you would expect to be on the right hand side of Newton's equations

Therefore the right hand side of Einstein's equations must involve not a particular component of a tensor, but it must generalize to something that involves all components of a tensor.

That means Einstein's generalization of Newton must be as follows.

The right hand side ought to be

$$\ldots = 8\pi G T^{\mu\nu}$$

A special case of it is when both μ and ν are time.

Then it becomes the energy density.

But if the equation that we are constructing is to be true in every frame, it has to be a tensor equation involving all the μ ν components.

What has to be on the left hand side?

It must also be a rank 2 tensor.

Otherwise the equation would not make sense.

It must be symmetric because the right hand side is symmetric.

It must have whatever other properties the right hand side has.

But the left hand side will not be something made up out of matter.

It will be made up out of the metric.

It has to do with geometry and not masses and sources.

So we shall call the left hand side $G^{\mu\nu}$, and we shall pursue our guess work on the equation

$$G^{\mu\nu} = 8\pi G T^{\mu\nu} \tag{24}$$

The only thing we know about $G^{\mu\nu}$ is that it is made up out of the metric.

It probably has second derivatives in it to be the analog of the laplacian of Newton equation (4).

In other words, it will involve the metric in some form or other and very likely second derivatives of the metric.

Ricci tensor and curvature scalar

We have begun to see what kind of object we would like to find and put on the left hand side of equation (24), which will be the generalization of Newton's field equation in the general relativity setting of Einstein.

When we find a good candidate for it, in order to evaluate it we can ask the following question : when we are in a situation where non-relativistic physics should be a good approximation, does this $G^{\mu\nu}$ reduce to just $\nabla^2 g_{00}$?

Perhaps it does.

Perhaps it doesn't.

If it doesn't, then we throw it away and try to find a different candidate.

Let's explore the possibilities for $G^{\mu\nu}$.

It is a tensor made up out of the metric.

It has second derivatives, or at least it must have some terms which have second derivatives.

So it is not the metric by itself.

What kind of tensor can we make out of a metric and second derivatives?

We have already met one : it is the curvature tensor.

Let's recall what is the curvature tensor.

It is made up out of the Christoffel symbols, which themselves are functions of the metric.

Let's start with the Christoffel symbols

$$\Gamma^{\sigma}_{\nu\tau} = \frac{1}{2}\, g^{\sigma\delta}\ [\, \partial_\tau g_{\delta\nu} + \partial_\nu g_{\delta\tau} - \partial_\delta g_{\nu\tau}\,] \tag{25}$$

The only important thing in this equation is that the right hand side involves the first derivatives of g terms.

Next, let's write the curvature tensor in all its glory(expression (25) in chapter 3, with different dummy indices).

$$\mathcal{R}^{\ \sigma}_{\mu\nu\tau} = \partial_\nu \Gamma^{\sigma}_{\mu\tau} - \partial_\mu \Gamma^{\sigma}_{\nu\tau} + \Gamma^{\lambda}_{\mu\tau}\Gamma^{\sigma}_{\lambda\nu} - \Gamma^{\lambda}_{\nu\tau}\Gamma^{\sigma}_{\lambda\mu} \tag{26}$$

This rank 4 tensor is the object – the "mathematical probe" – which tells us whether there is real curvature in space-time, not just curvature due to curvilinear coordinates that could be flattened (like a uniform acceleration can be flattened, see lesson 1 fig 4, and lesson 4 fig 17).

If any component of the curvature tensor is non-zero at any point or in any region of space, then the space is curved.

The right hand side of equation (26) uses the summation convention for the index λ in the last two terms.

Again, the only important point in the expression of $\mathcal{R}^{\sigma}_{\mu\nu\tau}$ is that it μντ involves another differentiation.

The Christoffel symbol, which is not a tensor, involves first derivatives of the metric.

And the first two terms of the curvature tensor are first derivatives of the Christoffel symbols.

Therefore the curvature tensor has second derivatives of the metric in its first two terms, and squares of first derivatives in its last two terms.

Therefore $\mathcal{R}$ is a candidate, or various functions of it are candidates, to appear on the left hand side of equation (24).

But wait!

The Riemann curvature tensor $\mathcal{R}^{\sigma}_{\mu\nu\tau}$ has four indices, whereas the energy-momentum tensor has only two.

What can we do with the curvature tensor to transform it into a thing with only two indices?

We can contract it.

Remember the rule : if you set the upper index σ equal to one of the lower indices and apply the summation convention, we eliminate two indices and get a rank 2 tensor.

If we used τ for the lower index in the contraction, we would discover that we get zero.

The various symmetries and minus signs in the expression of $\mathcal{R}^{\sigma}_{\mu\nu\tau}$ would wind up yielding 0.

However if we set σ and ν equal, we won't get zero.

We will get something that we can call the tensor $\mathcal{R}_{\mu\tau}$.

So from a tensor with four indices, by contraction we can build a tensor with two indices.

But we have to be careful not to get zero.

We won't get the zero tensor when we contract σ with ν or σ with μ.

And in fact the two tensors we get by contracting σ with ν and σ with μ happen to be the same tensor except for the sign.

So there is only one tensor that we can build out of two derivatives acting on the metric and which has only two indices.

It is actually a well-known theorem.

This tensor is called the *Ricci tensor*.

And it is a contraction of the Riemann tensor.

The Riemann tensor has a lot more components.

The Ricci tensor has less information.

As a consequence, it can be zero while the Riemann tensor is not zero.

As always, if you have a tensor you can raise and lower its indices.

That means there are also things denoted $\mathcal{R}^{\mu}_{\tau}$, $\mathcal{R}^{\tau}_{\mu}$ and $\mathcal{R}^{\mu\tau}$.

You can raise and lower indices using the metric tensor.

Another fact about the Ricci tensor, is that it happens to be symmetric.

And this is true of its version with upper indices as well

$$\mathcal{R}^{\mu\tau} = \mathcal{R}^{\tau\mu} \tag{27}$$

Now, this you just check by using its definition.

I don't know any simple quick argument.

All these tensors have fairly complicated expressions, but checking the facts we stated is most of the time pretty mechanical.

The Ricci tensor $\mathcal{R}^{\mu\tau}$ is symmetric.

The right hand side of equation (24) is a rank 2 tensor, with two contravariant indices, and is symmetric.

So the Ricci tensor $\mathcal{R}^{\mu\tau}$ is a candidate for the left hand side :

The Big Question:

$$\mathcal{R}^{\mu\nu} = ? \ 8\pi G T^{\mu\nu} \tag{28}$$

Now there is another tensor that we can make out of the Ricci tensor.

It is by contracting its two indices (after having raised or lowered one of them).

We define

$$\mathcal{R} = \mathcal{R}^{\mu}_{\mu} = \mathcal{R}^{\mu\tau} g_{\mu\tau} \tag{29}$$

The $\mathcal{R}$ defined by equation (28) is called the curvature scalar.

It is a scalar.

It has no indices left at all.

So it's not something we want on the left hand side of equation (24).

But we can multiply it.

For instance we can multiply it like this $g^{\mu\nu}\mathcal{R}$.

This does give us again a tensor of the type we are looking for.

That is another possibility.

$$g^{\mu\nu}\mathcal{R} \; = ? \;\; 8\pi G T^{\mu\nu} \tag{30}$$

I'm not recommending either of these at the moment.

I'm just saying from what we have said up till now, either of these could be possible laws of gravitation.

The left hand sides of equations (28) or (30) both involve second derivatives of the metric tensor, equated to something on the right hand side which looks like a density and flow of energy and momentum.

Which one should we pick?

Well, we know one more thing.

That is the conservation of energy and momentum or better yet the continuity equation for energy-momentum.

If we believe that energy and momentum have the property that they can only disappears by flowing through walls of systems, then we are forced to the conclusion that

$$D_\mu T^{\mu\nu} = 0$$

That is the continuity equation.

From that, it follows that we must also have

$$D_\mu G^{\mu\nu} = 0 \tag{31}$$

So the first thing we could do is check whether either of the candidate solutions to equation (24), that is, those in equations (28) or (30) satisfy condition (31).

If not then the left hand side simply cannot equal the right hand side, unless we give up local continuity of the energy and momentum.

Let's check the left hand side of equation (30).

We want to calculate

$$D_\mu \ (g^{\mu\nu}\mathcal{R})$$

Covariant derivatives satisfy the usual product rule of differentiation.

So we have

$$D_\mu \ (g^{\mu\nu}\mathcal{R}) = (D_\mu \ g^{\mu\nu})\mathcal{R} + g^{\mu\nu}(D_\mu \ \mathcal{R}) \tag{32}$$

First of all, the covariant derivative of the metric tensor is zero.

It is a consequence of the definition of covariant derivative, see lesson 3.

Covariant derivatives are by definition tensors which in the special good frame of reference are equal to ordinary derivatives.

But the good frame of reference is by definition the frame of reference in which the derivative of g is 0.

So the first term on the right hand side of equation (32) is eliminated.

Secondly, $\mathcal{R}$ is a scalar.

The covariant derivative of a scalar is just the ordinary derivative.

So we can rewrite

$$D_\mu \left(g^{\mu\nu}\mathcal{R}\right) = g^{\mu\nu}\ \partial_\mu \mathcal{R} \tag{33}$$

Certainly, in general, the derivative of the curvature is not zero.

We know that there are geometries which are more curved in some places and less curved in others.

So it cannot be the case that $\partial_\mu \mathcal{R}$ is identically equal to zero.

And the factor $g^{\mu\nu}$ doesn't change anything.

So we are led to eliminate the candidate $g^{\mu\nu}\ \mathcal{R}$.

What about the other candidate, the Ricci tensor with upper indices?

We do the same thing.

We calculate

$$D_\mu\ \mathcal{R}^{\mu\nu}$$

The calculation is a little harder, but not much.

The answer is

$$D_\mu\ \mathcal{R}^{\mu\nu} = \frac{1}{2}\ g^{\mu\nu}\ \partial_\mu \mathcal{R} \tag{34}$$

Again, it cannot be zero, for the same reason that the right hand side of equation (33) can't be zero.

The covariant derivative of the Ricci tensor happens to be *exactly one half* that of the other candidate, $g^{\mu\nu}\ \mathcal{R}$.

But now we know the answer !

Einstein tensor and Einstein field equations

We now know which tensor on the left hand side of equation (24) will satisfy all the conditions, including having covariant derivative zero.

The left hand side of the gravity equation (24) that will work is

$$G^{\mu\nu} = \mathcal{R}^{\mu\nu} - \frac{1}{2}\, g^{\mu\nu}\, \mathcal{R} \tag{35}$$

And a theorem can be proved, which says that there is no other tensor (up to a multiplicative factor) made up out of two derivatives acting on the metric that is covariantly conserved.

Of course, we could have twice $G^{\mu\nu}$ or half of it, or seventeen times it.

But now it just becomes a question of matching equation (24), with the $G^{\mu\nu}$ we found, to Newton's equations in the appropriate approximation, that is, where everything is moving non-relativistically.

Either there is some correct numerical multiple that will ensure the match or there isn't.

If there isn't, then we're in trouble.

$$G^{\mu\nu} = 8\pi G T^{\mu\nu}$$

When we say that we want to match formulas, we mean that we take equation (24), we look at the time-time component of $G^{\mu\nu}$ given by equation (35), and we want, with some appropriate multiplicative factor in front of $G^{\mu\nu}$, in the non-relativistic limit – that is, everybody moving slowly, not too strong a gravitational field – to find Newton's equation (4).

The answer is that it works.

And the multiplicative factor is simply 1.

It is just a piece of luck that it is not some other number.

There is nothing deep about this factor.

So, we finally arrived at the equation we were looking for, generalizing Newton's equation.

$$\mathcal{R}^{\mu\nu} - \frac{1}{2}\, g^{\mu\nu}\, \mathcal{R} = 8\pi G\, T^{\mu\nu} \qquad (36)$$

The path we followed in this lesson is *essentially* Einstein's calculations.

He knew pretty much what was going on.

But he didn't quite know what the right equation linking the energy-momentum tensor and the geometry was.

I believe in the beginning he actually did try $\mathcal{R}^{\mu\nu} = T^{\mu\nu}$.

But he eventually realized that it didn't work.

And he searched hard for a suitable left hand side.

I don't know how many weeks of work it took him to do all of this.

But in the end he discovered $G^{\mu\nu}$.

In early summer 1915, at the invitation of Hilbert, Einstein gave a series of lectures in Göttingen, exposing his work of the past few years and his present problem to link sources and geometry.

This lead, during the summer and fall of 1915, to an exchange of friendly letters, but also to a race, between Einstein and Hilbert.

Noether would have gotten there first if she cared at all about the physics!

Einstein published his final correct equations on November 25.

Simultaneously Hilbert published the same equations obtained from the lagrangian approach.

That is the reason why the name of Hilbert is sometimes also attached next to Einstein's to general relativity.

However Hilbert himself never denied that all the credit for general relativity should go to Einstein.

$G^{\mu\nu}$ is the Einstein tensor.

$\mathcal{R}^{\mu\nu}$ is the Ricci tensor.

$\mathcal{R}$ is the curvature scalar.

Equation (36) is now known as Einstein field equations.

$$\mathcal{R}^{\mu\nu} - \frac{1}{2}\, g^{\mu\nu}\, \mathcal{R} = 8\pi G\, T^{\mu\nu} \qquad (36)$$

They generalize Newton equations.

And in the appropriate limit, they do reduce to Newton equations.

The continuity equation has played a fundamental role in this derivation.

It is what leads us to look only for candidates, on the left hand side, with are covariantly conserved.

Notice that the principle is as fundamental in classical non-relativistic physics as in relativity.

It says that conservation of energy or of momentum is a local property.

Things cannot disappear here and immediately materialize in Alpha Centauri.

They have to travel through intermediate locations.

There is another way to derive Einstein field equations, using the action principle.

It is much more beautiful and much more condensed.

We introduce the principle of least action for the gravitational field.

The calculations are harder than what we did.

If interested see standard class notes on the website.

But in the end the field equations just pop out.

Equation (36) reveals something interesting.

We see that in general the source of the gravitational field is not just energy density.

It can involve energy flow.

It can involve momentum density and it can even involve momentum flow.

Now as a rule the momentum components, and even the energy flow, but certainly the momentum flow and the momentum density are much smaller than the energy density.

Why can we say that?

It has to do with the impact of the speed of light in the formulas.

When we put back the speed of light (that is, when we use ordinary units), the energy density is always huge because it gets a factor c^2, like in $E = mc^2$.

On the other hand momentum is typically not huge because it is just mass times velocity.

When you are in a non- relativistic situation, when velocity is slow, energy density is by far the biggest component.

The other components of the energy-momentum tensor are much smaller.

They are typically smaller by powers of v/c.

In other words, in a frame of reference where the sources of gravitational field are moving slowly, the only important thing in the right hand side of equation (36) is T^{00}, that is, the energy density ϱ.

It is also true that, in the same limit, the only important thing on the left hand side is the second derivative of G^{00}.

So in non-relativistic limit these things match and reduce to Newton's equations.

But if you are outside the non-relativistic limit, in places where sources are moving rapidly, or even places where the sources are made up out of particles which are moving rapidly, even though the whole system may not be moving so much, other components of the energy-momentum tensor do generate gravitation.

Since we now know that, in a sense, gravitation is just geometry, we see that in relativistic situations, all the components of the energy-momentum tensor do participate in the generation of the curvature.

It is not just energy – or mass – that causes it.

Let's now consider a special case.

Just as in Maxwell's equations there are solutions which involve no sources, the same situation is true here.

So let's consider the case with either no sources or when we are in a region of space where there are no sources.

Then, on the right hand side of equation (36), we have zero.

It is just a little simplification, but the equation does become simpler in that case.

Suppose that $T^{\mu\nu} = 0$.

In other words

$$\mathcal{R}^{\mu\nu} = \frac{1}{2}\, g^{\mu\nu}\, \mathcal{R} \tag{37}$$

Let's calculate R by, on both sides, contracting μ and ν.

Contraction is done on one upper and one lower index.

So let's first lower the ν index on both sides (by multiplying by the appropriate version of the metric tensor).

We get

$$\mathcal{R}^{\mu}_{\nu} = \frac{1}{2}\, g^{\mu}_{\nu}\, \mathcal{R} \tag{38}$$

Then contraction of the left hand side gives precisely $\mathcal{R}$.

On the right hand side, g^{μ}_{ν} is the Kronecker delta.

Its contraction yields 4.

We arrive at $\mathcal{R} = 2\mathcal{R}$.

Therefore $\mathcal{R} = 0$.

In the case of no sources, the curvature scalar is zero.

Therefore in equation (36), where the right hand side is already zero, we can also drop the $\mathcal{R}$ term on the left hand side.

We arrive at

$$\mathcal{R}^{\mu\nu} = 0 \tag{39}$$

Einstein equation becomes simpler.

It is called the *vacuum* case.

In this case the Ricci tensor is zero.

Notice, as we have already said, that it *doesn't* imply that the Riemann curvature tensor be zero.

The solutions to equation (39) are non trivial.

They contain gravitational waves with no sources, just like there can be electromagnetic waves even in space without electric charges.

Another example is the Schwarzschild metric.

It is roughly speaking analogous to a point mass.

Outside the point mass there is no matter, nothing.

So just like Newton's equations, everywhere outside the mass, are the same as for empty space, the Schwarzschild metric, outside the singularity, is a solution to the vacuum Einstein equation.

It is conceptually a simple check, although the calculations are a real nuisance.

You start from the Schwarzschild metric, sit down and spend the rest of the day calculating a ream of Christoffel symbols, get to the curvature tensor, contract it to arrive at the Ricci tensor, $\mathcal{R}^{\mu\nu}$, and observe that it is equal to zero.

Metrics for which the Ricci tensor is zero are called Ricci flat.

So anywhere except at the singularity – where every- thing is crazy and meaningless anyway, just like for Newton's point masses – the Schwarzschild metric is Ricci flat.

It is not the same as flat.

So gravitational waves satisfy equation (36).

Schwarzschild metric satisfies it, except at the singularity.

And that is the basic fact about Einstein field equations.

Some Interesting Questions and Answers

Q1: Shouldn't Einstein equations contain a cosmological constant ?

We haven't said anything about the cosmological constant – whether it exists or not – because it can be thought of as part of $T_{\mu\nu}$.

From this point of view, the "cosmological constant" is an extra tensor term on the right hand side of equation (36).

We could denote it

$$T_{\mu\nu}^{\ cosmological}$$

And that would not change the look of Einstein equation.

If it was indeed a scalar, we could write it

$$T_{\mu\nu}^{\ cosmological} = \Lambda \, g_{\mu\nu}$$

Then it would make sense to shift it on the left hand side, because it would only be a part of the geometry.

This yields the following equation frequently met in the literature

$$\mathcal{R}^{\mu\nu} - \frac{1}{2} \, g^{\mu\nu} \, \mathcal{R} + \Lambda \, g_{\mu\nu} = 8\pi G \, T^{\mu\nu}$$

Here is a brief history of this added term.

When Einstein wrote his equation, the common view among cosmologists was that the universe was the Milky Way, and that it was stable.

But Einstein soon realized that his equation implied that the universe could not be stable.

So he added an extra term to counter the instability.

In 1929, Hubble discovered that the universe was expanding.

Then the cosmological became irrelevant.

Later in his life Einstein called it "the biggest blunder of my career".

In fact, to this day, the cosmological constant is the subject of a debate because it can frequently be called to the rescue as a "explanatory factor" in various riddles of the cosmos.

However most theoretical physicists dislike ad hoc factors or procedures to solve (or fudge) a problem.

Yet it is sometimes fruitful if one thinks of the introduction of quantas by Max Planck as an ad hoc procedure to overcome a problem with black-body radiations.

Q2 : Did Einstein use the equation we derived in this lesson to explain the precession of the perihelion of the orbit of Mercury?

It is certainly in one of Einstein's papers.

Let's talk about the two great successes that general relativity quickly achieved : the prediction of the bending of light-ray from distant stars by the Sun, and the explanation of the oddities in the orbit of Mercury.

Concerning the bending of light-rays, he did not have the Schwarzschild solution.

But he did have the approximation to the Schwarzschild solution at fairly large distances.

Of course the Sun is not a black hole.

But outside the solar Schwarzschild radius the geometry is exactly the same as that of Schwarzschild.

Far away from the Schwarzschild radius, Einstein knew how to make a good approximation.

The fact that the Sun is so big means that the corrections from Newton are small.

They can be done using the techniques of perturbation theory : you add some small corrections to something you already know, and fit them so that the equations still hold.

For a light-ray you start from a nice straight line, do a little bit of perturbation theory and work out the modification of the trajectory.

For the orbit of Mercury, most likely what he did was just start from the Keplerian orbits.

That is, he started from Newton solution.

Then he fitted the small correction on the left hand side to the small correction on the right hand side.

And he found exactly the right discrepancy in the perihelion of Mercury.

But then Schwarzschild, within a few weeks after Einstein first publication, calculated his exact solution of the equation.

And from them you could do much better than perturbation theory.

Notice one last point concerning the orbits of planets.

The ordinary gravitational potential in $1/r$, leading to a centripetal force law in $1/r^2$, is pretty much the only law that leads, in the simplest case, to elliptical closed orbits without precession.

It is a curious and somewhat accidental fact.

Notice, though, that "accidental facts" are precisely the kind of facts that Einstein did not take for granted.

It is the rejection of the accidental character of the equality between inertial mass and gravitational mass which ultimately lead him to the theory of general relativity.

Lesson 10 : Gravity waves

Introduction

The questions we want to address in this lesson are weak gravitational fields, linearity versus non-linearity, gravitational waves (aka gravity waves).

We will finish with an overview of how Einstein field equations can be derived from the action principle.

Again working out the equations of general relativity is always unpleasant.

We are not going to do it here.

The calculations would fill pages even for simple things.

And they probably would not be terribly illuminating.

To learn the subject you really have to compute yourself and solve the equations and so forth.

You just have to sit down and do it on your own.

On the other hand the principles are straightforward.

And it is easy enough to explain what you get when you do solve equations.

So that is the way we will talk about gravity waves : by writing down the equations and then writing their solutions.

To start with, we are interested in what could be called *weak* gravitational waves.

Weak gravitational fields

We talk of weak gravitational field when the gravitational variations over space-time are small enough that we can make approximations, such as the variations squared of the gravitational field can be taken to be zero.

When a quantity is small and we expand an equation about the smallness of that quantity, the usual rule is to ignore things which are of higher order than one in that small quantity.

The typical example is using the beginning binomial theorem

$$(1+\epsilon)^n \approx 1 + n\epsilon$$

Okay so we start from an equilibrium.

And we are going to talk about fluctuations, or perturbations, about the equilibrium situation.

We will look at the simplest equilibrium solution of Einstein equations.

Let's write Einstein equations first.

And let's take the case without any matter – no energy-momentum tensor.

Then the equations of motion are just

$$\mathcal{R}_{\mu\nu} - \frac{1}{2}\, g_{\mu\nu}\, \mathcal{R} = 0 \tag{1}$$

It can be simplified.

If we take the trace of both sides, after having raised one index upstairs, since $\mathcal{R}$ is the trace of the Ricci tensor $\mathcal{R}^{\mu}_{\nu}$, and 4 is the trace of g^{μ}_{ν} (which is the Kronecker delta), this gives $\mathcal{R} = 0$.

Therefore we can ignore the second term in equation (1).

And it becomes simply

$$\mathcal{R}_{\mu\nu} = 0 \tag{2}$$

It doesn't really matter because we are not going to write down the details anyway.

But equation (2) is Einstein's equation in a context where there is no energy-momentum tensor on the right hand side of the full equation (see equation (36) of lesson 9).

What is an equilibrium situation?

It is, first of all, a solution which has no time dependence.

And the solution also doesn't have any matter on the right hand side – matter being another word for the energy-momentum tensor.

There is really only one equilibrium situation.

It is just empty space that is time independent.

By empty space we mean empty flat space, no curvature, no interesting gravitational field.

And in that case, the metric $g_{\mu\nu}$ has a simple form.

Remember that it is improper to say "the metric is equal to such and such matrix", because the metric is a tensor.

In each coordinate system, it has a different expression with components.

It is the same for vectors in 3D Euclidean space.

A vector is not a set of three numbers.

It is a direction in space, which, in a given basis, has an expression with three numbers.

And in another basis, the collection of three numbers will be different.

The same is true of tensors.

They are abstract creatures.

And they have expressions with multi-indexed collections of components, which depend on the coordinate system chosen to describe space-time and the objects in it.

If I wrote that the metric of a flat plane, in Euclidean geometry, is just the Kronecker delta symbol, you would correct me : "no, the metric of a flat plane is not the Kronecker delta."

What is true is that a possible expression of the flat metric, with an appropriate coordinate system, is the Kronecker delta.

But if I used other coordinates, the expression of the metric would not be the Kronecker delta.

We could use curved coordinates, use polar coordinates, use any other kind of coordinates.

What is special about flat space, in Euclidean geometry, or in Riemaniann geometry, is that you can find coordinates in which the metric has a nice simple form.

The same is true in general relativity – in which, as we know, the geometry is Minkowski-Einstein geometry.

At any point in space-time the signature of the metric at that point is three plus signs for the three spatial coordinates and a minus sign for the time coordinate.

For flat space-time with no gravitational field, there is a choice of coordinates in which the metric $g_{\mu\nu}$ has a simple form

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3}$$

So the four columns correspond to (t, x, y, z).

Same for the four rows.

This is the metric of flat space-time in the most appropriate coordinate system.

And it is an equilibrium solution.

That is, it is a solution of the Einstein field equations, and this solution doesn't depend on time.

Remember that the complete description of motion is : "the field tells the sources how to move" (they move along geodesics) and "the sources tell the field what to be" (it is a solution to Einstein field equations, which generalize Poisson's equation of the Newtonian case, see equation (4) of lesson 9).

In general the solution is time dependent of course.

But here we are interested in an equilibrium solution – therefore independent of time.

Secondly we are interested in the equilibrium solution for the simplest type of source configuration, namely, no sources.

So we are looking at the case where Einstein field equations reduce to $\mathcal{R}_{\mu\nu} = 0$.

They say that certain components of curvature are equal to 0.

It is a purely geometric condition.

And we have a solution which is particularly simple : no curvature at all anywhere.

It is easy the check that it satisfies $\mathcal{R}_{\mu\nu} = 0$.

This concludes the first step toward weak gravitational waves.

Now let’s think about a space-time which is close to empty space-time.

Suppose that, very far away from where we are, some complicated phenomenon is happening.

For example a binary pulsar is rotating rapidly.

And it emits some complicated gravitational waves.

Near the pulsar the gravitational field may be very strong.

Even the gravitational waves (—> the evolution of the field over space-time, like the moving ripples created by a stone thrown in a pond) might be rather strong.

But if you go far enough away the gravitational radiation, that is, the gravitational waves that are produced by this thing, are going to be very weak.

But if you go far enough away the gravitational radiation, that is, the gravitational waves that are produced by this thing, are going to be very weak.

What does weak mean?

Weak means that the metric can be chosen – again, I emphasize, in the appropriate coordinate system – to be equal to $\eta_{\mu\nu}$ plus something small.

Small means that the components of the small thing are much smaller than those of η.

So we write

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \tag{4}$$

As far as I know, it is called h because it is the letter that comes after g.

Unlike $\eta_{\mu\nu}$, the perturbation term $h_{\mu\nu}$ is in general a function of position.

When we say "position" we mean the four coordinates of an event in space-time of course.

$h_{\mu\nu}$ varies from place to place and also from time to time, and it might describe a wave.

We will come back to waves in a moment.

But let's first ask ourselves : what are we going to do with this $h_{\mu\nu}$?

Answer : we will take the metric given by equation (4), calculate its Ricci tensor $\mathcal{R}_{\mu\nu}$, and set it equal to zero.

That will give us an equation for h.

The equation that we actually get for h is not big enough to fill a whole page, but it is sufficiently big to be unpleasant.

So I'm just going to be schematic. I'm going to show you what goes into it

First of all, let's look a the Ricci tensor $\mathcal{R}_{\mu\nu}$.

It is really a combination of components of the Riemann curvature tensor $\mathcal{R}^{\sigma}_{\mu\nu\tau}$.

I'm not going to write the full curvature tensor, I'm just going to remind you what it contains, that is, its structure.

So the $\mathcal{R}$ below means the full rank 4 curvature tensor.

It has the following structure

$$\mathcal{R} = \partial\Gamma + \Gamma\Gamma \tag{5}$$

It contains first derivatives of the Christoffel symbols, and it contains the Christoffel symbols quadratically, that is products of two Christoffel symbols.

I could put the indices everywhere, spell out the various terms which look like $\partial\Gamma$, and the various terms which look like $\Gamma\Gamma$.

But expression (5) giving the structure is really all we need.

What about the Christoffel symbol Γ?

Likewise its structure is

$$\Gamma = \frac{1}{2} g^{-1} \partial g \tag{6}$$

The symbol g^{-1} means the inverse matrix of g, or equivalently the metric tensor with two contravariant indices.

Just like g with covariant indices can be expanded as in equation (4), g^{-1} can also be expanded in powers of h.

The first contribution to it is just the inverse of the η symbol.

But that is η itself, because η is its own inverse.

And then we simply have the same h term but with a minus sign (just like $(1/(1+\varepsilon) \approx (1-\varepsilon))$.

So the expansion of g^{-1} up to the first perturbation term is, in matrix notation,

$$g^{-1} = \eta - h \tag{7}$$

where actually the same matrix h appears with a minus sign.

Now, dropping the 1/2 because we are only looking at structures, equation (6) becomes

$$\Gamma \sim (\eta - h)\partial h \tag{8}$$

The wiggly equal sign ~ is used to equate things which are not equal at all...

Equation (8) is the structure of the Christoffel symbol.

It contains one term which is the power one of h.

And this is multiplied by the first derivative of h.

We are also going to assume that ∂h is small.

This would make a mathematician jump.

But physically it means that, far away from the pulsar, not only the ripple h itself is small, in the sense that it has a small amplitude, but that its variations are attenuated too.

So the derivative of h is also considered to be a small thing.

η times derivative of h is one order of magnitude small.

h times the derivative of h is two orders of magnitude small.

It is quadratic in the fluctuation, or in the small gravitational field, so we ignore it.

For instance, if $h = .01$ and $\partial h = .01$, then their product is $.0001$.

In this approximation procedure $\eta\partial h$ is just ∂h.

So from equation (8) we deduce that, in the approximation of weak gravitational radiation, the Christoffel symbol is simply proportional to some collection of derivatives of h.

What about the Riemann tensor then?

It will contain derivatives of Γ and products of two Γ's.

That will be various kinds of $\partial^2 h$, and various products of the form $\partial h \partial h$, that is second derivatives of the gravitational field and products of first derivatives of the gravitational field.

Incidentally, h is called the gravitational field, or the field of gravitational waves.

$$\mathcal{R} = \partial^2 h + \partial h \partial h \tag{9}$$

So immediately, we say that $\partial h \partial h$ is an order of magnitude smaller than $\partial^2 h$ and again we ignore it.

Finally the Ricci tensor has the same form as the curvature tensor.

So whatever the devil the Ricci tensor $\mathcal{R}_{\mu\nu}$ is, it is composed out of simple second derivatives of the metric tensor.

From that we can conclude that Einstein equations have a relatively simple form.

There is still plenty of indices around, and the number of different terms involving second derivatives of $h_{\mu\nu}$ is significant.

It is complicated enough.

So we don't want to write it out here, but it is basically built out of second derivatives – second derivatives with respect to position, second derivatives with respect to time, and maybe even some terms combining a derivative with respect to position and a derivative respect to time.

Furthermore there are several components of it - how many components of it?

Well, the Ricci tensor has a μ and a ν, each running from 0 to 3.

That makes 16 components.

But the whole works, whatever it is, is composed of second derivatives of h.

So our equation of motion is some kind of equation that looks like this

$$\partial^2 h = 0 \tag{10}$$

This leads us naturally to the next topic.

Gravitational waves

$$\partial^2 h = 0$$

Equations like equation (10), especially in relativity, are usually wave equations.

Let's recall what the wave equation for an ordinary wave looks like.

Let's say we look at a wave that is moving along the z-axis as time goes by.

So consider a wave field that we can call $\phi(t, z)$.

It will satisfy

$$\frac{\partial^2 \phi}{dt^2} = \frac{\partial^2 \phi}{dz^2} \tag{11}$$

This is the *simplest* wave equation that you can imagine.

And simple solutions to it are waves which move either to the right or to the left, figure 1.

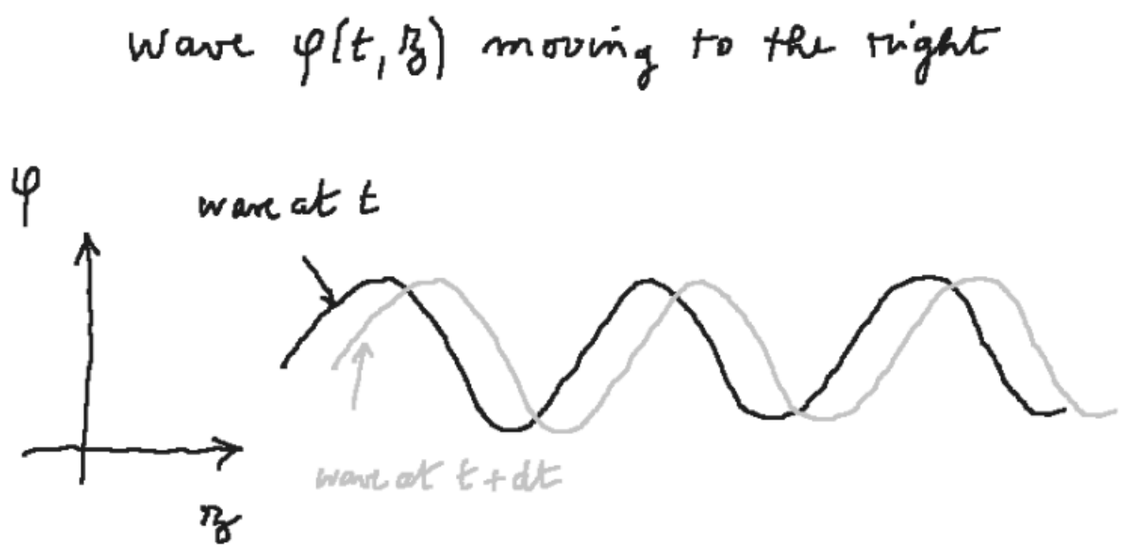


Figure 1 : Wave moving to the right.

You can add a wave which goes to the right and a wave which goes to the left.

It is still a solution because the equation is *linear*.

If we had more directions of space the structure of the equation would be a little more complicated.

Instead of just derivative of ϕ with respect to z we would also have derivatives with respect to x and to y.

The equation could be rewritten

$$\frac{\partial^2\phi}{dt^2} - \frac{\partial^2\phi}{dx^2} - \frac{\partial^2\phi}{dy^2} - \frac{\partial^2\phi}{dz^2} = 0 \tag{12}$$

But it is evident that there is a family similarity between the kind of equation like (10), and the kind of equation like (12).

And in fact by clever manipulation you can make these equations look exactly alike.

Incidentally equation (10) is not just one equation.

There is an equation for each μ and ν.

Let's put brackets around it to remind ourselves that there are sixteen components of $\mathcal{R}_{\mu\nu}$:

$$\left[\partial^2 h\right]_{\mu\nu} = 0 \tag{13}$$

These are not all independent.

It is really quite a bit simpler than that.

But still in principle there are 16 components, therefore 16 equations.

The collection of equations (13) is somewhat similar to Maxwell's equations in that respect.

Maxwell's equations have the form of wave equations, for example for the electric and magnetic field.

But there are several components.

There are three components of electric field, three components of magnetic field.

It sounds like there is only six equations but in fact there are eight equations in the complete set of Maxwell's equations.

It is the same sort of pattern with equations (13) : several equations, but all of similar form, and not all independent.

Pay attention to the fact that equations (13) are only suggestive.

But we proved that they involve only second derivatives.

In the process of producing equations (12) anything that involves h, or derivative of h or second derivative of h, is small, so when it is multiplied by another small quantity (as is the case when we do contraction, which introduces multiplicative terms of the form $\eta + h$) we omit the result because it becomes of a higher order of magnitude in smallness.

Remember that what we do is an approximation.

But it is a well-defined approximation.

Technically, it is the linearization of Einstein's equations.

It simply means that in Einstein field equations we throw away everything of higher power that h.

It is a good approximation when the gravitational radiation, or the gravitational wave is weak.

So we have equations of form $[\partial^2 h]_{\mu\nu} = 0$, which we are not going to specifically write down because they are kind of messy.

Before we discuss their solutions, let's go back to the observation we made earlier – and in fact we repeated many times since the beginning of the class – that for the space- time to be flat, the metric doesn't have to be necessarily η.

But with a proper choice of coordinates it can be made to be η.

That means that there are other ways to represent flat space-time.

We can make a coordinate transformation on $\eta_{\mu\nu}$.

The metric will change, as well as the form of the solution to Einstein field equation, but it will still be exactly the same physical solution.

Therefore there must be solutions of equations (13) which look like they are non trivial.

But they really represent flat space, in coordinates, however, with little ripples in them.

So let's do that in Euclidean geometry to understand this point.

Let's consider the flat page in Euclidean geometry.

It is a flat space.

So one of its metrics is simply δ_{mn}, where m stands for the X-axis, and n for the Y-axis.

So we start with "good" coordinates X and Y .

Then we introduce coordinates X' and Y' which incorporate little disturbances.

And we express the coordinates (X, Y) in the coordinates (X', Y')

$$X = X' + f(X',\ Y')$$
$$Y = Y' + g(X',\ Y') \tag{14}$$

It is a coordinate change.

We *haven't* changed the space in any way.

All we have done is change coordinates.

The new slightly wiggly coordinates represented in the old "good" ones are shown in figure 2.

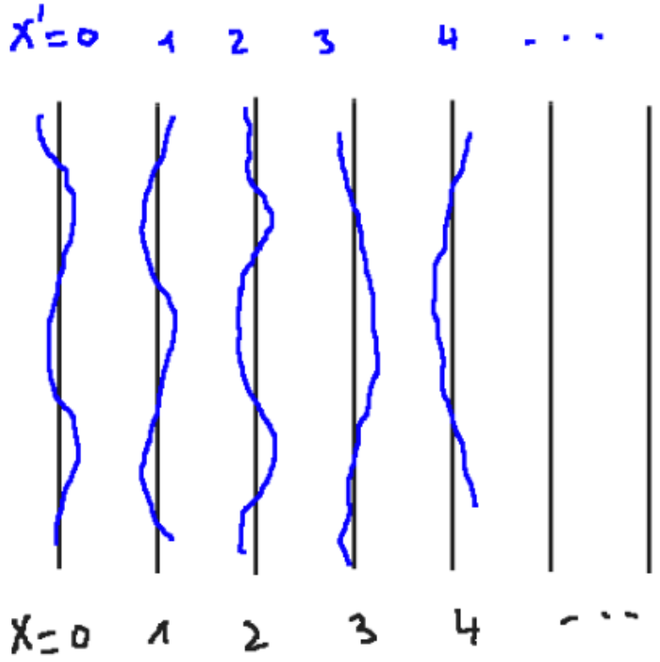


Figure 2 : Flat space with "good" coordinates and slightly wriggly coordinates.

The metric in the "good" coordinates is

$$dS^2 = dX^2 + dY^2 \tag{15}$$

Now what is the metric in the (X',Y') coordinates?

Remember, we are just in a Euclidean space, we are not doing space-time.

So let's work out, just for fun, the metric in (X',Y').

We can write

$$dX = dX' + \frac{\partial f}{\partial X'^{\,m}} dX'^{\,m} \tag{16a}$$

likewise

$$dY = dY' + \frac{\partial g}{\partial X'^{\,m}} dX'^{\,m} \tag{16b}$$

Our notations are a little awkward, because X'^{m} represents the collection X'^{1}, X'^{2}, that is simply X' and Y'.

And equations (16) incorporate a sum with the summation convention.

Then we plug equations (16) into equation (15).

We find out that dS^2 is not $(dX')^2 + (dY')^2$.

It contains $(dX')^2 + (dY')^2$ plus some cross-terms.

If you work it out – assuming that f and g, and their derivatives, are small so that you can make the appropriate approximations – you will find that

$$dX^2 + dY^2 = (dX')^2 + (dY')^2 + h_{mn}\, dX'^{\,m}\, dX'^{\,n} \tag{17}$$

The right hand side is the square of the infinitesimal distance between two points calculated with the metric tensor expressed in the new wiggly coordinates.

Does this small correction to the metric tensor mean that the page is not flat anymore?

Of course not.

It just means that the coordinates we use have wiggles in them (when we represent them in "good" coordinates, fig. 2).

It is a nice little exercise to compute, in equation (17), what the correction h is.

You will get

$$h_{mn} = \frac{\partial f}{\partial Y'} + \frac{\partial g}{\partial X'} \tag{18}$$

That will mean nothing to you until you try to work out an example, or until you try to prove it.

It is a small perturbation on the metric.

It has the form that we already met in equation (4).

But it doesn’t represent anything physical. It is just a somewhat trivial change of coordinates.

In other words, there exists perturbations on the metric which don't correspond to any physical effect on the space-time – it can remain flat even though the metric has the form of equation (4).

Once again a flat space-time is one not where the metric is necessarily η but where we can find coordinates such that the metric in those coordinates becomes η.

Likewise, in equations (13), there are small perturbations we can write down which just represent curvilinear coordinates of space-time but don't change its geometry or its physics.

How can we eliminate those phony solutions which automatically solve the equation because they are just flat space-time, in curvilinear coordinates, and don't represent any real physics ?

You do it by imposing more equations on the metric.

I could write down the equations but it is not important.

What is important is that they will wipe out the unwanted spurious solutions, the ambiguities on the coordinates, in other words the unphysical meaningless solutions of Einstein's equations.

And you can do that in a wide variety of ways.

Once you have done that the equations become pretty simple.

They become wave equations.

The components of h then satisfy perfectly ordinary wave equations, like equation (12).

It becomes

$$\frac{\partial^2 h_{\mu\nu}}{dt^2} - \frac{\partial^2 h_{\mu\nu}}{dx^2} - \frac{\partial^2 h_{\mu\nu}}{dy^2} - \frac{\partial^2 h_{\mu\nu}}{dz^2} = 0 \tag{19}$$

Each component of the fluctuation satisfies a wave equation.

It *means* that all the components of the metric just move down the axis so like waves, linear waves.

On the other hand there are also some constraints.

There are more equations than one for each μ and ν.

The reason is that we have these extra equations to eliminate the spurious fake solutions, that we saw because of the coordinate ambiguity.

Once we do that, we find out the physical solutions, the ones that really have meaning.

We will classify them.

Suppose we have a wave moving down the z-axis.

Here is its equation

$$\frac{\partial^2 h_{\mu\nu}}{dt^2} - \frac{\partial^2 h_{\mu\nu}}{dz^2} = 0 \tag{20}$$

What does it look like?

The simplest solution has the form

$$\phi = \sin k(z - t) \tag{21}$$

It is a wave that at a fixed instant of time is just a sine wave, see figure 1.

And it moves to the right down the axis with unit velocity.

Unit velocity here means the speed of light.

k is the frequency of the wave, also called the wave number.

It is the number of complete oscillations per unit length.

And of course it can be any number.

Short wave lengths have large k, and long wave lengths have small k.

Each component of h will have a solution like that.

It will be proportional to $\sin k(z - t)$.

So let's write

$$h_{\mu\nu}(t,\ z) = h^0_{\mu\nu} \sin k(z - t) \tag{22}$$

The coefficient $h^0{}_{\mu\nu}$ is not a function of position.

It is just a numerical coefficient that multiplies the sine of $k(t - z)$.

And it is about all that we can write down.

Equation (22) is the nature of a gravitational wave, each component of the metric behaving like a wave moving down the axis.

However, as we said, there are more equations, to filter out spurious solutions.

When we impose them all, we find out something interesting.

To understand it, let's first of all remember some facts about electromagnetism.

In electromagnetism we do the same thing : we solve Maxwell's field equations.

We can solve them either for the vector potential or for the electric and magnetic fields.

But when we write down all the equations, we find some constraints.

The constraints are called *transversality of the of the field*.

It means that the electric and magnetic fields always point in directions perpendicular to the motion of the wave, figure 3.

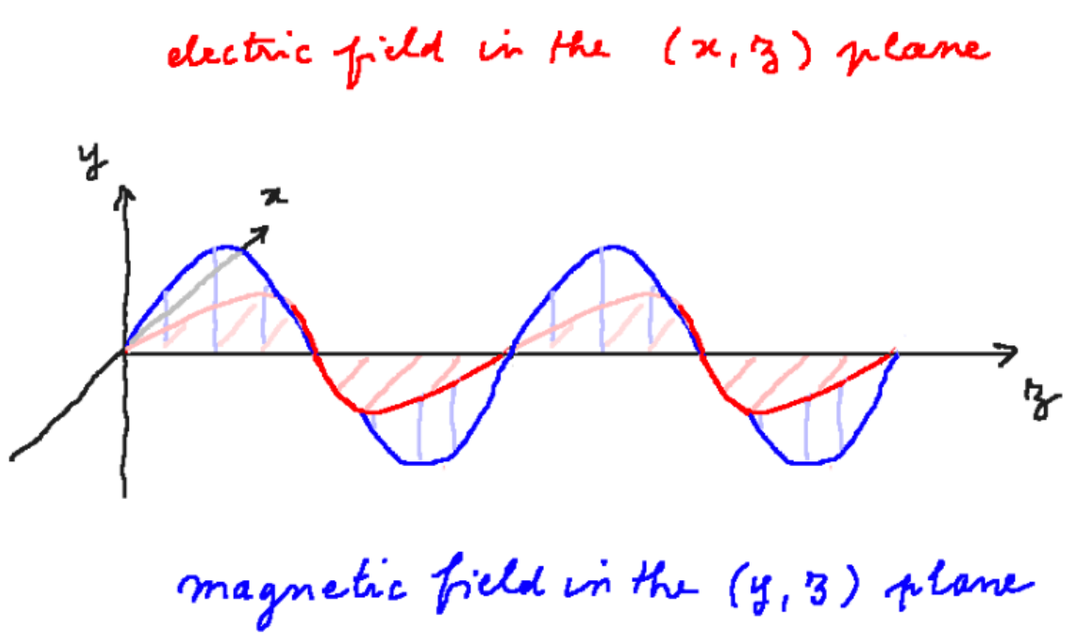


Figure 3 : Electric and magnetic fields photographed at one instant t.

Not only the electric field and magnetic field solutions are waves, but they are transverse waves.

That is, the are perpendicular to the axis of progression of the wave, which is the z-axis.

And they are perpendicular to each other.

All the little horizontal and vertical sticks shown in figure 3 oscillate over time (not in unison) in such a way that it suggests a movement down the z axis.

When talking about transverse moving waves, usually nothing physically moves in the longitudinal direction.

Think for instance of ripples moving on the surface of a water pond.

If you look at a cork floating somewhere, it will move up and down as the wave passes it, but it won't move longitudinally.

Very similar things happen here with gravitational waves.

The consequence of the constraints added to filter out spurious solutions is that the waves have to be transverse.

To say that the waves are transverse means that the components $h_{\mu\nu}$ involving t or z are zero.

The only components of h which are allowed to be none zero are the components in the plane perpendicular to the direction of the wave.

If we incorporate this fact into the full set of equations – equations (12) plus the equations removing the fake spurious fluctuations –, we find that a gravitational wave has a very simple form.

The only non zero components of the perturbation term added in the metric are

$$h_{ij}(t,\ z) = h^0_{ij} \sin k(z - t) \tag{23}$$

where we use the dummy variables i and j, which can each vary over x and y.

Let's think of the spatial part of space-time at time t.

And let's slice it into planes at different z, figure 4.

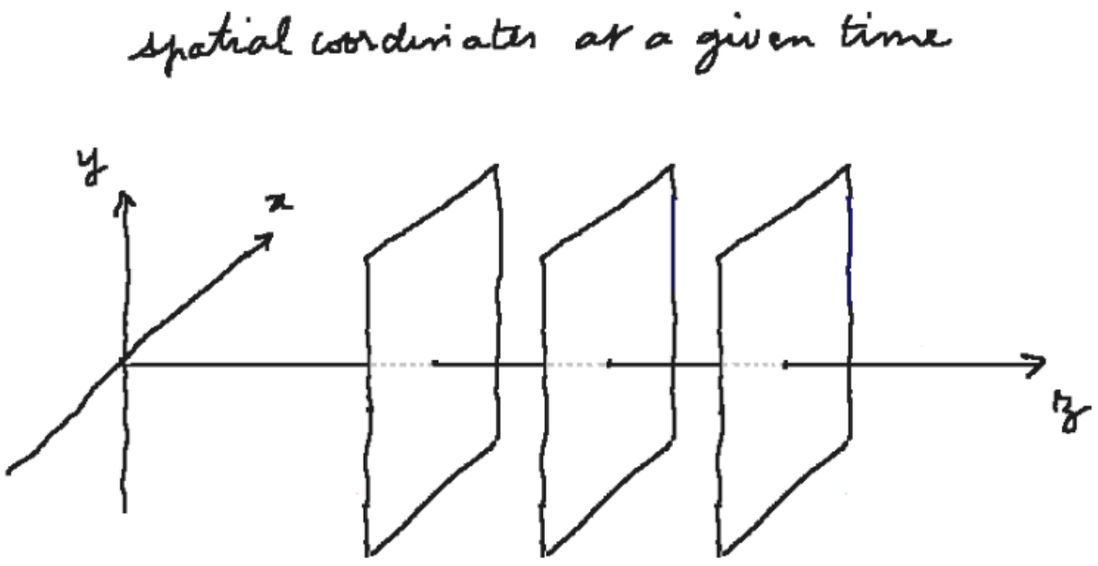


Figure 4 : Slices of x y planes.

At a given location z and given time t the metric h_{ij} is the metric in one of the two-dimensional planes shown in figure 4.

The components are simply numbers.

At each z and t, the metric is simply a set of numbers.

There is one more equation.

It comes from Einstein's field equations.

And it says that the trace of the metric h_{ij} is equal to zero.

In equation form, this is

$$h_{xx} + h_{yy} = 0 \tag{24}$$

That is it.

That is the whole set of equations.

And what it tells us is that the metric of a gravitational wave is

$$
\begin{aligned}
h_{00} &= 0 \\
h_{0x} &= 0 \\
h_{0y} &= 0 \\
h_{0z} &= 0 \\
h_{zx} &= 0 \\
h_{zy} &= 0 \\
h_{zz} &= 0
\end{aligned}
$$

That is to say, any component involving either t or z is equal to zero.

The only components which are non zero are

$$
\begin{aligned}
h_{xx} \\
h_{yy} \\
h_{xy}
\end{aligned}
$$

They are given by equation (3).

Furthermore, from equation (24), h_{xx} and h_{yy} are opposite to each other.

The metric tensor is symmetric so for any μ and ν, $h_{\mu\nu} = h_{\nu\mu}$.

That is pretty simple.

But what does it mean?

First of all all of these components of the perturbation term vary with t and z, and only with t and z.

If the wave is going down the z-axis, then by definition the variation is along the z-axis.

So h_{xx} is equal to some coefficient $h^0{}_{xx}$ multiplied by the sine of $k(z - t)$.

We can write the formulas for the three non zero terms.

$$\begin{aligned} h_{xx} &= h^0_{xx} \sin k(z-t) \\ h_{yy} &= h^0_{yy} \sin k(z-t) = -h_{xx} \\ h_{xy} &= h^0_{xy} \sin k(z-t) \end{aligned} \tag{25}$$

Let's look at it at a fixed instant of time, for example time t equal 0.

As we move down the z-axis, there is a perturbation on the metric.

The metric tensor is a little bit different from just a flat space metric – and it is a real curvature not just a coordinate effect.

It oscillates as we go down the z-axis, figure 4.

The components that oscillate are the components that have to do with the metric of the plane perpendicular to the wave.

What does it mean to have an h_{xx} added to η_{xx}, which has value one in the most natural coordinates.

We write it

$$g_{xx} = 1 + h^0_{xx} \sin k(z - t) \tag{26}$$

It means that proper distances along the x-axis are a little bit different than what they would be if there was no perturbation term.

Suppose we are talking about the distance between the point $x = -1/2$ and the point $x = +1/2$ (and $y = 0$).

The proper distance is the real distance between the two points.

Remember, the coordinates x's are just labels.

The proper distance is given by the metric.

And this proper distance may be either a little bit more than 1 meter or a little bit less than 1 meter, if we are working in meters.

That depends on where we are on the z-axis.

For instance, if we are at a point z where $h^0_{xx} \sin kz$ is positive (remember, $t = 0$), that would mean that a stick of length 1 meter would be a little bit shorter than the distance between the two points.

It would not be because the meter stick would be shorter – a meter stick is a meter stick –, but because the plane would be stretched in the x direction.

Now what about the other dimension, at the same position z (and the same time $t = 0$)?

We have

$$g_{yy} = 1 + h^0_{yy} \sin k(z - t) \tag{27}$$

But h_{yy} is minus h_{xx}, from equation (24).

So we can write

$$g_{yy} = 1 - h^0_{xx} \sin k(z - t) \tag{28}$$

In other words, at a point z where the x direction is stretched, the y direction is compressed.

A meter stick oriented along the y-axis would exceed a little bit the distance between the points with y coordinates 1/2 and −1/2.

And conversely, when x is compressed, y is stretched.

This is shown in figure 5.

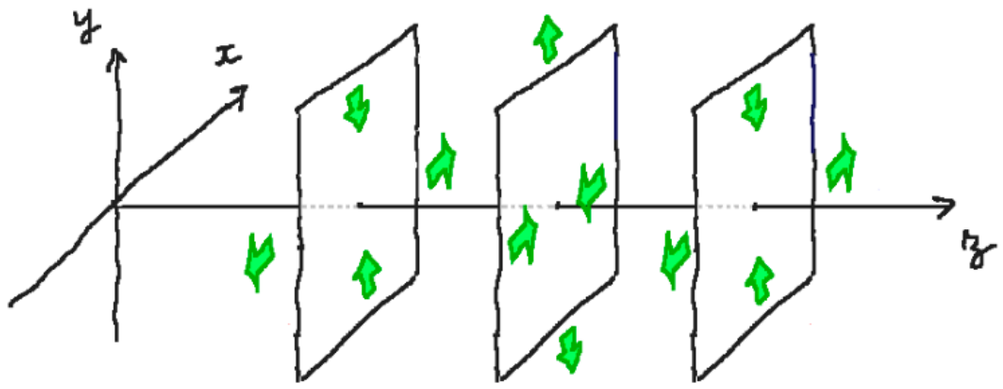


Figure 5 : Alternative stretching and compression.

We can also reason at one fixed point z and see what happens over time.

As time flows, the wave passes us.

So alternatively our x-axis is stretched or compressed, and our y-axis is compressed or stretched.

This is a *real* physical effect.

The gravitational wave *creates* curvature.

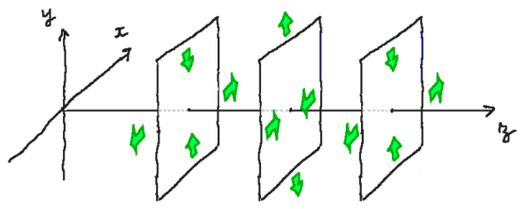
The wave is, if you like, a *ripple* of curvature of space-time moving with the speed of light.

At a given point on the z-axis, when the wave passes it, that is when time flows, it *creates* a kind of tidal force there.

The nature of the tidal force is to actually *cause* a meter stick in the plane at that position to be compressed and stretched.

When it is compressed horizontally, it is stretched vertically, and when it is compressed vertically, it is stretched horizontally.

On the other hand if, at a given time, we look at all the points along the axis, we also see this alternation of compression and stretching, figure 5.



At a given point, if you took a piece of plywood, figure 6, you would see it alternatively be compressed horizontally and stretched vertically, and stretched horizontally and compressed vertically.

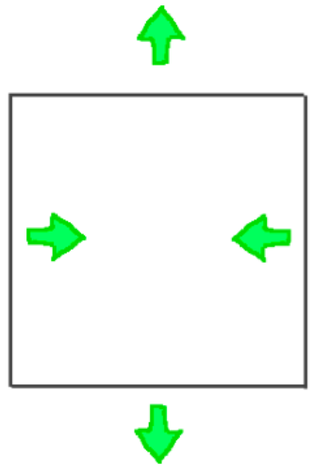


Figure 6 : Tidal forces on a piece of plywood.

It is interesting to observe that there is also another solution.

Then it is not h_{xx} and h_{yy} which are non-zero, but h_{xy}.

$$h_{xy} = h^0_{xy} \sin k(z-t) \tag{29}$$

And h_{yx} is the opposite.

The metric would look like

$$\delta_{xy} + \begin{pmatrix} 0 & h_{xy} \\ -h_{xy} & 0 \end{pmatrix} \tag{30}$$

where δ_{xy} is the Kronecker symbol, or equivalently the unit matrix.

It still has $h_{xx} + h_{yy} = 0$, because both are zero.

It is the same phenomenon as in figure 6, except that the oscillating squeezing and stretching are along the 45° axes.

It is not a new solution.

It is just the initial solution rotated by 45°.

Then any linear superposition of these two solutions is also a solution.

That is all the gravitational waves there are.

You first pick the longitudinal direction in which the wave is moving, then you pick a set of perpendicular axes in the perpendicular plane, and you construct the wave alternatively stretching and compressing as we described.

Q. : How can we measure this squeezing and compressing?

We can use a strain gauge.

The wave will create honest real stresses in the piece of plywood.

A strain gauge will register them when the wave is going past.

If the wave were static instead of moving, then the piece of plywood would really be unaffected.

Everything would be simultaneously squeezed, or stretched, the same way, the rulers which measure the plywood, the strain gauges which measure the stress, etc.

But it is the oscillating character of the solution which really does create real honest stresses and strains in it.

It does have real curvature.

Q. What triggers the gravitational wave ?

In electromagnetism a moving charge, for example an electron rotating around a proton, or a current alternating in an antenna, creates electromagnetic waves.

Similarly in gravitation, a moving mass will create a gravitational wave, for instance a star or a black hole or a pulsar or whatever, in orbit around another one.

For example the famous binary pulsar of Hulse and Taylor, discovered in 1974, is made of two very concentrated masses, which rotate around each other at a rather small distance.

Neither one is a black hole.

The whole thing is not a black hole.

But it has a very strong varying gravitational field and it produces gravitational waves.

Now near the binary pulsar, $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ is a bad approximation.

The gravitational field is too strong to linearize the equation this way.

But if you get at some distance away, then the wave spreads out, dilutes itself and gets weak enough that it becomes a good approximation.

If you are far away, and you take the z-axis to be the line between you and the pulsar, you will feel a gravitational wave passing you, which has the expression (26).

There will be stresses and strains in the plane perpendicular to the line of sight to the pulsar.

Q. : How to detect deformation if the meter sticks are deformed too?

Meter sticks made of matter would indeed be affected.

But a wooden meter stick and a steel meter stick would probably react differently.

So in theory that is one way to detect the tidal forces created by the wave.

Another way to detect the gravitational wave is to create some system which has a resonance at the frequency of the gravitational wave.

A system usually has its own natural frequency of oscillation.

If it is reinforced by an oscillating force of the same frequency, bringing in some energy to compensate that lost in the vibrations, it enters into a sustained or so-called driven resonance.

Under those circumstances the response will be particularly big.

Notice that the gravitational waves we expect from various sources in astronomy are extremely weak.

The additional terms $h_{\mu\nu}$ are extremely small, so small that the dimensionless effect on a rod would be something like 10^{-21}.

The fact that people even contemplate measuring them is quite astonishing.

However they can be measured.

That is the purpose of the LIGO (Laser Interferometer Gravitational-Wave Observatory) experiments conducted by Caltech and MIT in Livingston, Louisiana, and Hanford, state of Washington.

LIGO is a gravitational wave detector. It is not a steel rod.

It is a pair of mirrors, plus laser beams, like a modern version of the Fabry-Perot interferometer.

And an interference effect, due to the relative motion in the x and y directions, is produced and measured.

But basically a gravitational detector is a system which is allowed to be set into oscillation.

The gravitational collapse of a black hole or the collision of two black holes happen from time to time in the universe.

We can calculate how many such events happen per unit of time, that are within range of the best imagined detectors.

I think we could detect, in theory, one black hole collision per year.

It is a lot.

The wave moves at the speed of light.

Therefore, if we had at our disposal other means of detection of the black hole collision than the gravitational wave produced, we ought to see the signal at the same time as the wave.

Gravitational radiation is a pretty weak effect when you are far back.

From the collision of two black holes, it is an enormous effect when you are close to it.

It is then much bigger than any other kind of radiation that is emitted.

But if the collision is taking place at large cosmological distances, it can be the case that the only way to detect it would be through gravitational radiation.

That is why LIGO is interesting : for astronomy it is a new instrument opening a new window on the universe.

Before the LIGO detection of 2015, gravitational waves had never been observed "directly".

But they had already been observed "indirectly" through other physical effects.

Just like an orbiting system of electric charges emits electromagnetic radiations which carry off energy, masses which are rotating about each other emit waves and lose energy.

Their rotation will speed up a little bit due to this loss of energy.
The study of the binary pulsar of Hulse and Taylor – which are two stars going around each other – showed a perfect match between the gravitational waves it should emit, the energy loss, and the rotation acceleration.

This last effect can be measured by the change of frequency of the pulsating light received.

Gravitational waves is the main thing which causes the energy of the pulsar to decrease.

The orbital period the Hulse-Taylor binary pulsar is 7.75 hours.

Since its discovery in 1974, we have observed its period decrease.

As of 2016, it had lost a little less than a minute, exactly as predicted.

Q. : The components h00, h0x, h0y, h0z are all zero. Does this mean there is no curvature along the time axis?

No.

There is curvature in the time direction.

The first derivatives $\partial h_{mn}/\partial t$ are not necessarily equal to zero.

The curvature tensor has a whole bunch of products of such first derivatives and also has second derivatives.

So it will have non zero components involving time.

But still it is a very special kind of curvature that is not generic.

Q. : Considering that the Einstein equations are non linear, how did we end up with solutions which are linear?

This is only because we assumed that the discrepancy between flat space-time and the actual (truly non flat) solution to equation (1) was very small.

The linearity is because the wave is weak.

Generally speaking, small oscillations about an equilibrium, in a first approximation, can be taken to be linear.

In the Taylor expansion

$$f(x+ h) = f(x) + hf'(x) + h^2/2\, f''(x) + h^3/6\, f'''(x) +...$$

if h is small we can take h^2 and the higher powers of h to be zero.

Then we have

$$f(x+ h) \approx f(x) + hf'(x)$$

That is the kind of approximation we make, for instance, when we assimilate a swinging pendulum(with smalls wings) with a mass attached to a spring whose restoring force is an exact linear function of the displacement around the rest position.

This is also the approximation we make when we linearize any type of vibration around a static equilibrium.

Q. : How far off from an accurate solution are we? How much error do we make?

It depends how big is h. If h is 1/10, and we wanted to add terms in h^2, the corrections would be 1%.

Q. : If h is large enough that we have to add higher order terms, will that affect the transversality of the waves?

No. The solutions will still be transverse, but won't simply add to give new solutions. For instance two waves in opposite direction will no longer go through each other, they will scatter

That is it for gravitational waves.

You can learn more about them by any good book on general relativity
or the extra notes on the website.

They will write down all the details of the solutions to the equations.

These will fill a page or two, and won't be terribly illuminating.

But we covered the basic principles.

Now for a quick exposition of the full derivation……..

Einstein-Hilbert action for general relativity

To finish this part of the class, let me present another way of thinking about the equations of general relativity.

Again, solving the equations is too complicated, and we won't do it here.

Nevertheless, the basic idea is easy to grasp.

It is to apply to general relativity the most central principle of physics that we have, that is, the principle of least action.

Indeed, all the systems that we know about, mechanical systems, electromagnetic radiation, quantum field theory, the standard model of particle physics, etc., are governed by an action principle.

In ordinary particle mechanics, the action attached to a trajectory is an integral over time along the trajectory of the particle.

In field theory, the action attached to a field is an integral over space and time.

A field configuration is a value of the field at every point of space and time.

We also saw that the first case is a special instance of the more general second case.

The trajectory of a particle can be seen as a field whose substrate space-time is reduced to simply one time axis, instead of a space-time with spatial and time coordinates, see figure 7.

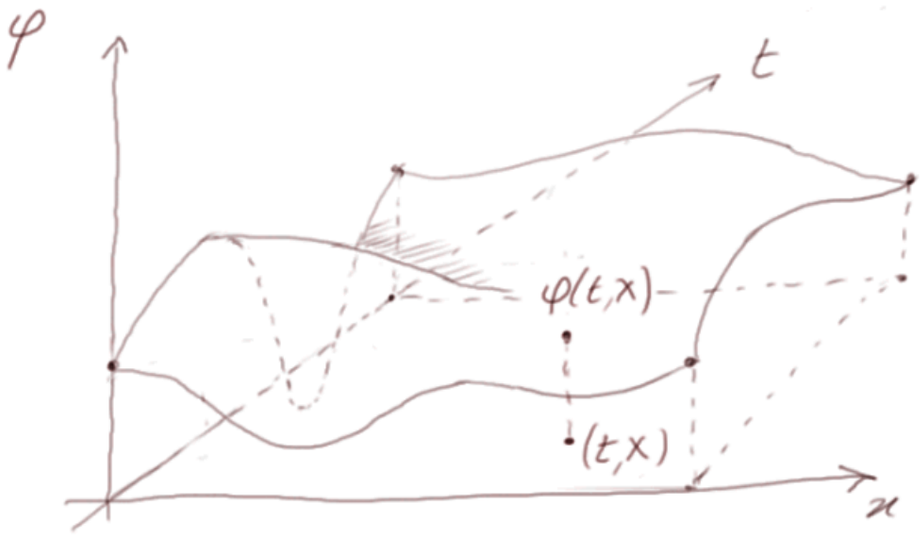


Figure 7 : Field $\phi(t, x)$. If we omit the spatial dimension x, we have, in the ϕ t plane, the trajectory of a particle over time.

Not every configuration of the field is a solution of the equations of the theory, just like not every trajectory is a solution of the equations of motion.

Every trajectory is a thinkable trajectory, but not every trajectory satisfies the equations of motion.

In the same way every configuration or, better yet, every history, $\phi(t, x)$ is a thinkable possible value of the field in space and time.

The variable t allows us to view $\phi(t, x)$ as the history of a spatial field evolving over time, just like the trajectory of a particle can be viewed as the history of its position.

But they are not all solutions.

In both cases, the behavior of the particle (its actual trajectory), or the behavior of the field (its actual configuration), is that which minimizes the action.

More accurately, as we stressed many times, it is the configuration which makes the action stationary.

In the case of a particle, the quantity integrated over is called the lagrangian.

In the simplest situations, it is the kinetic energy minus the potential energy of the system.

In the case of a field, the quantity integrated over is called the lagrange density (because there is a spatial integration as well as an integration over time).

Let's focus on the case of a field with time and space in its substrate set, that is $\phi(t, x)$.

And x is a generic variable, which we could also denote X, standing for x, y and z.

When we write X^{m} we mean x, y and z.

And when we write X^{μ} we mean t, x, y and z.

The action attached to this field has the form

$$A = \int dX \; dt \; \mathcal{L}(\phi, \; \phi_\mu) \tag{31}$$

where $\mathcal{L}$ is the lagrange density.

It depends on ϕ as well as on its derivatives with respect to the four X^μ, which we denote ϕ_μ.

What do we do with this A?

We minimize it, or we make it stationary.

We find the field configuration which minimizes (we won't repeat each time "or make it stationary") the action subject to a constraint.

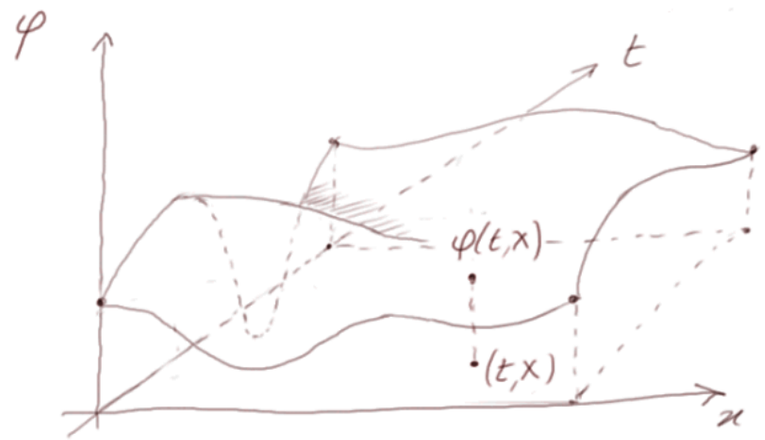
The constraint is that we are given the values of the field on the boundary of a region of space-time.

And the minimization process will yield the value of the field in the whole region.

This region boundary plays the same role as the initial and final positions of a particle, at time t_1 and time t_2, in the case of particle motion.

And the solution of the minimization of the action yielded the whole particle trajectory between t_1 and t_2.

In the case of our field $\phi(t, x)$, or $\phi(t, X)$, the boundary can be the rectangular region in the $x\ t$ plane of figure 7, or it can be a more elaborate boundary.



Einstein field equations can be derived from an action principle.

It is a long computation but the action given by equation (31) is simple.

From the action, or more accurately from the lagrange density, we write the Euler-Lagrange partial differential equations in the multidimensional space-time case which generalize

$$\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{X}} = \frac{\partial \mathcal{L}}{\partial X} \tag{32}$$

And after a few hours of tedious but straightforward calculations we land on the Einstein field equations.

Let's look at what the lagrange density is.

It has some interesting pieces in it.

But first of all let's do a simpler thing : let's review what the space-time volume is and how it is calculated.

Going one more step back, let's consider first an ordinary space, say a two dimensional Riemannian variety.

The index m of the coordinates runs over 1 and 2, and so does n.

And when convenient, the first coordinate is simply denoted x and the second y.

The variety has a metric

$$dS^2 = g_{mn}\, dX^m\, dX^n \tag{33}$$

The square of the distance between two neighboring points is given by this formula (33).

Now suppose we are interested in the volume of a small region of space, let's say a small rectangle with sides x to $x+dx$, y to $y+dy$, see figure 8.

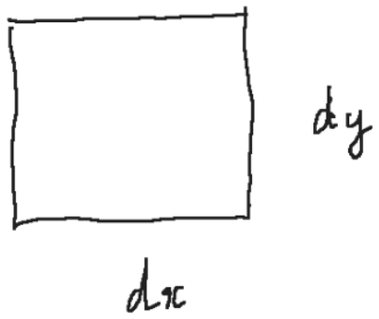


Figure 8 : Small area, also called small "volume".

And suppose the metric has only two diagonal components g_{xx} and g_{yy}.

We want to know the area.

We used the generic name "volume", but in two dimensions, it is customarily called the area, of course.

We would know it if we knew the actual proper distance associated with *dx* and the actual proper distance associated with *dy*.

What is the proper distance associated with *dx*?

It is given by equation (33) : it is $\sqrt{g_{xx}}\ dx$.

So the area, or "volume", of the rectangle is

$$volume = \sqrt{g_{xx}\ g_{yy}}\ dx\ dy \tag{34}$$

The important thing is that it is *not* just *dx dy*.

No more so than the proper distance between two neighboring points on the *x* axis would be just *dx*.

The coordinates are only labels.

It is the metric which gives the distances – that is why it is called the metric.

It warps everything if you like.

Only in Euclidean geometry do the labels correspond directly to distances.

If you take a flat rectangular piece of rubber, print on it Euclidean coordinates axes, and then deform the piece of rubber with bumps, throughs, stretching and compressing, the coordinates axes no longer give directly distances, but they remain useful labels to mark points and coordinate axes.

We used the case where the metric only has diagonal terms, because it is easier to calculate the volume, and to illustrate the difference between *dxdy* and the volume.

In other words, equation (34) is just a special case when the metric has the form

$$g_{mn} = \begin{pmatrix} g_{xx} & 0 \\ 0 & g_{yy} \end{pmatrix} \tag{35}$$

There could be, however, other off diagonal components to the metric.

Its general form is

$$g_{mn} = \begin{pmatrix} g_{xx} & g_{xy} \\ g_{yx} & g_{yy} \end{pmatrix} \tag{36}$$

In that case, the area, or "volume", is obtained from the determinant of the matrix on the right hand side of equation (36).

For a 2 by 2 matrix, you remember that the determinant is $g_{xx}g_{yy} - g_{xy}g_{yx}$.

For 3 by 3, it is a bit more complicated. It is given by Cramer's rule.

For 4 by 4 it is obtained as a linear combination of 3 by 3, etc.

But we don't need to know the formulas.

The notation for the determinant is |g|.

And the general formula for the volume in figure 8 is

$$volume = \sqrt{|g|}\ dx\ dy \tag{37}$$

This formula holds in any space curved or not.

If we consider a whole region on the variety, and we want to know its area, or "volume", you guessed how we are going to calculate it.

We integrate formula (37).

It yields

$$volume = \int_{region} \sqrt{|g|}\ dx\ dy \tag{38}$$

Now if we are talking about the volume of a three-dimensional space, the formula is the straightforward generalization of equation (38), with a *dz* appearing in the differential element.

$$volume = \int_{3-dim\ region} \sqrt{|g|}\ dx\ dy\ dz \tag{39}$$

If we are in a four-dimensional space-time, it is useful to define a similar concept, which would now involve the fourth component *dt* as well.

The rank two metric tensor now corresponds to a 4 by 4 matrix.

In general relativity the metric is not Riemannian but Minkowskian, i.e. its signature is − + ++, nevertheless at each point in space-time it has a value and a determinant.

And the volume can be defined generalizing again equation (39).

$$volume = \int_{4-dim\ region} \sqrt{|g|}\ dt\ dx\ dy\ dz \tag{40}$$

It is called the space-time volume, and it is an interesting quantity.

It is an invariant.

It is the same in every coordinate system.

That is an important characteristic of the space-time volume.

If you "recoordinatize" the region, that is, if you change coordinates, in some sense the physical space-time volume of any region stays the same.

Other quantities which will be invariant under change of coordinates are based on |g|.

If we take, for instance, the volume element $\sqrt{|g|}\, dt\, dx\, dy\, dz$ and integrate it multiplied by any scalar $S(t, X)$, where X stands for $x\ y\ z$,

$$\int \sqrt{|g|}\ S(t,\ X)\ dt\ dX \tag{41}$$

we also get an invariant quantity which doesn't depend on the coordinates.

Indeed, it is just a sum of little volume elements, each multiplied by the value of a scalar field.

Since the scalar is, by definition of a genuine physical scalar value, an invariant, the integrand in formula (41) is invariant.

Therefore expression (41) is invariant, even though writing it down explicitly necessitates the use of a set of coordinates.

In relativity theory the action is always supposed to be the same in every coordinate system.

Otherwise the laws of physics which came out of it would depend on the coordinate system.

If the laws of physics are supposed to be the same in every coordinate system, one way of ensuring that is to make the action an invariant.

So it is natural to take for the action

$$\int dX^4 \sqrt{|g|} \text{ x } some\ scalar \tag{42}$$

where dX^4 stands for the complete differential element.

Let's take the Einstein theory without any energy-momentum tensor.

No sources, just a pure gravitational field, nothing else in space-time.

What kind of scalar can we introduce in expression (42)?

How many scalars are there that we could make up only out of the metric ?

There is for instance the curvature scalar.

Anything else?

Answer : well, simply any number, 7, 4, or 16, or any number you like.

So we can use for the scalar

$$S(t,\ X) = \mathcal{R} + a\ number$$

because numerical numbers are always regarded as invariant.

If you see the number 7 at some point in space-time (not the component of something, but a pure number), everybody else sees the same number 7.

The number we add next to the curvature scalar depends on the laws of physics.

It is itself a law of physics.

It has a standard notation : it is denoted Λ (read lambda).

And the formula for the action is

$$\int dX^4 \sqrt{|g|}\ \ [\mathcal{R} + \Lambda] \tag{43}$$

This additional number Λ is called the cosmological constant.

It creates a term in the lagrange density which is just proportional to the volume itself.

In the last lesson, we did not introduce the cosmological constant.

We just discussed it briefly at the end.

If we don't introduce it, the action reduces to

$$\int dX^4 \sqrt{|g|}\, \mathcal{R} \tag{44}$$

Einstein's field equations are what we obtain when we minimize expression (44), or make it stationary.

Therefore at the solution we must have

$$\delta \int dX^4 \sqrt{|g|}\, \mathcal{R} = 0 \tag{45}$$

Where would the energy-momentum tensor appear in the lagrange density $\sqrt{|g|}\mathcal{R}$?

The energy-momentum tensor would come in as additional terms in the lagrange density which would depend on the material and other fields.

For example there could be electromagnetic energy.

There could be other kinds of energy.

They would come in as additional terms under the integral sign in equation (45).

Equation (45) is the vacuum case. It is what governs gravitational waves.

Other things could indeed enter into the lagrange density, to start with, sources of course.

But they could also be made up from different fields.

However, the lagrange density will always have the square root of determinant of g in factor.

Its general form will be

$$\sqrt{|g|}\ [\ \mathcal{R} + other\ terms\ ] \tag{46}$$

For instance, in the case of the electromagnetic field is described by an anti symmetric tensor $F_{\mu\nu}$.

Expression (45) would then be

$$\sqrt{|g|}\ [\ \mathcal{R} + F_{\mu\nu}\ F^{\mu\nu}\ ] \tag{47}$$

That would govern the electromagnetic field.

But let's leave that out.

And let's just finish with the Hilbert-Einstein action for the vacuum case

$$\mathcal{A} = \int dX^4 \sqrt{|g|}\ \mathcal{R} \tag{48}$$

It was discovered, I think, just about simultaneously by Einstein and Hilbert.

But Einstein already had the field equations which he had derived following the method presented in lesson 9 :

$$\mathcal{R}^{\mu\nu} - \frac{1}{2}\ g^{\mu\nu}\ \mathcal{R} = 8\pi G\ T^{\mu\nu} \tag{49}$$

He already had the whole works, and knew how it all fitted together.

Hilbert and Einstein independently sat down and asked themselves : is it possible to derive equations (49) from an action principle?

Both came up with the same answer that it is.

They derived equations (49) from the action given by an extension of equation (48) when there are sources in space-time, that is, when elements from the energy-momentum tensor complete the lagrange density.

So equation (5) is a physical principle that contains all the Einstein field equations.

To clarify further what we said in this section about fields and applying the principle of least action to them, what is the field $\phi(t, X)$ in general relativity, which appears in equation (31)?

It is the metric.

$$\phi(t,\ X) = g_{\mu\nu}(t,\ X) \tag{50}$$

Equation (48) describes an action that is made up out of a metric.

The principle to find the actual metric created by the physics is to vary the metric a little bit in equation (48) and look for that which minimizes the action.

If we move away from the solution g, the action given by equation (48) should grow bigger than at the solution.

The mathematics is to solve the Euler-Lagrange equation for a field

$$\sum_{\mu} \frac{\partial}{\partial X^{\mu}} \frac{\partial \mathcal{L}}{\partial \frac{\partial \phi}{\partial X^{\mu}}} = \frac{\partial \mathcal{L}}{\partial \phi} \tag{51}$$

where

$$\mathcal{L}(\phi, \phi_{\mu}) = \sqrt{|g|}\mathcal{R}$$

.

And where $\mathcal{R}$ is itself a function of g.

In conclusion, the Einstein-Hilbert form of general relativity is equation (44) which we reproduce below

$$\delta \int dX^4 \sqrt{|g|}\, \mathcal{R} = 0$$

Steps to derive the field equations from the action

Here is the complete procedure, starting from the last equation, to arrive at Einstein field equations :

1. Express the Christoffel symbols in terms of the metric. It involves first derivatives of the metric components.
2. Express the curvature tensor in terms of the Christoffel symbols. It involves first derivatives of the Christoffel symbols and products of Christoffel symbols.
3. Express the curvature scalar by contracting the curvature tensor.
4. Multiply it by the square root of the determinant of the metric. It is a matrix of four rows and four columns. The determinant is a great big thing made up out of lots of components of the metric.
5. Apply the Euler-Lagrange variational equations (50) to the lagrange density that you wrote.

After a few hours of calculations, you will end up with Einstein field equations.

We are finished with general relativity.