Contents

Prologue: The Shape Of Earth

Some people think that Earth is round while some think it is flat. Who is right? Obviously, the scientific community, schools and probably your parents support the theory of round Earth, which is a fair reason for believing that maybe Earth really is round.

Anyway, believing is not fun, and it's definitely not what theoretical physicists should do, no matter how credible sources we are talking about. Theoretical physics means understanding. Being right and knowing the truth is not the point, and understanding is not a means to achieve the truth, whatever that would mean; instead, it is the goal itself.

In theoretical physics we prefer simple, incorrect models over correct, complex models, because complex models cannot be understood. For a theorist Earth is lifeless, universe is homogeneous, only male and female genders exist and a function is equal to its second-order Taylor polynomial.

Complex models are not wortlhess anyway. Observational data is an indispensable means to achieve understanding, and it is an utterly complex, utterly true model. Complex models are raw data, to be mined and hammered into simpler descendants, without caring too much about pieces of truth that get destroyed along the way.

The beautiful, and completely unexplained thing is that in the end the simplest models tend to turn out to be the most truthful ones. QED is defined by one formula,

$$\mathcal{L} = \bar{\psi}(\mathrm{i}\not{D} - m)\psi - \frac{1}{4}F^2,$$

and it constains almost all phenomenology we know of, including for example all chemistry, and gives us predictions that are correct to the eight decimal. The standard model is even more truthful, but it is also more complex, so we like QED more.

So let me explain why I think, and not believe, that Earth is round, without caring about whether I'm right or wrong about that. First, let's be clear about what we're talking about. It's not the shape of ground: since there are mountains and valleys, ground is obviously not flat nor round. We're talking about the surface to which the gravitational force is perpendicular. Flat Earth would mean that gravity pulls the same direction everywhere, and round Earth would mean that gravity pulls towards the center of Earth.

The question of the shape of Earth is therefore deeply connected to the nature of gravity. One theory, which is consistent with flatness but not roundness of Earth, is that there is a rocket engine on the other side of the pancake Earth accelerating it and causing things fall down to ground, just like acceleration of a car pulls things to the back seat.

This theory doesn't satisfy me. How does it work? How does the pancake Earth withstand the forces the rocket engine exerts on it? Why don't we observe the tremor of the engine? Why there is a rocket engine? The "Why?" question must be asked about any theory, yet with this one it is particularly important, since rocket engines are extremely complicated systems.

Another theory of gravity says that any two pieces of matter attract each other. The attraction causes large, coarse clouds of matter to collapse towards smaller size and higher density. At some point the density is so high that pressure is enough to cancel the attraction. A state of equilibrium is reached and the cloud has become a dense, round, perhaps solid object. Round, because the attractive force doesn't care about directions, and round is the only finite-sized shape that doesn't care about directions.

So theory suggests that our universe should be filled with round-shaped dense chunks of matter. Look up the sky and see Sun, Moon and stars. If we now put some little test particles around some collapsed cloud, all the pieces that make up the cloud pull the test particles towards the cloud. The test particles also attract each other, but their mutual attraction is negligible when compared to the combined pull of all the matter in the cloud. A gravitational force that pulls everything towards the center of the round, collapsed cloud of matter has emerged.

These considerations are enough to make me think that pieces of matter attract each other and that Earth is a collapsed cloud of dust. It doesn't matter where the idea came from; the important thing is that it makes sense. Anyway, if someone's interested, I can reveal where the idea came into my mind: it came from Isaac Newton.

Newton formulated his universal law of gravitation during the 17. century. It states that any two bodies of masses m_1 and m_2 attract each other with a force of magnitude

$$|F| = G\frac{m_1 m_2}{r^2},$$

in which G is a universal constant and r the distance between the bodies, and that forces that act on a body are combined simply by adding. Together with Newton's second law

$$F = m\frac{\mathrm{d}^2 x}{\mathrm{d}t^2}$$

it predicts the movements of our Solar System as accurately as humans can measure them—except for a little anomaly in Mercury's motion.

Now, the question "why?" The $\frac{1}{r^2}$-dependence is the same as how the intensity of radiation, for example light of fire or volume of speech, depends on distance to the source. This is because all the radiaton gets distributen on a sphere of radius r, and the area of the sphere is proportional to r^2. So if the force is mediated by some kind of radiating waves or particles, "gravitons," this could be the reason why it is proportional to $\frac{1}{r^2}$.

Why the proportionality $m_1 m_2$? If all the forces are combined by addition, then two similar particles put together must create twice the pull of a single particle. So doubling the mass of the pulling particle must lead to doubling of the force. Also, if we double the particle that is being pulled, both of them are pulled by the same force. If we now put them together and consider them as one particle, the force must get doubled, because according to

Newton's second law it takes twice as much force to move the double particle.

Why gravitational forces are combined by simple addition? If the real rule for combining the forces is something else than simple addition, it is likely that for weak forces addition may be accurate enough. This is what linear approximation means. Gravity is weak—the gravitational constant G has the value of about 10^{-10} in SI units, and the gravitational force between an electron and a proton is about $10^{-??}$ times the electric force, which also has a $\frac{1}{r}$-dependence—so it's possible that we only observe the regime in which the linear approximation is sufficient.

These are not exhaustive answers, and also other questions remain. Why is gravitational force universal? Why it exists? Does it even exist? Despite the magnificent success of Newton's law of gravitation and everybody's concrete experience of falling down to ground, in this book we will discover that gravitational forces are not real but fictional, arising from a peculiar perspective, like an optical illusion. As we will see, in some sense the rocket engine theory of gravity is better than Newton's theory.

Chapter 1

Spacetime

In physics the space of all events—which includes your birth, your death, and the moment you're reading these words—is called spacetime. It is, or at least seems to be, four-dimensional: it takes four coordinates—for example the Greenwhich Mean Time, the elevation, the latitude and the longitude—to label the events. The first of these four is obviously different to the rest: it is timelike, while the other three are spacelike.

1.1 Mathematical foundations

For simplicity, let's ignore the latitude and longitude and use only one timelike and one spacelike coordinate, denoted here by t and r. We may draw a diagram of events by simply plotting them on paper. Conventionally t is taken to increase vertically and r horizontally.

The life of a particle appears as a continuous line of subsequent events in the diagram. The line is called the world line of the particle. The length of the world line, denoted by τ, is defined to be measured by a real or imagined clock carried by the particle. As I'm writing this, my world line is 26.3 years long.

There is nothing subjective or relative about the length of a world line. It measures how much progress local physical processes and natural phenomena have made along it. Ticking of a clock is one process, as are decay of alcohol levels in blood, learning to walk and growth of a cancer tumor.

Everyday experience and the choice of the Greenwich Mean Time as t suggest that if we take two near events separated by $\mathrm{d}t$ and $\mathrm{d}r$ and draw a world line between them, the length of the world line is $\mathrm{d}\tau = \mathrm{d}t$, and that any two world lines that begin and end at the same events have the same length, which is the difference Δt of the time coordinate values of the endpoint events. If I meet today a person who was born in the same hospital at the same day as i was, he or she will also turn out to be 26.3 years old.

This Galilean or Newtonian view of time is not exactly true. For example in 1971 researchers observed that identical atomic clocks carried by different airplanes taking different routes made slightly different numbers of ticks. Therefore the equality $\mathrm{d}\tau = \mathrm{d}t$ can only be approximately true.

Could we find a more accurate formula for $\mathrm{d}\tau$? A simple attempt would be to consider $\mathrm{d}\tau$ as the hypothenuse of a right-angled triangle with $\mathrm{d}t$ and $\mathrm{d}r$ as the other two sides. Assuming that units of measurement are chosen suitably, the Pythagorean theorem would give

$$\mathrm{d}\tau^2 = \mathrm{d}t^2 + \mathrm{d}r^2.$$

Now, if $\mathrm{d}r$ is much smaller than $\mathrm{d}t$, which corresponds to a worldline of a slowly moving particle, we get $\mathrm{d}\tau \approx \mathrm{d}t$. If the particle moves fast, then $\mathrm{d}\tau > \mathrm{d}t$; in other words, fast moving particles experience more time than those at rest.

This is not correct. According to experimental evidence, fast moving particles experience less, not more, time than those at rest. If we change the plus sign in the Pythagorean theorem into a minus one, we get

$$\mathrm{d}\tau^2 = \mathrm{d}t^2 - \mathrm{d}r^2.$$

Now fast moving particles experience less time.

If t and r are measured in suitable units, then $\mathrm{d}\tau^2 = \mathrm{d}t^2 - \mathrm{d}r^2$ is actually correct. If we would also want to include longitude and latitude, denoted by θ and ϕ, we would have to take something like

$$\mathrm{d}\tau^2 = \mathrm{d}t^2 - \mathrm{d}r^2 - r^2\,\mathrm{d}\theta^2 - r^2\sin^2\theta\,\mathrm{d}\phi^2,$$

or if we used Cartesian x, y, z space coordinates instead of polar ones,

$$\mathrm{d}\tau^2 = \mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2.$$

The signature $+---$ manifests the fact that spacetime has one timelike and three spacelike dimensions.

Let's continue with only t and r for clarity. There are a number of ways to "derive" $\mathrm{d}\tau^2 = \mathrm{d}t^2 - \mathrm{d}r^2$, yet I find them not satisfactory. Science advances by taking random ideas, testing them, and keeping only the ones that work. It's usually not possible to logically advance from observations to universal laws. Therefore I find it best to just present the core of the theory, proceed from that, and let you see if that works.

So let's see. If $\mathrm{d}r = 0$, we have $\mathrm{d}\tau = \mathrm{d}t$, which means that particles at rest can measure their worldine length, in other words their own time, correctly with t. If we increase $\mathrm{d}r$, then $\mathrm{d}\tau$ gets smaller. Time of a moving particle seems to tick slower than t.

If $\mathrm{d}t = \mathrm{d}r$, then $\mathrm{d}\tau = 0$. In other words, time of a particle moving with the speed $\frac{\mathrm{d}r}{\mathrm{d}t} = 1$ freezes. If $\mathrm{d}t < \mathrm{d}r$, then $\mathrm{d}\tau$ becomes imaginary, which does not make sense. Let's assume that for all realistic particles $\mathrm{d}t \geq \mathrm{d}r$: no particle moves with a speed greater than 1.

Since we need atomic clocks and airplanes to observe differences in worldline length, the speed limit, which in our units is simply equal to one, must be a very high speed. It must be greater or equal to the speed of anything we know of, including light.

Observations suggest that light actually travels with exactly the maximum speed, which is why the maximum speed is usually called the speed of light. That means that if t is measured in years, then r must be measured in lightyears.

For an aircraft travelling with the speed $\frac{\mathrm{d}r}{\mathrm{d}t} = v$, we have

$$\frac{\mathrm{d}\tau}{\mathrm{d}t} = \frac{\sqrt{\mathrm{d}t^2 - \mathrm{d}r^2}}{\mathrm{d}t^2} = \sqrt{1 - \left(\frac{\mathrm{d}r}{\mathrm{d}t}\right)^2} = \sqrt{1 - v^2}.$$

Light travels about a million times faster than sound travels in air, so if the aircraft travels with the speed of sound, we have

$$\frac{\mathrm{d}\tau}{\mathrm{d}t} = \sqrt{1 - (10^{-6})^2} \approx 1 - \frac{1}{2}\left(10^{-6}\right)^2 = 0.999,999,999,999,5,$$

in which the approximation is just the first-order Taylor expansion of square root at 1. The error made with using $\mathrm{d}t$ as $\mathrm{d}\tau$ is ridiculously small. But if the aircraft could move with half the speed of light, which would take it to the Moon and back in less than five seconds, we would get

$$\frac{\mathrm{d}\tau}{\mathrm{d}t} = \sqrt{1 - \left(\frac{1}{2}\right)^2} = \sqrt{\frac{3}{4}} \approx 0.86.$$

Now the error would be significant.

1.2 Lorentz transformations

Our t and r are one choice of coordinates. They define what it means for a particle to be at rest, but we know that resting is a very relative concept. In everyday situations we change the velocity of a coordinate system by simply taking $r \to \tilde{r} = r - v\,t$ with v the change of velocity and using the original clock t as the new one $\tilde{t}$ without any modifications. Particles that were rest in the original coordinate system appear to move with the velocity $-v$ in the new one. To change back, we take $\tilde{r} \to r = \tilde{r} + v\tilde{t}$ and $\tilde{t} \to t = \tilde{t}$.

This is not satisfactory, since $\mathrm{d}\tau$ cannot be calculated by the same formula as in the original coordinate system: by using the chain rule of calculus, we get

$$\begin{aligned}
\mathrm{d}\tau^2 &= \mathrm{d}t^2 - \mathrm{d}r^2 \\
&= \left(\frac{\partial t}{\partial \tilde{t}}\,\mathrm{d}\tilde{t} + \frac{\partial t}{\partial \tilde{r}}\,\mathrm{d}\tilde{r}\right)^2 - \left(\frac{\partial r}{\partial \tilde{t}}\,\mathrm{d}\tilde{t} + \frac{\partial r}{\partial \tilde{r}}\,\mathrm{d}\tilde{r}\right)^2 \\
&= \left(1\,\mathrm{d}\tilde{t} + 0\,\mathrm{d}\tilde{r}\right)^2 - \left(v\,\mathrm{d}\tilde{t} + 1\,\mathrm{d}\tilde{r}\right)^2 \\
&= \mathrm{d}\tilde{t}^2 - v^2\mathrm{d}\tilde{t}^2 - 2v\,\mathrm{d}\tilde{t}\,\mathrm{d}\tilde{r} - \mathrm{d}\tilde{r}^2.
\end{aligned}$$

For the new coordinates $\tilde{t}$ and $\tilde{r}$ to be of the same type as the original ones, we should have just $\mathrm{d}\tau^2 = \mathrm{d}\tilde{t}^2 - \mathrm{d}\tilde{r}^2$.

We can find such coordinates most easily by first transforming into so-called lightcone coordinates σ_+, σ_-, which means visually rotating the t, r coordinates by 45 degrees counter-clockwise. Mathematically we take

$$t \to \sigma_- = \frac{t - r}{\sqrt{2}} \qquad \text{and} \qquad r \to \sigma_+ = \frac{t + r}{\sqrt{2}}.$$

The $\sigma_\pm$ coordinate axes are the left- and right-moving light rays that pass through the event at $t = 0, r = 0$. Neither of the $\sigma_\pm$ is spacelike nor timelike; it can be said that they are both lightlike coordinates.

Recalling the binomial identity $(a+b)(a-b) = a^2 - b^2$, we get

$$\mathrm{d}\sigma_- \,\mathrm{d}\sigma_+ = \frac{\mathrm{d}t - \mathrm{d}r}{\sqrt{2}} \frac{\mathrm{d}t + \mathrm{d}r}{\sqrt{2}} = \frac{\mathrm{d}t^2 - \mathrm{d}r^2}{2}.$$

This is why lightcone coordinates are useful to us: the worldline element $\mathrm{d}\tau$ obeys the simple expression

$$\mathrm{d}\tau^2 = 2\,\mathrm{d}\sigma_-\mathrm{d}\sigma_+.$$

Now it is easy to obtain new lightcone coordinates $\tilde{\sigma}_\pm$ in which $\mathrm{d}\tau^2$ has the same form: we take $\sigma_\pm \to \tilde{\sigma}_\pm = \xi^{\pm 1}\sigma_\pm$ with ξ a positive-valued parameter. Then $\mathrm{d}\sigma_\pm = \xi^{\mp 1}\mathrm{d}\tilde{\sigma}_\pm$ and

$$\mathrm{d}\tau^2 = 2\,\mathrm{d}\sigma_-\mathrm{d}\sigma_+ = 2\,\xi\,\mathrm{d}\tilde{\sigma}_- \frac{1}{\xi}\,\mathrm{d}\tilde{\sigma}_+ = 2\,\mathrm{d}\tilde{\sigma}_-\mathrm{d}\tilde{\sigma}_+.$$

Now we just rotate back by

$$\tilde{\sigma}_- \to \tilde{t} = \frac{\tilde{\sigma}_- + \tilde{\sigma}_+}{\sqrt{2}} \qquad \text{and} \qquad \tilde{\sigma}_+ \to \tilde{r} = \frac{\tilde{\sigma}_- - \tilde{\sigma}_+}{\sqrt{2}}.$$

If we put all this together, we get

$$\begin{aligned}
\tilde{t} &= \frac{1}{\sqrt{2}}\left(\xi^{-1}\sigma_- + \xi\sigma_+\right) \\
&= \frac{1}{2\xi}(t-r) + \frac{\xi}{2}(t+r) \\
&= t\,\frac{1}{2}\left(\xi^{-1} + \xi\right) - r\,\frac{1}{2}\left(\xi^{-1} - \xi\right) \\
&= t\cosh\theta - r\sinh\theta
\end{aligned}$$

with $\theta = -\log\xi$. Similarly

$$\tilde{r} = r\cosh\theta - t\sinh\theta.$$

This is almost the same as an ordinary rotation of plane, which is understandable: we changed the plus sign in $a^2 + b^b = c^2$, which would have lead to ordinary rotations, into a minus one to get the spacetime version of the Pythagorean theorem. This is a useful notion, since it means that we can expect many similarities between rotations and Lorentz transformations and between ordinary space and spacetime.

The inverse transformation is the same transformation into the opposite direction:

$$\tilde{t} \to t = \tilde{t}\cosh(-\theta) - \tilde{r}\sinh(-\theta) = \tilde{t}\cosh\theta + \tilde{r}\sinh\theta$$

and

$$\tilde{r} \to r = \tilde{r}\cosh(-\theta) - \tilde{t}\sinh(-\theta) = \tilde{r}\cosh\theta + \tilde{t}\sinh\theta$$

We may now calculate $\mathrm{d}\tau^2$. We get

$$\begin{aligned}
\mathrm{d}\tau^2 &= \mathrm{d}t^2 - \mathrm{d}r^2 \\
&= \left(\frac{\partial t}{\partial \tilde{t}}\,\mathrm{d}\tilde{t} + \frac{\partial t}{\partial \tilde{r}}\,\mathrm{d}\tilde{r}\right)^2 - \left(\frac{\partial r}{\partial \tilde{t}}\,\mathrm{d}\tilde{t} + \frac{\partial r}{\partial \tilde{r}}\,\mathrm{d}\tilde{r}\right)^2 \\
&= \left(\cosh\theta\,\mathrm{d}\tilde{t} + \sinh\theta\,\mathrm{d}\tilde{r}\right)^2 - \left(\sinh\theta\,\mathrm{d}\tilde{t} + \cosh\theta\,\mathrm{d}\tilde{r}\right)^2 \\
&= (\cosh^2\theta - \sinh^2\theta)\,\mathrm{d}\tilde{t}^2 + (\sinh^2\theta - \cosh^2\theta)\,\mathrm{d}\tilde{r}^2.
\end{aligned}$$

Since $\cosh^2\theta - \sinh^2\theta = 1$, we get $\mathrm{d}\tau^2 = \mathrm{d}\tilde{t}^2 - \mathrm{d}\tilde{r}^2$.

The transformation we found is called Lorentz transformation. It relates two coordinate systems that correspond to two observers that move uniformly with respect to each other.

Lorentz transformation stretches the t, r spacetime diagram in the direction of the other light ray and shrinks by the same factor in the direction of the other one. It keeps the spacetime volume element $\mathrm{d}t\,\mathrm{d}t$ intact and tilts the t axis in the opposite direction it tilts the r axis.

1.3 Absolute and subjective

While others may disagree with me about a certain fact, none can disagree with me about what my view is on that fact. The length of particles world line, τ, is equal to the time experinced by the particle, and is by its very definition an absolute quantity. This is just repeating the same I stated in the beginning just using different words.

Another absolute thing is the $+---$ (or $-+++$; which one is used is just a notational convention) signature of spacetime. No matter what perspective we choose, there will always be one timelike dimension and three spacelike dimensions.

World lines of particles can be divided into two types: those with $\mathrm{d}\tau^2 > 0$, called timelike, and $\mathrm{d}\tau^2 = 0$, called lightlike. We can also imagine spacelike world lines with $\mathrm{d}\tau^2 < 0$, although such worldlines would have imaginary length, which does not make much sense. Type of a world line is an absolute quantity.

For any two events with a timelike or lighlike worldline between them, their temporal order defined by the time coordinate t remains invariant under a Lorents transformation. This makes sense: the arrow of time points from cause to effect. The arrow of time is at least for any practical purpose irreversible, a fact visible in the Second Law of thermodynamics. Causal order of events with a timelike or lightlike worldline connecting them is an absolute quantity.

These are the most important absolute quantities in spacetime. Most quantities that common sense regards as absolute are in reality subjective. The most important of them is of course the time elapsed between two events: different world lines may have different lengths, even if they happen to begin and end at the same two events.

This subjectivity suggests that simultaneity is also subjective: if one observer feels that one year has passed and anotherone feels that two years has passed, can we decide which events on their world lines are simultaneous? We could, yet the decision would be arbitrary.

A Lorentz transformation maps two events with the same values of t but different values of r into events with different values of $\tilde{t}$. Therefore events that appear simultaneous in the original coordinates do not map into simultaneous events in the new coordinates. Simultaneity of two separate events is a subjective concept.

If two events appear simultaneous in one coordinate system, their separation is spacelike. For any two events A and B with spacelike separation we can find a coordinate system in which they are simultaneous. We can also find a coordinate system in which A

occurs before B according to the timelike coordinate, and a coordinate system in which B occurs before A. The temporal order of events with spacelike separation is a subjective concept, and therefore it's not possible for such events to be causally related.

In everyday setting distance can be measured with rigid rods, but absolutely rigid rods cannot exist: if there was such a rod, then moving its other end would immediately cause also the other end to move. The causally related events would have a spacelike separation, which is not possible.

The generally most useful definition of distance is the result of radar ranging measurement. Radar ranging works by flashing a light and measuring how long it takes for the reflection to come back. The time is then divided by 2, because the pulse makes a round trip, and by the speed of light (which in our units is just 1).

The spatial distance of two world lines is not an absolute concept. Above I meant the distance that an observer at α measures to another observer at $\alpha + \mathrm{d}\alpha$ using radar. Radar works by flashing a light and measuring how long it takes for the reflection to come back. The time is then divided by 2, because the pulse makes a round trip, and by the speed of light (which in our units is just 1). Radar ranging is a very concrete measure of how far something is: if it takes more than two seconds for light to bounce back from the surface of the Moon, then Moon really is far away.

Because the reflection cannot come back at the same event it was sent, it doesn't make any absolute sense to talk about distances of events. Instead we have to talk about distances of world lines. Distance may depend on when the light pulse is sent, but because it comes back at another event, it does not make sense to talk about the distance between two world lines at some specific time or event.

Distance is subjective, which makes spatial sizes and shapes of objects subjective too.

1.4 Causal structure of spacetime

Given any event E, we can draw two light rays coming to it and two light rays emerging from it. What results is so-called light cone. All events inside the light cone can be connected to E by a timelike worldline and can therefore be causally connected to it. Events in the upper part of the cone are in the future of E, and events in the lower part in its past.

Events outside the light cone cannot be connected to E by timelike or lightlike worldlines. They have a spacelike separation to E, and cannot be causally connected with it. Observers at E cannot see nor affect anything outside the cone. It doesn't matter how insane or violent things are happening outside the cone. They are invisible and untouchable to E, period. The spacetime outside the cone is completely oblivious to E.

Of course an observer at E may live long enough for a specific event outside E's cone to enter into his or her current cone and become visible. Anyway, human lifetime less than a millionth of typical relevant worldline lengths in cosmology. We could try to circumvent this by hibernation or by choosing an accelerating worldline, yet biology severely limits our possibilities. On cosmological scale we cannot wait.

According to standard cosmology, unverse was once opaque. If we look far enough down our light cone, our sight hits this wall called recombination. What we concretely observe is cosmic microwave radiation. Because of this, we only see a finite portion of spacetime (at least if we only detect electromagnetic radiation). Furthermore, since light always travels along the lightcone's border, we do not see the interior of the past cone, but only the border.

The causal structure manifested in cones drawn by light rays is the most fundamental structure of spacetime. It defines which events can be in causal connection to each other, and which are nonexistent to each other. It may be impossible to change in anyway, no matter what we tried to do.

Chapter 2

Gravity

The coordinates we have used correspond to observers that move uniformly. They are not very convenient for accelerating observers. An accelerating observer would prefer accelerating coordinate system in which unchanging value of spacelike coordinate corresponds to a curved worldline of an accelerating particle.

2.1 Uniformly accelerating coordinates

Acceleration experienced by a particle is an absolute quantity. It can be calculated by Lorentz transforming into a coordinate system in which the particle is temporarily at rest and then calculating $\frac{\mathrm{d}^2 r}{\mathrm{d}t^2}$ just as we would in Newtonian mechanics. If the acceleration experienced by the particle does not change along its worldline, the particle is said to be accelerating uniformly.

In Newtonian-Galilean world a uniformly accelerating observer draws a parabola on t, r coordinate system. This cannot be exactly true, since it means that speed increases without any bound. In reality speed cannot exceed the speed of light and a uniformly accelerating observer draws some kind of hyperbola that asymptotically approaches the 45 degree slope of a light ray.

This makes heuristic sense, since as the observer's speed approaches the speed of light, its time slows down and the acceleration it experiences will have less and less real time per a time coordinate step to raise the speed.

The shape of the world line of a uniformly accelerating observer is an absolute concept. It looks the same in every (t, r) coordinate system in which $\mathrm{d}\tau^2 = \mathrm{d}t^2 - \mathrm{d}r^2$. The shape is therefore invariant in Lorentz transformation.

For any straight, finite line that begins at the event at $(0, 0)$ and ends at the event at (t, r), Lorentz transformations keep $t^2 - r^2$ invariant. This is an immediate consequence of the invariance of $\mathrm{d}\tau^2 = \mathrm{d}t^2 - \mathrm{d}r^2$. Therefore any line defined by constancy of $t^2 - r^2$ has a Lorentz-invariant shape. If $r^2 - t^2 = \alpha^2$ with real α, the line is timelike and could be a world line of a uniformly accelerating observer. Let's take such lines as the lines of constant spacelike coordinate of our accelerating coordinate system and use α as the spacelike coordinate.

All such lines correspond to a particle temporarily at rest when $t = 0$. If we solve the line's equation for r as a function of t and α, we get

$$r(t, \alpha) = \sqrt{t^2 + \alpha^2}.$$

Differentiating this twice with respect to t gives

$$\frac{\mathrm{d}^2 r}{\mathrm{d}t^2} = \frac{\alpha^2}{(t^2 + \alpha^2)^{3/2}}.$$

Setting $t = 0$ gives $\frac{1}{\alpha}$. In other words, observers corrseponding to smaller α experience greater accelerations, and as α approaches zero, the acceleration approaches inifity.

It may seem strange to use a coordinate system in which acceleration depends on the spacelike coordinate, but if we want that the radar distance between two nearby lines of constant spacelike coordinate does not depend on the time coordinate, we have no other choice. This can be easily seen from a picture.

Next we have to choose a timelike coordinate. Lorentz transformation changes the velocity of a reference frame, which for a uniformly accelerating frame should mean just translating the time coordinate. In other words a Lorentz transformation should map a line of constant timelike coordinate into another line of constant timelike coordinate.

We therefore define the lines of constant timelike coordinate as the lines we get when we inverse-Lorentz transform the r axis, and use the transformation parameter θ as the new timeline coordinate. We use inverse transformation in order to get a time coordinate that increases as t increases and not the opposite.

To get the transformation formula, we note that inverse-Lorentz transforming the point $(0,1)$ by θ gives $t = \sinh\theta$ and $r = \cosh\theta$, so

$$\frac{t}{r} = \frac{\sinh\theta}{\cosh\theta} = \tanh\theta.$$

We can now write the complete transformation rule, which is

$$t \to \theta = \operatorname{arctanh}\left(\frac{t}{r}\right) \qquad \text{and} \qquad r \to \alpha = \sqrt{r^2 - t^2}.$$

The θ, α coordinates are called Rindler coordinates. They cover only one quarter of the whole spacetime which will turn out to make sense.

The inverse transformation is simpler: we Lorentz transform $(0, \alpha)$ by θ, which gives

$$\theta \to t = -\alpha\sinh\theta \qquad \text{and} \qquad \alpha \to r = \alpha\cosh\theta.$$

For the world line element $\mathrm{d}\tau$ we get

$$\begin{aligned}
\mathrm{d}\tau^2 &= \left(\frac{\partial t}{\partial\theta}\,\mathrm{d}\theta + \frac{\partial t}{\partial\alpha}\,\mathrm{d}\alpha\right)^2 - \left(\frac{\partial r}{\partial\theta}\,\mathrm{d}\theta + \frac{\partial r}{\partial\alpha}\,\mathrm{d}\alpha\right)^2 \\
&= (-\alpha\cosh\theta\,\mathrm{d}\theta - \sinh\theta\,\mathrm{d}\alpha)^2 - (\alpha\sinh\theta\,\mathrm{d}\theta + \cosh\theta\,\mathrm{d}\alpha)^2 \\
&= \alpha^2(\cosh^2\theta - \sinh^2\theta)\mathrm{d}\theta^2 + (\sinh^2\theta - \cosh^2\theta)\mathrm{d}\alpha^2.
\end{aligned}$$

Since $\cosh^2\theta - \sinh^2\theta = 1$, we get the simple formula

$$\mathrm{d}\tau^2 = \alpha^2\mathrm{d}\theta^2 - \mathrm{d}\alpha^2.$$

2.2 Ramifications of acceleration

The formula $\mathrm{d}\tau^2 = \alpha^2\mathrm{d}\theta^2 - \mathrm{d}\alpha^2$, or the metric as it is often called, has many good properties. First, it is very simple. Second, it does not depend on the timelike coordinate θ, so phenomenology remains the same as we move along θ; in particular, the constant α lines really are world lines of uniform acceleration, and the radar distance of neighboring constant α lines does not change we move along θ. Third, it has no term with $\mathrm{d}\theta\,\mathrm{d}\alpha$.

Every accelerating coordinate system gives rise to fictious forces that seem to affect universally to everything. In rally car the force seem to randomly shake the driver and the kartanlukija in varying directions, and in a carousel there is a centrifugal force that tries to rip everything away from the centre. If angular momentum is fixed, then the centrifugal force becomes stronger near the center, and if the angular velocity is fixed, then the force weakens when moving towards the center.

In the Rindler coordinates the fictious force points towards lower values of α. It does not depend on θ, but it gets greater as α decreases. It pulls everything towards $\alpha = 0$, including light.

For constant α we have simply $\mathrm{d}\tau = \alpha\,\mathrm{d}\theta$. If two light rays are sent from $\alpha = \alpha_0$ and received at $\alpha_1 > \alpha_0$, both rays take the same amount $\Delta\theta$ of coordinate time to travel. Anyway, the greater the α, the faster the observer's own time ticks compared to the coordinate time θ, so the time experienced by the sender between the events of sending is less than the time experiend by the receiver between the events of receiving.

If many subsequent light rays are being sent, the frequency by which the receiver observes them is less than the frequency by which they are being sent. This holds for any frequency of signals sent from lower α to higher, including the frequengy of light, which consists of electromagnetic waves.

Therefore the color of visible light shifts towars the lower frequency end of the spectrum, that is, towars red. It can be viewed simply as a Doppler shift: in terms of the original t, r coordinate system the sender moves slower at the time of sending than the receiver moves at the time of receiving.

As $\alpha \to 0$, interesting things seem to happen. First, as we noted earlier, acceleration becomes infinite, and it becomes infinitely difficult to hang on the accelerating frame. On the other hand, since $\mathrm{d}\tau \to 0$, particle's time freezes. Redshift diverges and all signals coming from the vicinity of $\alpha = 0$ get redshifted to infrared and ultimately to black. All timelike and lightlike worldlines approach asymptotically the $\alpha = 0$ line. Motion stops. No worldline crosses the line, which is, according to the radar definition of distance, infinitely far away from any line of constant $\alpha > 0$, since no light ray will ever return from $\alpha = 0$.

The meaning of the line $\alpha = 0$ can be most easily understood in terms of t, r coordinates. It is a lightlike limiting case of the uniformly accelerating timelike worldlines. It consists of a light ray that comes from right, bounces at $0, 0$ and goes back to right; in other words it consists of the right side of the light cone of the event $0, 0$.

All events on the Rindler coordinate system are located on the right outside the $0, 0$ light cone. No event in the future cone is a cause to an event in the Rindler frame, and no event in the past cone is a consequence of an event in the Rindler frame.

Therefore letting go off the Rindler frame and falling past the $\alpha = 0$ line means leaving the events in the Rindler frame for good. If you do that, no observer that stays in the Rindler frame will ever hear from you again. This is is reflected to the Rindler coordinate system: it appears that nothing can pass the $\alpha = 0$ line, because signals coming from a falling particle will arrive later and later at any fixed $\alpha > 0$. The signals that the falling particle sends at the event of crossing will be received at $\theta = \infty$, that is, never.

This is why the lightlike $\alpha = 0$ line is called an event horizon. It marks a moment in time after which you have left a certain region of spacetime for good. Those who stay in the region cannot see past the event horizon, much like sailors cannot see past the fsfds horizon.

2.3 The equivalence principle

Gravity seems to affect everything in the same way, just like a fictious forces affect particles in accelerating coordinate systems. According to so-called equivalence principle, gravity *is* a fictious force, and we fall towards the ground because the ground is accelerating towards us.

If this is true, then the Rindler coordinate system must be at least approximately right for us here on Earth, and the phenomena we found in the Rindler frame must also be observable in terrestrial laboratories. This is how it has turned out to be; for example in 1959 researchers observed redshift of light that climbed 22.5 meters up from the basement of their laboratory.

Therefore we don't need a separate theory of gravity. Apparent gravitational forces are a by-product of the nature of spacetime, whatever it is.

A puzzling thing about gravity on Earth is that Earth is round. Every point of the surface of Earth seems to accelerate outwards from Earth's center. Gravity of the Rindler coordinate system can be made disappear by just transforming back to the original coordinates where $\mathrm{d}\tau^2 = \mathrm{d}t^2 - \mathrm{d}r^2$, but in the Rindler coordinates gravity points always to the same direction.

The answer is that the true geometry of spacetime is such that there is no spacetime coordinate system in which

$$\mathrm{d}\tau^2 = \mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2$$

would hold everywhere. More precisely, we can make gravity disappear in some small volume somewhere around Earth by choosing a freely falling coordinate system, but if we try to extend the coordinate frame to include larger portions of spacetime around Earth, effects of gravity will be unavoidable.

The situation is analogous to the two-dimensional surface of Earth. We want to draw maps of the surface, and if possible, we want the Pythagorean theorem to hold, of course. It would be nice if we could calculate distances by $\mathrm{d}d^2 = \mathrm{d}x^2 + \mathrm{d}y^2$, in which x and y are our surface coordinates measured in kilometers, for example. We would want to be able to draw the map on flat paper so that x and y form a square grid and distances can be measured by a ruler.

For any small region, like one tiny country, this can be done with reasonable accuracy. For a continent like Asia it's impossible, and maps that try to cover the whole Earth look ridiculous with no exception. Such maps typically use polar coordinates θ, ϕ for which

$$\mathrm{d}d^2 = r^2 \left(\mathrm{d}\theta^2 + \sin^2\theta \, \mathrm{d}\phi^2\right)$$

in which r is the radius of Earth. The polar coordinates look like square-gridded x, y coordinates only if $\theta \approx 0$, that is, on the equator.

The reason why mapping Earth is difficult is that the surface of Earth is curved. Because of this analogy, also spacetime is said to be curved.

2.4 Motion in curved spacetime

Freely falling observers do not experience any forces in the Rindler frame, since the spacetime it describes is flat. Let's modify the Rindler metric a bit and write

$$\mathrm{d}\tau^2 = \left(1+\alpha^2\right)\,\mathrm{d}\theta^2 - \mathrm{d}\alpha^2,$$

and this time let α range all the way down to $-\infty$. Like the Rindler metric, this metric does not depend on θ, and therefore lines of constant α are lines of constant acceleration, and neighboring lines of constant α have constant radar distance.

For large values of α this metric approaches the Rindler meric. Worldlines arc towards $\alpha = 0$ and acceleration goes to zero as $\alpha \to \infty$. For large negative α the metric approaches the mirror image of the Rindler metric, and again world lines arc towards $\alpha = 0$. For $\alpha \approx 0$ the metric differs from the Rindler metric. The coefficient of $\mathrm{d}t^2$ is not close to zero but close to one. Time does not freeze, and the constant line $\alpha = 0$ is timelike world line.

Therefore worldlines oscillate around $\alpha = 0$, just like a particle would oscillate in a tunnel carved straight through the center of the Earth. The metric gives rise to a gravitational force that points everywhere towards the same point. This metric resembles the spacetime geometry around Earth, although is not exactly the same.

Now imagine a finite-sized object sitting at $\alpha = 0$. It falls freely while it does not move, but its different parts experience forces that point towards its center. These forces compress the object, and if they are strong enough, the object gets crushed. The appearance of these tidal forces is a manifestation of curvature. Curvature is an absolute property of spacetime that cannot be made to disappear by changing perspective.

Uniform, freely falling motion follows a worldline with a maximal length compared to other worldlines between the same endpoint events. In coordinates in which $\mathrm{d}\tau^2 = \mathrm{d}t^2 - \mathrm{d}r^2$ such a wordline appears as a straight line. Such lines of extremal length are called geodesics. In spacetime timelike geodesics have maximal length, while in ordinary space any geodesic has minimal length. The difference arises from the difference in the metric signature.

In the Rindler metric worldlines of particles bend because clocks tick quicker at higher α altitudes. A straight worldline can make itself longer by visiting higher values of α and collect ticks with a higher frequency. On the other hand, navigating to higher α requires high speeds, which slows the clock down. The actual worldline of a particle is a smooth arc that makes an optimal compromise between these two time dilation effects.

At low values of α even a small angle between a constant α line and the worldline means high speed and low clock frequency. This makes the worldlines to reverse their curvature when they are close enough to $\alpha = 0$. The result is the more or less (exactly?) Gaussian-shaped worldlines that visit a high α altitude and asyptotically approach $\alpha = 0$ in the past and the future.

We may calculate freely falling worldlines by solving the variational equation

$$\delta\,\tau = \delta \int \mathrm{d}\tau = \int \delta \mathrm{d}\tau = \int \delta \frac{\mathrm{d}\tau}{\mathrm{d}\sigma}\,\mathrm{d}\sigma,$$

in which σ is a variable that parameterizes the worldine, for small variations that vanish at the endpoint events.

Let's solve the equation generally for the metric

$$\mathrm{d}\tau^2 = A(r)\,\mathrm{d}t^2 - B(r)\,\mathrm{d}r^2$$

which includes freely falling, Rindler and our modified Rindler metrics as special cases. The solution will be two functions $t(\sigma)$ and $r(\sigma)$, but we can use t as the parameter, which makes the solution just one function $r(t)$. We have

$$S = \int_{t_i}^{t_f} \frac{\mathrm{d}\tau}{\mathrm{d}t}\,\mathrm{d}t = \int_{t_i}^{t_f} L(r, \tfrac{\mathrm{d}r}{\mathrm{d}t})\,\mathrm{d}t$$

with

$$L(r, \tfrac{\mathrm{d}r}{\mathrm{d}t}) = \frac{\mathrm{d}\tau}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\sqrt{A\,\mathrm{d}t^2 - B\,\mathrm{d}r^2} = \sqrt{A - B\left(\frac{\mathrm{d}r}{\mathrm{d}t}\right)^2}.$$

If $A(r) \ll B(r)\,\frac{\mathrm{d}r}{\mathrm{d}t}$??? which corresponds to particles moving slowly in our coordinate system, we can write

$$L(r, \tfrac{\mathrm{d}r}{\mathrm{d}t}) = A^2\sqrt{1 - \left(\sqrt{\frac{B}{A}}\frac{\mathrm{d}r}{\mathrm{d}t}\right)^2} \approx \sqrt{A} - \frac{1}{2}\frac{B}{A}\left(\frac{\mathrm{d}r}{\mathrm{d}t}\right)^2.$$

This is (minus) the Lagrangian of a Newtonian particle of mass $\frac{B}{A}$ moving in potential $\sqrt{A(r)}$. We can think of $\sqrt{A(r)}$ as a gravitational potential that gives rise to gravitational forces. In our modified Rindler metric the world lines oscillate around $\alpha = 0$ because it is a local minimum of gravitational potential. The gravitational potential is the same function (up to the square root) that determines how fast clocks tick. “Gravity slows time down.”

To calculate the equation of motion, we change $r(t)$ to $r(t) + \delta r(t)$, in which $\delta r(t)$ is a small variation for which $\delta r(t_i) = \delta r(t_f) = 0$. We get

$$\delta S = \int_{t_i}^{t_f} \delta L\,\mathrm{d}t = \int_{t_i}^{t_f}\left(\frac{\delta L}{\delta r}\,\delta r + \frac{\delta L}{\delta \frac{\mathrm{d}r}{\mathrm{d}t}}\,\delta\frac{\mathrm{d}r}{\mathrm{d}t}\right)\mathrm{d}t.$$

Because

$$\delta\frac{\mathrm{d}r}{\mathrm{d}t} = \frac{\mathrm{d}r + \delta r}{\mathrm{d}t} - \frac{\mathrm{d}r}{\mathrm{d}t} = \frac{\mathrm{d}\delta r}{\mathrm{d}t},$$

we have

$$\delta S = \int_{t_i}^{t_f}\left(\frac{\delta L}{\delta r}\,\delta r + \frac{\delta L}{\delta \frac{\mathrm{d}r}{\mathrm{d}t}}\,\frac{\mathrm{d}\delta r}{\mathrm{d}t}\right)\mathrm{d}t.$$

We can write

$$\frac{\delta L}{\delta \frac{\mathrm{d}r}{\mathrm{d}t}}\,\frac{\mathrm{d}\delta r}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\delta L}{\delta \frac{\mathrm{d}r}{\mathrm{d}t}}\,\delta r\right) - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\delta L}{\delta \frac{\mathrm{d}r}{\mathrm{d}t}}\right)\delta r$$

and therefore

$$\delta S = \int_{t_i}^{t_f} \left(\frac{\delta L}{\delta r}\,\delta r + \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\delta L}{\delta \frac{\mathrm{d}r}{\mathrm{d}t}}\,\delta r\right) - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\delta L}{\delta \frac{\mathrm{d}r}{\mathrm{d}t}}\right)\delta r \right) \mathrm{d}t.$$

The middle term is a total time derivative, and thus gives contribution only at the endpoints. The contribution is proportional to δr, and since it vanishes at the endpoints, the contribution is zero. We are left with

$$\delta S = \int_{t_i}^{t_f} \left(\frac{\delta L}{\delta r} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\delta L}{\delta \frac{\mathrm{d}r}{\mathrm{d}t}}\right)\right) \delta r \, \mathrm{d}t.$$

Since this holds for any small $\delta r(t)$ that vanishes at the endpoints, for $\delta = 0$ to hold the stuff inside the parenthesis must vanish at every point of the worldline. We are left with the Euler-Lagrange equation

$$\frac{\delta L}{\delta r} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\delta L}{\delta \frac{\mathrm{d}r}{\mathrm{d}t}}\right) = 0.$$

For

$$L(r, \tfrac{\mathrm{d}r}{\mathrm{d}t}) = \sqrt{A(r) - B(r)\left(\frac{\mathrm{d}r}{\mathrm{d}t}\right)^2}$$

we get

joo joo joo

2.5 Spacetime around Earth

Common knowledge gives us many hints about the real geometry of spacetime around Earth. First, all gravitational effects seem to be spherically symmetric around Earth, so let's use spherical r, θ, ϕ coordinates as the spacelike coordinates and assume that the metric does not depend on θ or ϕ (except for the standard $\sin^2\theta$ in the spherical line element part). Second, gravitational forces around Earth seem to not change along time, so let's assume that the metric does not depend on the timelike coordinate t.

Based on these assumptions and the fact that the line element is always quadratic in the coordinate differentials, we write down an ansatz for the metric:

$$\begin{aligned} \mathrm{d}\tau^2 = {} & A(r)\,\mathrm{d}t^2 - B(r)\,\mathrm{d}r^2 \\ & -C(r)\mathrm{d}r\,\mathrm{d}t - D(r)\,r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right). \end{aligned}$$

We can eliminate? two of the functions A, B, C and D by changing $t \to t(t,r)$ and $r \to r(t,r)$ appropriately. Let's change them so that $C(r) \equiv 0$ and $D(r) \equiv 1$. We get

$$\mathrm{d}\tau^2 = A(r)\,\mathrm{d}t^2 - B(r)\,\mathrm{d}r^2 - \, r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right).$$

If a particle moves only in the up-down direction, we have $\mathrm{d}\theta = \mathrm{d}\phi = 0$ and therefore $\mathrm{d}\tau^2 = A(r)\,\mathrm{d}t^2 - B(r)\,\mathrm{d}r^2$, and we can directly apply the discussion of the previous section. As we calculated, the gravitational force is proportional to?? $-\frac{\mathrm{d}\sqrt{A}}{\mathrm{d}r}$. Because gravitational force vanishes at the center of the earth, the function $A(r)$ must start with zero r derivative at $r = 0$. Because the force points towards the center of Earth, the derivative cannot be negative, so $A(r)$ must begin to smoothly increase from $r = 0$.

The function $B(r)$ can be thought of as causing some kind of spatial deformation. At this point we cannot say much about it.

If we assume that the gravitational force vanishes as $r \to \infty$, then $A(r)$ must approach a constant value as $r \to \infty$. It is reasonable to assume that both $A(r)$ and $B(r)$ approach the value 1, as that would mean that the metric approaches the flat metric as we move away from Earth, and earth's effects on spacetime geometry should be non-negligible only close to it.

If a particle moves arbitrarily around Earth, then we generally have $\mathrm{d}\theta \neq \mathrm{d}\phi \neq 0$. Anyway, a freely falling particle always moves in a plane with the center of Earth located on it, and we can choose the orientation of the coordinate system so that $\mathrm{d}\phi = 0$. Then we have

$$\mathrm{d}\tau^2 = A(r)\,\mathrm{d}t^2 - B(r)\,\mathrm{d}r^2 - \, r^2\,\mathrm{d}\theta^2.$$

Chapter 3

Dynamics

So far we have only discussed flow of time, relations of events and motion of particles in invented spacetimes. Now we will figure out how spacetime actually curves.

3.1 The action principle

Very often motion of a physical system can be described by writing down an action S, which is a function that takes a whole motion as its argument and produces a number, and by stating that for a realistic motion the action has an extremal value with respect to all other motions restricted by some boundary conditions.

The motion of a freely falling particle from the previous chapter is an example of this action principle: the action of a motion was chosen to be the worldline length, and the boundary conditions were that the compared worldlines should start and end at the same endpoint events.

It's not obvious why the action principle should be used. Sure, it is simple, but do we have any reason to believe that it leads to correct equations of motion? One reason for favoring the action principle is the path integral formulation of quantum mechanics.

According to the path integral formulation, the probability amplitude for one particular motion is proportional to $\mathrm{e}^{i\hbar S}$, in which S is the action of the motion and $\hbar$ is a small connstant called Planck's constant. The amplitude for the event that one of several different motions takes place is the sum of amplitudes for all those motions. Now, if the motions are located away from an extremum of S, and they differ from each other much more than by $\hbar$, they oscillate wildly in the complex plane and more or less cancel each other, resulting in a very small amplitude.

If the motions are located around an extremum, then there are many motions with approximately the extremum value of S even if their actions differed much more than by $\hbar$, and their amplitudes point approximately in the same direction in the complex plane. They do not cancel each other, and we get a large amplitude. Therefore, if the scale of actions we consider is much more than $\hbar$, which corresponds to macroscopic motion, we should expect to see only motions with extremal action.

It is not at all clear how spacetime and gravity works at the quantum level. Anyway, we can try to use the action principle for spacetime geometry. The first attempt is to just use the spacetime volume as the action, just like we used spacetime length for the action of a particle. This does not work, because the only metric that extremizes spacetime volume is $\mathrm{d}\tau^2 \equiv 0$.

The difference to the particle case is that the particle moves in a background metric—the spacetime metric—relative to which we could fix the endpoints of the worldline and which is used to calculate the worldline length, but the spacetime metric itself doesn't.

The next option is to find some kind of absolute number that measures curvature locally. By integrating that number over spacetime, with the integration measure calculated from the metric, we get a potentially useful spacetime action function.

3.2 A jungle of indices

In order to measure curvature, we need to study differential geometry of Riemannian manifolds. Let's start from $x^\bullet$ ($\bullet = 0, 1, 2, 3$) coordinates for which

$$\mathrm{d}\tau^2 = \left(\mathrm{d}x^0\right)^2 - \left(\mathrm{d}x^1\right)^2 - \left(\mathrm{d}x^2\right)^2 - \left(\mathrm{d}x^3\right)^2 .$$

We may write this as

$$\mathrm{d}\tau^2 = \sum_{\bullet=0}^{3}\sum_{{\color{orange}\bullet}=0}^{3} g_{\bullet{\color{orange}\bullet}}\,\mathrm{d}x^{\bullet}\mathrm{d}x^{{\color{orange}\bullet}} \doteq g_{\bullet{\color{orange}\bullet}}\,\mathrm{d}x^{\bullet}\mathrm{d}x^{{\color{orange}\bullet}},$$

in which the numbers $g_{\bullet{\color{orange}\bullet}}$ are the elements of a diagonal matrix with the numbers $1, -1, -1, -1$ on the diagonal. We use the summation convention, which means that whenever there is a repeated index in an expression, a summation over it is performed.

I use colored bullets as indices because greek indices that are traditionally used for typesetting make formulas difficult to read. If you don't like the colored bullets, you can download the LaTeX source code of this book. By uncommenting the line below the line `% TRADITIONAL GREEK INDICES` and compiling you get a version with traditional greek indices. Formulas change for example by

$$g_{\bullet{\color{orange}\bullet}}\,\mathrm{d}x^{\bullet}\mathrm{d}x^{{\color{orange}\bullet}} \to g_{\alpha\beta}\,\mathrm{d}x^{\alpha}\mathrm{d}x^{\beta}.$$

If we now change to an arbitrary coordinate system $\tilde{x}^{\bullet}$, a simple application of the chain rule gives

$$\mathrm{d}\tau^2 = g_{\bullet{\color{orange}\bullet}}\,\frac{\partial x^{\bullet}}{\partial \tilde{x}^{{\color{cyan}\bullet}}}\mathrm{d}\tilde{x}^{{\color{cyan}\bullet}}\,\frac{\partial x^{{\color{orange}\bullet}}}{\partial \tilde{x}^{{\color{red}\bullet}}}\mathrm{d}\tilde{x}^{{\color{red}\bullet}} = \tilde{g}_{\bullet{\color{orange}\bullet}}\,\mathrm{d}\tilde{x}^{\bullet}\mathrm{d}\tilde{x}^{{\color{orange}\bullet}},$$

in which

$$\tilde{g}_{\bullet{\color{orange}\bullet}} = \frac{\partial x^{{\color{cyan}\bullet}}}{\partial \tilde{x}^{\bullet}}\,\frac{\partial x^{{\color{red}\bullet}}}{\partial \tilde{x}^{{\color{orange}\bullet}}}\,g_{{\color{cyan}\bullet}{\color{red}\bullet}}.$$

This is obvious: if we change coordinates, we must also transform the numbers $g_{\bullet{\color{orange}\bullet}}$ in an appropriate way; otherwise $\mathrm{d}\tau^2$ would depend on which coordinate system is used. When we transformed for example into the Rindler coordinates, we first had $x^0 = t$, $x^1 = r$, $g_{00} = 1$, $g_{11} = -1$ and $g_{01} = g_{10} = 0$, and after transforming $\tilde{x}^0 = \theta$, $\tilde{x}^1 = \bullet$, $\tilde{g}_{00} = \bullet$, $\tilde{g}_{11} = -1$ and $\tilde{g}_{01} = \tilde{g}_{10} = 0$.

The separation $\mathrm{d}x^{\bullet}$ of two nearby events is essentially a small arrow located somewhere in spacetime. All small separations located at the same event form a vector space, called the tanget space of the event. In our notation one upper index after a quantity indicates it is a vector.

The numbers $g_{\bullet\textcolor{orange}{\bullet}}$ define a bilinear map that takes two vectors $A^{\bullet}$ and $B^{\bullet}$ from the tangent space and produces the number $g_{\bullet\textcolor{orange}{\bullet}}A^{\bullet}B^{\textcolor{orange}{\bullet}}$. Such a linear map is called a tensor. All linear maps are specific to the vector space they act in, and if we say we have a tensor, we also have to specify the event at which it is located.

It's customary to refer to the numeric representation of a tensor as tensor, just like it is customary to refer to coordinates of an event as event, or components of a vector as vector, even though in all three cases the entity and the numbers representing it are not the same thing: the numbers depend on the coordinate system while the entity doesn't. Sometimes this induces ambiguity, but usually the context makes it clear what the notation means. It is very difficult to find tensor notation that would be unambiguous and readable at the same time.

The tensor $g_{\bullet\textcolor{orange}{\bullet}}$ is called metric tensor. It is said to define an inner product on spacetime. It is symmetric in the indices $\bullet\textcolor{orange}{\bullet}$, which means that $g_{\bullet\textcolor{orange}{\bullet}} = g_{\textcolor{orange}{\bullet}\bullet}$. The worldline element $\mathrm{d}\tau^2$ is calculated by taking the inner product of $\mathrm{d}x^{\bullet}$ with itself.

We can also have tensors with different index structures. Any tensor with one lower index, which is a linear map that takes one vector and produces a number, is called one-form or a dual vector. One-form $O_{\bullet}$ transforms by

$$O_{\bullet} \to \tilde{O}_{\bullet} = \frac{\partial x^{\textcolor{orange}{\bullet}}}{\partial \tilde{x}^{\bullet}} O_{\textcolor{orange}{\bullet}},$$

that is, exactly the opposite way as $\mathrm{d}x^{\bullet}$. That ensures that $O_{\bullet}\mathrm{d}x^{\bullet}$ does not change along the coordinate system.

We can also think of a vector as a tensor that takes one one-form and produces a number. In general a tensor has n lower indices and m upper indices, and takes n vectors and m one-forms and produces a number. The transformation law of the numeric representation of a n, m tensor features n instances of $\frac{\partial x^{\bullet}}{\partial \tilde{x}^{\textcolor{orange}{\bullet}}}$ and m instances $\frac{\partial \tilde{x}^{\bullet}}{\partial x^{\textcolor{orange}{\bullet}}}$ in such a way that the number it produces out of the vectors and one-forms does not depend on the coordinate system we use. For example

$$T^{\bullet}{}_{\textcolor{orange}{\bullet}\textcolor{cyan}{\bullet}} \to \tilde{T}^{\bullet}{}_{\textcolor{orange}{\bullet}\textcolor{cyan}{\bullet}} = \frac{\partial \tilde{x}^{\bullet}}{\partial x^{\textcolor{red}{\bullet}}} \frac{\partial x^{\textcolor{magenta}{\bullet}}}{\partial \tilde{x}^{\textcolor{orange}{\bullet}}} \frac{\partial x^{\textcolor{green}{\bullet}}}{\partial \tilde{x}^{\textcolor{cyan}{\bullet}}} T^{\textcolor{red}{\bullet}}{}_{\textcolor{magenta}{\bullet}\textcolor{green}{\bullet}}.$$

We can think of a n, m tensor not only as a linear map that takes n vectors and m one-forms and produces numbers, but also as a linear map that takes $n-p$ vectors and $m-q$ one-forms and produces a $n-p, m-q$ tensor.

From a n, m tensor and a p, q tensor we can form a new $n+p, m+q$ tensor by so-called outer product. It works just as you could guess: the numerical representation of the outer product of $A_\bullet$ and $B^\bullet$ is $A_\bullet B^{\textcolor{orange}{\bullet}}$.

Actually the operation of plugging vectors and one-forms into a tensor and out a number can be thought of as first forming an outer product of the tensor and all the vectors and one-forms and then contracting all the lower-upper index pairs, that is, setting the index pair equal and summing over. When we calculate the line element, we first form $g_{\bullet\textcolor{orange}{\bullet}}\mathrm{d}x^{\textcolor{cyan}{\bullet}}\mathrm{d}x^{\textcolor{red}{\bullet}}$ and then contract the ${}_\bullet{}^{\textcolor{cyan}{\bullet}}$ and ${}_{\textcolor{orange}{\bullet}}{}^{\textcolor{red}{\bullet}}$ pairs to get $g_{\bullet\textcolor{orange}{\bullet}}\mathrm{d}x^\bullet\mathrm{d}x^{\textcolor{orange}{\bullet}}$.

There is still one more way to think of the $\mathrm{d}\tau^2$ and other similar inner products: we first define $\mathrm{d}x_\bullet = g_{\bullet\textcolor{orange}{\bullet}}\mathrm{d}x^{\textcolor{orange}{\bullet}}$ and then calculate $\mathrm{d}\tau^2 = \mathrm{d}x_\bullet \mathrm{d}x^\bullet$. We can use the metric tensor to lower any upper index. To raise indices we define a symmetric upper-index version of the metric tensor $g^{\bullet\textcolor{orange}{\bullet}}$ by the condition $g^{\bullet\textcolor{orange}{\bullet}}g_{\textcolor{orange}{\bullet}\textcolor{cyan}{\bullet}} = \delta^\bullet_{\textcolor{cyan}{\bullet}}$, in which $\delta^\bullet_{\textcolor{cyan}{\bullet}}$ is the Kronecker delta, which is the numeric representation of the unit tensor that does nothing.

The essence of this jungle of indices is that we form outer products of tensors by writing them in a row and destroy index pairs by contracting upper-lower index pairs. If an index is in the wrong floor, we can raise or lower it with either version of the metric tensor. If we manage to destroy all index pairs, what is left is just a number which does not depend on the coordinate system in any way. Technically the number is a $0, 0$ tensor for which the numeric representation and the tensor itself happens to be the same thing.

3.3 Curvature

Vectors (or tensors) located at different events in spacetime cannot be directly added or compared, because they do not live in the same vector space. For this reason it's not obvious what a "derivative of a vector" would mean, for example. Anyway, we can paraller transport vectors from tangent space to tangent space and then add or compare them.

Parallel transport means moving the vector without changing its direction or length. In a flat space it is trivial: it doesn'r depend on the path along which it is performed, and if we parallel transport a vector back into its own tangent space, it comes back as unchanged. For this reason we can effectively think of all vectors at all points belonging to the same vector space.

In curved spaces the situation is different. Take the surface of Earth as an example. If we have somewhere at the equator a vector that points up North and we parallel transport it straight to the North pole and then parallel transport it back to the equator along a straigh line that makes a right angle with the first line of parallel transport, we get a vector that is parallel to the equator. If we now parallel transport it back to the original point along the equator, we find that it has got rotated by 90 degrees.

Inner product of a vector with itself measures length, so parallel transport keeps of a vector $A^\bullet$ keeps $A_\bullet A^\bullet$ intact. Inner product of two vectors divided by their length characterizes the difference in the directions they point in, so parallel transport keeps also that intact. Since it already kept the lengths intact, it must keep the inner product of any two vectors intact.

If we have a geodesic parameterized by τ, then it's natural to assume that the velocity vector $\frac{\mathrm{d}x^\bullet}{\mathrm{d}\tau} \doteq v^\bullet$ doesn't change its direction nor length along it; in other words, it gets parallel transported along it. This together with the condition that inner products remain intact is almost enough for defining how vectors parallel transport along a geodesic, and also along an arbitrary worldline, since we may since we can think of any worldline as being composed of many small geodesic segments.

The condition that $v^\bullet$ parallel transports along a geodesic doesn't fix the parallel transport completely, because it doesn't say anything about how vectors rotate in planes orthogonal to $v^\bullet$. We need another condition.

Consider any two near geodesics $x^\bullet(\tau$ and $\bar{x}^\bullet(\tau)$ that start near each other, are initially parallel, and are parameterized by their length. The separation $\mathrm{d}x^\bullet$ of their points is a function of τ and is located on $x^\bullet(\tau)$ (and also on $\bar{x}^\bullet(\tau)$ since the geodesics are close to each other). The vector $\mathrm{d}x^{\bullet}$ parallel transports along $x^{\bullet}(\tau)$. This condition fixes the rest of the parallel transport.

Curvature of spacetime can be measured by studing how a vector changes by parallel transporting it along a loop that comes back to itself. Let's take two separation vectors, $\mathrm{d}x^{\bullet}$, $\mathrm{d}y^{\bullet}$ and a third vector $V^{\bullet}$, and parallel transport $V^{\bullet}$ first by $\mathrm{d}x^{\bullet}$, then by $\mathrm{d}y^{\bullet}$, and then back by $-\mathrm{d}x^{\bullet}$ and $-\mathrm{d}y^{\bullet}$. The linear approximation to how $\mathrm{d}z^{\alpha}$ changes can be written as

$$V^{\bullet} \to V^{\bullet} + R^{\bullet}{}_{{\color{orange}\bullet}{\color{cyan}\bullet}{\color{red}\bullet}} V^{\color{orange}\bullet} \mathrm{d}x^{\color{cyan}\bullet} \mathrm{d}y^{\color{red}\bullet}$$

with $R^{\bullet}{}_{{\color{orange}\bullet}{\color{cyan}\bullet}{\color{red}\bullet}}$ a tensor that takes three vectors and produces another vector.

The tensor $R^{\bullet}{}_{{\color{orange}\bullet}{\color{cyan}\bullet}{\color{red}\bullet}}$ is called the Riemann tensor. It encapsulates all the important local curvature properties of spacetime. It is antisymmetric in the last two indices, $R^{\bullet}{}_{{\color{orange}\bullet}{\color{cyan}\bullet}{\color{red}\bullet}} = -R^{\bullet}{}_{{\color{orange}\bullet}{\color{red}\bullet}{\color{cyan}\bullet}}$, since switching the last two indices amounts to reversing the path along with the vector to be contracted with the second index is parallel transported.

In order to have an action for gravity, we need to turn the Riemann tensor into a number. We can do that by contracting two pairs of its indices. It wouldn't be sane to contract the last pair, since $g^{\bullet{\color{orange}\bullet}}$ is symmetric, and the Riemann tensor is antisymmetric in the last pair. Such a contraction would lead to identically zero tensor.

If we lower the the first index of the Riemann tensor, it can be shown (I hope I find a way to do that in an understandable manner) that it is also antisymmetric in the first two indices: $R_{\bullet{\color{orange}\bullet}{\color{cyan}\bullet}{\color{red}\bullet}} = -R_{{\color{orange}\bullet}\bullet{\color{cyan}\bullet}{\color{red}\bullet}}$. Therefore we also should not contract the first index pair. Therefore the only options are to contract one index from the first pair with one index from the second pair. Because of the antisymmetries, these options differ only by sign.

Let's define

$$R_{\bullet{\color{orange}\bullet}} = R_{\bullet{\color{cyan}\bullet}{\color{red}\bullet}{\color{orange}\bullet}} g^{{\color{cyan}\bullet}{\color{red}\bullet}}.$$

The tensor $R_{\bullet{\color{orange}\bullet}}$ is called the Ricci tensor. Now we just contract its indices and get

$$g^{\bullet{\color{orange}\bullet}} R_{\bullet{\color{orange}\bullet}} \doteq R.$$

This is a number called Ricci scalar.

3.4 Einstein field equation

We can try to use the Ricci scalar in spacetime action S. If spacetime is empty, the action is just

$$S = \int_{\mathcal{V}} R \, \mathrm{d}^4\tau,$$

in which $\mathcal{V}$ is the spacetime regime we are considering, and $\mathrm{d}^4\tau$ is the integration measure calculated from the metric.

If the spacetime is not empty, the total action will be the action of spacetime plus the action of matter, plus possibly some additional interaction terms. The action of matter is usually taken as a spacetime integral over a Lagrangian $\mathcal{L}$. This integral must be integrated using the same measure, which is calculated from the spacetime metric. Without additional interactions the total action is

$$S = \int_{\mathcal{V}} (R + \mathcal{L}) \, \mathrm{d}^4\tau.$$

The motion of matter obviously depends on spacetime geometry. Since $\mathcal{L}$ is integrated using the measure $\mathrm{d}^4\tau$ calculated from the metric, the dependency goes also the other way around: matter and its motion affects spacetime geometry. Mathematically this is because varying the metric varies not only R but also the measure $\mathrm{d}^4\tau$.

So there is an interaction between spacetime and matter, in other words gravity and matter, even though we didn't include any explicit interaction between gravity and matter. All matter shape spacetime, and all matter react to spacetime geometry. Gravity is universal.

To find the differential equation that locally quantifies the condition $\delta S = 0$, we first have to choose a coordinate system that covers $\mathcal{V}$. At this first step the coordinates just label events, and the only relation the events It can be shown (I'll show later) that the correct metric integration measure is

$$\mathrm{d}^4\tau = \sqrt{|g|} \, \mathrm{d}^4 x$$

in which g is the determinant of the metric. The action becomes

$$S = \int_{\mathcal{V}} (R + \mathcal{L})\sqrt{|g|}\,\mathrm{d}^4x = \int_{\mathcal{V}} \left(\sqrt{|g|}R + \sqrt{|g|}\mathcal{L}\right)\,\mathrm{d}^4x.$$

Here we see the matter-spacetime coupling explicitly as $\sqrt{|g|}$ multiplies $\mathcal{L}$.

Let's vary the action. We get

$$\begin{aligned}
\delta S &= \int_{\mathcal{V}} \delta\left(\sqrt{|g|}R + \sqrt{|g|}\mathcal{L}\right)\,\mathrm{d}^4x \\
&= \int_{\mathcal{V}} \left(\frac{\delta\sqrt{|g|}R}{\delta g^{\bullet{\color{orange}\bullet}}} + \frac{\delta\sqrt{|g|}\mathcal{L}}{\delta g^{\bullet{\color{orange}\bullet}}}\right)\delta g^{\bullet{\color{orange}\bullet}}\,\mathrm{d}^4x \\
&= \int_{\mathcal{V}} \left(\sqrt{|g|}\frac{\delta R}{\delta g^{\bullet{\color{orange}\bullet}}} + \frac{\delta\sqrt{|g|}}{\delta g^{\bullet{\color{orange}\bullet}}}R + \frac{\delta\sqrt{|g|}\mathcal{L}}{\delta g^{\bullet{\color{orange}\bullet}}}\right)\delta g^{\bullet{\color{orange}\bullet}}\,\mathrm{d}^4x \\
&= \int_{\mathcal{V}} \left(\frac{\delta R}{\delta g^{\bullet{\color{orange}\bullet}}} + \frac{1}{\sqrt{|g|}}\frac{\delta\sqrt{|g|}}{\delta g^{\bullet{\color{orange}\bullet}}}R + \frac{1}{\sqrt{|g|}}\frac{\delta\sqrt{|g|}\mathcal{L}}{\delta g^{\bullet{\color{orange}\bullet}}}\right)\delta g^{\bullet{\color{orange}\bullet}}\,\sqrt{|g|}\,\mathrm{d}^4x.
\end{aligned}$$

We can write this as

$$\int_{\mathcal{V}} \left(G_{\bullet{\color{orange}\bullet}} + \alpha_{\bullet{\color{orange}\bullet}} - T_{\bullet{\color{orange}\bullet}}\right)\delta g^{\bullet{\color{orange}\bullet}}\,\sqrt{|g|}\,\mathrm{d}^4x$$

in which $T_{\bullet{\color{orange}\bullet}}$ is the last term in the parenthesis of the second-last equation,

Because δS is a scalar, the expression in the parenthesis times $\delta g^{\bullet{\color{orange}\bullet}}$ must also be a scalar, and since $\delta g^{\bullet{\color{orange}\bullet}}$ is a tensor, the quantity in the parenthesis is a tensor.

What we would need to do next is to figure out how to calculate the numerical components of the Riemann tensor given the components of the metric and then calculate how they (or the contractions $R_{\bullet{\color{orange}\bullet}}$ and R) vary as the metric is varied. It is a long road, and I do not inted to do it here (perhaps later in an appendix). Instead of that I quote some results from Wikipedia.

Varying the first term in the parenthesis gives, after dropping a term that becomes a total derivative,

$$\frac{\delta R}{\delta g^{\bullet{\color{orange}\bullet}}} = R_{\bullet{\color{orange}\bullet}}.$$

The second term gives

$$\frac{1}{\sqrt{|g|}}\frac{\delta\sqrt{|g|}}{\delta g^{\bullet\bullet}}R = -\frac{1}{2}R\,g_{\bullet\bullet}$$

Because δS is a scalar, the expression in the parenthesis times $\delta g^{\bullet\bullet}$ must also be a scalar, and since $\delta g^{\bullet\bullet}$ is a tensor, the quantity in the parenthesis is a tensor.

also the Lagrangian $\mathcal{L}$ of matter must be taken into consideration. The total action of spacetime filled with matter is a spacetime integral of $R + \mathcal{L}$.

volume over which we are integrating.

, also the Lagrangian $\mathcal{L}$ of matter must be taken into consideration. The total action of spacetime filled with matter is a spacetime integral of $R + \mathcal{L}$.

Volume arises from distance, and therefore the integration measure $\mathrm{d}^2\tau$ must be calculated from the metric. It can be shown that in coordinates x^α the correct integration measure is

$$\sqrt{|g|}\,\mathrm{d}x^0\mathrm{d}x^1\mathrm{d}x^2\mathrm{d}x^3 \doteq \sqrt{|g|}\,\mathrm{d}^4x$$

in which g is the determinant of the metric.

The complete action is now

$$\int_{\mathcal{V}}(R+\mathcal{L})\sqrt{|g|}\,\mathrm{d}^4x \doteq S$$

in which $\mathcal{V}$ is the region of spacetime we integrate over.

To extremize this, we calulate the difference of the integral's value for two slightly different metrics, but if we do not restrict the metrics in any way, an extremum value does not exist. If we take both metrics to smoothly become the same at the border of the integration region $\mathcal{V}$, an extremum value can be found.

In practice we compare $g_{\alpha\beta}$ to $g_{\alpha\beta} + \delta g_{\alpha\beta}$, in which $\delta g_{\alpha\beta}$ is the small difference of our metrics, and calculate δS to first order in $\delta g_{\alpha\beta}$. All surface terms vanish because $\delta g_{\alpha\beta}$ vanishes smoothly at the border of $\mathcal{V}$.

The result of the calculation will be that metric $g_{\alpha\beta}$ corresponds to an extremum value of spacetime action if and only if the Einstein field equation holds at every point in $\mathcal{V}$. The Einstein field equation reads

$$R_{\alpha\beta} - \frac{1}{2} R\, g_{\alpha\beta} = T_{\alpha\beta},$$

in which

$$T_{\alpha\beta} = \frac{-2}{\sqrt{|g|}} \frac{\delta(\sqrt{|g|}\,\mathcal{L})}{\delta g_{\alpha\beta}}$$

is so-called energy-momentum tensor.

In vacuum (and with zero cosmological constant) $T_{\alpha\beta} = 0$. Einstein field equation becomes

$$R_{\alpha\beta} = \frac{1}{2} R\, g_{\alpha\beta}.$$

Multiplying by $g^{\alpha\beta}$ and contracting both indices gives $R = 2\,R$, which implies $R = 0$. Therefore the Einstein field equation for vacuum can be written as just $R_{\alpha\beta} = 0$.

Chapter 4

Almost flat spacetime

Things to consider: mass, energy, force

initial-value problem

electromagnetism

prologue: the shape of earth (why earth is round? the answer is newton's universal gravitation)(cavensish experiments demonstrates it)(Newton's theory explains motion of every planet in our solar system except mercury's)

Contents

Kaarevuudessa ei ole mitään ihmeellistä. Ei ole mitään syytä olettaa että jokin avaruus josta havaitaan vain pieni osa olisi isossa skaalassa laakea. Ihmiset jotka olettivat maan pinnan (se taso jota vastaan kohtisuorassa painovoima osottaa) olevan laakea tekivät virheen.

There are not many such numbers. One is the so-called Ricci curvature scalar R. If we use it, solving the variational equation will yeld $R_{\alpha\beta} = 0$, in which $R_{\alpha\beta}$ is so-called Ricci tensor. I will explain in the next section what R and $R_{\alpha\beta}$ are. The important, and rather miraculous thing is that $R_{\alpha\beta} = 0$ is the correct field equation for the geometry of empty spacetime.

We may now sketch how we could figure out the geometry of spacetime outside Earth. First, spacetime seems to be spherically symmetric around Earth. Therefore, let's use spherical r, θ, ϕ coordinates as the spacelike coordinates and assume that the metric does not depend on θ or ϕ (exept for the standard $\sin^2\theta$ in the spherical line element part). Second, gravitational forces around Earth seem to not change along time, so let's assume that the metric does not depend on the timelike coordinate t. Third, let's anticipate that the coordinate system can be chosen so that there are no cross terms (for exampe a term of the form $\mathrm{d}t\,\mathrm{d}r$) in the metric.

Based on these assumptions we write down an ansatz for the metric:

$$\mathrm{d}\tau^2 = A(r)\,\mathrm{d}t^2 - B(r)\,\mathrm{d}r^2 - r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right).$$

Next we read from Wikipedia how to calculate $R_{\alpha\beta}$ of the ansatz. The formula is awful. We set $R_{\alpha\beta}$ to zero and get some coupled differential equations for $A(r)$ and $B(r)$. We solve them and get

$$A(r) = 1 - \frac{r_s}{r} \qquad \text{and} \qquad B(r) = \frac{1}{A(r)}$$

with r_s a positive-valued parameter. Putting everything together we have

$$\mathrm{d}\tau^2 = \left(1 - \frac{r_s}{r}\right)\mathrm{d}t^2 - \frac{1}{1 - \frac{r_s}{r}}\mathrm{d}r^2 - r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right).$$

This metric is called Schwartzchild metric. It describes the geometry of spacetime around planets, nonorotating black holes and many other astronomical bodies.

To find out the spacetime geometry inside Earth, we need an equation for nonempty spacetime geometry. We can do this by extremizing the combined action of matter and spacetime. In other words, we solve the variational equation

$$\delta \int (R + \mathcal{L}) \, \mathrm{d}^4\tau = 0$$

in which $\mathcal{L}$ is somehow calculated from local motion of matter in spacetime. Solving the equation will yeld

$$R_{\alpha\beta} - \frac{1}{2} R \, g_{\alpha\beta} = T_{\alpha\beta},$$

an equation known as the Einstein field equation. Tensors $g_{\alpha\beta}$ and $T_{\alpha\beta}$ are so-called metric tensor and energy-momentum tensor. Unfortunately the Einstein field equation is too difficult for us to solve inside Earth.

The variational equation and the Einstein field equation are the global and local version of one single condition that the spacetime geometry obeys. It is wrong to say that they dictate how spacetime geometry evolves, since evolution happens as a function of one-dimensional time, and here we do not have any time at all before the field equation has done its work.

This is reflected on how we found the Schwartschild solution: we did *not* start from some initial conditions and then let the equations evolve them into the future. Instead we started from an ansatz for the whole spacetime geometry and let the field equation fix the two unknown functions we had left in the ansatz.

It is difficult to say what the variational equation or the Einstein field equation means. They are a condition that spacetime geometry and matter obey. No more can be definitely said about them. $g_{\theta\theta}$

The above

Differential geometry is challenging—either extremely messy or extremely abstract—

Let's assume that freely falling motion has an extremal worldline length also when spacetime is curved.

We could calculate freely falling worldlines exactly by solving the variational equation

$$\delta\tau = \delta \int \mathrm{d}\tau = \int \delta \, \frac{\mathrm{d}\tau}{\mathrm{d}\sigma} \, \mathrm{d}\sigma = 0$$

in which σ is a worldline parameter and the variation is small and vanishes at the endpoint events.

Let's take for example the metric

$$\mathrm{d}\tau^2 = A(t,r)\,\mathrm{d}t^2 - B(t,r)\,\mathrm{d}r^2.$$

We have

$$\delta\tau = \int \delta \mathrm{d}\tau$$

We can therefore speculate with any metrics without worrying about wether

When large portions of Earth must be covered, polar coordinates θ and ϕ are usually used. For them

$$\mathrm{d}d^2 = r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right),$$

where r is the radius of Earth. Because the spacetime around Earth seems to be spherically symmetric, we should use polar coordinates for spacetime. Because gravitational effects on Earth seem to be independent of time, we should expect the spacetime metric to be independent of time.

We may now write down an ansatz for the spacetime metric around Earth:

$$\mathrm{d}\tau^2 = A(r)\,\mathrm{d}t^2 - B(r)\,\mathrm{d}r^2 - r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right).$$

Chapter 5

Our universe

5.1 Expanding universe

5.2 Field equations

Chapter 6

Black holes

6.1 The Schwartzchild solution

motion around black hole
how black hole looks like

6.2 Black holes merge

6.3 dsfs

$$\mathrm{d}\tau^2 = \mathrm{d}t^2 - A(t)^2 \mathrm{d}r^2$$

$$\sigma_+ \to \tilde{\sigma_+} = \eta\sigma_+ \qquad \text{and} \qquad \sigma_- \to \tilde{\sigma_+} = \eta^{-1}\sigma$$

6.4 Accelerating observers

Rindler metric:

6.5 Equivalence of gravity and acceleration