Following recommendations of Bell "how to teach special relativity", we give an introduction into the theories of relativity which introduces not only the space-time interpretation. It starts, instead, with the Lorentz ether interpretation of special relativity as well as its generalization to relativistic gravity – an ether interpretation of the Einstein equations of general relativity in harmonic coordinates. The impossibility to identify, by observation, absolute time in the Lorentz ether, and the preferred background coordinates in its generalization is, then, used to introduce and justify the space-time interpretation.

The differences between Lorentz ether and space-time interpretation are considered and discussed. The final decision which interpretation is preferable is left to the reader.

An Introduction into Relativity based on the Lorentz ether

I. Schmelzer

July 8, 2017

Chapter 1

Introduction

The aim of this text is to give an introduction into special and general relativity which gives not only the standard space-time interpretation of relativity, as proposed by Minkowski for special relativity and developed into the curved space-time interpretation of general relativity by Einstein. It also gives an introduction into the Lorentz ether interpretation of special relativity, which can be seen as the original interpretation of the theory developed by Lorentz, Poincare and Einstein. The failure of the Lorentz ether to extend to relativistic gravity was, of course, a strong argument against the Lorentz ether, and has probably played a decisive role in the rejection of the ether concept. But it was only a historical accident.

Today it is known that there exists an extension of the Lorentz ether interpretation to the Einstein equations of general relativity in harmonic gauge. This fact alone does not mean that this interpretation has any advantages in comparison with the space-time interpretation. But even for those who reject the ether completely it makes sense not to use a historical accident which is unfortunate for the ether interpretation, but, instead, to present their arguments against the best imaginable defense of an ether interpretation.

Moreover, to have several interpretations is a value in itself, because it allows, in a much more objective way, to distinguish those parts of the theory which are really physical from those which are metaphysical. Indeed, the criterion to distinguish the two is simple: The really physical results should be the same in all interpretations. Instead, those statements which differ in different interpretations, or make sense only in one of the interpretations, have to be classified as metaphysical. Having only a single interpretation, it would be much more difficult to make this distinction.

The ability to identify correctly the metaphysical aspects of the space-time interpretation may become essential for the development of quantum gravity or a theory of everything, because the metaphysical elements can be given up without any problem, while the physical parts have to be recovered accurately even if the more fundamental theory uses completely different metaphysics.

Chapter 2

Special Relativity

2.1 How to construct Doppler-shifted solutions of a wave equation

The wave equation is the equation

$$\Box u(\vec{x},t) = \left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2\right) u(\vec{x},t) = 0$$

The operator ∇^2 is the Laplace operator, which acts on the spatial variables, so that in the three-dimensional case we have

$$\nabla^2 = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right).$$

The constant c is the speed of the wave. In relativity, c is the speed of light waves, but such a wave equation can be used also to describe sound waves, or waves on the surface of water.

Let's assume that we have found a solution of this equation u(x,t). Then there exists a surprisingly simple method to construct other solutions of the same wave equation. We can choose an arbitrary parameter $|v| < c$, and define the following coordinates:

$$t' = \gamma\left(t - \frac{v}{c^2}\,x\right), \quad x' = \gamma(x - v\,t), \quad y' = y, \quad z' = z, \quad \text{with} \quad \gamma = \frac{1}{\sqrt{1 - v^2/c^2}}.$$

Then, all we have to do is to replace in the solution $u(x,y,z,t)$ all the x, y, z, t by x', y', z', t', and use the formula above to obtain another, different function $u_v(x,y,z,t)$:

$$u_0(x,y,z,t) \to u_v(x,y,z,t) = u_0(x'(x,y,z,t), y'(x,y,z,t), z'(x,y,z,t), t'(x,y,z,t))$$

This function $u_v(x,y,z,t)$ is also a solution of the same wave equation. You can simply try it out.

$$
\begin{aligned}
\frac{\partial}{\partial t}u_v &= \frac{\partial t'}{\partial t}\frac{\partial}{\partial t'}u_0 + \frac{\partial x'}{\partial t}\frac{\partial}{\partial x'}u_0 = \gamma\left(\frac{\partial}{\partial t'} - v\frac{\partial}{\partial x'}\right)u_0,\\
\frac{\partial^2}{\partial t^2}u_v &= \gamma^2\left(\frac{\partial^2}{\partial t'^2} - 2v\frac{\partial}{\partial t'}\frac{\partial}{\partial x'} + v^2\frac{\partial^2}{\partial x'^2}\right)u_0\\
\frac{\partial}{\partial x}u_v &= \frac{\partial t'}{\partial x}\frac{\partial}{\partial t'}u_0 + \frac{\partial x'}{\partial x}\frac{\partial}{\partial x'}u_0 = \gamma\left(-\frac{v}{c^2}\frac{\partial}{\partial t'} + \frac{\partial}{\partial x'}\right)u_0,\\
\frac{\partial^2}{\partial x^2}u_v &= \gamma^2\left(\frac{v^2}{c^4}\frac{\partial^2}{\partial t'^2} - 2\frac{v}{c^2}\frac{\partial}{\partial t'}\frac{\partial}{\partial x'} + \frac{\partial^2}{\partial x'^2}\right)u_0\\
\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2}\right)u_v &= \gamma^2\left(\left(1-\frac{v^2}{c^2}\right)\frac{1}{c^2}\frac{\partial^2}{\partial t'^2} + \frac{v^2-c^2}{c^2}\frac{\partial^2}{\partial x'^2}\right)u_0\\
&= \left(\frac{1}{c^2}\frac{\partial^2}{\partial t'^2} - \frac{\partial^2}{\partial x'^2}\right)u_0 = 0.
\end{aligned}
$$

The transformation of the coordinates which we have used here to create the new solution is, in the case of c being the speed of light, named Lorentz transformation. For other wave equations, like sound waves or water waves, the name "Lorentz transformation" is not used, and the Lorentz transformations are seldom used. But, nonetheless, the mathematics of the Lorentz transformation works in the same way for these equations too.

This new solution of the same wave equation is known as the Doppler-shifted solution. It has a well-defined physical meaning: If the source of the wave is, in the original solution, at rest, then the Doppler-shifted solution is the solution where the same source is moving with the velocity v.

Note that above solutions are clearly physically different. The source is, in the first solution, at rest, while it moves in the second solution. And if the first solution has spherical symmetry. The second solution does no longer have such a symmetry: If the train moves in your direction, you hear a different sound than if the train moves away. Despite these differences, some properties remain unchanged an symmetric - namely the speed of the wave.

2.2 Relativistic denotations

It seems time to introduce a lot of relativistic conventions, conventions which allow to write memorize quite short formulas for otherwise quite long and boring sets of equations.

The first idea is to handle time - an a priori something qualitatively quite different from spatial coordinates - like a spatial coordinate. This coordinate is denoted $x^0 = ct$. The introduction of the factor c into the definition of x^0 allows to get rid of a lot of factors c in the formulas, and to recover them correctly all one has to remember is that formula $x^0 = ct$. The spatial coordinates obtain indices from 1 to 3. A typical relativistic formula contains the space and time coordinates in an equivalent way, thus, is a formula for some "four-dimensional space-time". But there will be, often enough, also formulas which are not fully relativistic, formulas which handle spatial indices separately. Now, there is a convention which handles this: Whenever an index runs from 0 to 3, thus, is a space-time index, one uses a Greek letter to denote it. If the index is, instead, a spatial index only, one uses Latin indices for it. So, one can write a function which depends on all four coordinates as $f(x^\mu)$, or as $f(x^0, x^i)$.

A nice convention if one considers transformations of coordinates, like the Lorentz transformation, is to use the same index for the new coordinates, only primed. So, one writes the quite short formula $x^{\mu'} = x^{\mu'}(x^\mu)$ to describe the four new coordinates, as functions depending on the four old coordinates. The rule how to transform partial derivatives between different coordinates are, obviously, the following:

$$\frac{\partial}{\partial x^\mu} = \sum_{\mu'=0}^{3} \frac{\partial x^{\mu'}}{\partial x^\mu} \frac{\partial}{\partial x^{\mu'}}, \qquad \frac{\partial}{\partial x^{\mu'}} = \sum_{\mu=0}^{3} \frac{\partial x^\mu}{\partial x^{\mu'}} \frac{\partial}{\partial x^\mu}.$$

The next simplification is to write partial derivatives $\frac{\partial}{\partial x^\mu}$ simply as ∂_μ. This gives:

$$\partial_\mu = \sum_{\mu'=0}^{3} \frac{\partial x^{\mu'}}{\partial x^\mu} \partial_{\mu'}, \qquad \partial_{\mu'} = \sum_{\mu=0}^{3} \frac{\partial x^\mu}{\partial x^{\mu'}} \partial_\mu.$$

Above formulas are sufficiently easy to remember, and have something in common: The index of the summation appears two times - once as an upper index, once as a lower index. This rule applies so often, that it can be considered as a general rule: Whenever there is a summation, it has to be over one upper and one lower index. And whenever there is an index used twice, as an upper index as well as a lower index, there is summation over this index. The second part became simply a convention, so that the summation sign can be omitted if the index is used twice, as an upper and a lower index. So, the formula becomes:

$$\partial_\mu = \frac{\partial x^{\mu'}}{\partial x^\mu} \partial_{\mu'}, \qquad \partial_{\mu'} = \frac{\partial x^\mu}{\partial x^{\mu'}} \partial_\mu.$$

In relativistic physics, there are a lot of objects with various indices, and for each index one has to remember how it has to be transformed if the coordinates are transformed. The formula above describes two possibilities - one which uses the partial derivatives of the new coordinates as functions depending on the old coordinates, and one which uses the partial derivatives of the old coordinates as functions of the new coordinates. Above four times four matrices are inverse to each other. To make remembering easy, there has been a simple convention: the upper vs. the lower position of the index defines which is the transformation rule. And which rule has to be used corresponds to the rule above: the sum has to be about one upper and one lower index. So, a lower index transforms in the same way as the partial derivatives. Say, for some a_μ we obtain the transformation rule:

$$a_\mu = \frac{\partial x^{\mu'}}{\partial x^\mu} a_{\mu'}, \qquad a_{\mu'} = \frac{\partial x^\mu}{\partial x^{\mu'}} a_\mu.$$

Instead, if we have some upper index, we have to use the other, inverse rule:

$$a^\mu = \frac{\partial x^\mu}{\partial x^{\mu'}} a^{\mu'}, \qquad a^{\mu'} = \frac{\partial x^{\mu'}}{\partial x^\mu} a^\mu.$$

2.3 The Lorentz group

With these denotations, let's rewrite now the operator $\Box$ for the wave equation:

$$\Box = \partial_0^2 - \sum_i \partial_i^2 = \eta^{\mu\nu}\partial_\mu\partial_\nu$$

where $\eta^{\mu\nu}$ is a 4×4 matrix where only the diagonal entries are nonzero, and have the values $\eta^{00} = 1, \eta^{ii} = -1$. Now, let's assume we have a general coordinate transformation which is linear in the coordinates, thus, $x^{\mu'} = a^{\mu'}_{\mu} x^\mu$. This gives $\partial_\mu = \frac{\partial x^{\mu'}}{\partial x^\mu}\partial_{\mu'} = a^{\mu'}_{\mu}\partial_{\mu'}$ and

$$\Box = \eta^{\mu\nu}\partial_\mu\partial_\nu = \left(\eta^{\mu\nu}a^{\mu'}_{\mu}a^{\nu'}_{\nu}\right)\partial_{\mu'}\partial_{\nu'} = \eta^{\mu'\nu'}\partial_{\mu'}\partial_{\nu'}.$$

Thus, the property of the Lorentz transformation we need to obtain another solution of the same wave equation is

$$\eta^{\mu\nu}a^{\mu'}_{\mu}a^{\nu'}_{\nu} = \eta^{\mu'\nu'}.$$

These special coordinate transformations form a group: We can apply two such transformations, one after the other, and the result would be the same, yet another solution of the wave equation. And we can invert it, to get the original solution back - and the inverted transformation would be also, yet, another example of such a transformation. This group is named the Lorentz group.

2.4 An important difference: Active vs. passive coordinate transformations

The mathematics of coordinate transformations can be used in two from a physical point of view very different ways.

The first way is named ¡b¿passive¡/b¿ or ¡b¿alias¡/b¿ coordinate transformations. In this case, one and the same physical solution is described in different ways, using different coordinates. This is nothing but an application of pure mathematics, without any physical importance. If the mathematics are used correctly, it does not matter at all which system of coordinates you use to describe the solution. Your choice of coordinates is arbitrary. You can check this, by trying to compute something measurable using different coordinates. The final result should be the same. If not, you have made a mathematical error. But you are not obliged to do such things at all. You can, as well, choose one system of coordinates, and refuse even to look at any other one, but you are nonetheless able to compute everything physical using only that single system of coordinates.

The second way is named ¡b¿active¡/b¿ or ¡b¿alibi¡/b¿ coordinate transformation. In this case, the coordinates remain unchanged. What changes is the solution. This is what we have used here to obtain a new, Doppler-shifted solution of the wave equation out of a given one.

Note the difference: We could have used any other coordinate transformation to get another description of the same solution. If that other coordinate transformation would not be a Lorentz transformation, we would have to rewrite the same wave equation in these other coordinates. It would have been a different-looking equation, but the different-looking solution of the different-looking equation would have been simply another, equally valid, description of the same solution of the same equation.

Instead, a different coordinate transformation would not have allowed us to construct a new, different solution of the same equation. For this trick, it was essential that the transformed equation looked, by accident, like the original equation. In general, the transformation would have given us some solution of some other equation, something in no way useful to study the solutions of the wave equation. But in our case, the Lorentz transformation has given us, by happy accident, another, new solution, with different properties, on the same wave equation.

Could we have used the Lorentz transformations to describe the same wave equation in other coordinates? Yes, of course. But these other coordinates would have been a quite unnatural, strange choice. Of course, one is free to use whatever coordinates one likes. But, in the case of water waves, would be the point of using a "time coordinate" which have nothing to do with real time?

So, once we have understood this important difference between active (alibi) and passive (alias) transformations, we have found a useful application of the Lorentz transformations as active transformations, transformations which create new, different solutions of the wave equation.

2.5 Other equations where Lorentz transformations allow to find new solutions

Once the Lorentz transformations allow to create new solutions of the wave equation, another question appears: Are there other interesting equations with the same property, so that we can use Lorentz transformations to find new, different solutions? The answer is positive. There are a lot of other, more complex but very interesting equations where the Lorentz transformations allow to find new solutions.

2.5.1 Mass terms

First of all, one can add a mass term. This gives a Klein-Gordon equation, which describes a massive scalar particle:

$$\frac{\partial^2}{\partial t^2}u(\vec{x},t) - c^2\nabla^2 u(\vec{x},t) + \left(mc^2\right)^2 u(\vec{x},t) = 0.$$

In the more economic relativistic denotations, with the units taken in such a way that $c = 1$, this equation looks like

$$\Box u(x^\mu) + m^2 u(x^\mu) = 0.$$

This equation is also linear. The solutions of this equation are waves, but their velocity is already lower than c.

2.5.2 Interaction terms of different fields

The field itself is not obliged to be a simple real field, it can have many components. And these many components can interact in quite complex ways. All one needs to preserve the property that the Lorentz transformation of a solution gives another solution is that the interaction term does not contain any spatial derivatives. Thus, the general form of the equation would be the following: Some number of fields $u^\alpha(x^\mu)$ and the following equation:

$$\Box u^\alpha(x^\mu) + V^\alpha(u^\beta(x^\mu)) = 0.$$

These interaction terms make the system of equations itself nonlinear. This makes it usually difficult, or even impossible, to find exact solutions. But, despite this, the basic property of the Lorentz transformation remains - if we have an exact solution, the Lorentz transformation creates a new, different solution of the same equation.

2.5.3 The electromagnetic field

Another very interesting example is the electromagnetic field. The simplest way to describe it, and to see how the Lorentz transformation can be used, is to consider the equation for the electromagnetic potentials $\phi(x,t), \vec{A}(x,t)$ in a gauge condition

$$\frac{1}{c}\frac{\partial}{\partial t}\phi + \frac{\partial}{\partial x}A^x + \frac{\partial}{\partial y}A^y + \frac{\partial}{\partial z}A^z = 0,$$

which is named Lorenz gauge. Note that it is not "Lorentz gauge", because it is not named after Hendrik Lorentz, but after another physicist, Ludvig Lorenz.

In relativistic denotations, the electric scalar potential ψ and the magnetic vector potential $\vec{A}$ are combined into an electromagnetic four-potential A^μ, with $A^0 = \phi$. Then, the Lorenz gauge obtains the much simpler form

$$\partial_\mu A^\mu = 0.$$

Let's first check that the Lorenz gauge remains the Lorenz gauge condition We also have to check what happens with the Lorenz gauge during the Lorentz transformation:

$$\partial_{\mu'} A^{\mu'} = \left(\frac{\partial x^\mu}{\partial x^{\mu'}}\partial_\mu\right)\left(\frac{\partial x^{\mu'}}{\partial x^\mu}A^\mu\right) = \left(\frac{\partial x^\mu}{\partial x^{\mu'}}\right)\left(\frac{\partial x^{\mu'}}{\partial x^\mu}\right)\partial_\mu A^\mu = \partial_\mu A^\mu.$$

Here we have used that the coefficients $\frac{\partial x^{\mu'}}{\partial x^\mu}$ are constants, so can be taken out of the partial derivative, and that $\frac{\partial x^{\mu'}}{\partial x^\mu}$ is the inverse matrix of $\frac{\partial x^\mu}{\partial x^{\mu'}}$.

In the Lorenz gauge, the Maxwell equations become simply wave equations for the electromagnetic potential, so that we have:

$$\Box A^\mu(x^\nu) = 0.$$

This is already a good starting point, given that we already know that the operator $\Box$ behaves as necessary. But the Lorentz transformation of this equation also has to do a transformation of the vector index. So, the result of the Lorentz transformation will be a solution of the following equation:

$$\Box A^{\mu'} = \Box\left(\frac{\partial x^{\mu'}}{\partial x^\mu}A^\mu\right) = \left(\frac{\partial x^{\mu'}}{\partial x^\mu}\right)\Box A^\mu(x^\nu) = 0.$$

This is, fortunately, also equivalent to the original equation. Again, we can take the constant coefficients out of the differential operator. And the resulting system of equations is equivalent to the original one, because the matrix $\frac{\partial x^{\mu'}}{\partial x^{\mu}}$ is invertible, with $\frac{\partial x^{\mu}}{\partial x^{\mu'}}$ being the inverse matrix. So, the system is equivalent to the original system of equations $\Box A^{\mu} = 0$.

The Maxwell equations have been, in fact, the equations where the first equations where the Lorentz transformations have been found.

2.5.4 The Dirac equation

The Dirac equation is another important equation in modern physics. It is the equation used to describe fermion fields. It is an equation for a system of four complex (or, equivalently, eight real) fields. The equation was obtained by Dirac in an attempt to take a square root out of the Klein-Gordon equation. But the Klein-Gordon equation is already among the equations for which the Lorentz transformation allows to construct new solutions. So, it is not really a surprise that this works for the Dirac equation too.

In relativistic notations, the equation looks like

$$i\gamma^{\mu}\partial_{\mu}\psi(x^{\mu}) = m\psi(x^{\mu}).$$

The square of the operator $\gamma^{\mu}\partial_{\mu}$ is, by construction of the matrices γ^{μ}, the Laplace operator $\Box$. So, it follows from this equation that $-\Box\psi = m^2\psi$, thus, the equation is a square root of the Klein-Gorden equation.

It also follows immediately that the Lorentz-transformed equation fulfills the same basic property, namely that its square gives the Klein-Gordon equation. Unfortunately, this is nonetheless not exactly the same equation, but a different representation. Fortunately, all these different representations are equivalent, thus, one can find a transformation U so that $\psi'(x^{\mu}) = U\psi(x^{\mu})$ is already again a solution of the Dirac equation in its original representation.

The subtle point is that this operator U is not uniquely defined. The operator -U would do it too.

2.5.5 And, again, various interaction terms

What has been said about point-wise interaction terms for the scalar wave equation holds also for all the other equations considered - it is possible to add various point-wise interaction terms. The freedom of choice is somewhat restricted, not completely without any restrictions (except for containing no derivatives) as in the scalar case - as the EM field, as the Dirac operator follow some transformation laws, and these transformation laws have to fit each other.

The interaction terms which are important in the standard model of particle physics are:

1.) Non-abelian gauge fields: This is a generalization of the EM field, and formally looks like several such EM fields which additionally interact with each other. 2.) The interaction of these gauge fields with Dirac fermions. The interaction term has a quite special form, namely a replacement of the partial derivative by an additional term:

$$i\partial_{\mu} \to i\partial_{\mu} + gA_{\mu},$$

with the charge of this fermion field being g.

2.5.6 The equation of the Standard Model of particle physics

So, we have found a lot of wave equations which all share the same nice property: If we have a solution of these equations, we can, using a Lorentz transformation, create new, different solutions of these equations.

How important is this class of wave-like equations? The surprising news is that all the fundamental fields, all the fields used in the Standard Model (SM) of modern particle physics, are described by such equations. And they all share the same constant c - the speed of light. In principle, this SM can be considered as a single big equation, let's name it the SM equation, containing many different parts, which interact with each other. But all these parts, and all the interaction terms, fit into the list of equations above. So that the SM equation is also of this type. If we have one solution of the SM equation, we can apply a Lorentz transformation, and will obtain another, different solution of the SM equation, a solution which will be, in comparison with the original solution, Doppler-shifted.

2.6 Relativistic effects

What is the consequence of the fact that all fields, the whole SM, follow an equation where we can apply a Lorentz transformation to obtain a new solution of the same equation? Some surprising results about the behavior of clocks follow.

2.6.1 Time dilation

Let's construct, out of what we have, namely of of things described by the SM, a clock. This clock will be described by some trajectory, with some numbers on it which denote the time measured at this point. Let's simply assume, as an example, that this trajectory is, for the initial solution with the clock at rest, a line with the spatial coordinates (0,0,0) and with a result 0 at time 0 and result 1 at time 1.

What happens now with this solution, if we apply some Lorentz transformation? We obtain another solution. This other solution describes a clock of the same construction, but moving. The Lorentz transformation is linear, thus the point $(x^\mu) = (0,0,0,0)$ remains the same. But the point where the clock shows 1, which was originally at $(x^\mu) = (1,0,0,0)$, will now be in $(x^{\mu'}) = (\gamma, -\frac{v}{c}\gamma, 0, 0)$. The line remains a line, so that the clock is moving with the speed v (which is $\frac{v}{c}$ in terms of the time coordinate $x^0 = ct$. But the clock shows the result 1 at real time γ instead of real time 1. But we have

$$\gamma = \sqrt{1-\frac{v^2}{c^2}}^{-1} > 1$$

for every $0 < |v| < c$. That means, the moving clock shows clock time 1 only at real time $\gamma > 1$, thus, is dilated.

2.6.2 Length contraction

What happens with a ruler? The ruler at rest is some solution of the SM, but we cannot idealize it as a single point, we need two points of it. The begin we put, again, at $(x^\mu) = (0,0,0,0)$, the mark of the ruler with length 1 will, if we measure the x-direction, be at $(x^\mu) = (0,1,0,0)$. At rest, the begin moves toward $(x^\mu) = (1,0,0,0)$ and the mark with the 1 toward $(x^\mu) = (1,1,0,0)$.

We apply the same procedure and obtain a solution of the same type, thus, also a ruler, but moving. The four resulting points will be the following: The begin will be, like for the clock, moving along the line from $(x^{\mu'}) = (0,0,0,0)$ to $(x^{\mu'}) = (\gamma, -\frac{v}{c}\gamma, 0, 0)$. The point with mark 1 will, instead, move from $(x^{\mu'}) = (-\gamma\frac{v}{c}, \gamma, 0, 0)$ to $(x^{\mu'}) = (\gamma(1-\frac{v}{c}), \gamma(1-\frac{v}{c}), 0, 0)$. Where is the mark 1 of this moving ruler at $t = 0$? We have to compute where the line between the two points intersects the line $t' = 0$. The result is $(x^{\mu'}) = (0, \gamma(1-\frac{v^2}{c^2}), 0, 0)$. Thus, the mark 1 of the moving ruler will be, at time $t = 0$, at a distance $\gamma^{-1} < 1$ from the origin. The moving ruler is shorter.

2.6.3 Impossibility to identify global contemporaneity

A similar fate waits for devices which allow to measure absolute contemporaneity. Such a device would tell us that the two events $(x^\mu) = (0,0,0,0)$ and $(x^\mu) = (0,1,0,0)$ have happened at the same moment of time. But now we apply a Lorentz transformation to this device, and what is the result? The same device, only in a moving state, claims that $(x^\mu) = (0,0,0,0)$ and $(x^\mu) = (-\gamma\frac{v}{c}, \gamma, 0, 0)$ have happened at the same time.

2.6.4 Impossibility to measure absolute rest

And now it is already clear what happens if we try to measure, with some physical device, what is absolute rest. Suppose we have such a device. This device measures that the line from $(x^\mu) = (0,0,0,0)$ to $(x^\mu) = (1,0,0,0)$ is at absolute rest. We do the same trick again, and apply the Lorentz transformation to this measurement of absolute rest. What do we obtain? A solution which describes the same measurement device, which claims that the line from $(x^\mu) = (0,0,0,0)$ to $(x^\mu) = (\gamma, -\frac{v}{c}\gamma, 0, 0)$ is at absolute rest. Which is, of course, wrong.

The consequence is the ¡b¿relativity principle¡/b¿: Once all our physical equations allow the application of Lorentz transformations to obtain a new solution of our equations, only moving relative to the original solution, there cannot exist a physical device which identifies the own state of movement.

2.7 The Lorentz ether vs. Minkowski space-time

What follows from the relativity principle?

2.7.1 The Lorentz ether

In the Lorentz ether interpretation, nothing follows. Wave equations appear, most natural, as a sort of sound wave equations of some material. Once we observe something nicely described by wave equations, the reasonable hypothesis is that they are all similar to sound waves of some ether.

The relativity principle is, then, only an unfortunate consequence of the fact that all what we can actually observe are sound waves of the same ether. For usual condensed matter, there exist other things, like light rays, which are not sound waves of this condensed matter. With these additional possibilities accessible to use, the relativity principle would become invalid, and we would be able to measure as absolute rest, as absolute contemporaneity, and measure absolute lengths and time. Without them, we are simply unable to distinguish things which, in reality, are different.

In the Lorentz ether, time dilation and length contraction are distortions of our measurement devices caused by their motion relative to the ether. True distances and true time could be measured - if our measurement devices would be, by accident, really at rest. Once we cannot know if they are really at rest, one cannot be sure that our measurements are undistorted by the ether.

2.7.2 The Minkowski space-time

There is another interpretation of special relativity - the space-time interpretation proposed by Minkowski. It is the one accepted by the mainstream of physics. At this place I have to admit that, given that I'm not a proponent of the space-time interpretation, the argumentation in its favor will be suboptimal.

From point of view of the space-time interpretation, once there is no possibility, by any measurement, to distinguish if we are at rest, then no such animal like "absolute rest" exists in reality. The same for absolute contemporaneity: Once there is no possibility to establish it, uniquely, by measurement, there is no absolute contemporaneity in our real world.

This is, of course, in sharp contradiction with classical common sense. Common sense has no problem with the mathematical possibility of a space-time. But this would be simply a collection of the states of space, in various moments of absolute time. It would not be a description of what exists, but of all what has existed in the past, exists now, and will exist in the future. All this would have to be taken together, and given a common status of existence, without any reference of when it existed. This would be, in fact, a new concept of timeless existence. There is no more any difference between events which have happened and those which will happen, they all simply exist.

In this four-dimensional world there would be no present at all. What is the present which we experience? It can be only something particular, derivative, restricted to the particular trajectory in this space-time which describes our own life stream, our world-line, which contains all the events of our own life, from our birth to our death.

Chapter 3

Relativistic Gravity

Special Relativity, in the space-time interpretation, does not suggest any modifications. There is a need to make Newtonian gravity compatible with relativity. Such a proposal has been already made by Poincare in this 1905 paper. It failed. But the failure was a failure of the theory to agree with observation, not an internal, theoretical problem. One may reject the space-time interpretation as some artificial, mystical object, acceptable only to fatalists who think that our future is predefined, but this does not change the fact that it is, in itself, a consistent theory, which does not have any internal problems, which would be a starting point for developing general relativity.

One may object that Einstein, developing general relativity, was following philosophical ideas which were quite similar to those he used to develop special relativity. So, in above cases, some equivalence principle played a central role, and, in general, relativism and observer-dependence of many things. But to follow Einstein may not be a good idea. The ideas Einstein has used to find general relativity were in part simply wrong in the final theory (Mach's principle is wrong in GR), in part only approximately true (a non-trivial gravitational field has non-zero curvature, while pure acceleration gives zero curvature, so that there is no exact equivalence between gravity and acceleration in GR), in part misguided (Kretschmann's objection: every physical theory can be presented in a general covariant form, thus, this cannot be a physical property of GR).

And, in fact, the resulting theory is qualitatively very different from special relativity. In particular, the central, distinguishing property of GR is background independence, its completely local character. It does not contain any global object. As Newtonian theory, as special relativity contain such absolute, global objects: Newtonian space and time, or the Minkowski space-time. Above are predefined global objects, with permanent, fixed properties which do not change in time, and cannot be influenced in any way by matter – they define a fixed stage.

So, if we look at special relativity and general relativity, we see two conceptually very different theories, and there is no natural connection between them, no open problem in SR so that its solution would lead us to GR.

So, essentially one of the simplest ways to introduce GR is simply to define it as it is, as a covariant, background-independent metric theory of gravity.

3.1 Making the Lorentz ether compressible and transient

Surprisingly, this is different for the Lorentz ether.

One could think about an interpretation slightly different from the Lorentz ether, let's name it Lorentz space-time, which would be equally sterile as the Minkowski space-time. It would be simply the Minkowski space-time with an unobservable preferred frame. This preferred frame would define the global notion of "now", but would not have any meaning in terms of some ether. There would be no ether, so that the wave equation would describe sound waves of the ether. There would be simply some universal abstract absolute wave operator $\square = c^{-2}\partial_{\mathfrak{t}}^2 - \Delta$ without further physical interpretation. This Lorentzian space-time would have absolute space and absolute time, and no Galilean invariance, thus, conceptually even closer to Newton's space-time. Nothing would lead from this interpretation toward another theory.

The Lorentz ether is different. The Lorentz ether assumes that this Newtonian background space is filled with some ether, and that the wave equations describe various sound waves of the ether. Now, this attempt to explain the universal wave equation as some physical equation, analogical to those we already know from condensed matter theory, leads immediately to some problems, inconsistencies of the explanation, which tell us that this ether interpretation can be only an approximation. Indeed, the ether is assumed to be ideally rigid, incompressible, homogeneous and static. But, of course, a wave of the ether defines some compression, some deformation, some inhomogeneity, some change in time. So, there has to be some more fundamental theory, which defines how the Lorentz ether deforms and changes in time.

So, the Lorentz ether contains, in itself, acceptance that the theory as it is is only an approximation of a different, more complex theory, which contains an inhomogeneous, transient, compressible ether.

At a first look, this seems to be not much information, not enough to hope that this allows to extract sufficient information to derive some fundamental ether theory. Surprisingly, it appears quite sufficient. There is a quite straightforward way from what we have found here toward that more fundamental ether theory, and the resulting ether theory of gravity appears to be very close to GR, so close that we obtain the Einstein equations of GR in a natural limit of this ether theory.

3.1.1 Continuity and Euler equations

The natural way to start to look for a dynamic ether theory is to use standard equations of standard condensed matter theory. The most general equations are the continuity and Euler equations. In the preferred coordinates $\mathfrak{t}, \mathfrak{x}^i$ the continuity equation would be:

$$\partial_{\mathfrak{t}}\rho + \partial_{\mathfrak{x}^i}(\rho v^i) = 0, \tag{3.1}$$

and the Euler equations would be in the most general form, with a stress tensor σ^{ij}) instead of a scalar pressure:

$$\partial_{\mathfrak{t}}(\rho v^j) + \partial_{\mathfrak{x}^i}(\rho v^i v^j - \sigma^{ij}) = 0. \tag{3.2}$$

Now one can rewrite these condensed matter theory equations in four-dimensional denotations, introducing the four-dimensional fields $\mathfrak{g}^{00} = \rho, \mathfrak{g}^{0i} = \rho v^i, \mathfrak{g}^{ij} = \rho v^i v^j - \sigma^{ij}$. Then, the continuity and Euler equations obtain the form

$$\partial_{\mathfrak{x}^\mu} \mathfrak{g}^{\mu\nu}(\mathfrak{x}) = 0.$$

The next interesting observation is that, if we introduce the four-dimensional wave operator

$$\square = \partial_{\mathfrak{x}^\mu} \mathfrak{g}^{\mu lambda}(\mathfrak{x}) \partial_{\mathfrak{x}^\lambda},$$

then the equations ca be rewritten as a simple wave equation for the four preferred coordinates $\mathfrak{x}^\nu$:

$$\partial_{\mathfrak{x}^\mu} \mathfrak{g}^{\mu\nu}(\mathfrak{x}) = \square \mathfrak{x}^\nu = 00.$$

In particular, the continuity equation is equivalent to $\square \mathfrak{t} = 0$, and the Euler equations to $\square \mathfrak{x}^i = 0$.

3.1.2 The harmonic equation

Let's now look at the properties of the general wave equation $\square u(\mathfrak{x}, \mathfrak{t}) = 0$. An important property of this equation is that it appears, in a natural way, as an Euler-Lagrange equation of the following action:

$$S = \frac{1}{2} \int \mathfrak{g}^{\mu\nu} \partial_\mu u \partial_\nu u d^3 \mathfrak{x} d\mathfrak{t}.$$

Variation for $u(\mathfrak{x}, \mathfrak{t})$ gives

$$\frac{\delta S}{\delta u} = \square u.$$

This action allows to consider the question of how to reach general covariance of this expression. To be general covariant, the expression under the integral should remain unchanged under coordinate transformations. Here, we have to apply the general rule how partial derivatives transform, that means,

$$\frac{\partial}{\partial x^\nu} = \frac{\partial x^{\nu'}}{\partial x^\nu} \frac{\partial}{\partial x^{\nu'}}.$$

The other rule we have to apply is the rule of transformation for the measure $d^4 x$, which gives a determinant in the transformation rule:

$$d^4 x = \left| \frac{\partial x^\kappa}{\partial x^{\kappa'}} \right| d^4 x'.$$

Taking all this together gives

$$\mathfrak{g}^{\mu\nu} \frac{\partial u}{\partial x^\mu} \frac{\partial u}{\partial x^\nu} d^4 x = \mathfrak{g}^{\mu\nu} \frac{\partial x^{\mu'}}{\partial x^\mu} \frac{\partial x^{\nu'}}{\partial x^\nu} \frac{\partial u}{\partial x^{\mu'}} \frac{\partial u}{\partial x^{\nu'}} \left| \frac{\partial x^\kappa}{\partial x^{\kappa'}} \right| d^4 x' = \mathfrak{g}^{\mu'\nu'} \frac{\partial u}{\partial x^{\mu'}} \frac{\partial u}{\partial x^{\nu'}} d^4 x'$$

so that we need the following transformation rule:

$$\mathfrak{g}^{\mu'\nu'} = \mathfrak{g}^{\mu\nu} \frac{\partial x^{\mu'}}{\partial \mathfrak{x}^\mu} \frac{\partial x^{\nu'}}{\partial \mathfrak{x}^\nu} \left| \frac{\partial x^\kappa}{\partial \mathfrak{x}^{\kappa'}} \right|,$$

Thus, the rule of transformation one needs for covariance is a little bit more difficult than that for a tensor field with two upper indices.

Fortunately, if we have a tensor field with two upper indices $g^{\mu\nu}$, there is a well known way to define another field which transforms like the field we need. We have to multiply it with the square root of the determinant of its inverse matrix $g_{\mu\nu}$. Indeed,

$$\sqrt{|g_{\mu'\nu'}|} = \sqrt{\left|\frac{\partial x^\mu}{\partial x^{\mu'}}\frac{\partial x^\nu}{\partial x^{\nu'}}g_{\mu\nu}\right|} = \sqrt{\left|\frac{\partial x^\mu}{\partial x^{\mu'}}\right|\left|\frac{\partial x^\nu}{\partial x^{\nu'}}\right||g_{\mu\nu}|} = \left|\frac{\partial x^\kappa}{\partial x^{\kappa'}}\right|\sqrt{|g_{\mu\nu}|}$$

So, it appears that the expression $g^{\mu\nu}\sqrt{|g_{\mu\nu}|}$ has exactly the transformational behavior we need. So, we define

$$\mathfrak{g}^{\mu\nu}(\mathfrak{x},\mathfrak{t})d^3\mathfrak{x}d\mathfrak{t} = g^{\mu\nu}\sqrt{|g_{\mu\nu}|}d^3\mathfrak{x}d\mathfrak{t}.$$

Now, for the determinant one uses the abbreviation $g = |g_{\mu\nu}|$. It appears (even if the formulas considered up to now do not require this) that this determinant is usually negative. So, to avoid an extra factor i, one prefers to take the square root not of g, but of -g. The tensor field $g_{\mu\nu}(\mathfrak{x},\mathfrak{t})$ is, then, the gravitational field. Using the usual transformation rules for tensor fields, it can be defined in arbitrary systems of coordinates. The action can, then, be rewritten in arbitrary coordinates as

$$S = \frac{1}{2}\int g^{\mu\nu}\partial_\mu u\partial_\nu u\sqrt{-g}d^4x$$

and the resulting Euler-Lagrange equation is, also in arbitrary coordinates, defined by

$$\Box u = \partial_\mu g^{\mu\nu}\sqrt{-g}\partial_\nu u = 0.$$

3.1.3 Conservation laws as Euler-Lagrange equations for the preferred coordinates

Given that the action for the scalar field $u(x)$ is covariant, thus, does not depend on the choice of the coordinates, we can now look how the action which gives the harmonic coordinate condition depends on the preferred coordinates. The action itself can be written as the sum of the terms for four preferred coordinates $\mathfrak{x}^\alpha$, which look, in general coordinates, like four "scalar fields" $\mathfrak{x}^\alpha(x)$, with some nonzero but otherwise arbitrary constants Ξ_α:

$$S = \frac{1}{2}\int \Xi_\alpha g^{\mu\nu}\partial_\mu\mathfrak{x}^\alpha\partial_\nu\mathfrak{x}^\alpha\sqrt{-g}d^4x.$$

Now, the harmonic coordinate conditions appear, in a quite covariant way, as Euler-Lagrange equations for the preferred coordinates $\mathfrak{x}^\alpha(x)$:

$$\Box\mathfrak{x}^\alpha(x) = \frac{\partial}{\partial x^\mu}g^{\mu\nu}(x)\sqrt{-g}\frac{\partial}{\partial_x^\nu}\mathfrak{x}^\alpha(x) = 0 \tag{3.3}$$

This observation is, in fact, a particular example of the Noether theorem. We have a Lagrangian which depends, in a quite explicit way, on the preferred coordinates $\mathfrak{x}^\alpha(x)$. But it depends only on partial derivatives of these preferred coordinates, thus, it does not change if we shift the preferred coordinates: $\mathfrak{x}^\alpha(x) \to \mathfrak{x}^\alpha(x) + c^\alpha$. Thus, the theory has translational invariance in space and time. The Noether theorem tells us that this leads to conservation laws for energy and momentum. But this is, in this case, a trivial consequence of the Euler-Lagrange equations, and the fact that it depends not on the $\mathfrak{x}^\alpha(x)$ themselves, but only on its partial derivatives:

$$\frac{\delta S}{\delta \mathfrak{x}^\mu} = \frac{\partial L}{\partial \mathfrak{x}^\mu} - \partial_\kappa \frac{\partial L}{\partial_\kappa \mathfrak{x}^\mu} + \ldots.$$

Given that the first term is zero, the Euler-Lagrange equation obtains automatically the form of a conservation law.

The energy-momentum tensor which is conserved here depends, obviously and explicitly, on the preferred coordinates. It is the energy-momentum density

$$T^{\mu\nu}(x) = \Xi_\alpha g^{\mu\nu}(x)\sqrt{-g}\partial_\nu \mathfrak{x}^\alpha(x),$$

which, in the preferred coordinates, reduces to

$$T^{\mu\nu}(\mathfrak{x}, \mathfrak{t}) = \Xi_\nu g^{\mu\nu}(\mathfrak{x}, \mathfrak{t})\sqrt{-g},$$

so that the conservation laws $\partial_\mu T^{\mu\nu}(\mathfrak{x}, \mathfrak{t}) = 0$ reduce to the harmonic conditions, that means, the continuity and Euler equations. In particular, the energy conservation law becomes the continuity equation, and the momentum conservation laws become the Euler equations.

3.1.4 The Lagrangian of the general Lorentz ether

With the action

$$S_{harm} = \frac{1}{2}\int \Xi_\alpha g^{\mu\nu}\partial_\mu \mathfrak{x}^\alpha \partial_\nu \mathfrak{x}^\alpha \sqrt{-g} d^4x.$$

we have a nice equation for the preferred coordinates, which appears to be the classical conservation laws of condensed matter theory – the continuity and Euler equations. But to define how everything moves, we need more equations. In particular, we need equations for all components of the gravitational field $g_{\mu\nu}$ as well as equations for the various material properties of the ether, which we will denote now by $\varphi^m(x)$. So, let's look for a general Lagrangian of the following form:

$$S_{gen} = \int L_{gen}(g_{\mu\nu}, \varphi^m, \mathfrak{x}^\mu) d^4x.$$

Now, let's note the main difference between a Lagrangian for an ether which interacts with some external, different material, and for an ether which does not have interactions with anything external, but is influenced only by its own material properties. In the first case, we would have to modify energy and momentum conservation, because only the whole energy, and the whole momentum would be conserved, not the parts of it which can be attributed to the ether only. There would be some exchange of energy and momentum between the ether and all the other matter. In the latter case, the conservation laws remain unchanged, uninfluenced by the material properties. We assume here that there is no such external matter, thus, the harmonic equation has to remain unchanged.

That means, we have to assume that

$$\frac{\delta S_{gen}}{\delta x^\alpha} = \Xi_\alpha \partial_\mu (g^{\mu\alpha}\sqrt{-g}) = \frac{\delta S_{harm}}{\delta x^\alpha}.$$

From this condition it immediately follows that

$$\frac{\delta(S_{gen} - S_{harm})}{\delta x^\alpha} = 0.$$

In other words, the difference has to be covariant, should not depend in any way on the preferred coordinates $\mathfrak{x}^\mu(x)$.

Covariance is the most important, distinguishing property of the Lagrangian of General Relativity (GR), as for the Lagrangian of the gravitational field itself, which is

$$L_{GR} = \int R\sqrt{-g} - \Lambda\sqrt{-g} d^4x,$$

as well as of the Lagrangian of the various matter fields of GR. There are, of course, more general possibilities to define covariant expressions, in particular any function $f(R)$ of the scalar curvature would be covariant, as well as, say, other possibilities to obtain scalar expressions out of the curvature tensor like $R^{\mu\nu}R_{\mu\nu}$. Here, the Lagrangian of GR is simply a restriction to the lowest order terms, which makes sense, because these are the terms most relevant in a large distance approximation. This restriction to lowest order terms makes sense in our ether theory too, thus, we can restrict ourselves to the GR Lagrangian too. Thus, we have obtained now, as the natural Lagrangian for an ether theory, the Lagrangian of GR, together with a simple additional term which enforces a preference for harmonic coordinates.

$$L = L_{harm}(g_{\mu\nu}, \mathfrak{x}^\mu) + L_{GR}(g_{\mu\nu}) + L_{matter}(g_{\mu\nu}, \varphi^m).$$

3.2 Basic properties of the Lorentz ether

The general Lorentz ether we have derived above is obviously a theory different from GR, given that it has a different Lagrangian, and, as a consequence, different equations. Nonetheless, it shares some key properties with GR.

3.2.1 The Einstein Equivalence Principle

The first remarkable observation is what the additional term does not influence at all – namely the equations for the matter fields. Given that the harmonic part of the Lagrangian does not depend on the matter fields $\varphi^m(x)$ at all, we have

$$\frac{\delta S}{\delta \varphi^m} = \frac{\delta S_{matter}}{\delta \varphi^m}.$$

Thus, the equations for the matter fields are the same as in GR. And that means that the equations for the matter fields do not depend on the preferred coordinates, that means, the Einstein Equivalence Principle (EEP) holds exactly for the Lorentz ether.

The EEP includes the Weak Equivalence Principle, that the local effects of motion in a curved space-time (gravitation) are indistinguishable from those of an accelerated observer in flat space-time, without exception. Moreover, the outcome of any local non-gravitational experiment in a freely falling laboratory is independent of the velocity of the laboratory and its location in space-time.

But the EEP is restricted to non-gravitational experiments. The Strong Equivalence Principle (SEP), which covers also gravitational experiments, does not hold: The equations for the gravitational field contain terms which depend on the preferred coordinates.

Above results are consequences of the "action equals reaction" principle, which is a consequence of the Lagrange formalism. Formally, the "action equals reaction" principle is simply the consequence that the functional derivatives do not depend on the order:

$$\frac{\delta}{\delta u}\frac{\delta}{\delta v}S(u,v) = \frac{\delta}{\delta v}\frac{\delta}{\delta u}S(u,v).$$

So, if the Euler-Lagrange equation for a variable u depends on v, then the Euler-Lagrange equation for v also depends on u, and both dependencies are equal.

This can be applied here too. The equation for the preferred coordinates, which is the harmonic equation, depends only on the gravitational field – this is how we have defined the gravitational field in this theory, namely as the field which defines the coefficients $\mathfrak{g}^{\mu\nu}(\mathfrak{x}, \mathfrak{t})$ of the harmonic equation. It follows, that only the equations of the gravitational field, but not the equations for matter fields, can depend on the preferred coordinates.

3.2.2 How the equations of the gravitational field are influenced

Let's look now at the type of influence which the preferred coordinates have on the equations of the gravitational field. Let's start with looking at the local behavior, ignoring global things. From a local point of view, and assuming that we use the correct sign for the $\Xi_\alpha > 0$, the additional terms of L_{harm} are the same as those we would obtain if we would simply add four scalar matter fields. These additional fields would not interact in any way with any type of matter, thus, this would be a classical example of dark matter. It would be not of the form we urgently need in cosmology, because what is necessary there is cold (massive) dark matter, while this would be hot (massless) dark matter, which would radiate away with the speed of light wherever it is. Nonetheless, these additional terms would not add anything qualitatively different.

Let's also note that the influence which the additional terms have on the equations of the gravitational field depends in their size on the parameters Ξ_α. In particular, it decreases proportional to the Ξ_α. That means, in the limit $\Xi_\alpha \to +0$ we obtain the undistorted equations of GR for the gravitational field, the Einstein equations.

If the coefficients are sufficiently small, $1 \gg \Xi_\alpha > 0$, we obtain, therefore, the Einstein equations, together with a negligible contribution of dark matter which can be ignored. Nonetheless, the equation for the preferred coordinates remains unaffected, the size of the Ξ_α, as long as they are non-zero, does not change anything for the harmonic equation. It remains an equation of the theory.

3.2.3 Global and qualitative restrictions following from the geometric character of the preferred coordinates

An important point is that, despite the equations are the same, there exist important qualitative differences between the ether theory, where the $\mathfrak{x}^\alpha(x)$ define a set of preferred coordinates and a theory which adds four scalar fields, denoting them $\mathfrak{x}^\alpha(x)$.

First of all, every solution of the ether theory defines a solution of the field theory. But this solution of the field theory is a quite strange one, because the four scalar fields $\mathfrak{x}^\alpha(x)$ do not follow the usual boundary conditions for matter fields, but have to become infinite in infinity. Locally, this may seem irrelevant, but one can expect some differences if one looks at the global, cosmological picture. One can distinguish the effects related with the boundary conditions by fixing one system of coordinates $\mathfrak{x}_0^\alpha(x)$ and describing all other solutions in form of a decomposition $\mathfrak{x}^\alpha(x) = \mathfrak{x}_0^\alpha(x) + \delta\mathfrak{x}^\alpha(x)$. Then, the $\mathfrak{x}_0^\alpha$-dependent terms would be responsible for effects related with these boundary conditions, but, once fixed, not describing any local physical degrees of freedom, and the $\delta\mathfrak{x}^\alpha(x)$ part as a really scalar field, with boundary conditions appropriate for usual scalar fields.

What would be the influence of the $\mathfrak{x}_0^\alpha(x)$-depending part? Qualitatively we can use the same argument used to classify Einstein's cosmological constant term as cosmological. The terms in the Lagrangian do not depend on partial derivatives of the metric, but only on the metric itself. So, their importance decreases locally, where derivatives of the metric may be large, and they become important only at large, cosmological distances, where the local oscillations of the metric have been averaged out. This argument has a weak point, because there is another possibility for the terms to become relevant, namely if the gravitational field becomes singular in the preferred coordinates. This can happen also locally, and in the case of the gravitational collapse it happens, indeed, near the horizon.

Another, much more important difference is that, even if every solution of the ether theory defines a solution of the theory with scalar fields, this is usually simply wrong in the reverse direction. Four arbitrary functions rarely define a valid system of coordinates. Fortunately, the decomposition $\mathfrak{x}^\alpha(x) = \mathfrak{x}_0^\alpha(x) + \delta\mathfrak{x}^\alpha(x)$ allows to specify the restrictions which follow. Let's consider waves with well-defined momentum, without restriction of generality $\delta\mathfrak{x}^\alpha(\mathfrak{x}) = a\sin(k_i\mathfrak{x}^i - |k|\mathfrak{t}))$. Does this function define a valid coordinate?

The answer is a quite general one: They define a valid system of coordinates if the constant a is sufficiently small. In this case, the transformation $\mathfrak{x}_0^\alpha(x) \to \mathfrak{x}_0^\alpha(x) + \delta\mathfrak{x}^\alpha(x)$ is only a small deformation, too small to violate the condition that the determinant of the transformation has to be positive everywhere, and also too small to change the character of the coordinates themselves, so that $\mathfrak{t}$ remains time-like and the $\mathfrak{x}^i$ remain space-like.

But if the deformation becomes large, the resulting four functions may not define a valid system of coordinates. First, four arbitrary functions do not necessarily define a global system of coordinates at all. Second, the resulting coordinates may be incompatible with the ether interpretation, if the resulting preferred time coordinate is space-like instead of time-like, which would make the ether density negative.

3.2.4 The equation for sound waves and translational symmetry

If we have some quite general material, there may be more than density, velocity, and stress tensor which defines its complete state. It can have a lot of other, different material properties. There may be temperature, various chemical properties, various types of lattice defects and corresponding defect densities, whatever. Changes in all these properties may influence others, and distribute in the medium following various equations. What is it what distinguishes the particular type of waves we name “sound waves”? It is, first of all, their quite general character. Whatever the medium, we would expect some sort of sound waves. So, one can reasonably expect that they do not depend on the various material properties which are, in the general case, not even known to exist.

This will be, obviously, reached by waves where the material moves as a whole, with all parts of the material moving locally at the same, so that there are no local distortions, which would depend on the various material properties. But if we assume, as the characteristic property of sound waves, that the material moves as a whole, then translational symmetry adds a new element to the consideration: Once the oscillation can be, locally, described by a translation of the material as a whole, translational symmetry makes this sound wave locally unobservable. If the local environment moves or not is something one cannot observe if the theory has translational invariance, because in this case the moved state is indistinguishable from the unmoved one.

Let’s look at it from the other side, and let’s simply define the sound waves as the waves which are undetectable locally, because the material moves as a whole. If it is undetectable, then the solution with the sound wave have to fulfill the same equations as the solution without sound waves. Now, let’s consider the equations we have already fixed. This is the harmonic equation for the preferred coordinates. A solution with such waves would be indistinguishable from one without such waves if one could, erroneously, assume that the preferred coordinates are not the true $\mathfrak{x}^\mu(x)$, but, instead, some other $\tilde{\mathfrak{x}}^\mu(x) = \mathfrak{x}^\mu(x) + \delta\mathfrak{x}^\mu(x)$, which fulfill the same equations for the preferred coordinates. So, we obtain an equation for the dislocation of the material which gives the sound waves:

$$\Box\delta\mathfrak{x}^\mu(x) = 0.$$

So, we can identify the waves of the ether which are equivalent to a change of the preferred coordinates with the sound waves of the ether.

3.3 The Lorentz ether interpretation of the Einstein equations

As we have seen, the waves of the ether which can be described as equivalent to a change of the coordinates, and which we have identified with the sound waves of the ether, appear to be almost invisible. They behave like dark matter. Nonetheless, they are, at least in principle, observable.

But it makes sense to consider an approximation of the Lorentz ether, where these waves become completely invisible.

3.3.1 The harmonic non-cosmological limit

So, the full Lagrangian of our ether theory is

$$L = L_{harm}(g_{\mu\nu}, \mathfrak{x}^{\mu}) + R\sqrt{-g} - \Lambda\sqrt{-g} + L_{matter}(g_{\mu\nu}, \varphi^m)$$

with

$$L_{harm} = \frac{1}{2}\Xi_{\alpha} g^{\mu\nu}\partial_{\mu}\mathfrak{x}^{\alpha}\partial_{\nu}\mathfrak{x}^{\alpha}\sqrt{-g}.$$

The constant Λ is named "cosmological constant", because the term $\Lambda\sqrt{-g}$ becomes important only for cosmological considerations. There is a good and simple argument for this, namely that the term does not depend on partial derivatives of the metric. But the same argument holds also for the harmonic term. It also does not depend on derivatives of the metric. It depends on derivatives of the preferred coordinates, and this means that there is another reason why the term may become important: If the harmonic coordinates start to degenerate and become infinite. This happens near the black hole horizons. But, beyond this possibility, the harmonic terms have also only a cosmological character.

But that means that we can consider a harmonic non-cosmological approximation of the theory. In this harmonic non-cosmological approximation, we assume harmonic coordinates, but omit all the cosmological terms – by definition those who do not depend on derivatives of matter fields or gravity. The result is the Lagrangian of general relativity:

$$L \to L_{GR} = R\sqrt{-g} + L_{matter}(g_{\mu\nu}, \varphi^m).$$

So, the equations of the resulting limiting theory will be the Einstein equations of GR, together with the harmonic condition. Let's now consider the properties of this limit.

3.3.2 The failure of derivation of the harmonic equation

Outside the limit, the harmonic equation was an Euler-Lagrange equation. In the limit, the Lagrange equation degenerates, and the harmonic equation is no longer an Euler-Lagrange equation. Indeed, the harmonic condition has been obtained as the Euler-Lagrange equation for the preferred coordinates, which had, because of the translational invariance, the form of a conservation law. But in the limit, we have no longer any dependence of the Lagrangian on the preferred coordinates. The Euler-Lagrange equation for the preferred coordinates therefore simply disappears.

Of course, if we consider such a limit as $\Xi_{\alpha} \to 0$, where the resulting harmonic equation depends on the Ξ_{α} only in the trivial way of a constant factor $\Xi_{\alpha}\Box\mathfrak{x}^{\alpha} = 0$, it makes no sense to remove the equation in the limiting procedure. But it is no longer an Euler-Lagrange equation, but becomes an independent equation, without any justification based on the Lagrange formalism.

With the Lagrangian being fully independent on the preferred coordinates, and being the same as the Einstein-Hilbert Lagrangian of GR, the remaining equations are now exactly the Einstein equations of GR. Once we leave the harmonic coordinates as an equation, the harmonic equation becomes now simply a harmonic coordinate condition, which allows to make a choice of the preferred coordinates of the Lorentz ether.

3.3.3 The failure of the Noether theorem to define local energy-momentum conservation

Closely connected with the harmonic condition no longer being an Euler-Lagrange equation is that the Lagrange formalism no longer gives any conservation law. This is a quite general consequence of the general covariance of the Lagrangian. Instead of defining the Noether conservation law, the Euler-Lagrange equation for the preferred coordinates now defines nothing.

A not very satisfactory partial solution of this problem in GR are various proposals for energy-momentum pseudo-tensors, which fulfill conservation laws, but do not transform like real tensor fields, except for linear transformations.

Once we leave the harmonic equation, we do not need such a pseudo-tensor, because we know from the theory before the limit that the harmonic condition is the energy-momentum conservation law.

3.3.4 The Strong Equivalence Principle for the Lorentz ether

Already in the full ether theory it is not easy to observe the preferred system of coordinates. In particular, the EEP already holds exactly, so that no local non-gravitational observation allows to identify the preferred frame. Moreover, the local observable effects are similar to those of massless completely dark matter, so that they will be hidden behind a lot of other (massive) dark matter.

But in the harmonic GR limit even these remaining possibilities no longer exist, and the Strong Equivalence Principle holds. That means, the preferred coordinates are now completely unobservable.

Nonetheless, even in this limit the preferred coordinates exist, and follow the harmonic equation.

That means, solutions with different choices of preferred coordinates are, even if indistinguishable by observation, nonetheless physically different states of the ether.

Let's illustrate this with the example of the Minkowski metric, which would be a solution of the Einstein equations of GR for the vacuum. One set of preferred coordinates would be, of course, those of an inertial frame. In these coordinates, the metric would be constant, and the ether interpretation would describe it as a homogeneous, stationary ether.

But there would be other, non-trivial solutions of the harmonic equation on the Minkowski background. In these other solutions, the gravitational field would not be constant. And that means that the ether interpretation would describe it as a non-constant solution of the ether equations, a solution which would be neither homogeneous nor static. These solutions would fulfill the continuity and Euler equations, and they would describe non-trivial sound waves of the ether.

3.4 The space-time interpretation

A situation where different physical states – like the Lorentz ether being at rest or moving – cannot be distinguished by observation seems unacceptable to many physicists. So, in particular, Einstein argued:

> For the theoretician such an asymmetry in the theoretical structure, with no corresponding asymmetry in the system of experience, is intolerable. If we assume the ether to be at rest relatively to K, but in motion relatively to K', the physical equivalence of K and K' seems to me from the logical standpoint, not indeed downright incorrect, but nevertheless unacceptable.

The situation in GR is even more serious. Here, we have not only a single global fixed velocity of the ether which appears unobservable. We have, instead, states of the ether with or without sound waves which appear indistinguishable. Thus, not only the ether at rest and a moving ether would be indistinguishable, but even a homogeneous static ether and an inhomogeneous oscillating ether have to be indistinguishable. So, the argument becomes even stronger.

Moreover, GR appears to be much closer to a philosophy named "relationalism", which goes back to Descartes [4]:

> According to Descartes, there is no "space" at all, but only physical objects which can be in touch with each other. The "position" or location, respectively, of an object is only defined by the naming of other physical objects close to it, i.e. the position of a body is the set of those objects to which the body is contiguous. Equally important is the concept of "motion", which is defined as the change of position. Thus motion is determined by the change of contiguity, i.e. only in relation to other objects. This point of view is denoted as relationalism. [3]

Let's compare this with the concept of absolute space and time proposed by Newton:

> According to Newton, "space" exists by itself, independently of the objects in it. Motion of a body can be defined with respect to space alone, irrespectively whether other objects are present. . . . according to Newton, space exists independently of objects, weather they are present or not. The location of objects is the part of space that they occupy. This implies that motion can be understood without regard to surrounding objects. Similarly, Newton uses absolute time, leading to a space–time picture which provides an always present fixed background over which physics takes place. Objects can always be localized in space and time with respect to this fixed non-dynamical background. [3]

This absolute space and time is realized in the Lorentz ether. In the Minkowski space-time interpretation of SR, we have yet an absolute space-time. And this absolute space-time defines some global properties. In particular, Newton's rotating bucket argument remains valid, and shows that at least acceleration is absolute. Similarly, light rays follow straight lines. But what is a straight line is defined by some absolute, global geometry, which is not influenced by physics. And this property of light rays could be used, via the synchronization procedure known as "Einstein synchronization", to define also a notion of global contemporaneity. It was not uniquely defined, but it depended only on a single parameter – the velocity of the observer who decided to use this "Einstein synchronization".

But in GR we do not have any such global aspects of absolute space-time. Say, assume that there are no sound waves of the ether near you. So, in your environment everything looks like in flat Minkowski space. But this does not tell us anything about the presence of ether sound waves outside your environment. Thus, even if you know the true preferred coordinates in your environment, you don't know how to extend them to infinity. In SR, knowing the ether velocity locally would have been sufficient to know it globally, and to know how space-time splits into space and time. In GR, this is no longer possible.

So, special relativity has removed from space only one single information – the velocity of the ether in a single point. This has caused some confusion with common sense ideas about space and time, but at least their combination, the Minkowski space-time, remained untouched. SR was far away from the philosophy of Descartes that there is no space at all, only physical objects which can be in touch with each other. Instead, GR is quite close to this.

The philosophy of the space-time interpretation is, essentially, this philosophy of Descartes that space does not exist. In the harmonic GR limit of the Lorentz ether, absolute space and time yet exist, we are only, by unfortunate circumstances that we are unable to see anything beyond this approximation, unable to observe the differences between really different states of the ether. The space-time interpretation goes beyond this. Not only absolute space and time, but even absolute Minkowski space-time does not exist. What exists is the gravitational field. And this gravitational field defines all what locally looks like space and time. But the absolute Minkowski space-time of SR does not exist.

3.4.1 The mathematical formalism of getting rid of the background

So, all the properties of the background of space and time which are yet present in the GR limit of the Lorentz ether (like, in particular, the translational symmetry in space and time) but unobservable, have to be removed from the theory. On the other hand, some of the properties remain. For example, that we can, at least approximately, describe the solutions as functions of four real continuous parameters named "space-time.

From a mathematical point of view, this is not difficult, because the mathematical apparatus to describe similar things is well-known, and it was necessary to develop an adequate mathematical apparatus anyway, simply because (as usual for mathematics) a similar problem existed in some completely different situation. One such situation was the use of curved coordinates. Curved coordinates were useful already in classical theory. If one describes a spherically symmetric solution, it makes sense to describe it in spherical coordinates, for the simple reason that the solution will look much simpler in these coordinates. But this does not mean that anything of fundamental importance will look simple in these coordinates. One can define distances, volumes and so on in spherical coordinates too, but this becomes a quite complex mathematical object, which has to be defined separately, and nothing as simple as the same objects of Euclidean geometry in Cartesian coordinates. Another situation was the geometry of curved surfaces. A curved surface will be, locally, described by two local coordinates, which are two real numbers. But the distances on this surface will be, clearly, something which depends on the particular properties of the surface, on the embedding of this surface into three-dimensional space. Here, the formula for the distance will not only look more ugly than that of Euclidean distance in Cartesian coordinates, but there is no such Euclidean distance defined on the surface.

So, the mathematical apparatus to handle such things exists. It is the apparatus of differential geometry, which describes the techniques how to use general systems of coordinates. This apparatus allows to describe theories which are defined using a particular special system of coordinates in a covariant way, which no longer depends on this special choice of coordinates. Einstein has initially thought that the possibility to describe a theory in such a covariant way is a physical restriction. But it is not. As was argued by Kretschmann [5], all physical theories can be written in a covariant form. Einstein has accepted this point, and with our presentation of the Lorentz ether we have, essentially, given an illustration. We have transformed the continuity and Euler equations from their classical form (3.1), (3.2) in the preferred coordinates into a covariant equation (3.3) for the preferred coordinates $\mathfrak{x}^\alpha(x)$. We have also transformed the condition that the ether density has to be positive, $\rho(\mathfrak{x}) > 0$, which is defined only in the preferred coordinates, into a covariant condition for the preferred time coordinate $\mathfrak{t}(x)$ itself, namely that it has to be time-like, $g^{\mu\nu}(x)\partial_\mu \mathfrak{t}(x)\partial_\nu \mathfrak{t}(x) > 0$.

So, we have applied the mathematical apparatus of general covariance to the Lorentz ether, and the result was a formulation of the theory which is covariant, that means, can be written in arbitrary coordinates in the same way. But, on the other hand, in our variant the preferred coordinates of the theory remain in the theory even if we use other coordinates to describe it − as four scalar functions $\mathfrak{x}^\alpha(x)$, which follow a covariant equation $\Box\mathfrak{x}^\alpha(x) = 0$.

On the other hand, we can get now rid of the preferred coordinates, as well as the absolute space-time which they define. All we have to do is to modify the theory in such a way that they do no longer appear in the equations. The first step was the GR limit. The consequence was that the Einstein equations, as the Euler-Lagrange equations of a Lagrangian which does not depend on the $\mathfrak{x}^\alpha(x)$, also do not depend on the $\mathfrak{x}^\alpha(x)$. The next step, which is done by the space-time interpretation, is to remove the preferred coordinates completely from the theory. As a consequence, the harmonic equation disappears too.

3.4.2 Mathematical consequences of the space-time interpretation

The consequence of this is that there is no longer anything which requires that the space-time has the usual form $\mathbb{R}^3 \times \mathbb{R}$. It has to be some four-dimensional manifold, that's all. For example, the space could be a three-dimensional sphere, so that the space-time would be $S^3 \times \mathbb{R}$. In this case, there would be no single set of three spatial coordinates which could describe, uniquely, all points. At least at one point the coordinates would degenerate.

How to find such coordinates one can illustrate using the surface of a sphere S^2, like the surface of the Earth. Analogical mathematical formulas would, then, work for S^3 too. One choice of coordinates would be, say, to remove the North pole, and to project all other points to a plane through the equator, using a line through that point and the North pole. Or you could use the same scheme, only with the South pole instead of the North pole. The complete solution of GR for this space-time would consist of the solution for above choices, together with a proof that the solution is the same on the part covered by above sets of coordinates – that means, all except the two poles. Neither the part of the solution without the North pole nor the part without the South pole would be the complete solution. But only such an incomplete part would allow the introduction of coordinates which could be used as preferred coordinates. So, no Lorentz ether interpretation for this solution is possible – only one for a part of it.

So, together with the possibility of solutions with non-trivial topologies, we see also that the notion of what is a complete solution differs. A complete solution of the Lorentz ether is defined for all values $-\infty < \mathfrak{x}^\alpha < \infty$. The complete solution of GR may contain more, with the other parts described by other sets of coordinates, which cover these other parts of the manifold.

What also changes is the notion of symmetry. A solution of the Einstein equations on $S^3 \times \mathbb{R}$ can have rotational symmetry. But what we would obtain if we would use some coordinates, say, those without the North pole, and use the corresponding ether interpretation? The ether would, quite obviously, have maximal density at the South pole, and decreasing density if one moves away from the South pole. The notion of symmetry of the Lorentz ether is much more restricted, because it requires also that the symmetry applies to the Newtonian background space too.

And, last but not least, there also exists no preferred time coordinate. So, the space-time is not even obliged to split into something like $M^3 \times \mathbb{R}$ for some arbitrary three-dimensional manifold M^3. And even if it does, the removal of a global time coordinate from the theory has important consequences. Remember that the preferred time coordinate has to be time-like, to give a positive ether density. Once no global time coordinate is obliged to exist, there may be solutions which do not have a global time-like coordinate.

The first solution of GR of this type is quite well-known, it is Gödel's rotating universe. From a topological point of view, it is trivial $\mathbb{R}^3 \times \mathbb{R}$. So, you can introduce there preferred coordinates. But none of these coordinates would be time-like everywhere. So, an attempt to find an ether interpretation ends fatally with negative ether densities at some parts of the solution.

3.4.3 Physical consequences of the space-time interpretation

The main difference is that the Strong Equivalence Principle obtains a completely different meaning.

In the GR limit of the Lorentz ether, solutions with different choices of the preferred coordinates were interpreted as different solutions (they had, for example, different ether densities and velocities), but they were indistinguishable for observation.

In the space-time interpretation, these same solutions are no longer considered as different at all. The name "equivalence principle" suggests only some equivalence, which implicitly presupposes some difference. But this is misleading – the different solutions are not considered to describe different physical situations at all. They are considered as different descriptions, using different coordinates, of the same physics.

Bibliography

[1] Schmelzer, I.: A generalization of the Lorentz ether to gravity with general-relativistic limit, Advances in Applied Clifford Algebras 22, 1 (2012), p. 203-242, arXiv:gr-qc/0205035

[2] A. Einstein, Ether and the Theory of Relativity, an address delivered on May 5th, 1920, in the University of Leyden

[3] M. Gaul, C. Rovelli: Loop Quantum Gravity and the Meaning of Diffeomorphism Invariance, Lect.Notes Phys. 541 (2000) 277-324, arxiv:gr-qc/9910079

[4] R. Descartes: Dioptrique, Les Météores. In: Discours de la méthode (1637)

[5] E. Kretschmann, Über den physikalischen Sinn der Relativitätspostulate, A. Einsteins neue und seine ursprüngliche Relativitätstheorie, Ann. Phys. 53, 575-614 (1917)