Lecture Notes PX421: Relativity and Electrodynamics

Y.A. Ramachers, University of Warwick
Draft date April 11, 2014

Contents

Chapter 1

Full Special Relativity

This chapter should expose the foundations of Special Relativity including all assumptions entering the construction of the theory. Subsequently, the theory and its application to classical mechanics is demonstrated, mainly as a reminder of previous lessons on the topic, however, using more advanced mathematical tools. These are being introduced step-by-step either in separate excursions or along the way of progressing with topics from Special Relativity.

The content of this chapter draws mainly from [1] with some input from [2]. For additional material to any of the sections in this chapter, please see those two textbooks unless you have a different favoured textbook.

1.1 Homogeneity and isotropy of space

The two topics in the section title represent far reaching concepts in physics. A truly fundamental start to the lecture would attempt to rather 'derive' these concepts instead of starting with them in mind. In order to get anywhere in this lecture let's rush through such fundamental ideas. This has the advantage of at least naming the keywords along the way for anyone interested to delve deeper later on.

One possible start to get to grips with homogeneity and isotropy of space would be to consider the idea of **motion**.

Motion: Any definition of movement needs an answer to: 'relative to what'.

The line of thought then proceeds naively as follows: If you need a reference, introduce a coordinate system. Then immediately you have a definition of motion according to

Motion: Not all coordinates of all points of a body remain constant.

Let's untangle that big step of thought from above into its more fundamental parts.

Manifold: First of all, you need a continuous set of points. That is called a manifold, just a set, nothing much more to it.

Metric: This is less fundamental. Define a distance measure on the manifold. That is then called a metric.

Reference frame: This is the least fundamental. In fact, it's nothing else then an arbitrary labelling of points which is called a coordinate system. Still, these can be quite complex constructions and are therefore a perfect tool to confuse students.

The most simple metric space in the sense of simplicity according to human experience is the Euclidean space. A quick demonstration of metric space and coordinate system confusions would be: Describe a point in Euclidean space using Cartesian coordinates, (x,y,z). Then change to polar/spherical coordinates, (r,θ, ϕ), and watch students struggling for the first two years on maths for scientists courses. The swap itself is irrelevant for the point, it's still the same point in the manifold, merely its labels have changed.

Considering that more and more complex structures will have to be scrutinised and used to do physics in or with them, a clearer mathematical description of points and metrics and 'things' in spaces would be highly welcome. Such tools are all headed as coordinate independent descriptions. You use them since many years in the form of **vectors**. Using vectors makes no difference to the physics description other than it's more convenient.

However, with respect to describing physics, we are not quite there yet. Vectors can change quite easily:

1. Translate the entire set of coordinate labels:

$$\boldsymbol{r'} = \boldsymbol{r} + \boldsymbol{\alpha}$$

2. Rotate a reference direction:

$$\boldsymbol{r'} = \mathcal{R}\boldsymbol{r}$$

with $\mathcal{R}$ being a rotation matrix.

The idea is to get rid of **ALL** dependencies on labelling a point. Hence, you require **invariance** under translation and rotation. That now has big(!) consequences for physics. Why? Here is a reminder:

Noether Theorem essence: Every invariance (symmetry operation) implies a conserved quantity in nature and vice versa.

Here this would translate as: Translation invariance implies momentum conservation, whereas rotation invariance implies angular momentum conservation.

Another way to put it is to define homogeneity and isotropy as results of symmetry operations by requiring invariance under translation and rotation transformations. This specific requirement of invariance is also called **Euclidean principle of relativity**. Another way to express it could be to require that physics is the same under variation of 6 parameters. These are the 3 components of $\boldsymbol{\alpha}$ and 3 rotation angles which build the Euclidean group (see Sec. 1.2.1).

1.2 Introducing the time parameter

This is heading towards the **Galilean principle of relativity** which could be expressed as a relativity of velocity.

Let $\boldsymbol{r} = \boldsymbol{r}(t)$; define the term velocity as

$$\boldsymbol{v} = \frac{d}{dt}\boldsymbol{r}(t)$$

and acceleration as

$$\boldsymbol{a} = \frac{d^2}{dt^2}\boldsymbol{r}(t).$$

A convenient path towards the Galilean relativity principle emerges following Newton's work on classical mechanics.

Newton starts with:

1. Time intervals between any 2 events are invariant for any 2 observers, i.e. there is a universal time.

2. Spatial separation of 2 events is invariant for any 2 observers, i.e. there is a universal space

Now describe the dynamics between any 2 bodies:

1. Law of inertia: $\boldsymbol{a}(t) = 0$ if and only if the force between the bodies $\boldsymbol{F} = 0$, i.e. vanishes.

2. Second law: $\boldsymbol{F}(t,\boldsymbol{r}) = m\,\boldsymbol{a}(t)$.

Newton's laws claim that acceleration and force are absolute, i.e. any transformation from absolute space S (with position vector $\boldsymbol{r}(t)$) to a reference system in uniform motion ($\boldsymbol{a}(t) = 0$) relative to S, say S' (with $\boldsymbol{r'}(t)$) leaves the dynamics (the description of the physics under investigation) invariant if

$$\frac{d^2\boldsymbol{r}(t)}{dt^2} = \frac{d^2\boldsymbol{r'}(t)}{dt^2}.$$

Let's see how that works using Galilean transformations:

$$\boldsymbol{r'}(t) = \boldsymbol{r}(t) - \boldsymbol{u}\,t$$

where $\boldsymbol{u}$ is the relative velocity between 2 reference frames. Caution here! The vector $\boldsymbol{r}(t)$ is the position vector between 2 points interacting but each point has its own position vector in absolute space. This can be expressed as Newton's third law:

$$\boldsymbol{r} = \boldsymbol{P}_2 - \boldsymbol{P}_1 \quad \text{or (third law)} \quad \boldsymbol{r} = \boldsymbol{P}_1 - \boldsymbol{P}_2$$

The Galilean transformation changes the $\boldsymbol{P}$'s first of all. So

$$m\boldsymbol{a}(t) = m\boldsymbol{a'}(t) = \boldsymbol{F'}(t, \boldsymbol{r'}(t))$$

assuming invariance of mass. Plug in the transformation and get

$$= \boldsymbol{F'}(t, \boldsymbol{P}_2 - \boldsymbol{u}\,t - \boldsymbol{P}_1 + \boldsymbol{u}\,t) = \boldsymbol{F'}(t, \boldsymbol{r}(t)) \equiv \boldsymbol{F}(t, \boldsymbol{r})$$

where the last equivalence statement originates from the concept that any force function is completely determined by its arguments. After all, a force taken at an identical point in space and time should be identical whether it was previously transformed or not, it's still the same point in space and time.

The lesson is: Newton dynamics is invariant under Galilean transformations which is the the Galilean principle of relativity.

Now consider how many numbers determine the transformation: All 6 parameters from the Euclidean group plus 3 numbers for the relative velocity vector plus the time translation invariance (the freedom to choose the origin in time). That sums up to 10 numbers. Ergo, the Galilean group has 10 parameters free to choose and leave the physics unchanged.

1.2.1 Excursion: 'What is a group and why is it interesting here?

(Caution: This is not a complete definition, mathematically sound. It covers merely what we need here and might give you an idea what this term means. For anyone interested in proper definitions, you will have to look up a maths textbook on group theory.)

Have a set of transformations G between reference frames (translations, rotations, etc.). Then any principle of relativity requires that

1. Two members of G put together must give again an element of G.
2. Every transformation has an inverse (Galilean: $\boldsymbol{u} \rightarrow -\boldsymbol{u}$)
3. There is an identity transformation (Galilean: $\boldsymbol{u} = 0$)

Then a set G is called a group.

Why would that be of any interest?

Two steps, really. First, given a group of transformations, one would naturally have to ask for: "transform what?"

Here (relativity for example) a set of reference frames. However, this could just as well deal with something else:

- event coordinates
- events
- four-vectors
- tensors
- spinors
- fields
- state vectors in a Hilbert space
- ...

Step 2: If an object of a physical theory turns out to be invariant under a group G then the theory is said to be invariant under this 'symmetry group'. This is the most successful way in the history of physics to build a theory.

Why some groups are 'realised' in nature and (most) others are not, is not known. Once you have found one though, you're on to a winner.

1.3 The way beyond Galilean relativity

The fastest argument I've seen to motivate physics beyond Galilean relativity is quoted in [1]: Consider Maxwell's electrodynamics.

Waves in a vacuum are predicted to move with a constant speed

$$v = \frac{1}{\sqrt{\epsilon_0 \mu_0}} \approx \text{speed of light as it was known at the time}$$

Problem: Invariance under Galilean transformations should be true for electromagnetism, which implies a constant v (Maxwell) must imply variable ϵ_0, μ_0 due to the observer dependence of the constant value under Galilean transformations.

However, the numbers ϵ_0, μ_0 are not permitted to vary since forces like the Coulomb force are directly proportional to ϵ_0 and they must be invariant in Galilean relativity. This is inconsistent and did provide a major headache for decades to most physicists at the time.

One way out was the invention of an 'ether', i.e. a preferred reference frame. In such a frame it was postulated that $v = c =$ constant holds for electromagnetism.

Another related effort was undertaken by Lorentz. He derived the set of transformations, which now carry his name, directly from Maxwell's equations. The idea was to understand better which properties of Maxwell's theory were at odds with Galilean transformations. Maxwell electromagnetism was not as established as it is today and Galilean transformations worked perfectly fine with the prime theory at the time, Newton mechanics. Therefore the unusual transformations derived by Lorentz were considered a hint towards understanding the ether theory rather than anything displaying a challenge to Galilean transformations.

Once the Lorentz transformations were derived from first principles, it changed all of physics. That is the paradigm shifting impact of Einstein's contribution and the basis of special relativity

1.4 Derivation of the Lorentz transformation

The first principles mentioned above can be cast into the following two items:

EP1 All inertial frames are equivalent with respect to the laws of physics (relativity principle)

EP2 The speed of light in empty space, c, is independent of the state of motion of its source (constancy of speed of light)

However, there are quite a few additional assumptions which enter the theory and therefore its derivation. In the following the Lorentz transformations will be derived with explicit reference to these additional assumptions.

1.4.1 Linearity and absence of change of direction

Start with using EP2: Let's choose 2 reference frames, a frame S and a primed frame S', with the following properties

$$x = x' = 0 \quad ; \quad t = t' = 0$$

This translates as having S and S' synchronised at the start of a light signal. Also we describe here a one-dimensional, x only, light signal for now. The generalisation to three spatial dimensions follows later.

Direction preservation now means for the analysis of the light signal that

$$\frac{\triangle x}{\triangle t} = \pm c \Rightarrow \frac{\triangle x'}{\triangle t'} = \pm c$$

combine both expressions to

$$(\triangle x)^2 - c^2\ (\triangle t)^2 = 0$$

Then synchronisation gives

$$x^2 - c^2 t^2 = x'^2 - c^2 t'^2 = 0$$

hence get for **any** interval

$$(\triangle x)^2 - c^2\ (\triangle t)^2 = (\triangle x')^2 - c^2\ (\triangle t')^2 = 0$$

only if the transformation is linear.

(Otherwise the following could be true which would spoil the sequence above)

$$(\triangle x)' = (x_2 - x_1)' \neq (x_2' - x_1')$$

and the same for $(\triangle t)'$.

Now assume linearity of a transformation then the following is true:

If a body K in a reference frame S moves with constant velocity u then this implies that K moves in a different reference frame S' with u' **but** also constant!

So in general you get:

$$x' = a_{11}\, x + a_{12}\, t + const1 \tag{1.1}$$

$$t' = a_{21}\, x + a_{22}\, t + const2 \tag{1.2}$$

Let S and S' be synchronised at some point $t_0' = t_0 = 0$; $x_0' = x_0 = 0$ then it follows immediately that $const1 = const2 = 0$.

Then any one-dimensional movement with speed u in x-direction would result in the following derivation:

Examine the origin of S', i.e. $x' = 0$:

$$a_{11}\, x + a_{12}\, t = 0$$
$$\Rightarrow \quad u = \frac{x}{t} = -\frac{a_{12}}{a_{11}}$$

insert this back in eqn. 1.1 to get

$$x' = a_{11}\,(x - ut) = a_{11}(u)\,(x - ut) \tag{1.3}$$

The parameter a_{11} could be a function of u since the speed u is a constant for any given reference frame.

1.4.2 Invariance under time reflection

Now let's apply EP1: Relativity says S and S' are equally valid descriptions for the body K. Let's assume S moves relative to S' with speed -u and plug this into eqn.1.3.

$$x = a_{11}(-u)\,(x' - (-u)t') \tag{1.4}$$

Here enters the invariance under time reflection $(t \to -t)$ since that implies $u \to -u$ with x unchanged. Let's apply this to eqn.1.1, then we get

$$\begin{aligned} x' &= a_{11}(u)\, x + a_{12}(u)\, t \\ t' &= a_{21}(u)\, x + a_{22}(u)\, t \end{aligned} \tag{1.5}$$

transforms to

$$\begin{aligned} x'' &= a_{11}(-u)\,x - a_{12}(-u)\,t \\ t'' &= a_{21}(-u)\,x - a_{22}(-u)\,t \end{aligned} \tag{1.6}$$

Invariance now requires that $x'' = x'$ and $t'' = -t'$. This implies the following set of relations between the coefficients:

$$\begin{aligned} a_{11}(-u) = a_{11}(u) \qquad & a_{12}(-u) = -a_{12}(u) \\ a_{21}(-u) = -a_{21}(u) \qquad & a_{22}(-u) = a_{22}(u) \end{aligned} \tag{1.7}$$

The coefficients $a_{11,22}$ therefore are supposed to be even functions of u which is easily implemented by assuming them to be functions of u^2.

1.4.3 Identity assumption

Referring back to eqns.1.3 and 1.4, when considering a one-dimensional light ray: $x = ct$ and $x' = ct'$ (using EP2) then we get:

$$\begin{aligned} x' &= a_{11}(u^2)\,x\left(1 - \frac{u}{c}\right) \\ x &= a_{11}(u^2)\,x'\left(1 + \frac{u}{c}\right) \end{aligned} \tag{1.8}$$

The identity assumption requires the following to be true: any transform from system S to S' and back to S should yield the identity transformation.

Simple as that may seem, realising this requirement using eqns.1.8 gives:

$$1 = a_{11}(u^2)\,a_{11}(u^2)\left(1 - \frac{u}{c}\right)\left(1 + \frac{u}{c}\right)$$

$$\Leftrightarrow \quad a_{11}^2(u^2) = \frac{1}{1 - \frac{u^2}{c^2}}$$

Hence we encounter the (in)famous special relativity 'gamma' factor:

$$a_{11} = \gamma = \frac{1}{\sqrt{1 - \beta^2}}$$

using $\beta = \frac{u}{c}$.

1.4.4 Finish deriving the one-dimensional Lorentz transformations

All the main ingredients have been identified hence finishing the derivation is straightforward. Let's see how it goes:

From

$$-\frac{a_{12}}{a_{11}} = \frac{x}{t} = u$$

get

$$a_{12} = \frac{-u}{\sqrt{1-\frac{u^2}{c^2}}}$$

hence

$$x' = \gamma\,(x - u\,t)$$

and for $a_{21,22}$ take eqn.1.4 and solve for t':

$$x = a_{11}\,(x' + u\,t') \;\Rightarrow\; t' = \frac{1}{u}\,\left(\frac{x}{a_{11}} - x'\right)$$

plug-in a_{11} and x':

$$t' = \frac{1}{u}\,\left[\frac{x}{\gamma} - \gamma\,(x - u\,t)\right] = \frac{1}{u}\,\left[\left(\frac{1}{\gamma} - \gamma\right)\,x + \gamma\,u\,t\right]$$

Use this rather useful **identity relation**:

$$\frac{1}{\gamma} - \gamma = \sqrt{1-\beta^2} - \frac{1}{\sqrt{1-\beta^2}} = \frac{1-\beta^2-1}{\sqrt{1-\beta^2}} = -\gamma\,\beta^2$$

to get to

$$t' = \frac{1}{u}\,\left(\gamma\,u\,t - \gamma\beta^2\,x\right) = \gamma\left(t - \frac{u}{c^2}\,x\right)$$

from which you can read off

$$a_{22} = \gamma \quad ; \quad a_{21} = -\frac{u}{c^2}\gamma$$

That's it!

So, here are the complete one-dimensional Lorentz transformations which should look utterly familiar to you.

$$\begin{aligned} x' &= \frac{x - u\,t}{\sqrt{1-\frac{u^2}{c^2}}} \\ t' &= \frac{t - \frac{u}{c^2}\,x}{\sqrt{1-\frac{u^2}{c^2}}} \end{aligned} \tag{1.9}$$

The inverse transformation is just as important. You can always get it though by inserting the following replacements into eqns.1.9: $x' \leftrightarrow x$, $t' \leftrightarrow t$ and $u \leftrightarrow \; -u$.

$$\begin{aligned} x &= \gamma\,(x' + u\,t') \\ t &= \gamma\left(t' + \frac{u}{c^2}\,x'\right) \end{aligned} \tag{1.10}$$

Eqns.1.9 and 1.10 will be referred to as LT1 and LT2, respectively, for the entire rest of the lecture (unless I use the equation numbers in this write-up).

1.4.5 What about other velocities than c?

The shortest expression of the Einstein principle 2, EP2, would probably be $c = c'$, but what about other velocities?

Have a body K move in S with velocity v, then $x = vt$. Now what is v in S', i.e. a system moving relative to S with speed u? The Lorentz transformations suggest to simply transform and have a look at the outcome. So that's what we'll do. Starting from simply stating $v' = x'/t'$, use LT1 to get from primed coordinates to un-primed and therefore express v' with known numbers from system S:

$$v' = \frac{x - u\,t}{t - \frac{u\,x}{c^2}} = \frac{v\,t - u\,t}{t - \frac{u\,v\,t}{c^2}} = \frac{v - u}{1 - \frac{u\,v}{c^2}}$$

hence if $v = c$ you get

$$v' = \frac{c - u}{1 - \frac{u}{c}} = c\,\frac{c - u}{c - u} = c$$

Therefore whichever way you look at a body K now moving with c, it always appears to be moving with c, from S as well as S', hence c is the limiting speed by construction and the whole concept is self consistent. No big deal but reassuring nevertheless.

1.4.6 Three-dimensional Lorentz transformations

There is a bit more to the Lorentz transformations than what you have seen so far. Some important properties can only be revealed when generalising them to three spatial dimensions. That's what we'll do next.

Starting the discussion, let's assume again motion in x-direction for one reference frame, S', relative to another frame, S, and let them be synchronised, $t = t' = 0$. An observer stationary in S' and holding a mirror at a distance Y' flicks a switch to start a light flash at the origin O. The quick analysis of the picture would give the following set of statement: The diagram shows 2 events in reference frames S',

Figure 1.1: Observer in S', diagram of light flash in S'. Insert the drawing during the lecture.

1. Event O happens at $t' = 0$; $x' = 0$
2. Event P happens at $t' = T'$; $x' = 0$

Hence the mirror in S' is at the position $Y' = \frac{c}{2}\, T'$.

That was the trivial bit. Now look at the analysis of the identical system from reference frame S, moving relative to S' with speed u. The diagram again shows 2

Figure 1.2: Observer in S, diagram of light flash in S'. Insert the drawing during the lecture.

events taking place:

1. Event O happens at $t = 0$; $x = 0$ by construction
2. Event P happens at $t = T = \gamma T'$ (This follows using LT2 and $x' = 0$) and $x = X = \gamma\, u\, T'$ (again, LT2 and $x' = 0$).

Now the question is what happens to Y in comparison to the Y' as derived before? Look at the distance from O to P first:

$$d = 2\sqrt{Y^2 + \left(\frac{X}{2}\right)^2}$$

and also (we know it's a light flash)

$$d = cT = c\gamma T'$$

hence

$$\begin{aligned}
(c\gamma T')^2 &= 4\left[Y^2 + \left(\frac{\gamma u T'}{2}\right)^2\right] \\
\Rightarrow\ c^2\gamma^2 T'^2 &= 4Y^2 + u^2\gamma^2 T'^2 \\
\Rightarrow\ 4Y^2 &= c^2\frac{c^2-u^2}{c^2-u^2}T'^2 \\
\Rightarrow\ Y &= \frac{cT'}{2} = Y'
\end{aligned}$$

and the same argument works for the z-direction. The message is that directions perpendicular to the direction of relative motion are simply not affected in any way by a Lorentz transformation, i.e. nothing happens. This is a very useful and practical insight into Lorentz transformations and has profound implications for many practical applications of the theory, say in electromagnetism to name just one example. Trivial as it may seem, it's rather important and you would do well to remember it when it comes to exercises and exam problems.

1.4.7 First glimpse at a compact way of writing Lorentz transformations

After repeating and reminding you of all the basics of Lorentz transformations, it's probably not a big surprise to see them written down in more compact form using vectors and matrices. You probably saw this several times before and for all you care at this stage, it's really just convenient without any consequences (or specific purpose). Let's have a look at writing LT2 in this proposed compact form and you'll see what I mean:

$$\begin{aligned}
\begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} &= \begin{pmatrix} \gamma & \gamma\beta & 0 & 0 \\ \gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} \\
&= L(\beta)\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} \qquad (1.11)
\end{aligned}$$

In order to write LT1, simply substitute $\beta \leftrightarrow -\beta$ and interchange primed with un-primed coordinates.

Comments:

- The suggestive short-hand: Column vector=LT-matrix times Column vector is not only convenient. This is the entry point of geometry into physics! This is the main reason why we will have to deal with this in a little more detail later on, i.e. specifically the topics of tensors and four-vectors etc.
- Caution: The transformations and their corresponding groups have a priori nothing to do with the objects they act (operate) on.
- The strategy in physics therefore is typically: (A) Describe the maths for the transformation (groups!) first and then (B) come to create suitable objects to describe the physics consistently. This would be the way to construct any Lorentz-invariant theory.

Here in this lecture, we will follow the strategy from above and apply it to classical mechanics and electromagnetism (which, as you know, is already Lorentz-invariant - it's just not so obvious). The next step in this programme could be to follow Dirac's work and apply the strategy to Quantum mechanics.

1.4.8 Exercises

Collection sheet [2.1] - Derive the identities:

- $\gamma^2 - 1 = \gamma^2 \, \beta^2$
- $\gamma - 1 = \frac{\gamma^2}{\gamma+1} \, \beta^2$

1.5 Excursion Tensor Analysis

1.5.1 Introduction

As a warm up, let me showcase the most important elements of special relativity and their typical mathematical expressions. This serves either as a reminder for some of you or simply as name-dropping for everyone else. Afterwards, let's dissect all this and learn a few things along the way.

Recalling absolute time and space in Sec. 1.2, it should be clear from eqns. 1.9 and 1.10 that for Lorentz transformations in general there is no absolute time and space anymore. A time interval changes from one frame to another, $\triangle t' \neq \triangle t$, and similarly the space intervals.

However, Minkowski showed in 1908 that a new invariant (or absolute) exists: The line element of 4D spacetime! This looks like this:

$$s^2 \equiv c^2t^2 - x^2 - y^2 - z^2$$

Quick check: Assume $y = z = 0$ and get

$$s'^2 = c^2t'^2 - x'^2 = \gamma \left[\left(ct - \frac{u}{c}x \right)^2 - (x - ut)^2 \right]$$

$$= \gamma^2 \left(c^2t^2 - 2tux + \frac{u^2x^2}{c^2} - x^2 + 2xut - u^2t^2 \right)$$

$$= \frac{\left(1 - \frac{u^2}{c^2}\right) c^2t^2 - \left(1 - \frac{u^2}{c^2}\right) x^2}{1 - \frac{u^2}{c^2}} = c^2t^2 - x^2 = s^2$$

which means this line element s is indeed invariant. What's a line element? What does this all mean? I hope that can all be clarified below.

Another interesting keyword in connection to the above is the term Poincare transformation. These are all transformations which leave the line element invariant. Yes, there are indeed more than the Lorentz transformations from above which do that. Again, nothing more about it at this point.

Now have a look at the differential form of the line element:

$$(ds)^2 = (dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2$$

and notice the small change in notation, from classical t, x , y, z to more mathematical $x^{0,1,2,3}$ labels for coordinate names. Also the brackets are optional for many textbooks. I'll drop them soon but for now let's be as precise as reasonable.

This little change in notation is rather important and will have many practical consequences. It is **index notation**, dreaded by many students but eminently useful, so you will have to learn to cope with it (see exercises in Sec. A). This notation enables one to write the line element as follows:

$$(ds)^2 = \eta_{\mu\nu}\, dx^\mu\, dx^\nu \tag{1.12}$$

The customary notation counts Greek letter indices (such as μ, ν) from 0 to 3, while Latin letter indices (such as i,j) from 1 to 3. They are often used to address purely spatial coordinates. However, usually, you will mainly see Greek letter indices labelling all four coordinates in spacetime. Again, at this point this notation doesn't mean much to you and we will examine and train it in much more detail later (keyword: Einstein summation convention).

This notation also introduces an interesting matrix

$$\boldsymbol{\eta} = (\eta_{\mu\nu}) = (\eta_{\nu\mu}) \equiv \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{1.13}$$

which you will get to know a little better later on. All of this comes down to Minkowski's insight that this line element introduces geometry into physics. It might not be immediately obvious but this is how it goes: The line element defines an invariant length (invariant with respect to Lorentz transformations, that is, by construction). What this means is that we now have a method to **define** distance in spacetime in a unique (invariant) way.

The Minkowski statement from above enters through this object $\boldsymbol{\eta}$, which receives a physical meaning in special relativity. It's a geometrical object from maths, a special tensor of rank 2 (see later), a **metric**.

So here is the point where one can turn around and unravel the whole line of naming strange concepts from above. The fundamental starting point for all these concepts is in fact the metric which Minkowski thankfully made clear to physicists at the time, setting all the discussions on a firm footing for the first time.

Let's therefore start again on special relativity and its foundations, step-by-step. What is a metric?

For those who took the course on metric spaces, apology for this physics definition. In short, a metric defines distances on a manifold (set of points) and represents an additional, independent property of the manifold.

Generally, we are not used to the concept of taking a metric into account explicitly when calculating distances. That becomes clearer when considering the metric we usually take for distance calculation, i.e. the Euclidean metric. It's not explicitly named as such since it's trivial. The matrix corresponding to the Euclidean metric is simply the unit matrix, i.e. all one's on the diagonal and zeros elsewhere.

It's not exactly an intellectual challenge to put the above in explicit terms but I find it helps as a little aside. Also, it's another good warm-up for later. Therefore, consider calculating the distance between the origin and a point P on the x-axis, at some time t in the future (we are dealing with spacetime after all, using only space would be the same for Euclidean metrics). That gives a distance squared s^2 according to (take $\boldsymbol{E}$ as the Euclidean metric symbol, $\mathbf{x}$ as the vector symbol and $\boldsymbol{x}^T$ as the transposed vector - transpose here = turn column vector to row vector and vice versa)

$$s^2 = (P(x,t) - O(x=0,t=0))^2 = \boldsymbol{x}^T \boldsymbol{E}\boldsymbol{x} = (t-0, x-0, 0, 0) \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} t-0 \\ x-0 \\ 0 \\ 0 \end{pmatrix}$$

remembering that P and O are points described by vectors. No surprise, the distance squared is simply $s^2 = t^2 + x^2$. The metric in between is trivial in this case, hence you never see it. The Minkowski metric doesn't look like much either but all the strangeness of special relativity can be traced back to that switch in sign.

It should be stressed here that the inversion of the signature is completely equivalent, hence some textbooks have the time coordinate with the negative sign and the spatial coordinates all positive. That is merely a sign convention and has no impact on the physics. The convention above is the one used in this lecture. It's worth remembering when comparing explicit calculations between textbooks, i.e. a common cause of confusion.

Repeating the exercise above using the Minkowski metric, you can quickly spot an oddity which is very important and will be mentioned many times more during the lecture. The squared distance using Minkowski for the example above would look like this: $s^2 = c^2t^2 - x^2$. Really, the 'c' is merely there to give us a consistent, no headache, unit system. Nothing special about it.

But look at the distance formula - the negative sign implies that something really strange is predicted by special relativity if this were true. Squared distances could vanish and even turn negative. Consider a light ray from the origin going to P, then $x = ct$ and the theory says that the distance measured between O and P is in fact zero. The Euclidean metric might put that at, say, a light-year distance while the Minkowski metric says zero. No wonder that special relativity wasn't an immediate hit after 1905. This needs some serious explanations and to get anywhere near them, we need some mathematical tools first.

1.5.2 Working with tensors

So, what is a tensor?[1]

Tensor A linear operator acting on another tensor of equal or lower rank to produce a third equal or lower rank tensor.

That's one way to put it. Certainly not a definition, using the very term to define a term is never a good idea. We will end up defining a tensor using its properties under transformations but that is quite a way ahead yet. Let's start a little more pedestrian and just use the word 'tensor' without knowing really what it is and similarly for the 'rank of a tensor'. Let's see how to get some intuition going on these things.

Do you know any tensors already? Most certainly you do.

Zero rank tensors This is simply a number, any number, often rather termed scalars. These you know quite well.

Rank 1 tensors The classic example for these are vectors, or 'numbers in one column'.

Rank 2 tensors You have seen the Minkowski metric which is such a thing. These tensors can always be represented by a matrix. This is often a big hazard to students since it's very natural to identify matrix equal tensor but that is plain wrong. It's a one-way street: Any rank 2 tensor can be represented by a matrix but not every matrix is a tensor.

A rank 3 tensor or higher would by analogy need to be represented by higher dimensional arrays of numbers (a block of numbers for rank 3) but that's clearly not very practical. Better to just use abstract symbols for those. Fortunately, we will not need to go beyond rank 2 tensors at any point in the lecture. Those of you planning to attend the general relativity lecture though need to get used to the idea of rank 4 tensors.

Pedestrian style, let's get started with vectors. Taking the pseudo definition of a tensor from above literally, we can state: A vector operating on another vector (for instance) can yield (a) another vector or (b) a scalar.

That sounds familiar: case (b) would be your usual dot product operation between 2 vectors, case (a) the cross product. Simple, and yet, a plain cheat. You use this since years and yet the description of it is simply wrong. Have a closer look. This will become a little tedious and boring but it's well worth the effort.

[1]Most of this material comes from [3]

Vectors and One-forms

Vectors have components once you specify a reference frame. You can always fully define any reference frame by defining its **basis vectors** which then form a vector space (I hope this reminds you of your linear algebra lessons). Have some general basis with vectors $\boldsymbol{g}_1$, $\boldsymbol{g}_2$ which span a two-dimensional space. Then any vector, $\boldsymbol{v}$, can be written as

$$\boldsymbol{v} = \sum_{i=1}^{2} v^i \boldsymbol{g}_i$$

All the sum's will become very tedious indeed when using the index notation, hence a convention was implemented, the **Einstein summation convention** to ease calculations and confuse students, apparently[2]. Here is the definition for this case. It doesn't matter which case we consider, it's always the same consequence: Drop the explicit summation sign and imagine always to sum over repeated indices occurring in a diagonal position with respect to each other.

$$\boldsymbol{v} = \sum_{i=1}^{2} v^i \boldsymbol{g}_i = v^i\, \boldsymbol{g}_i = v^1 \boldsymbol{g}_1 + v^2 \boldsymbol{g}_2$$

The index 'i' in the example above is 'summed over'. The operation is also called 'contracted', you contract over 'i', signalling that the index 'i' is not available for any other manipulation you might come up with.

Moving on, let's make it explicit that the v^i are the components of $\boldsymbol{v}$ and $\boldsymbol{g}_i$ are the basis vectors, i.e. two completely different things (numbers, vectors).

So, how to form the dot-product? Assume you have two vectors

$$\boldsymbol{u} = u^i \boldsymbol{g}_i \quad ; \quad \boldsymbol{v} = v^i \boldsymbol{g}_i$$

You write the dot-product using the new index notation:

$$\boldsymbol{u} \cdot \boldsymbol{v} = u^i\, v^i\, \boldsymbol{g}_i \cdot \boldsymbol{g}_i = u^1\, v^1\, \boldsymbol{g}_1 \cdot \boldsymbol{g}_1 + u^2\, v^2\, \boldsymbol{g}_2 \cdot \boldsymbol{g}_2$$

which is plain nonsense.

The lesson here is that indices matter when using the Einstein summation convention. Each and every pair of contracted indices requires its own summation independent of other contraction operations. That means it gets its own index!

Try again:

$$\boldsymbol{u} \cdot \boldsymbol{v} = u^i\, v^j\, \boldsymbol{g}_i \cdot \boldsymbol{g}_j$$

[2]Please have a look at Appendix A.1 for another definition from [6].

$$= u^1\, v^1\, \boldsymbol{g}_1 \cdot \boldsymbol{g}_1 + u^1\, v^2\, \boldsymbol{g}_1 \cdot \boldsymbol{g}_2 + u^2\, v^1\, \boldsymbol{g}_2 \cdot \boldsymbol{g}_1 + u^2\, v^2\, \boldsymbol{g}_2 \cdot \boldsymbol{g}_2$$

This still looks odd.

This is the point at which you can't get any further without a bit of a radical choice to make. You either introduce a new mathematical object, a new kind of vector, or you change the familiar definition of the dot-product. Of course, both has happened and been explored but for relativity theory (and differential geometry etc) the choice is to introduce a new kind of vector, something we will call a **one-form**.

Here is how it goes: Let's represent $\boldsymbol{u}$ as before in the basis $(\boldsymbol{g}_1, \boldsymbol{g}_2)$ but choose for $\boldsymbol{v}$ a different, yet unknown basis $(\boldsymbol{g}^1, \boldsymbol{g}^2)$. The components of $\boldsymbol{v}$ are merely numbers, so it will not matter a thing whether we label them with index up or low, so write

$$\boldsymbol{v} = v_i\, \boldsymbol{g}^i = v_1\, \boldsymbol{g}^1 + v_2\, \boldsymbol{g}^2$$

Now the strange expression for the dot-product from above looks a little different (not much, granted):

$$\boldsymbol{u} \cdot \boldsymbol{v} = u^1\, v_1\, \boldsymbol{g}_1 \cdot \boldsymbol{g}^1 + u^1\, v_2\, \boldsymbol{g}_1 \cdot \boldsymbol{g}^2 + u^2\, v_1\, \boldsymbol{g}_2 \cdot \boldsymbol{g}^1 + u^2\, v_2\, \boldsymbol{g}_2 \cdot \boldsymbol{g}^2$$

Now all left to do (remember the index up basis 'vectors' are not yet defined, hence take your pick on what seems best in order to keep the familiar dot-product, i.e. the reason to go to all this trouble) is to choose $\boldsymbol{g}^1$, $\boldsymbol{g}^2$ such that the expression above reduces to the known dot-product

$$\boldsymbol{u} \cdot \boldsymbol{v} = u^1\, v_1 + u^2\, v_2$$

Therefore require the following relations to hold:

$$\boldsymbol{g}_i \cdot \boldsymbol{g}^i = 1$$

and

$$\boldsymbol{g}_1 \cdot \boldsymbol{g}^2 = \boldsymbol{g}_2 \cdot \boldsymbol{g}^1 = 0$$

Put more concisely, as a definition of a one-form in this case, one can write

$$\boldsymbol{g}_i \cdot \boldsymbol{g}^j = \delta_i^j = \begin{cases} 1 & \forall i = j \\ 0 & \forall i \neq j \end{cases} \tag{1.14}$$

where the δ–symbol is called the Kronecker delta.

Eqn. 1.14 simply defines the inverse basis. All the elements of the inverse basis of a vector space are called **one-forms**. They are not vectors! Constructed graphically in 3-D, for instance, they look like planes and a dot-product with an 'arrow'-like vector gives a scalar which would be the 'number of piercings' of the arrow through the set of planes.

In many textbooks as well as traditionally, a vector with an upper index on its components, by convention, is called a **contravariant** vector and lower index vectors, **covariant** vectors.

This translates to: Contravariant vectors are vectors. Covariant vectors are one-forms.

Why would anyone bother? You didn't learn the dot-product for vectors correctly as you now know but it really didn't matter! In Euclidean space it's all the same and there are no consequences mixing up the concepts. However, in Minkowski spacetime it makes a difference and even more so in general relativity.

If that all appears to come a little fast, there are two excellent textbooks going to great length in motivating and explaining all of this again [4] and [5]. The point, again, is that in all aspects of physics involving a Euclidean space, the distinction between vectors and one-forms simply does not matter. They become identical objects. Once you leave the confines of 'normal' space and have to deal with non-trivial space-times then the distinction makes all the difference. It's good to learn about all this in special relativity first since here it is only a change of sign going from a one-form representation of a rank-1 tensor to a vector representation and vice versa. In general relativity such a switch can involve heavy computations.

The gradient as a one-form

The reason to single out the gradient of something (it's an operator, i.e. it acts on something where the 'something' doesn't need to be specified as long as that operation is defined) is that it will yield the prototype for a one-form, something that is more naturally a one-form than a vector.

Mathematically, that's a silly thing to say since these are equivalent expressions, i.e. once you have one, you have the other. However, in physics we start from describing the location of something more naturally as a vector, an arrow pointing towards a mass point for instance. If you do that, then the gradient is the template for a one-form. Again, you used it all through Maths for Scientists lectures and never knew what you are dealing with. The simplicity of Euclidean space makes that possible without a hitch.

A gradient represents a rate of change along 'something', i.e. it introduces a direction to something, a function for instance. The easiest path to see a gradient as a one-form is to take the gradient along a vector (its counterpart in the world of differential geometry) rather than a scalar function. Let me remind you that you have done that many times and called it a 'divergence', i.e. a change in a vector(-field). You might not have called it a vector field explicitly but that's what you have done. You needed it all the time in the Electrodynamics lecture where you computed change in the electric field for instance. That translates into calculating the divergence of the vector field, called the Electric field.

Let's do that most explicitly and learn some useful things about vectors along the way[3].

Define a general vector, i.e. one that is valid in all metric spaces (attention, this is something new, we haven't had such a general definition yet - it is very, very important in general relativity, less so for us). Connect two points A and B and describe the line from A to B using a parameter λ. In plain algebra instead of words:

Figure 1.3: Line between any two points A and B and the parameter λ to describe it. Insert the drawing during the lecture.

$$P(\lambda) = A + \lambda\,(B - A)$$

where $P(\lambda)$ could be a line or any curve. Then define the vector from A to B:

$$\boldsymbol{v}_{AB} = \left.\frac{dP(\lambda)}{d\lambda}\right|_{\lambda=0}$$

$$\Rightarrow \frac{d}{d\lambda}\,(A + \lambda\,(B - A)) = B - A = \text{Tip-Tail}$$

This then becomes a **local definition** of a vector, i.e. at a point. Now go backwards, take any vector $\boldsymbol{v}$ and construct the curve $P(\lambda)$.

Assume a function f, taking points as arguments, and pick a fixed point P_0 (remember, this is by construction a point along the vector $\boldsymbol{v}$). Then write to first order

$$f(P(\lambda)) = f(P_0) + (P(\lambda) - P_0) \cdot \tilde{\boldsymbol{df}}$$

Now let's dissect this expression. Starting from a fixed point on the vector, P_0, we add another term. The first part on the left has the appearance of a vector ('Tip-Tail'), the difference between some point on the curve P, the Tip, and the

[3]This part attempts to give a short version of the extensive discussion in [4]

fixed point, the Tail. This term is multiplied with something new. First of all, the multiplication must(!) yield a number since the function consists of numbers and this expression is to yield a function value. So the candidate for the multiplication operation is the dot product, connecting a vector with this new term. This now already defines $\tilde{\boldsymbol{df}}$ as a one-form (contracted with a vector gives a number).

Is it possible to gain a little more understanding? Geometrically, we have a fixed point and an arrow pointing along $\boldsymbol{v}$, embedded in a diffuse 'ether' of function values (the function f is defined everywhere in that space). Zoom in into a small region (first order Taylor series) around the fixed point with a small arrow pointing in some direction. There is another geometrical structure evident, i.e. the planes (if it all were in 3D space, of course) of constant function values around the fixed point. The vector $\boldsymbol{v}$ clearly must pierce some of these planes, since otherwise there would be no change and the derivative would be identically zero. The planes, being pierced by the vector, are the one-form $\tilde{\boldsymbol{df}}$. They have to be planes (in 3D) since they represent the first order change **in any direction**. The vector then realises a specific direction and contracting both gives the first order approximation value of the function around the fixed point in that chosen direction. A lot of text,

Figure 1.4: Attempt at drawing the contraction of a vector (arrow) with a one form (planes). Insert the drawing during the lecture.

but this might look a little more familiar when using coordinates or rather vector components instead of the abstract expression above. Therefore, let's assume we have a reference frame this time (note, for the above, we didn't really need one - it is all very general)[4].

Given a global reference frame and corresponding coordinates, defining a curve and tangent vectors along it is particularly straightforward. Any change along the curve $P(\lambda)$ with a curve parameter λ in, say a 2D orthonormal coordinate system

[4]This discussion follows [5], but could be looked up in [4] too.

using x and y coordinates, can be expressed as

$$\frac{dP(\lambda)}{d\lambda} = \frac{dx}{d\lambda}\frac{\partial P}{\partial x} + \frac{dy}{d\lambda}\frac{\partial P}{\partial y}$$

due to the chain rule. This just says, any change of the x-coordinate as you move along the curve times the rate of change of the curve with x (and the same for y added) gives the directional change along the curve. Graphically, particularly simple in 2D, this procedure yields the tangent vector at a fixed point (the point where you evaluate the derivative!) along the curve.

The key point now is to realise that this procedure works for any function or curve. You can change to an operator representation and write

$$\frac{d}{d\lambda} = \frac{dx}{d\lambda}\frac{\partial}{\partial x} + \frac{dy}{d\lambda}\frac{\partial}{\partial y}$$

As it says in [5]: The trick to seeing the connection between derivatives and vectors is to view this equation as a vector equation in which:

$$\text{Vector} = \text{x-component} \times \text{x basis vector} + \text{y-component} \times \text{y basis vector}$$

That's step one in getting the message that derivatives and vectors have something to do with each other. Now is not the time to see that such basis vectors actually aren't necessarily vectors but rather one-forms[5].

The local definition of a vector along any curve is important, hence let's get back to that picture, have a curve, fix a point P_0 along it and define a vector $\boldsymbol{v}$ as the tangent vector along the curve at the point P_0. Now if you can define a vector then you can always also define a set of basis vectors at that chosen point in which to represent that vector $\boldsymbol{v}$. That translates into chosing a **local reference frame** at P_0.

Let's repeat the discussion from above in this more general setting. We need to calculate a gradient, a derivative, at some point along the curve in the direction of $\boldsymbol{v}$. Note that $\boldsymbol{v}$ is the tangent vector by definition. How would one calculate the rate of change of the curve in the direction of $\boldsymbol{v}$?

Well, the point is that $\boldsymbol{v}$ is already the tangent vector, nothing more to calculate, really. If we choose a simple orthonormal basis, say in 3D the $[x, y, z] = [x^i]$ Cartesian coordinates, $[\boldsymbol{e}_1 = \boldsymbol{i}, \boldsymbol{e}_2 = \boldsymbol{j}, \boldsymbol{e}_3 = \boldsymbol{k}]$, then one could represent $\boldsymbol{v}$ as

$$\boldsymbol{v}|_{\text{at} P_0 \; along \; P(\lambda)} = \sum_{i=1}^{3} \frac{dx^i}{d\lambda} \frac{\boldsymbol{\partial}}{\boldsymbol{\partial x^i}} = v^i \, \boldsymbol{\partial_i} = v^i \, \boldsymbol{e}_i$$

[5]The reason is that in the cartesian frame both become identical. The proper distinction between vectors and one-forms emerges once we look at the transformation properties, see below. That will also be the point when all doubts about tensors will be put to rest.

where index notation comes into full fruition and it is made explicit that the $\boldsymbol{e}_i$ basis is defined using partial derivatives. Note the upper index on v^i, signalling v as a vector (components have upper index) and the derivative symbol $\boldsymbol{\partial_i}$ having lower index components, signalling a one-form. They are summed over according to Einstein summation convention which is nothing else than the dot-product as it should be.

This way of calculating a tangent vector to any curve and define a local reference frame is universally valid, even in general relativity, which is very useful indeed. Typically, one would write the expression as an operator expression, omitting the $P(\lambda)$ symbol since it's all valid for any curve. Likewise it's not necessary to choose Cartesian coordinates. In fact, later our reference frame will be based on Minkowski space-time rather than a Euclidean space and the relations between basis vectors will change accordingly. Another point to note here is that the above is also known as the **directional derivative** if in addition the vector v is normalised, i.e. a unit vector.

Finally, a last ditch attempt to yet show you the same again from a more hands-on point of view. Here a drawing will greatly assist in getting the message. Consider a given space-time curve $P(\lambda)$, draw a tangent at some fixed point λ_0. Then pick a point λ_1 a little further up the curve and call that $\lambda_1 = \lambda_0 + \epsilon$. So much for the curve, no reference frame needed. Now, still on the same drawing pick some point $t_0 + \delta$ and $x_0 + \delta$ as the endpoints to a vector along the tangent at $\lambda_0 = (x_0, t_0)$. That's it, no more drawing. Just realise that the slope of the tangent and your vector simply is given by

$$\text{slope} = \frac{t_0 + \delta - t_0}{x_0 + \delta - x_0}$$

which is valid in the reference frame but now you also have purely for the curve:

$$\text{slope} = \lim_{\epsilon \to 0} \frac{P(\lambda_0 + \epsilon) - P(\lambda_0)}{\lambda_0 + \epsilon - \lambda_0} = \left.\frac{dP(\lambda)}{d\lambda}\right|_{\lambda_0}$$

Hence, here you have a local tangent vector definition without referring to any reference frame.

As stated in the footnote, the proper definition of a one-form still awaits you when reaching the topic of transformations. Here, the issue was merely mentioned, motivated and then more or less elegantly side-stepped and declared pending. This leaves one more oddity to clear up, the position of indices on basis vectors and one-forms.

Figure 1.5: Attempt at drawing the hands-on, operational, definition of a local vector as a tangent vector. Insert the drawing during the lecture.

Why are the indices of basis vector and one-forms in the wrong place?

Looking back at the introduction of basis vectors $\boldsymbol{g}_i$, you might have noticed something odd, i.e. the index is in low position while they are called vectors. This reflects a slight inconvenience of index notation in the senses that it is just as important to consider the object labelled with an index as is the position of the index. Here, the index labels vectors, not components of vectors. That makes all the difference. The reason why people continue using this notation and not just despair is that it's consistent. We wrote contractions of vector components with basis vectors without blinking and all works out fine.

Here is an explanation on how the index on basis vectors as well as one-forms can end up in the 'wrong' position and still deliver consistent maths. Consider the local definition of a vector. At any point P_0 in a metric space one can define a basis $\boldsymbol{e}_\mu$ as unit derivatives at that point. (Note the gradual introduction of Greek indices and the familiar $\boldsymbol{e}$ for an orthonormal set of basis vectors for special relativity).

For example, one can write for the first basis vector:

$$\boldsymbol{e}_0|_{P_0} \equiv \left.\frac{\partial}{\partial x^0}\right|_{P_0} \equiv \boldsymbol{x}_{,0} \equiv \boldsymbol{\partial}_0$$

where the 'comma-derivative' notation is briefly mentioned. It's used abundantly in general relativity but we will not use it any further. In any case, strictly speaking, the basis vector $\boldsymbol{e}_0$ should actually be written as $e^\mu{}_{,0}$, i.e. derivatives with respect to coordinate zero, applied to all components, μ in the upper position, of the vector.

Tensors

Moving on from vectors and one-forms to more general objects, i.e. tensors of rank 2. How can such a thing be represented?

Assume a tensor $\mathcal{T}$ acts on a vector $\boldsymbol{v}$ to give a vector:

$$\mathcal{T} \otimes \boldsymbol{v} = \boldsymbol{T}$$

In abstract notation as above, it's a simple thing but transferring this to index notation, one has to keep in mind that this operation can give you N-different vectors, with N being the dimension of the vector space, i.e. the number of independent basis vectors. The vector $\boldsymbol{T}$ would be different each time you chose a different basis vector for $\boldsymbol{v}$, hence the result in index notation must be the vector $\boldsymbol{T}_j$.

Now express the vector $\boldsymbol{T}_j$ as linear combinations of the given basis vectors:

$$\boldsymbol{T}_j = T_{ij}\, \boldsymbol{g}^i$$

where the T_{ij} are simple numbers, the coefficients here and later the tensor components. Then you could calculate the T_{ij} from

$$T_{ij} = \boldsymbol{g}_i \mathcal{T} \boldsymbol{g}_j$$

Let's take a closer look and gain some more insight on the way. Take any vector $\boldsymbol{v}$ and compute

$$\begin{aligned}
\mathcal{T}\boldsymbol{v} &= \mathcal{T}\left(v^j\, \boldsymbol{g}_j\right) \\
&= \mathcal{T}\,\boldsymbol{g}_j\, v^j \\
&= \boldsymbol{T}_j\, v^j \\
&= T_{ij}\boldsymbol{g}^i\, v^j \\
&= T_{ij}\boldsymbol{g}^i\, \left(\boldsymbol{g}^j \cdot \boldsymbol{v}\right) \\
&= \left(T_{ij}\, \boldsymbol{g}^i \boldsymbol{g}^j\right)\, \boldsymbol{v}
\end{aligned}$$

Hence you can get a representation of a general tensor using basis one-forms in this case (could be vectors - same derivation but calculating the T^{ij} components) as:

$$\mathcal{T} = T_{ij}\boldsymbol{g}^i \boldsymbol{g}^j$$

This is the general representation of a rank 2 tensor with components T_{ij}. In a little more detail this is sometimes referred to as a $(0, 2)$ tensor rank where the first position numbers the indices in the upper position, the second slot the number of indices in the lower position. There can just as well be mixed rank, $(1, 1)$-rank, tensors with coefficients T^i_j. These are the most troublesome since it's sometimes, not always, important which index comes first, i.e. stands left-most.

It should be clear now why rank 2 tensors can always be represented in matrix form. The matrix collects the components in a number grid, one index the row, the other the column. Nevertheless, you would never know which type of rank a tensor has just from looking at the matrix.

1.5.3 Transformations

There are other important operations which are represented as a matrix and even mingle with tensors, yet they aren't tensors, or often they aren't. This can cause no end to misunderstandings. Knowing your transformations from your tensors will be important already in special relativity but even more so in general relativity. Therefore we will discuss transformations in some detail and this way finally close the loop to the missing definition of tensors. You will have acquired enough maths by the end to comfortably follow the return to special relativity.

Generally, transformations are defined by a change of basis in a given vector space (which is the same as as saying basis in a given one-form space - they are closely related). The clue to understanding transformation lies in the realisation that scalars, vector and generally all tensors are geometric invariants, i.e. their components change under a transformation, not the tensors themselves.

That sounds maybe more mysterious than it actually is. Quick example, consider a football in a stadium. At any moment in time, there are many thousand different points of view on where the ball is (our side, opposite side, left, right, etc) but that doesn't change the ball. Its position simply is transformed by each single observer (shift of origin, typically).

For the following, we will consider only linear transformations such as the Lorentz transformation as an example.

Any change of basis can then be written as

$$\boldsymbol{g'}_1 = A^1_1 \boldsymbol{g}_1 + A^2_1 \boldsymbol{g}_2 + A^3_1 \boldsymbol{g}_3$$

$$\boldsymbol{g'}_2 = A^1_2 \boldsymbol{g}_1 + A^2_2 \boldsymbol{g}_2 + A^3_2 \boldsymbol{g}_3$$

etc. A shorter version could be

$$\boldsymbol{g'}_j = A^i_j \boldsymbol{g}_i$$

In order to qualify as a valid transformation, it must have an inverse, i.e. $\det\mathcal{A} \neq 0$, hence

$$\boldsymbol{g}_j = \left(A^{-1}\right)^i_j \boldsymbol{g'}_i$$

Similarly have

$$\boldsymbol{g'}^i = \left(A^{-1}\right)^i_j \boldsymbol{g}^j$$

and

$$\boldsymbol{g}^i = A^i_j \boldsymbol{g'}^j$$

This means one can transform any vector according to the equations above, replacing the basis vector with any arbitrary vector component.

$$v'_j = A^i_j\, v_i \quad ; \quad v'^i = \left(A^{-1}\right)^i_j\, v^j$$

Likewise, similar expression can be written immediately for rank 2 tensors when reminding yourself that we wrote any rank 2 tensor as being constructed from 2 basis vectors. So, tensor transformations look like this:

$$\begin{aligned} T'_{ij} &= A^k_i\, A^l_j\, T_{kl} \\ T'^i_j &= \left(A^{-1}\right)^i_k\, A^l_j\, T^k_l \\ T'^j_i &= A^k_i\, \left(A^{-1}\right)^j_l\, T^l_k \\ T'^{ij} &= \left(A^{-1}\right)^i_k\, \left(A^{-1}\right)^j_l\, T^{kl} \end{aligned} \tag{1.15}$$

Last thing to clarify would be what these A's are. Say you have a basis $\boldsymbol{g}_i$ and coordinates x^i as well as transformed $\boldsymbol{g'}_i$ and transformed coordinates x'^i and a relation

$$x^i = f^i(x'^k)$$

Represent the basis vectors locally, i.e. $\boldsymbol{g}_i = \boldsymbol{\partial}_i$ and likewise for the transformed basis and use the chain–rule to get

$$\boldsymbol{g'}_j = \frac{\partial\, x^1}{\partial\, x'^j}\, \boldsymbol{\partial}_1 + \frac{\partial\, x^2}{\partial\, x'^j}\, \boldsymbol{\partial}_2 + \ldots$$

in short

$$\boldsymbol{g'}_j = \frac{\partial\, x^i}{\partial\, x'^j}\, \boldsymbol{g}_i$$

Hence the numbers A^i_j from the expressions above are simply

$$A^i_j = \frac{\partial\, x^i}{\partial\, x'^j} \tag{1.16}$$

similarly

$$\left(A^{-1}\right)^i_j = \frac{\partial\, x'^i}{\partial\, x^j}$$

The definition of tensors can now proceed by adopting eqns. 1.15 with transformations given as in eqn. 1.16. Any object which transforms under linear transformations (1.16) as in 1.15 is called a tensor.

Explicit one-form and vector definition

Back in section 1.5.2, I promised to come back to a proper one-form (and vector) definition. In principle that has happened already using the tensor definition above. After all, one-forms and vectors are rank one tensors. However, I find an explicit definition, in all detail, helps to get to the point rather than more abstract expressions and can remove confusion maybe left-over from previous sections[6].

If you look for a template for an object or term that wants to be a vector then a physicist would tell you the position observable would be just the thing (for a mathematician such a question doesn't make any sense in the first place). In order to make the following even a little easier, let's take the differential position, $\boldsymbol{dx}$, vector. If there is a transformation (for instance to spherical coordinates) given to primed coordinates, $\boldsymbol{dx'}$, then you know how to write that explicitly:

$$\begin{aligned} dx'^1 &= \frac{\partial x'^1}{\partial x^1}\,dx^1 + \frac{\partial x'^1}{\partial x^2}\,dx^2 + \frac{\partial x'^1}{\partial x^3}\,dx^3 \\ dx'^2 &= \frac{\partial x'^2}{\partial x^1}\,dx^1 + \frac{\partial x'^2}{\partial x^2}\,dx^2 + \frac{\partial x'^2}{\partial x^3}\,dx^3 \\ dx'^3 &= \frac{\partial x'^3}{\partial x^1}\,dx^1 + \frac{\partial x'^3}{\partial x^2}\,dx^2 + \frac{\partial x'^3}{\partial x^3}\,dx^3 \end{aligned}$$

This is nothing else than the transformation rule we've discussed above, just more explicit and specialised to 3D coordinates. In matrix notation this looks quite familiar:

$$\begin{pmatrix} dx'^1 \\ dx'^2 \\ dx'^3 \end{pmatrix} = \begin{pmatrix} \frac{\partial x'^1}{\partial x^1} & \frac{\partial x'^1}{\partial x^2} & \frac{\partial x'^1}{\partial x^3} \\ \frac{\partial x'^2}{\partial x^1} & \frac{\partial x'^2}{\partial x^2} & \frac{\partial x'^2}{\partial x^3} \\ \frac{\partial x'^3}{\partial x^1} & \frac{\partial x'^3}{\partial x^2} & \frac{\partial x'^3}{\partial x^3} \end{pmatrix} \begin{pmatrix} dx^1 \\ dx^2 \\ dx^3 \end{pmatrix}$$

or in full index notation

$$dx'^i = A^i_j dx^j$$

For a one-form we follow the motivation in section 1.5.2 and consider, finally, the gradient as the best template for a one-form, given that positions are defined as vectors (otherwise we would have just the opposite discussion).

It's the chain rule that will yield all the differentiation we need. If an operator such as $\partial/\partial x$ is transformed from one into a primed coordinate system, say in a 3D

[6]This follows [5].

space again, we get:

$$\begin{aligned}
\frac{\partial}{\partial x'^1} &= \frac{\partial x^1}{\partial x'^1}\frac{\partial}{\partial x^1} + \frac{\partial x^2}{\partial x'^1}\frac{\partial}{\partial x^2} + \frac{\partial x^3}{\partial x'^1}\frac{\partial}{\partial x^3} \\
\frac{\partial}{\partial x'^2} &= \frac{\partial x^1}{\partial x'^2}\frac{\partial}{\partial x^1} + \frac{\partial x^2}{\partial x'^2}\frac{\partial}{\partial x^2} + \frac{\partial x^3}{\partial x'^2}\frac{\partial}{\partial x^3} \\
\frac{\partial}{\partial x'^3} &= \frac{\partial x^1}{\partial x'^3}\frac{\partial}{\partial x^1} + \frac{\partial x^2}{\partial x'^3}\frac{\partial}{\partial x^2} + \frac{\partial x^3}{\partial x'^3}\frac{\partial}{\partial x^3}
\end{aligned}$$

In matrix representation:

$$\begin{pmatrix} \frac{\partial}{\partial x'^1} \\ \frac{\partial}{\partial x'^2} \\ \frac{\partial}{\partial x'^3} \end{pmatrix} = \begin{pmatrix} \frac{\partial x^1}{\partial x'^1} & \frac{\partial x^2}{\partial x'^1} & \frac{\partial x^3}{\partial x'^1} \\ \frac{\partial x^1}{\partial x'^2} & \frac{\partial x^2}{\partial x'^2} & \frac{\partial x^3}{\partial x'^2} \\ \frac{\partial x^1}{\partial x'^3} & \frac{\partial x^2}{\partial x'^3} & \frac{\partial x^3}{\partial x'^3} \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial x^1} \\ \frac{\partial}{\partial x^2} \\ \frac{\partial}{\partial x^3} \end{pmatrix}$$

or in full index notation

$$\partial'_i = (A^{-1})_i^j \partial_j$$

where $(A^{-1})_i^j$ is the inverse transformation to the vector transformation matrix. This constitutes a definition of one-forms and vectors using their transformation properties just as for any other tensor.

Please note that it is irrelevant which transformation matrix you call matrix and which the inverse matrix. The point is that one-forms use the inverse transformation relative to the vector transformation. Geometrically that makes them different objects compared to vectors, i.e. not arrows. That doesn't help much with defining one-forms as outlined in section 1.5.2 but helps in realising that there is a fundamental difference between vectors and one-forms and that is the lesson to take away.

1.5.4 Back to Lorentz transformations

Let's put all these new maths tools into practice by transferring to special relativity. The first thing is obviously the Lorentz transformations. We have just discussed linear transformations. Connecting the above to Lorentz transformations should be straightforward.

In order to clarify the transition from arbitrary transformations and basis to the Minkowski metric and Lorentz transformation (and 4D coordinates), I will introduce new symbols for the latter:

- Transformed terms receive a 'hat' instead of a prime: $\hat{x}^\mu$
- The Lorentz transformation coefficients will be labelled L_ν^μ with abstract matrix symbol: $\mathcal{L}$

- The Minkowski metric coefficients are $\eta_{\mu\nu}$: Matrix $\boldsymbol{\eta}$ (sorry for the bold font, but there is no calligraphic Greek eta), see eqn. 1.13.

Lorentz transformation The transformations

$$\hat{x}^\mu = L^\mu_\nu \, x^\nu$$

represent the Lorentz group, defined by the condition

$$\eta_{\rho\sigma} \, L^\rho_\mu L^\sigma_\nu = \eta_{\mu\nu}$$

(note the contraction operations on ρ and σ in order to train your eyes on relations like this) and

$$det(\mathcal{L}) = \pm 1$$

in order to leave the line element invariant:

$$\begin{aligned} ds^2 &= \eta_{\mu\nu} \, dx^\mu \, dx^\nu = dx_\nu \, dx^\nu \\ &= L^\nu_\sigma \, d\hat{x}_\nu \, L^\sigma_\nu \, d\hat{x}^\nu \\ &= L^\nu_\sigma \, L^\sigma_\nu \, d\hat{x}_\nu \, d\hat{x}^\nu \end{aligned}$$

that is, ds^2 is invariant if

$$L^\nu_\sigma \, L^\sigma_\nu = 1$$

which is equivalent to

$$\frac{\partial x^\nu}{\partial \hat{x}^\sigma} \frac{\partial \hat{x}^\sigma}{\partial x^\nu} = 1$$

which is true in general for all partial differentials.

More specifically, let's do the entire exercise again. One can write LT1 (see, eqn. 1.9) in this notation as

$$L(-\beta) = L^\mu_\nu = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Checking invariance of the metric explicitly is conveniently done in matrix notation:

$$\mathcal{L}^T \, \boldsymbol{\eta} \, \mathcal{L} = \hat{\boldsymbol{\eta}}$$

which is

$$\begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} \gamma^2 - \beta^2\gamma^2 & -\gamma^2\beta + \gamma^2\beta & 0 & 0 \\ -\gamma^2\beta + \gamma^2\beta & \beta^2\gamma^2 - \gamma^2 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

using the identity

$$\gamma^2 - \beta^2\gamma^2 = \frac{1}{1-\beta^2} - \frac{\beta^2}{1-\beta^2} = 1$$

gives $\hat{\boldsymbol{\eta}} = \boldsymbol{\eta}$ and hence that ds^2 is invariant.

This definition of a Lorentz transformation immediately offers the possibility to define another very important object, used throughout the lecture, i.e. a four-vector.

Four-vector A four-vector is a vector which transforms as

$$\hat{x}^\nu = L^\nu_\mu x^\mu$$

with L^ν_μ a Lorentz transformation.

Four-vectors build a vector space with a length:

$$\begin{aligned} \boldsymbol{x}^2 &\equiv \eta_{\mu\nu}\, x^\mu\, x^\nu \\ &= (x^0)^2 - (x^i)^2 \end{aligned}$$

and a scalar product

$$\boldsymbol{x} \cdot \boldsymbol{v} = \eta_{\mu\nu}\, x^\mu\, v^\nu = x^0 v^0 - x^i v^i$$

hence the length is **semi-definite**, i.e. can be positive, negative or zero. Such

Figure 1.6: Space-time diagram showing space-like and time-like regions. Insert the drawing during the lecture.

regions of space-time are called: (Caution: this depends on the sign convention in textbooks but the drawing above is always correct).

$$\begin{array}{ll} \boldsymbol{x}^2 > 0 & \text{time-like} \\ \boldsymbol{x}^2 = 0 & \text{null (light-like)} \\ \boldsymbol{x}^2 < 0 & \text{space-like} \end{array}$$

1.5.5 Exercises

(Collection sheet [1,1-13])

1. Index notation: Simplify the following expressions if possible:
   - $A_{\mu\nu}\,B^{\nu}$
   - $A^{\mu}_{\ \nu}\,B_{\sigma}$
   - $A_{\nu}\,B^{\nu}$
   - $A_{\nu}\,B^{\nu\sigma}\,C_{\sigma}$
   - $A_{\mu\nu}\,B^{\sigma}_{\ \rho}\,C^{\mu\nu}$
   - $A^{\mu\nu\sigma\rho}\,B_{\nu}\,C_{\mu}$
2. Write the following abstract matrix multiplication in index notation:
   - $\mathbf{A} = \mathbf{M}\,\mathbf{g}\,\mathbf{N}$
3. Given a basis in three-dimensional space by

$$\underline{g}_1 = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} \;;\quad \underline{g}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \;;\quad \underline{g}_3 = \begin{pmatrix} -1 \\ -2 \\ 1 \end{pmatrix},$$

   - calculate the components v^i for a vector

$$\underline{v} = \begin{pmatrix} 3 \\ 3 \\ 6 \end{pmatrix}$$

     according to the general representation $\underline{v} = \sum_{i=1}^{3} v^i\, g_i$.
   - Find the one-form basis vectors.
   - Compute the components of the one-form corresponding to $\underline{v}$.
4. Let a coordinate system in a 2-dimensional Euclidean space have a basis

$$\underline{e}_1 = 2\,\underline{e}_x + 3\,\underline{e}_y; \quad \underline{e}_2 = \underline{e}_x - \underline{e}_y.$$

   Find the expansion of the vector $\underline{A} = 5\,\underline{e}_x + 6\,\underline{e}_y$ in this basis and test your result.

5. Find the transformation matrix and the corresponding metric tensor for the rotation coordinates $(r,\,\theta,\,t_r)$ and ω a constant in a Galileian space-time:
$$x = r\,\cos(\theta + \omega\,t_r)\,; \quad y = r\,\sin(\theta + \omega\,t_r)\,; \quad t = t_r$$

6. Find the local basis vectors and metric tensor for spherical polar coordinates defined by
$$x = r\,\sin\theta\,\cos\phi\,; \quad y = r\,\sin\theta\,\cos\,\phi\,; \quad z = r\,\cos\theta.$$

7. Quotient rule: Show that if $A^\alpha = B^{\alpha\lambda}\,C_\lambda$ is a tensor whenever **C** is a tensor, then **B** must be a tensor.

8. Show that the determinant of the metric tensor $g \equiv \det(g_{\mu\nu})$ is not a scalar.

9. Show that the invariant proper volume element in four-dimensional space is given by $dV = (-g)^{1/2}\,d^4x$, where $d^4x = dt\,dx\,dy\,dz$ in the coordinate system of the metric $g_{\mu\nu}$.

10. If $\phi(x^\mu)$ is a scalar function of the coordinates, show that $\frac{d\phi}{dx^\mu}$ transforms as a one-form.

11. In a coordinate system with coordinates x^μ, the invariant line element is $ds^2 = \eta_{\mu\nu}\,dx^\mu\,dx^\nu$. If the coordinates are transformed, $x^\mu \to \hat{x}^\mu$, show that the line element is $ds^2 = g_{\hat{\mu}\hat{\nu}}\,d\hat{x}^\mu\,d\hat{x}^\nu$, and write $g_{\hat{\mu}\hat{\nu}}$ in terms of the partial derivatives $\partial\,x^\mu/\partial\,\hat{x}^\nu$. For two arbitrary four-vectors **U** and **V** show that $\mathbf{U}\cdot\mathbf{V} = U^\mu\,V^\nu\,\eta_{\mu\nu} = \hat{U}^\mu\,\hat{V}^\nu\,g_{\hat{\mu}\hat{\nu}}$.

12. Given that a rank two tensor transforms as $\hat{A}^{\mu\nu} = a^\mu_\sigma\,a^\nu_\rho\,A^{\sigma\rho}$, where a^μ_σ is a linear transformation, show that, given two rank one tensors **B** and **C**, the products of the components, $T^{\gamma\delta} = B^\gamma\,C^\delta$ form a rank two tensor.

13. Show that the coefficients a^i_k of a linear transformation between rank one tensors **A** and **B**, where $B^i = a^i_k A^k$, form a rank two mixed tensor.

    (Collection sheet [2,2-4])

14. Sketch a space-time diagram containing two events, O and P, which display a time-like distance interval and two events, Q and R, which display a space-like distance interval.

15. Determine whether the following matrix represents a Lorentz transformation matrix, justifying your answer.
$$A^{\alpha\beta} = \begin{pmatrix} \sqrt{2} & 0 & 1 & 0 \\ 0 & 2 & 0 & 0 \\ 1 & 0 & \sqrt{2} & 0 \\ 0 & 0 & 0 & 1/2 \end{pmatrix}$$

16. If two events, P and O, are separated by a space-like distance interval, show using diagrams that (a) there exists a Lorentz frame in which they are simultaneous and (b) in no Lorentz frame do they occur at the same point.

Chapter 2

Applications: Mechanics

This chapter will remind you of a lot of concepts of special relativity that you should be very familiar with already. These are the classic applications of the theory on concepts from mechanics. It should be clear though that these really are merely applications of the theory and not 'The Theory'. We will go through some of those and I'll display a few more, maybe less well known applications. Subsequently, the exercise is repeated by applying special relativity to classical electromagnetism. Again, this is merely an application but a rather strange one since, as you should know, electromagnetism has special relativity already built-in! All we need to do is re-phrase Maxwell's theory such that it becomes manifestly invariant, i.e. make it obvious that this theory has been united (become one) with special relativity. As I said, sounds grand but all it is, is a simple translation since the unification has already happened right at the start without us having to do anything.

Ideally, this exercise would continue and apply special relativity to quantum mechanics but that topic deserves and requires separate lectures on its own since it is too rich to simply browse. Likewise, the extension of the theory, from special relativity to general relativity deserves and requires a separate lecture. Both these extensions are offered in the curriculum and you are encouraged to try them.

The list of items to go through during the rest of the lecture is composed of:

- Mechanics:
  - collection of reminders, mainly on kinematics
  - various four-vectors
  - Doppler effect
  - Acceleration
- EM:

– invariant formulation of electromagnetism (EM)

– Radiation (briefly)

2.1 Proper Time

Proper time is always the time measured by the clock of **one** observer. It is therefore an entirely local property. Each and every observer has his/her own clock, measuring individual proper times. Let's take a look at the cornerstone of special relativity and see how that translates into something a little more tangible.

Consider the line element ds^2 at any moment of rest, an arbitrary moment in time. Then on the local world-line of the observer, the spatial parts dx, dy, dz are all identically zero (trivially, since it's the observer's own world-line) and

$$ds^2 = c^2 dt^2$$

Now the clue is that we can always define any moment in time as a momentary frame of rest along the world-line (for that one, single observer) where the relation above is always true, then one can define

$$ds^2 = c^2 dt^2 \equiv c^2 d\tau^2$$

where τ is now defined as the proper time, $cd\tau = ds$, for that particular observer (or clock).

An immediate consequence, a rather important one, can be derived from here: Consider a stationary clock in system S', i.e. not just momentarily at rest but always, then for this clock, it's always

$$ds^2 = c^2 dt'^2 \equiv c^2 d\tau^2$$

with $d\boldsymbol{x}'^2 = 0$. If this observer compares his/her clock to the clock of another observer in a moving frame, S, something remarkable happens. Let's have the system S moving relative to S' with speed v in x-direction. Then for the observer in S, one can write

$$ds^2 = c^2 dt^2 - dx^2 = c^2 dt^2 \left(1 - \frac{v^2}{c^2}\right)$$

However, since ds^2 is invariant(!), it is identical for each observer and one can write

$$c^2 d\tau^2 = c^2 dt^2 \left(1 - \frac{v^2}{c^2}\right)$$

and hence

$$d\tau = \frac{dt}{\gamma}$$

which relates the time interval measured locally (the proper time) and the time interval determined for the moving observer, dt. Clearly, dt is a longer time interval than the local time interval ($\gamma > 1$).

This translates into the moving clock goes slower (1 second on the local clock are only < 1 seconds on the moving clock). This is the famous **time dilation**. The fun part of relativity is that the very same argument is true for the observer in S. Both observers have equally valid descriptions of which clock moves slower (that's relativity). It's always(!) the moving one.

This caused all manner of confusion since a clock moving slow is acceptable but this was considered to be an absolute statement. How could a single clock go fast and slow at the same time. How does it know which observer to show the fitting time?

The clue is, it doesn't. Once you mix up absolute and relative statements (measurements) you get into all sorts of trouble and paradoxes. The clock has its own identity and rhythm, no need to 'react' to anything. It's the relative motion of any arbitrary observer which causes him or her to perceive (measure) our clock to tick differently to how we measure it. All of this is anything but arbitrary but perfectly defined by the Lorentz transformations. You only need to be complete in the description of the physics and no paradoxes emerge.

2.2 Four-velocity

The equations above allow to draw another important conclusion immediately, the definition of a four-velocity, our first derived four-vector (as opposed to the coordinate four-vector which was merely stated as a starting point, not derived). First note from the definition of proper time along a given world-line that

$$\frac{cdt}{ds} = \frac{dt}{d\tau} = \gamma$$

Then we can define the **four-velocity $\boldsymbol{u}$** as

$$u^\mu = \frac{dx^\mu}{d\tau} = c \begin{pmatrix} c\frac{dt}{ds} \\ \frac{d\underline{x}}{ds} \end{pmatrix} = c\,\frac{c\,dt}{ds} \begin{pmatrix} 1 \\ \underline{\beta} \end{pmatrix} = c\,\gamma \begin{pmatrix} 1 \\ \underline{\beta} \end{pmatrix} \tag{2.1}$$

with the three-vector $\underline{x}$ as abbreviation for the three spatial coordinates and similarly for the three vector $\underline{\beta} = \underline{v}/c$ and $\underline{v}$ being the relative velocity between S and S'.

Looking closely at eqn. 2.1 you can see that each and every term in it is Lorentz invariant, hence this velocity definition is Lorentz invariant. Therefore, this definition is valid in all reference frames! Very useful, indeed.

A few quick properties should be derived too. As you will see as the lecture progresses, once you have a four-vector, the first thing you will want to do with it is to compute an invariant from it, a number that is specific for this four-vector and does not change under transformations. The recipe is very simple: square the vector. Here is the (not terribly useful) example for the four-velocity:

$$\boldsymbol{u}^2 = u_\mu\, u^\mu = \frac{dx_\mu\, dx^\mu}{d\tau^2} = c^2\,\frac{ds^2}{ds^2} = c^2$$

So the magnitude of any four-velocity vector is always constant and equal to the speed of light. No surprise there, this is simply the case by construction. Other four-vectors will show more interesting magnitudes.

The four-velocity is also quite an important geometrical entity. Have a look at the definition and you can see that it's the derivative of the world-line with respect to its own proper time. For a given world-line, the proper time really is simply a number parameter along the curve. Take the derivative along the line at any point gives you the tangent vector at that chosen point on the curve. That tangent vector can easily be normalised since its magnitude is always the same and constant, see above, and since it's a positive number, the tangent vector must be a time-like vector. Normalised and always time-like is a lengthy description of a perfect candidate for a time-like, momentary basis vector for a given world-line. Momentary, since you have to take the derivative at some fixed point along the curve. That corresponds to a moment in time for that body (observer, clock, ...) and that in turn allows to always define a momentary rest frame along the curve.

With the four-velocity given as the time-like basis vector, it is simple to define three further orthogonal four-vectors in a rest frame and you a have built yourself a set of basis vectors **at that moment in time along a specific world-line**, i.e. a local reference frame. That is a most useful tool, a local reference frame, in order to analyse all sorts of kinematic exercises. It will come into full fruition as soon as we come to acceleration. For now let's define

$$\frac{u^\mu}{c} = (\boldsymbol{e}_0)^\mu$$

as the time-like, local basis vector along an arbitrary world-line with four-velocity $\boldsymbol{u}(\tau)$.

2.2.1 Exercises

(Collection sheet [2,5]) Let the four-velocity **u** contain the three-velocity $\underline{v}$, then

Figure 2.1: Space-time diagram showing the normalised four-velocity as a unit time-like basis vector along a world-line. Insert the drawing during the lecture.

1. express u^0 in terms of $|\underline{v}|$,
2. u^i in terms of $\underline{v}$,
3. u^0 in terms of u^i and
4. $d/d\tau$ in terms of d/dt and $\underline{v}$.

2.3 Relativistic kinematics essentials

Having mastered the Lorentz transformations and the four-velocity, you are in a position to build all of the rest of classical mechanics quite cleanly. The first and most important concept to introduce on this path is called a four-momentum.

Four-momentum Define the four-momentum in line with Newtonian mechanics as $\boldsymbol{p} = m\,\boldsymbol{u}$

and check whether this makes sense. First, build an invariant from this four-vector candidate.

$$\boldsymbol{p}^2 = (p^0)^2 - (p^i)^2 = m^2\,\boldsymbol{u}^2 = m^2\,c^2$$

This should be nicely invariant since m and c are assumed to be pure numbers. Likewise this is already a tantalisingly familiar form. All that we still need are the four-vector components and an interpretation of the number 'm'. Consider the 'm' therefore as preliminary for now.

Here I take a big shortcut. Assume(!) $E = m\,c^2$ (see later in the lecture) and use Newton's relation as guidance for interpretation (let's also work in only one-dimension to simplify the arithmetic a little). Note that even such a crude derivation as shown below is good enough simply to make an educated guess on the interpretation of terms. Take

$$F = \frac{dp}{dt} \quad ; \quad p = mv$$

as given and likewise take an infinitesimal change of kinetic energy according to

$$dE = F\,dx = \frac{dp}{dt}\,dx = dp\,\frac{dx}{dt} = v\,dp$$

and use $E = mc^2$ to get (using $p = mv$, beware the 'm' symbol here) additionally $E = \frac{p}{v}c^2$. Combine it all to

$$E\,dE = c^2 p\,dp$$

and integrate

$$E^2 = c^2p^2 + E_0^2$$

with E_0 an integration constant. Replace p with E to

$$E^2 = \left(\frac{E\,v}{c}\right)^2 + E_0^2$$

solve for E

$$E = \frac{E_0}{\sqrt{1 - \frac{v^2}{c^2}}}$$

or

$$m(v) = \frac{E_0/c^2}{\sqrt{1 - \frac{v^2}{c^2}}}$$

then an interpretation of terms might be a little easier. It looks like that E_0/c^2 is the rest mass $m_0 = m(0)$. This enables the interpretation of

$$m \equiv m(v) = \gamma m_0$$

as the 'mass' in $p = mv = m(v)v$. Combine this with the definition of the four-velocity in order to get to an interpretation of the 'm' in $\boldsymbol{p} = m\,\boldsymbol{u}$ from the three-vector relation $\underline{p} = m(v)\underline{v}$. The three-velocity component of $\boldsymbol{u}$ is $\gamma\underline{v}$, hence the 'm' is in fact the rest mass m_0, i.e. the proper definition of the four-momentum results as $\boldsymbol{p} = m_0\,\boldsymbol{u}$

Now we can immediately get to the zero-component of the four-vector $\boldsymbol{p}$. This argument works as follows: Pick a system at rest ($\underline{p} = 0; m(0) = m_0$) and get

$$(p^0)^2 - (p^i)^2 = (p^0)^2 = m_0^2c^2$$

hence

$$p^0 = m_0 c = \frac{E_0}{c}$$

using $E = mc^2$ again. Having assumed a system of rest, however, since the four-momentum is a four-vector, all relations valid in one frame are valid in all frames. How convenient is that.

Writing this four-vector explicitly for any frame gives the **fundamental relation of relativistic kinematics**:

$$E^2 - |\underline{p}|^2 c^2 = m_0^2 c^4 \tag{2.2}$$

This deeply rooted connection between energy and momentum often leads to calling this four-vector the energy-momentum four-vector.

2.3.1 Exercises

(Collection sheet [2,6-11])

1. Calculate the speed of a particle whose kinetic energy is equal to its rest mass energy.
2. A particle of rest mass m and four-momentum **p** is examined by an observer with four-velocity **u**. Show that:
   - the observer measured energy is $E = \mathbf{p}\,\mathbf{u}$.
   - The rest mass attributed to the particle is $m^2c^2 = \mathbf{p}^2$
   - The momentum measured has magnitude $c|\mathbf{p}| = [(\mathbf{p}\cdot\mathbf{u})^2 - c^2\,\mathbf{p}^2]^{1/2}$
   - the ordinary three-velocity $\underline{v}$ has magnitude

   $$\beta = \frac{|\underline{v}|}{c} = \left[1 - \frac{c^2\,\mathbf{p}^2}{(\mathbf{p}\cdot\mathbf{u})^2}\right]^{1/2}.$$
3. A hypothetical particle has negative squared mass such that one can write $m = i\mu$, where μ is a real quantity with units of mass. In all other aspects, the particle satisfies the rules of special relativity.
   - Show that the energy, E, of the particle is

     $$E = \frac{\mu\,c^2}{\sqrt{\beta^2 - 1}}$$

     and hence determine the momentum, $p = |\underline{p}|$, of the particle in terms of its normalised speed β.

- What is the possible range of speed, assuming real values for E and p? For any maximum or minimum speed state the corresponding values of E and p.
- Consider the free propagation of the hypothetical particle in the laboratory frame at some speed β_L. Define two events along the trajectory of the particle separated by a time $\triangle t$ as observed in the LAB frame. Find a second reference frame, S', in which the particle propagates backwards in time.

4. For a particle whose position and momentum four-vectors are given by $\mathbf{X}$ and $\mathbf{P}$, the angular momentum tensor is defined by

$$L^{\mu\nu} = X^\mu P^\nu - P^\mu X^\nu.$$

Show that any freely moving particle has constant angular momentum.

5. Assume an observer at a point O emits a signal which travels faster than the speed of light and triggers an event P distinct from O. Show that this implies the existence of another, second observer for whom causality is violated.

6. Consider a particle with rest mass m moving with velocity $\underline{u}$ relative to the laboratory frame when it collides elastically with a second particle, also of rest mass m, which is at rest in the laboratory frame. After the collision, the particles have velocities $\underline{v}$ and $\underline{w}$. Show that if θ is the angle between $\underline{v}$ and $\underline{w}$, then

$$\cos\theta = \frac{c^2}{|\underline{v}|\,|\underline{w}|}\,\frac{(\gamma(v)-1)\,(\gamma(w)-1)}{\gamma(v)\,\gamma(w)}$$

Also compare the result to the Newtonian prediction of $\theta = \pi/2$ for all $\underline{v}$ and $\underline{w}$, particularly for the limiting case of $|\underline{v}|$ and $|\underline{w}|$ tending towards c.

2.4 Energy-mass equivalence derivation, Einstein 1906

The original paper on the derivation of $E = mc^2$ does not always lend itself to an easy understanding of the reasoning behind the famous equation. Einstein published a second paper containing an alternative derivation 1906 which appears to be more elementary. Unfortunately, this article is not available in translation, free of charge. Therefore, let's repeat his work in the following, learning by doing.

Consider a box of length L, extending in the x-direction. This box shall contain a laser (not part of the original argument but easier here to imagine), mounted on the left wall of the box at $x = 0$. The laser can emit a light pulse in positive x-direction. Assume total absorption of the light on the opposite wall of the box, initially at $x = L$. You may assume that for light the relation between energy and momentum is $E = p\,c$. Take M as the mass of the box and treat the eventual motion of the box as strictly non-relativistic. This was more or less the text of one

Figure 2.2: Box configuration for Einstein argument on $E = mc^2$ derivation. Initial configuration first, then after firing the laser pulse. Insert the drawing during the lecture.

of the 2011 exam questions. The start of the derivation is, as so often, momentum conservation. Consider the reaction of the box on sending out a laser pulse at $t = 0$:

$$p_{\text{light}} = p_{\text{box}}$$

Treating the box all non-relativistic, translates to

$$v_{\text{box}} = -\frac{p_{\text{light}}}{M} = -\frac{E}{M\,c}$$

where the only external pre-knowledge $E = p\,c$ for light, enters the derivation. Light now arrives at the absorption wall after a time interval of $\triangle t$. In that time the box moves by

$$\triangle x = v_{\text{box}}\,\triangle t$$

to the left (negative speed) and the light traverses a distance

$$L + \triangle x = c\,\triangle t$$

Finally insert all you got previously

$$\triangle x = -\frac{E}{M\,c}\,\triangle t = -\frac{E}{M\,c^2}(L + \triangle x)$$

However, after absorption (hence after $\triangle t$) the box must stop and apparently the centre of gravity moved by $\triangle x$, see above, which would be wrong! Conservation of the centre of gravity requires therefore to assign some mass, m, to the light pulse in order to get (conservation of centre of gravity, explicitly):

$$\begin{aligned} 0 &= M\,\triangle x + m\,(L + \triangle x) \\ &= -\frac{E}{c^2}\,(L + \triangle x) + m\,(L + \triangle x) \\ &= \left(-\frac{E}{c^2} + m\right)(L + \triangle x) \\ \Rightarrow & \quad E = mc^2 \end{aligned}$$

2.5 Waves

Here, again, we will use a shortcut provided by Einstein (conveniently keeping the little cheats in the family) and access this topic by using quantum theory (his photo-effect paper is the basis for considering light being made of particles). Another reason to examine waves at this point is the immediate connection to the energy-momentum four-vector from above, see sec. 2.3.

The fundamental kinematics relation (eqn. 2.2) for light results in

$$p_\mu\,p^\mu = m^2c^2 = 0$$

a null-vector, since light moves on the light-cone. That trivially implies that light should be made of particles having zero rest mass. However, that does not imply zero energy-momentum four-vector components! Such a particle can have any amount of energy and momentum assigned to it as long as $E = p\,c$ and hence the magnitude of the four-vector is zero.

One may therefore give photons a momentum and using quantum knowledge assign a wave number $p_{\text{photon}} = \hbar k$ (one-dimensional). That is a good starting point to propose a new four-vector:

$$k^\mu = \begin{pmatrix} \omega/c \\ \underline{k} \end{pmatrix}$$

where $\underline{k}$ is the wave three-vector and $\omega/2\pi$ is the wave frequency and $E_{\text{photon}} = \hbar\omega$.

Any wave can be described by a wavefunction

$$\Psi(\underline{x}) \propto \cos(\omega t - \underline{k}\cdot\underline{x})$$

where the spatial direction of propagation would be $\underline{k}$ with a phase velocity

$$v_{\text{phase}} = \frac{\omega}{|\underline{k}|}$$

in one fixed reference frame.

A frame-independent form is proposed to look like

$$\Psi(\underline{x}) \propto \cos(k_\mu x^\mu)$$

using the proposed wave one-form and the coordinate four-vector.

These proposals require some testing. For a wave describing light, we know that the energy-momentum four-vector is a null vector. Considering the relation between the wave vector and momentum in non-relativistic physics, it appears reasonable to assume that for light, we can require

$$k_\mu k^\mu = 0 = \frac{\omega^2}{c^2} - |\underline{k}|^2$$

which results in an invariant phase velocity of

$$\frac{\omega}{|\underline{k}|} = c$$

Furthermore, using the wave four-vector to examine effects of Lorentz transformations on waves, say use LT2 (1.10):

$$\begin{aligned}\begin{pmatrix} \omega/c \\ \underline{k} \end{pmatrix} &= \begin{pmatrix} \gamma & \gamma\beta & 0 & 0 \\ \gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \omega'/c \\ \underline{k}' \end{pmatrix} \\ &= \begin{pmatrix} \gamma\,(\omega'/c + \beta k'^x) \\ \gamma\,(k'^x + \beta\,\omega'/c) \\ k'^y \\ k'^z \end{pmatrix}\end{aligned}$$

Now look at a plane wave in S' with wave vector $\underline{k}'$ at an angle $\cos(\theta') = \frac{k'^x}{|\underline{k}'|}$ to the x'-axis. This gives (take zero-component of proposed four-vector times c):

$$\omega = \gamma\,(\omega' + \beta c\,|\underline{k}'|\cos(\theta'))$$

using the phase velocity relation

$$\omega = \gamma\omega'\,(1 + \beta\cos(\theta')) \tag{2.3}$$

which is the relativistic Doppler effect!

For an angle θ in S, set $\underline{k'}$ into the $x' - y'$ plane, i.e.

$$\underline{k'} = \begin{pmatrix} \cos\theta' \\ \sin\theta' \\ 0 \end{pmatrix} \frac{\omega'}{c}$$

in S. That results in two useful relations directly from here: Take the 1-component to derive

$$\frac{\omega}{c}\cos\theta = \gamma\,\frac{\omega'}{c}\,(\cos\theta' + \beta)$$

using the Doppler effect formula, eqn. 2.3 then results in:

$$\cos\theta = \frac{\cos\theta' + \beta}{1 + \beta\cos\theta'} \tag{2.4}$$

and from the 2-component, get:

$$\sin\theta = \frac{\sin\theta'}{\gamma\,(1 + \beta\cos\theta')} \tag{2.5}$$

These equations 2.4 and 2.5 describe relativistic aberration.

This is all quite re-assuring and convenient but there is one more consequence from proposing the wave four-vector, i.e. a new effect, specific to special relativity and unknown to non-relativistic physics. That is the **transversal Doppler** effect.

For the longitudinal Doppler effect, take the Doppler effect formula, eqn.2.3, from above and insert $\theta' = 0$:

$$\omega = \gamma\,\omega'\,(1 + \beta) = \omega'\,\sqrt{\frac{1+\beta}{1-\beta}}$$

For the transversal Doppler effect, insert $\theta' = \pi/2$:

$$\omega = \gamma\omega'$$

This new effect was first measured in the, now, classic Ives-Stilwell experiment 1938.

2.5.1 Light rays and taking photos

There is one more peculiarity in special relativity in connection with waves and light propagation. Describing waves with the wave null vector in the section above resulted in a rich harvest of known and even one unknown effect and it all works rather well. However, when it comes to measuring light as it received by an observer on a camera a subtle challenge arises. This is best explored in an exercise (another previous exam question) which emphasises the importance of precision in the concepts when analysing a system.

A distant camera snaps a photograph of a bullet with length b in its rest frame and velocity $\underline{v}$ relative to the camera. The direction to the camera is at an angle α from the direction of motion of the bullet. Behind the bullet and parallel to its path is a metre rule, at rest with respect to the camera. The task would be to calculate the length of the bullet as seen in the photo, i.e. how much of the the metre rule is hidden. The crucial concept to take away from this is the following: Considering a

Figure 2.3: Sketch of taking a camera picture of a passing bullet. Insert the drawing during the lecture.

Lorentz contraction (length contraction), the measurement of the start and end of a rod have to be taken simultaneously. Transferred to this problem, photons from the start and end of the bullet would have to be **emitted** simultaneously in the lab-frame (the camera frame - sits at rest and observes the bullet passing by).

For a photo, however, the photons have to be **received** simultaneously, not emitted.

From the sketch, read off that photon 2 travels an extra-distance $b' \cos\alpha$, resulting in an **earlier** emission time

$$t_1' - t_2' = -\frac{b' \cos\alpha}{c}$$

Insert into the LT to get

$$b = \gamma \left(b' + \frac{v}{c} \left(-b' \cos\alpha \right) \right)$$

hence

$$b' = \frac{b}{\gamma \left(1 - \beta \cos\alpha \right)}$$

2.5.2 Exercises

(Collection sheet [2,12-15])

1. In its rest frame, a source emits light in a conical beam of width ±45 degrees. In a frame moving towards the source at speed v, the beam width is ±30 degrees. What is the speed v? Solve using (a) the Lorentz transformations directly and (b) the relativistic velocity transformation formula.

2. Consider the aberration of light formula:

$$\cos\hat{\theta} = \frac{\cos\theta + \beta}{1 + \beta\cos\theta}.$$

   - Given a radioactive source which emits neutrinos isotropically, how could you produce a narrow beam of neutrinos?
   - Calculate the value of β in the aberration formula if the task is to hit a neutrino detector of radius 10 m in France with half the flux of a neutrino beam made from a source at CERN at a distance of 1300 m.

3. For a Quasar with a redshift of $z = \frac{\lambda - \lambda_0}{\lambda_0} = 4$, calculate how fast the Quasar moves relative to Earth at the time of emission of its light.

4. A rectangle with sides a'_0 and b'_0 in x', y' directions, respectively, moves with speed v parallel to the x-axis of an observer at rest. The observer at rest takes a camera picture of the moving rectangle which shall have lights positioned at each of its corners. Calculate the shape of the rectangle on the picture, i.e. its apparent side lengths a_0 and b_0 and draw a sketch of the rectangle according to how it would appear in the picture.

2.6 Linear acceleration in special relativity

Acceleration is the one kinematic concept blatantly absent so far when considering that we transfer Newtonian mechanics to relativistic mechanics. For some reason, acceleration is occasionally perceived by students as a concept outside of special relativity, something that can't be dealt with in a relativistic theory unless one applies the general relativistic theory. Nothing could be further from the truth, in short, total nonsense.

Consider motion in x-direction only (for now). We met the four-velocity already:

$$u^\mu = \begin{pmatrix} \gamma\, c \\ \gamma\, v^x \end{pmatrix}$$

where we now count the index only from 0 to 1 and suppress the y and z components. Let's introduce an acceleration (four-)vector

$$a^\mu = \begin{pmatrix} a^0 \\ a^1 \end{pmatrix}$$

where we simply assume that we have a fixed reference frame since we have explicit components. Choose the rest frame of an object under consideration. Now simplify even further and assume constant acceleration, g, as an example. This is useful preparation since we will need a momentary rest frame in order to be able to differentiate (see the section on the four-velocity, 2.2).

Get

$$\begin{pmatrix} a^0 \\ a^1 \end{pmatrix} = \left.\begin{pmatrix} \frac{du^0}{d\tau} \\ \frac{du^1}{d\tau} \end{pmatrix}\right|_{v^x=0} = \begin{pmatrix} 0 \\ \frac{dv^x}{d\tau} \end{pmatrix} \equiv \begin{pmatrix} 0 \\ g \end{pmatrix}$$

The first equality assumes that the rest frame is taken at the moment corresponding to $v^x = 0$, i.e. some well defined point on the world-line. No need to confuse the differential at that moment with the value itself. A function can very well have a zero value and still have a finite tangent at that point.

Ok, the first test for any proposed (four-)vector is always to build the invariant, i.e. to square it. That should then be valid in all frames.

$$a^2 = a_\mu a^\mu = \eta_{\mu\nu} a^\nu a^\mu = a^0 a^0 - a^1 a^1 = -g^2$$

So that is not too bad. All in all, from kinematics we can summarise 3 equations which are all valid in all inertial reference frames:

1.
$$u^\mu u_\mu = c^2$$

2.
$$a^\mu u_\mu = 0$$

3.
$$a^\mu a_\mu = -g^2$$

This now helps to gain further insight into the four-acceleration. The quickest way to derive the second equation follows from the first relation when differentiated:

$$\frac{d\boldsymbol{u}^2}{d\tau} = 0 = \frac{d}{d\tau}\left(\eta_{\mu\nu} u^\mu u^\nu\right)$$
$$= \eta_{\mu\nu}\left(u^\mu a^\nu + a^\mu u^\nu\right) = 2u^\mu a_\mu = 0$$

Let's put those three equations to work for us then. Eliminate a^0 from the second and third equation and u^1 with equation 1 and get step-by-step, i.e. from (2):

$$a_0 = -a_1 \frac{u^1}{u^0}$$

into (3)

$$a_1^2 \left(1 - \frac{(u^1)^2}{(u^0)^2}\right) = g^2$$

multiply through by $(u^0)^2$ and using (1) then gives

$$(a^1)^2 = \frac{g^2}{c^2}\,(u^0)^2$$

let's take the acceleration pointing in the positive x-direction and get

$$a^1 = \frac{du^1}{d\tau} = \frac{g}{c}\,u^0$$

For a^0 we get similarly

$$a^0 = \frac{du^0}{d\tau} = \frac{g}{c}\,u^1$$

That looks very much like two coupled, linear differential equations and hence we know the solution. Solve for u^0 and u^1 using (take the rest frame as the easiest option)

$$\tau = 0; \quad u^0 = c; \quad u^1 = 0$$

using the decoupling trick of differentiating each equation like so

$$\frac{da^0}{d\tau} = \frac{d^2(u^0)}{d\tau^2} = \frac{g}{c}\,\frac{du^1}{d\tau} = \frac{g^2}{c^2}\,u^0$$

and get

$$u^0 = c \cosh\left(\frac{g\,\tau}{c}\right)$$
$$u^1 = c \sinh\left(\frac{g\,\tau}{c}\right)$$

For the last step, solve for the coordinates with the initial conditions

$$t = 0; \quad x = 0; \quad \text{at } \tau = 0$$

using

$$u^0 = c\,\frac{dt}{d\tau}; \quad u^1 = \frac{dx}{d\tau}$$

and get

$$t = \frac{c}{g}\,\sinh\left(\frac{g\,\tau}{c}\right)$$
$$x = \frac{c^2}{g}\,\left(\cosh\left(\frac{g\,\tau}{c}\right) - 1\right)$$

This could look a bit cleaner if we choose

$$\tau = 0; \quad x = \frac{c^2}{g}$$

since then it is

$$x = \frac{c^2}{g}\,\cosh\left(\frac{g\,\tau}{c}\right)$$

Having these expressions for the coordinates makes it immediately clear how constant acceleration looks like in special relativity:

$$x^2 - c^2 t^2 = \frac{c^4}{g^2} = \text{constant}$$

which is nothing other than the equation for a hyperbola. That means the world-line in a space-time for a body accelerating uniformly is a hyperbola. This is fundamentally different to Newton mechanics! If you recall your kinematics from elementary school, constant acceleration for Newton's theory results in a parabola. Let's have a brief pause here and list some observations:

- The acceleration is not bound from the top, i.e. we can have $g \to \infty$ which would lead to $x = c\,t$.
- An object in region I can send no signals to regions III and IV, ever. These regions are called 'causally disconnected'.

Figure 2.4: Space-time diagram of a constant acceleration world-line including the 4 special wedges of uniform acceleration movement. Insert the drawing during the lecture.

- There are causally disconnected regions, i.e. there are coordinate horizons in special relativity! You could consider them as 'poor man's' black holes! Quantum theory in such a space-time (Unruh radiation) was the precursor to attempts on quantising proper black holes (Bekenstein-Hawking radiation).
- Taking any momentary rest frame along the hyperbola, i.e. any point along that world-line, then at each such point a set of basis vectors can be defined. Such basis vectors (we dealt with the four-velocity as the time-like basis vector when normalised) would here be **proper–time dependent**. This is our first example of such a non-trivial set of basis vectors and opens the possibility to do physics in such a non-trivial basis.

2.6.1 Kinematics in coordinate language

This now summarises all we have seen about relativistic kinematics in the most basic representation, i.e. as coordinates in a fixed, given reference frame. For some of you this will be a life-line if you really don't like the more compact notation so far, for others this may look terribly cumbersome and obtrusive.

Let's state this explicitly: Assume a reference frame as for LT2, i.e. assume you are at rest in S', then transform to S where S' appears to move relative to S with $\underline{v}$. Consider the differential Lorentz transformations and remember that $\underline{v}$ is independent of the coordinate system variables, i.e. NO Differentiation.

$$\begin{aligned} dt &= \gamma(v)\left(dt' + \frac{v}{c^2}dx'\right) \\ dx &= \gamma(v)\left(dx' + v\,dt'\right) \\ dy &= dy' \quad ; \quad dz = dz' \end{aligned}$$

Then we can build

$$\begin{aligned}
u_x = \frac{dx}{dt} &= \frac{dx' + v dt'}{dt' + \frac{v}{c^2} dx'} \\
&= \frac{u'_x + v}{1 + \frac{vu'_x}{c^2}} \\
u_y = \frac{dy}{dt} &= \frac{u'_y}{\gamma(v)\left(1 + \frac{vu'_x}{c^2}\right)} \\
u_z = \frac{dz}{dt} &= \frac{u'_z}{\gamma(v)\left(1 + \frac{vu'_x}{c^2}\right)}
\end{aligned} \tag{2.6}$$

Finally, get for the acceleration

$$a_x = \frac{du_x}{dt} = \frac{\frac{du_x}{dt'}}{\frac{dt}{dt'}}$$

and use

$$\frac{dt}{dt'} = \gamma(v)\left(1 + \frac{vu'_x}{c^2}\right)$$

and

$$\begin{aligned}
\frac{du_x}{dt'} &= \frac{a'_x + a'_x \frac{vu'_x}{c^2} - a'_x \frac{vu'_x}{c^2} - a'_x \frac{v^2}{c^2}}{\left(1 + \frac{vu'_x}{c^2}\right)^2} \\
&= \frac{a'_x\left(1 - \frac{v^2}{c^2}\right)}{\left(1 + \frac{vu'_x}{c^2}\right)^2} \\
&= \frac{a'_x}{\gamma^2(v)\left(1 + \frac{vu'_x}{c^2}\right)^2}
\end{aligned} \tag{2.7}$$

hence

$$a_x = \frac{a'_x}{\gamma^3(v)\left(1 + \frac{vu'_x}{c^2}\right)^3}$$

For $a_{y,z}$ get

$$a_{y,z} = \frac{1}{\gamma^2(v)\left(1 + \frac{vu'_x}{c^2}\right)^2}\left[a'_{y,z} - \frac{\frac{vu'_x}{c^2} a'_x}{1 + \frac{vu'_x}{c^2}}\right]$$

2.6.2 Rindler coordinates

Ok, so this is a bit of advanced special relativity (finally). Whilst the preceding section showed Lorentz transformed velocity and acceleration components in minute detail, corresponding to the First-Year Lorentz transformations of coordinates, we also saw acceleration defined locally as a four-vector. Any local definition using coordinate independent objects is automatically the most general definition and always applicable, i.e. well defined. As you might see in the general relativity lecture, local definitions are in fact the only ones remaining as soon as general curved spaces are permitted.

The flat spacetime of Minkowski, however, offers yet a third possibility how to deal with a general concept such as a time-dependent basis, required to define accelerations and yet have the convenience of Euclidean space where a strictly local definition of everything is not required. It is possible to map the (almost) entire Minkowski spacetime with a global coordinate system such as it is done with Cartesian coordinates in Euclidean space.

This is quite a big leap since now one could transfer all the physics of simple spaces to a complicated spacetime as experienced by an accelerated observer, say a particle or a clock or an observer in a rocket, simply by a single(!) transformation, i.e. from one coordinate system to another. A bit like going from Cartesian to Polar coordinates. That would be quite something and it was achieved back in the 60's. The coordinates corresponding to a uniformly accelerated observer are called Rindler coordinates.

In the following, we have a short look at them and point out some features. The real strength of this description is realised when actually describing physics in such coordinates. That, unfortunately, would require another long lecture course altogether (which should then also feature coordinate system describing rotating reference frames - that's when the real fun starts). Physics in such time-dependent basis' can look rather strange and yet they are far more realistic then the static Minkowski spacetime (we are constantly accelerated quite uniformly on Earth due to local gravity and rotating too).

Rindler coordinates describe one global reference frame for **all** observers in hyperbolic motion (that is another way of saying uniform acceleration, see above, since such motion results in a hyperbola as world-line). Assume one-dimensional accelerated motion in x, then the transformations are quite simple:

$$\begin{aligned} x^0 &= X\sinh T \\ x^1 &= X\cosh T \\ x^2 &= Y \quad ; \quad x^3 = Z \end{aligned}$$

and correspondingly

$$X^2 = (x^1)^2 - (x^0)^2$$

for hyperbolic motion. The capital letters are the Rindler coordinates. The transformation matrix looks like

$$A^\nu_\mu = \begin{pmatrix} \frac{\partial x^0}{\partial T} & \frac{\partial x^0}{\partial X} \\ \frac{\partial x^1}{\partial T} & \frac{\partial x^1}{\partial X} \end{pmatrix} = \begin{pmatrix} X\cosh T & \sinh T \\ X\sinh T & \cosh T \end{pmatrix}$$

which gives us a metric (good exercise to calculate metric tensor coefficients, given a transformation)

$$\boldsymbol{\eta}^{\text{Rindler}} = \boldsymbol{A}^T \boldsymbol{\eta}\, \boldsymbol{A}$$

which is

$$\begin{aligned} \boldsymbol{\eta}^{\text{Rindler}} &= \begin{pmatrix} X\cosh T & X\sinh T \\ \sinh T & \cosh T \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} X\cosh T & \sinh T \\ X\sinh T & \cosh T \end{pmatrix} \\ &= \begin{pmatrix} X^2\cosh^2 T - X^2\sinh^2 T & X\cosh T\sinh T - X\cosh T\sinh T \\ X\sinh T\cosh T - X\sinh T\cosh T & \sinh^2 T - \cosh^2 T \end{pmatrix} \\ &= \begin{pmatrix} X^2 & 0 \\ 0 & -1 \end{pmatrix} \end{aligned}$$

hence the line element in Rindler coordinates reads

$$ds^2 = X^2\, dT^2 - dX^2 - dY^2 - dZ^2$$

However, this line element is not defined everywhere! The Rindler coordinates are not quite fully global. They are only valid in slices or specific regions of Minkowski

Figure 2.5: Space-time diagram of a constant acceleration world-line including the 4 special wedges of uniform acceleration movement and their relation to Rindler coordinates. Insert the drawing during the lecture.

spacetime. The line element above is valid in regions I and III (see figure).

So much for the coordinates as such. No let's have a closer look at instantaneous rest frames (or momentary rest frames, same thing). Suppose you define one for a single object in sector I. Then you can always define an orthonormal set of basis vectors at that point of the world-line. Such a set is also called a **tetrad**.

At a fixed moment of time, τ_0, a tetrad can always be built as

$$(\boldsymbol{e}_0)^\mu = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad ; \quad (\boldsymbol{e}_1)^\mu = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \quad ; \ldots$$

That is we choose $\boldsymbol{e}_0 = 1/c\,\boldsymbol{u}(\tau_0)$ using the local vector definition as derivative along a curve at τ_0. But here the basis vectors have to be time dependent along the world-line: $\boldsymbol{e}_0(\tau)$, $\boldsymbol{e}_1(\tau)$, etc. Now, this is always possible, no problem, as mentioned before. The trick is, however, to see whether it is possible to get a general, globally (or thereabout) valid set of basis vectors given a complex spacetime such as for a uniformly accelerated observer. Often that is indeed possible and the argument works as follows.

In order to define the general set of basis vectors, choose another (second) instantaneous rest frame at $\tau \neq \tau_0$. Again, set (hat's for transformed coordinates, basis vectors etc.)

$$(\boldsymbol{e}_{\hat{0}}(\tau))^\mu = \frac{1}{c}\,(u(\tau))^\mu = \begin{pmatrix} \cosh\left(\frac{g\tau}{c}\right) \\ \sinh\left(\frac{g\tau}{c}\right) \\ 0 \\ 0 \end{pmatrix}$$

where the Rindler transformations have been applied. For $\boldsymbol{e}_{\hat{1}}(\tau)$ get (orthogonal to $\boldsymbol{e}_{\hat{0}}(\tau)$!)

$$\boldsymbol{e}_{\hat{1}}(\tau) = \frac{1}{g}\,\boldsymbol{a}(\tau)$$

from the properties of the acceleration four-vector with respect to the four-velocity, i.e. they are always orthogonal. Hence

$$(\boldsymbol{e}_{\hat{1}}(\tau))^\mu = \begin{pmatrix} \sinh\left(\frac{g\tau}{c}\right) \\ \cosh\left(\frac{g\tau}{c}\right) \\ 0 \\ 0 \end{pmatrix}$$

We can write this now as a proper transformation of basis vectors

$$
\begin{aligned}
(\boldsymbol{e}_{\hat{0}}(\tau))^\mu &= (\boldsymbol{e}_0(\tau_0))^\mu \cosh\left(\frac{g\tau}{c}\right) + (\boldsymbol{e}_1(\tau_0))^\mu \sinh\left(\frac{g\tau}{c}\right) \\
(\boldsymbol{e}_{\hat{1}}(\tau))^\mu &= (\boldsymbol{e}_0(\tau_0))^\mu \sinh\left(\frac{g\tau}{c}\right) + (\boldsymbol{e}_1(\tau_0))^\mu \cosh\left(\frac{g\tau}{c}\right) \\
(\boldsymbol{e}_{\hat{2}}(\tau))^\mu &= (\boldsymbol{e}_2(\tau_0))^\mu \\
(\boldsymbol{e}_{\hat{3}}(\tau))^\mu &= (\boldsymbol{e}_3(\tau_0))^\mu
\end{aligned}
$$

which is nothing else than a Lorentz transformation between instantaneous rest frames. It's easier to see when identifying

$$\gamma = \cosh\left(\frac{g\tau}{c}\right) \quad ; \quad \beta\gamma = \sinh\left(\frac{g\tau}{c}\right)$$

and $g\tau/c$ would be called rapidity parameter for a Lorentz boost. You can see now that any tetrad transported in this way is purely boosted, which implies that these transformations represent a non-rotating transport of the tetrad. It can be written more concisely as

$$
\begin{aligned}
\frac{d}{d\tau}\boldsymbol{e}_0(\tau) &= \frac{g}{c}\boldsymbol{e}_1(\tau) \\
\frac{d}{d\tau}\boldsymbol{e}_1(\tau) &= \frac{g}{c}\boldsymbol{e}_0(\tau)
\end{aligned}
$$

A (even) more general way to describe this transport is

$$c^2\frac{d}{d\tau}(\boldsymbol{e}_\mu(\tau))^\rho = (a^\rho u^\sigma - u^\rho a^\sigma)(\boldsymbol{e}_\mu(\tau))_\sigma \tag{2.8}$$

which is called **Fermi-Walker transport** of a tetrad.

The terms in the bracket made of acceleration and velocity four-vectors builds a very general anti-symmetric tensor which fully describes Lorentz boosts. Quick test using our previous calculation of $\boldsymbol{u}$ and $\boldsymbol{a}$:

$$c^2\frac{d}{d\tau}(\boldsymbol{e}_0(\tau))^\rho = (g(\boldsymbol{e}_1)^\rho\, c(\boldsymbol{e}_0)^\sigma - c(\boldsymbol{e}_0)^\rho\, g(\boldsymbol{e}_1)^\sigma)\ (\boldsymbol{e}_0(\tau))_\sigma = g\,c\,(\boldsymbol{e}_1)^\rho$$

The last information to take away from this discussion is that Fermi-Walker transport is valid for **any** accelerated motion, not only uniform acceleration. That should not come as a surprise since we used the most general local basis definitions and merely compared two of those sets to each other. As soon as you come to general relativity though, it is this very comparison which will not work anymore due to the curvature of spacetime and that is where the machinery of general relativity has to kick in.

2.6.3 Exercises

(Collection sheet [2,16-19])

1. Show that an observer's four-acceleration has only 3 independent components and write the magnitude of the acceleration measured in the observer reference frame as an invariant.

2. Given a position four-vector x^μ, the corresponding four-velocity $u^\mu = \frac{dx^\mu}{d\tau}$ and four-acceleration $a^\mu = \frac{du^\mu}{d\tau}$, show that

$$a^\mu = \gamma^4 \left(\frac{\underline{v}\,\underline{a}}{c}, \left(\frac{\underline{v}\,\underline{a}}{c} \right) \frac{\underline{v}}{c} + \gamma^{-2} \underline{a} \right)$$

   Use the relation $\gamma = \frac{dt}{d\tau}$.

3. Consider a space ship moving with constant acceleration g in the x-direction. It starts at a point $x = 0$ from rest. Calculate the distance d travelled when the ship reaches a specific speed v as measured from the initial rest frame.

4. Given a world-line in Cartesian coordinates, $x = R\cos(\omega\, t)$, $y = R\sin(\omega\, t)$ and $z = 0$ with R and ω constant for a particle moving in a circle. Then with $v = R\,\omega$ and $\gamma = \left(1 - \frac{R^2\omega^2}{c^2}\right)^{-1/2}$:

   - Calculate the components of the four-velocity along the world-line.
   - Calculate the components of the four-acceleration and its invariant magnitude and comment on the relativistic modifications to the classical value of $a = R\,\omega^2$.

Chapter 3

Applications: Electromagnetism

Applying special relativity to classical electromagnetism essentially reduces to re-writing the Maxwell equations and clarifying how existing physics concepts carry over to a relativistically invariant description, i.e. finding appropriate four-vectors and tensors for known quantities in the field. As a consequence some calculations and insights will become much simpler and clearer, others might not.

Let's remind ourselves first on what this is all about in the shortest possible way, i.e. quoting the Maxwell equations in hopefully familiar form.

$$\begin{aligned} \operatorname{div} \underline{E} &= \frac{\rho}{\epsilon_0} && (3.1)\\ \operatorname{div} \underline{B} &= 0 && (3.2)\\ \operatorname{curl} \underline{E} &= -\frac{\partial \underline{B}}{\partial t} && (3.3)\\ \operatorname{curl} \underline{B} &= \mu_0 \left(\underline{j} + \epsilon_0 \frac{\partial \underline{E}}{\partial t} \right) && (3.4) \end{aligned}$$

and additionally there is the conservation law

$$\operatorname{div} \underline{j} + \frac{\partial \rho}{\partial t} = 0 \qquad (3.5)$$

where ϵ_0 is the permittivity, μ_0 the permeability, ρ the charge density, $\underline{j}$ the current density and finally the equation of motion (Lorentz force law) is

$$\frac{d\underline{p}}{dt} = q\,(\underline{E} + \underline{v} \times \underline{B}) \qquad (3.6)$$

One major aim of this chapter is to guide you to the following representation of the identical set of equations. We will have to tackle quite a few new things on the way as might be obvious from just glancing at these below.

$$\partial_\mu F^{\mu\nu} = \sqrt{\frac{\mu_0}{\epsilon_0}}\, j^\nu \tag{3.7}$$

$$\partial_\mu (*F^{\mu\nu}) = 0 \tag{3.8}$$

$$\partial_\mu j^\mu = 0 \tag{3.9}$$

and the equation of motion (Lorentz force law):

$$\frac{dp^\mu}{d\tau} = \frac{q}{c}\, F^{\mu\nu}\, u_\nu \tag{3.10}$$

That all looks rather more compact but also a little mysterious. First we will have to have a reminder on differential operators and how to make them Lorentz-invariant.

3.1 Reminder on differential operators

Let's repeat the four-vector definition in a concise form first: Any set of 4 numbers is called a Lorentz invariant four-vector if it transforms under Lorentz transformations as

$$A'^\mu = \frac{\partial x'^\mu}{\partial x^\nu}\, A^\nu \tag{3.11}$$

and a 1-form transforms as

$$B'_\mu = \frac{\partial x^\nu}{\partial x'^\mu}\, B_\nu \tag{3.12}$$

Now, the four-gradient with respect to variables which form a four-vector transforms as a 1-form (and vice versa):

$$\frac{\partial}{\partial x'^\mu} = \frac{\partial x^\nu}{\partial x'^\mu}\, \frac{\partial}{\partial x^\nu} \equiv \partial'_\mu \tag{3.13}$$

and

$$\partial'^\mu = \eta^{\mu\nu}\, \partial'_\nu$$

This results in a coordinate representation

$$\partial^\mu = \frac{\partial}{\partial x_\mu} = \begin{pmatrix} \frac{\partial}{\partial x_0} \\ -\underline{\nabla} \end{pmatrix}$$

and

$$\partial_\mu = \frac{\partial}{\partial x^\mu} = \begin{pmatrix} \frac{\partial}{\partial x^0} \\ \underline{\nabla} \end{pmatrix}$$

The four-divergence then is the result of a contraction of the gradient with a vector (1-form):

$$\partial_\mu A^\mu = \partial^\mu A_\mu = \frac{\partial A^0}{\partial x^0} - \underline{\nabla} \cdot \underline{A} \tag{3.14}$$

The 4-Laplacian operator then is the invariant formed from the gradient when using the self-contraction or 'square it' rule:

$$\partial_\mu \partial^\mu = \frac{\partial^2}{\partial (x^0)^2} - (\underline{\nabla})^2 \equiv \Box \tag{3.15}$$

As a short remark, it might be useful to remind you of the notation here according to which $x^0 = ct$ in order to get all the c numbers and therefore units correct. Lastly, there is the four-dimensional volume element to consider, particularly when defining integrals. Fortunately, the volume element

$$d^4x = d(x^0) d^3(\underline{x})$$

is a Lorentz-invariant. Here is the quick answer to a simple 'show ...' question:

$$d^4x' = \frac{\partial(x'^0, x'^1, x'^2, x'^3)}{\partial(x^0, x^1, x^2, x^3)}\, d^4x$$

where the long expression for the fraction represents the Jacobian for coordinate transformations. In case of Lorentz transformations as coordinate transformations, the Jacobian is

$$\frac{\partial(x'^0, x'^1, x'^2, x'^3)}{\partial(x^0, x^1, x^2, x^3)} = \det(L) = 1$$

by definition, hence the transformed volume element is identical to the previous volume element, i.e. invariant.

3.2 Back to the Maxwell equations

The homogeneous equations

$$\begin{aligned} \operatorname{div} \underline{B} &= 0 \\ \operatorname{curl} \underline{E} &= -\frac{\partial \underline{B}}{\partial t} \end{aligned}$$

are solved simultaneously by introducing the potentials ϕ and $\underline{A}$ according to

$$\begin{aligned} \underline{E} &= -\underline{\nabla}\, \phi - \frac{\partial \underline{A}}{\partial t} \\ \underline{B} &= \operatorname{curl} \underline{A} \end{aligned} \tag{3.16}$$

The potentials still have one degree of freedom which needs fixing. This procedure leads to a process called gauge transformation:

$$\phi \to \phi - \frac{1}{c}\frac{\partial \Lambda}{\partial t} \quad ; \quad c\underline{A} \to c\underline{A} + \underline{\nabla}\Lambda$$

hence potentials are only true solutions when taking into account a gauge condition, for instance the Lorenz (no 't') condition

$$\operatorname{div} c\underline{A} + \frac{1}{c}\frac{\partial \phi}{\partial t} = 0 \tag{3.17}$$

Now for the inhomogeneous equations, using the potentials and the expressions eqns. 3.16 for $\underline{E}$ and $\underline{B}$, get:

$$\Box\phi = \frac{\rho}{\epsilon_0} \quad ; \quad \Box\, c\underline{A} = \sqrt{\frac{\mu_0}{\epsilon_0}}\,\underline{j}$$

where 'this' $\Box$ is the three-dimensional Laplacian operator. Now using that the Laplacian in four dimensions is a Lorentz invariant operator suggests writing the expressions from above as

$$\Box \begin{pmatrix} \phi \\ c\underline{A} \end{pmatrix} = \sqrt{\frac{\mu_0}{\epsilon_0}} \begin{pmatrix} c \cdot \rho \\ \underline{j} \end{pmatrix}$$

(Reminder: $c = 1/\sqrt{\mu_0\epsilon_0}$ and $x^0 = ct$) Then the four-potential

$$A^\mu = \begin{pmatrix} \phi \\ c\underline{A} \end{pmatrix}$$

is a four-vector if the four-current

$$j^\mu = \begin{pmatrix} c \cdot \rho \\ \underline{j} \end{pmatrix}$$

is a four-vector. So, let's build a simple model and test this assertion. Assume there is a charge e moving on a world-line or trajectory $z^\mu(\tau)$. Then one gets for the total charge and current along the trajectory:

$$\begin{aligned} \rho(x^\mu) &= e \cdot \int \delta(x^0 - z^0)\, \delta^3(\underline{x} - \underline{z}) dz^0 \\ &= e \cdot \int \delta^4(x^\mu - z^\mu)\, d\tau\, \frac{dz^0}{d\tau} \end{aligned}$$

and

$$\underline{j}(x^\mu) = e \cdot \int \delta^4(x^\mu - z^\mu)\, d\tau\, \frac{d\underline{z}}{d\tau}$$

However, since

$$u^\mu(\tau) = \begin{pmatrix} \frac{dz^0}{d\tau} \\ \frac{d\mathbf{z}}{d\tau} \end{pmatrix}$$

with u^μ the four-velocity of the charge on its world-line, can write

$$j^\mu(x^\nu) = e \cdot \int \delta^4(x^\nu - z^\nu)\, u^\mu(\tau)\, d\tau$$

as a candidate for a current four-vector. A quick check reveals that $d\tau$ is a number, hence invariant, $\delta^4(x^\nu - z^\nu)$, the four-dimensional delta function, is invariant (A point $z^\mu(\tau_0)$ in space-time does not change due to a transformation, only gets re-labelled) and $u^\mu(\tau)$ is a four-vector. Therefore, j^μ is a four-vector, composed only of invariant quantities. Finally, if j^μ is a four-vector then also the four-potential A^μ is a four-vector as discussed previously.

Using the four-potential enables us to quickly write the gauge transformation in invariant form as well as the Lorenz condition:

$$A_\mu \to A_\mu - \partial_\mu \Lambda$$

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

Likewise we can now write the charge continuity equation in invariant form

$$\partial_\mu\, j^\mu = 0$$

which is the first equation from the list of strangely written Maxwell equations at the start of this section. Finally, to round things off, we can also write the wave equation in invariant, four-vector, form

$$\Box\, A^\mu = \sqrt{\frac{\mu_0}{\epsilon_0}}\, j^\mu$$

3.3 From potentials to fields

Once the potentials are present, the next step to deal with fields is fairly straightforward, in principle. All that is required is to come up with a Lorentz invariant concept capturing all the physics of electromagnetic fields. Fortunately, there is such a suggestive connection in component form when looking at the original Maxwell equations and the newly found four-potential. In detail:

$$E_i = -\partial_i\, A_0 + \partial_0\, A_i$$

and

$$B_i = -(\text{curl}\underline{A})_i$$

such as[1]

$$\begin{aligned} B_1 &= -\partial_2 A_3 + \partial_3 A_2 \\ B_2 &= -\partial_3 A_1 + \partial_1 A_3 \\ B_3 &= -\partial_1 A_2 + \partial_2 A_1 \end{aligned}$$

Putting all together in the candidate for a proper field description, the field-strength tensor (Faraday tensor) can be defined as:

$$F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu = -F_{\nu\mu} \tag{3.18}$$

This tensor is antisymmetric by construction! When you interchange the indices, the sign changes. In component form it is:

$$F_{\mu\nu} = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & -cB_3 & cB_2 \\ -E_2 & cB_3 & 0 & -cB_1 \\ -E_3 & -cB_2 & cB_1 & 0 \end{pmatrix} \tag{3.19}$$

Alternatively get

$$F^{\mu\nu} = \eta^{\mu\rho}\eta^{\nu\sigma} F_{\rho\sigma} = \begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & -cB_3 & cB_2 \\ E_2 & cB_3 & 0 & -cB_1 \\ E_3 & -cB_2 & cB_1 & 0 \end{pmatrix} \tag{3.20}$$

One nice consistency test consists of applying the four-gradient on the Faraday tensor:

$$\partial_\nu F^{\mu\nu} = \partial_\nu \partial^\mu A^\nu - \partial_\nu \partial^\nu A^\mu = -\Box A^\mu$$

where the Lorenz condition $\partial_\nu A^\nu = 0$ was used. Finally, applying the wave equation from before, we end up seeing the first of the Maxwell equations proper, the inhomogeneous Maxwell equation in invariant form:

$$\partial_\nu F^{\mu\nu} = -\sqrt{\frac{\mu_0}{\epsilon_0}}\, j^\mu \tag{3.21}$$

[1] As usual, indices on three-vectors(!) are really only labels. Their position, up or down, has no meaning other than labelling components. Why? They are purely spatial components from the point of view of space-time, i.e. all they see is a Euclidean space where the distinction between vectors and one-forms does not exist.

In order to finish this section properly, I will at least have to quote the homogeneous Maxwell equations. However, we are not quite in a position yet, mathematically, to motivate them as such. Suffices to say at this stage that the homogeneous Maxwell equations reduce to four Jacobi identities:

$$\partial_\alpha F_{\mu\nu} + \partial_\mu F_{\nu\alpha} + \partial_\nu F_{\alpha\mu} = 0$$

For explanations, see later.

3.3.1 Exercises

(Collection sheet [3,1-4])

1. Show that the Maxwell equations
$$\underline{\nabla} \cdot \underline{E} = \frac{\rho}{\epsilon_0}$$
$$\underline{\nabla} \times \underline{B} = \mu_0 \, \underline{j} + \mu_0 \epsilon_0 \frac{\partial \underline{E}}{\partial t}$$
result from the Maxwell equation
$$\partial_\mu F^{\mu\nu} = \sqrt{\frac{\mu_0}{\epsilon_0}} j^\nu,$$
where for the four-vector current take
$$j^\nu = \begin{pmatrix} c\,\rho \\ \underline{j} \end{pmatrix}$$

2. Show that the Faraday tensor satisfies the identity $\partial_\alpha F_{\beta\gamma} + \partial_\beta F_{\gamma\alpha} + \partial_\gamma F_{\alpha\beta} = 0$ and if $\alpha = 1$, $\beta = 2$, $\gamma = 3$, the Maxwell equation $\underline{\nabla} \cdot \underline{B} = 0$ can be derived.

3. Using the relativistic, inhomogeneous Maxwell equation and the fact that the Faraday tensor is anti-symmetric, derive the continuity equation $\partial_\mu j^\mu = 0$.

4. Show that if $\mathbf{N}$ is an eigenvector of the Faraday tensor $\mathbf{F}$ with a non-zero eigenvalue s, that is $F_{\mu\nu} N^\nu = s\, N_\mu$, then $\mathbf{N}$ is a null vector.

   (Collection sheet [1,14])

5. Show that a contraction of a vector $\mathbf{V}$ with the 'projection tensor' $P^{\alpha\beta} = \eta^{\alpha\beta} - u^\alpha u^\beta$ projects $\mathbf{V}$ into a 3-surface orthogonal to the normalised ($|\mathbf{u}|^2 = 1$) four-velocity vector $\mathbf{u}$. One application of the projection tensor is the translation of Ohm's law into a Lorentz invariant expression: Write Ohm's law $\underline{j} = \sigma \underline{E}$, with three-vectors $\underline{j}$ and $\underline{E}$ and scalar σ invariantly in terms of the four-vector current $\mathbf{j}$, the Faraday tensor $\mathbf{F}$, σ and the four-velocity $\mathbf{u}$ (the velocity of the conducting element).

3.4 Particle Dynamics

This section answers the question on how a charge moves in a given, external, field. This will lead to the relativistic Lorentz force expression. Take

$$p^\mu = m\, u^\mu,$$

i.e. the expression for the four-momentum as a starting point. Then educated guessing, using $dt = \gamma\, d\tau$, might lead us to a re-write of the non-relativistic Lorentz force according to

$$\begin{aligned}\frac{d\underline{p}}{d\tau} &= q\,\gamma\,(\underline{E} + \underline{v} \times \underline{B}) \\ &= \frac{q}{c}\,(u^0 \underline{E} + \underline{u} \times (c\underline{B}))\end{aligned}$$

Then we would still need a 0-component in order to build a four-vector. This 0-component would have to have the identical dimension as the momentum four-vector, hence a candidate could be the rate of energy change if we use the speed factor c wisely. In detail, let's write

$$\frac{dE}{dt} = q(\underline{E} + \underline{v} \times \underline{B}) \cdot \underline{v} = q\,\underline{E}\,\underline{v}$$

Therefore, a good guess for the appropriate four-vector 0-component could be

$$\frac{dp^0}{d\tau} = \frac{q}{c}\,\underline{E}\,\underline{u}$$

Let's see how these various guesses can be summarised:

$$\begin{aligned}\frac{dp^\mu}{d\tau} &= \frac{d}{d\tau}\begin{pmatrix} E/c \\ \underline{p} \end{pmatrix} \\ &= \frac{q}{c}\begin{pmatrix} 0 & E_x & E_y & E_z \\ E_x & 0 & cB_z & -cB_y \\ E_y & -cB_z & 0 & cB_x \\ E_z & cB_y & -cB_x & 0 \end{pmatrix} \cdot \begin{pmatrix} u^0 \\ \underline{u} \end{pmatrix} \\ &= \frac{q}{c}\, F^\mu_{\ \nu}\, u^\nu \end{aligned} \tag{3.22}$$

Here the momentum four-vector appears in units of momentum. We can also write it in units of energy which is sometimes more appropriate:

$$\frac{dp^\mu}{d\tau} = \frac{d}{d\tau}\begin{pmatrix} E \\ c\,\underline{p} \end{pmatrix} = q\; F^\mu_{\ \nu}\, u^\nu$$

A term like force times velocity represents a power = energy per time which lends itself to an interpretation of the expression above as work done per time on a charge q by the electromagnetic field. Note that nothing in this derivation suggests that the reverse, i.e the influence of that very charge back onto the field enters anywhere and rightly so. The treatment of a charge backreaction on a field is far more complex than the above and is explicitly not covered by the expression (one of the hidden assumptions about the 'external' field).

The argument here rather expresses a different aspect, i.e. the charge here acts like a test particle in mechanics. It's being used to examine the field, i.e learn something about the field. This is the essence of the Lorentz force law. It's all about examining fields. This suggests that it might be possible to derive conservation laws for fields from this starting point (like in mechanics when examining test particles in some potential - not quite the same but you get the idea).

3.4.1 Exercises

(Collection sheet [3,5-6])

1. A particle with charge q and mass m travels through a laboratory with velocity $\underline{v}$ in x-direction (laboratory frame) when it encounters a constant electric field $\underline{E} = (0, E_y, 0)$.
   - State the non-zero components of the Faraday tensor in the laboratory frame.
   - Derive the equations of motion given these initial conditions from the Lorentz force law (one equation for each component in the laboratory frame).
2. Show that the Lorentz force law $\frac{dp^\mu}{d\tau} = q\, F^\mu_{\ \nu}\, u^\nu$ contains the non-relativistic Lorentz force law $d\underline{p}/dt = q\,(\underline{E} + \underline{v} \times \underline{B})$ by explicitly calculating the components of $\underline{p}$.

3.5 Conservation laws for fields

First of all we need a description of the energy and momentum belonging purely to the electromagnetic fields. We have a relativistic description of electromagnetic fields themselves in the form of the Faraday tensor and we have a dynamic force law, the Lorentz force law, in order to describe the physics of a test particle (charge) in such fields.

However, at this point we require a new concept which is a major ingredient of Maxwell's theory, i.e. that electromagnetic fields contain energy and momentum all on their own. Note that this idea is quite a big step away from the conventional notion of matter 'bits and pieces' or particles, having energy and momentum. Fields are by definition something 'other' than matter and still we can indeed ascribe energy and momentum to them.

Nevertheless, in order to gain some understanding on the energy and momentum content of fields, we can derive a new tensor which will turn out to be specific to fields alone and contain all that information. The starting point of that derivation is the Lorentz force law:

$$\begin{aligned}\frac{dp^\mu}{d\tau} &= \frac{q}{c}\,F^\mu_{\ \nu}\,u^\nu \\ &= \frac{1}{c}\,F^\mu_{\ \nu}\,j^\nu\end{aligned}$$

Note the implicit assumption of a point charge as test charge, i.e. the current and charge density become equivalent to total charge and total current independent of the volume considered. Strictly speaking $q\,u^\nu$ should be replaced by an integral over j^ν times a delta function at the location of the charge. Since this integral is trivial, the integral over j^ν is replaced by the four-vector itself with the implicit assumption that the 'per volume' units have disappeared in what follows. Combined with the inhomogeneous Maxwell equation:

$$\partial_\nu F^{\mu\nu} = -\sqrt{\frac{\mu_0}{\epsilon_0}}\,j^\mu$$

when solved for the current

$$j^\nu = -\sqrt{\frac{\epsilon_0}{\mu_0}}\,\partial_\lambda F^{\nu\lambda}$$

Inserting the current into the force law, we obtain:

$$\frac{dp^\mu}{d\tau} = -\epsilon_0\,F^\mu_{\ \nu}\,\partial_\lambda F^{\nu\lambda}$$

or more conveniently

$$\begin{aligned}\frac{dp_\mu}{d\tau} &= -\epsilon_0\, F_{\mu\nu}\, \partial_\lambda F^{\nu\lambda} \\ &= -\epsilon_0\, \left[\partial_\lambda (F_{\mu\nu} F^{\nu\lambda}) - F^{\nu\lambda}\partial_\lambda F_{\mu\nu}\right]\end{aligned}$$

where the second line follows using the product rule of differentiation. Now that might seem a right clutter stepping from the first to the second line and really rather unnecessary but hold out a little longer. This derivation is anything but easy or straightforward.

In the following we inspect and work purely on the second term in the second line, i.e. the $F^{\nu\lambda}\partial_\lambda F_{\mu\nu}$. The strategy is to bring in the homogeneous Maxwell equations which explains also why they were quoted a bit out of context earlier on. So, hold on, the ride continues. Write:

$$F^{\nu\lambda}\partial_\lambda F_{\mu\nu} = \frac{1}{2}\, F^{\nu\lambda}\partial_\lambda F_{\mu\nu} + \frac{1}{2}\, F^{\lambda\nu}\partial_\nu F_{\mu\lambda}$$

where we split the original expression into two identical halfs but re-label the dummy indices. Note that none of the contracted index pairs change position, merely their label names change. Why do such a thing? Simply because at this stage we can factorise one Faraday tensor term out of the sum like so:

$$F^{\nu\lambda}\partial_\lambda F_{\mu\nu} = \frac{1}{2}\, F^{\nu\lambda}\left(\partial_\lambda F_{\mu\nu} + \partial_\nu F_{\lambda\mu}\right)$$

Note the subtlety in having to exchange index positions twice (one to factor out the Faraday tensor, secondly at the last Faraday tensor as an educated guess) which cancels the negative sign from swapping indices of the Faraday tensor. Looking up the homogeneous Maxwell equations above, one might note that the expression in the bracket is (as intended, obviously) identical to two terms in that three term formula, hence inserting the Maxwell equation simplifies the above to:

$$F^{\nu\lambda}\partial_\lambda F_{\mu\nu} = -\,\frac{1}{2}\, F^{\nu\lambda}\;\partial_\mu F_{\nu\lambda}$$

That now requires a brief but close look at the remaining indices and you can see identical ones. That comes in very handy indeed when applying the product rule again to simplify the expression to:

$$F^{\nu\lambda}\partial_\lambda F_{\mu\nu} = -\,\frac{1}{4}\,\partial_\mu\left(F^{\nu\lambda}\; F_{\nu\lambda}\right)$$

which is the final expression to insert above. We will separately examine this full contraction operation later.

Coming back to the start of the derivation process, we can now insert and see where we are:

$$\frac{dp_\mu}{d\tau} = -\epsilon_0\,\partial_\lambda \left(F_{\mu\nu}F^{\nu\lambda} + \frac{1}{4}\,\delta^\lambda_\mu \left(F^{\sigma\rho}\;F_{\sigma\rho}\right)\right) \tag{3.23}$$

where we could conveniently again factorise the partial differentiation and compensate for the missing free indices with the Kronecker delta in order to keep the validity of the equation. For the full contraction, dummy indices were deliberately chosen such that they obviously belong together and have nothing to do with the other indices in the equation, as they should.

Pause and reflect to see that we now obtained an expression which describes the change of an energy-momentum four-vector in time as a total derivative of a long(ish) expression in brackets. This is a situation well known from continuum mechanics and fluids. Accordingly, the terms in brackets were well motivated to serve as an energy-momentum tensor which describes a continuum state (like a fluid but here it's the electromagnetic fields). Hence we define:

$$T_{\beta\alpha} = \epsilon_0\,\eta_{\lambda\alpha} \left(F_{\beta\nu}F^{\nu\lambda} + \frac{1}{4}\,\delta^\lambda_\beta \left(F^{\sigma\rho}\;F_{\sigma\rho}\right)\right) \tag{3.24}$$

as the energy-momentum tensor. The explicit appearance of the Minkowski metric is not exactly necessary (and often avoided in textbooks) but rather subsumed into the final expression. It is left explicitly in this case since then we can write eqn. 3.23 as

$$\frac{dp_\mu}{d\tau} = -\partial^\nu\,T_{\mu\nu} = \frac{1}{c}\,F_{\mu\lambda}\,j^\lambda$$

As an example how to calculate anything practically for this tensor, let's inspect the 00-component, T_{00}:

$$T_{00} = \epsilon_0\,F_{0\nu}F^\nu_{\;0} + \frac{\epsilon_0}{4}\,\delta^0_0\left(F^{\sigma\rho}\;F_{\sigma\rho}\right)$$

where the last term, the full contraction, really amounts to something you would always have to calculate separately for each and every component along the diagonal (due to the Kronecker delta), hence it's worthwhile to calculate that term once and be done with it. Here is how that works. Clearly, this is merely one way to arrive at the final expression. You might wish to check your favourite textbook for alternative derivations:

$$\begin{aligned}
F_{\sigma\rho}\, F^{\sigma\rho} &= -\mathrm{Trace}\,(F_{\mu\sigma}\, F^{\sigma\nu}) \\
&= -\mathrm{Tr}\left(\eta_{\mu\alpha}F^{\alpha\beta}\eta_{\beta\sigma}F^{\sigma\nu}\right) \\
&= -\mathrm{Tr}\left[\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & -cB_3 & cB_2 \\ E_2 & cB_3 & 0 & -cB_1 \\ E_3 & -cB_2 & cB_1 & 0 \end{pmatrix}\right. \\
&\qquad \left.\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & -cB_3 & cB_2 \\ E_2 & cB_3 & 0 & -cB_1 \\ E_3 & -cB_2 & cB_1 & 0 \end{pmatrix}\right] \\
&= -\mathrm{Tr}\left[\begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ -E_1 & 0 & cB_3 & -cB_2 \\ -E_2 & -cB_3 & 0 & cB_1 \\ -E_3 & cB_2 & -cB_1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ -E_1 & 0 & cB_3 & -cB_2 \\ -E_2 & -cB_3 & 0 & cB_1 \\ -E_3 & cB_2 & -cB_1 & 0 \end{pmatrix}\right] \\
&= -\mathrm{Tr}\left[\mathrm{diag}(E_1^2 + E_2^2 + E_3^2, E_1^2 - c^2B_3^2 - c^2B_2^2,\right. \\
&\qquad \left. E_2^2 - c^2B_3^2 - c^2B_1^2, E_3^2 - c^2B_2^2 - c^2B_1^2)\right] \\
&= -2\,\left(|\underline{E}|^2 - c^2|\underline{B}|^2\right)
\end{aligned}$$

Therefore, for the T_{00} calculation, all we need now is the first term

$$F_{0\nu}F^{\nu}_{\ 0} = |\underline{E}|^2$$

which leads to

$$\begin{aligned}
T_{00} &= \epsilon_0\, |\underline{E}|^2 + \frac{\epsilon_0}{2}\,\left(c^2\,|\underline{B}|^2 - |\underline{E}|^2\right) \\
&= \frac{\epsilon_0}{2}\,\left(|\underline{E}|^2 + c^2\,|\underline{B}|^2\right)
\end{aligned}$$

which is nothing else than the electromagnetic energy density of the fields in the rest frame of the charge density.

The T_{0i} terms ($i = 1, 2, 3$) turn out to be proportional to $(\underline{E} \times \underline{B})_i$ and form the Poynting vector which describes the energy flux of the fields. The 3×3 submatrix T_{ij} finally forms the Maxwell stress tensor of the fields. Therefore the entire energy–momentum description of the electromagnetic field has been obtained using this tensor.

Coming closer now to conservation laws for the fields, we note that a free field does not have any sources, i.e. the source term, the current four-vector, vanishes and hence the divergence of the tensor should vanish which is consistent with the expression at the start of the derivation, i.e.

$$\partial_\alpha T^{\alpha\nu} = 0$$

if the four-current $j^\nu = 0$, i.e. this defines the free field description. Similar conservation laws can be formulated under the presence of sources but we leave this topic at this point and turn our attention more towards the invariants of the field.

3.5.1 Addendum on duals and the Hodge operator

First needed is the permutation symbol, also called the Levi-Civita tensor. This is the one exception from the rule of never needing more than rank 2 tensors during the lecture. It is defined as:

$$\epsilon^{\alpha\beta\gamma\delta} = \begin{cases} 0 & \text{if 2 or more indices are equal} \\ +1 & \text{if indices permute evenly} \\ -1 & \text{if indices permute odd} \end{cases}$$

and the definition $\epsilon^{0123} \equiv +1$. Its main application is to define oriented volumes, for examples for integrals. The purpose of introducing this symbol here, however is a different one. Let's look quickly at one more relevant property before coming to that purpose: The epsilon tensor is invariant under Lorentz transformations!

$$\epsilon^{\alpha\beta\gamma\delta} L^\sigma_\alpha L^\tau_\beta L^\rho_\gamma L^\omega_\delta = \epsilon'^{\sigma\tau\rho\omega} \det(\boldsymbol{L}) = \epsilon'^{\sigma\tau\rho\omega}$$

since the determinant of the Lorentz transformation is equal unity. You can then calculate the lower index version, $\epsilon_{\alpha\beta\gamma\delta}$ by lowering each index separately with the Minkowski metric and get for example

$$\epsilon_{\alpha\beta\gamma\delta} = \epsilon^{\sigma\tau\rho\omega}\eta_{\alpha\sigma}\eta_{\beta\tau}\eta_{\gamma\rho}\eta_{\delta\omega} = \epsilon^{\sigma\tau\rho\omega}\det(\boldsymbol{\eta}) = -\epsilon^{\sigma\tau\rho\omega}$$

i.e. the sign changes if all indices are changed in position. That determinant is far from being obvious, admittedly but it is a consequence of the very definition of this total antisymmetric tensor $\boldsymbol{\epsilon}$. If you remember your first year maths and the definition of any determinant, typically for a 2×2 or 3×3 matrix then it involved such a total antisymmetric 'symbol' in order to get all the signs right. This is what happens here. In order to convince yourself, you could go through the $(\alpha\beta\gamma\delta)$ permutations and the corresponding $(\sigma\tau\rho\omega)$ permutations.

Anyway, the purpose of introducing this tensor is that it will be used to define for every antisymmetric tensor of rank p the **dual** tensor of rank n-p (where here n= 4 always).

This is of interest in particular for the only antisymmetric tensor you have seen so far, the Faraday tensor. Let's define the Hodge operator $*$ by forming the dual of a rank p antisymmetric tensor to a rank n-p tensor and apply to the p= 2 case:

$$*F^{\mu\nu} = \frac{1}{2!}\,\epsilon^{\mu\nu\sigma\tau}\,F_{\sigma\tau} \tag{3.25}$$

Now what does that mean? In components, $*F^{\mu\nu}$ has $\underline{E}$ and $\underline{B}$ components swapped, such as:

$$\begin{aligned} *F^{01} &= \frac{1}{2}\,\epsilon^{01\sigma\tau}F_{\sigma\tau} \\ &= \frac{1}{2}\,\left(\epsilon^{0123}F_{23} + \epsilon^{0132}F_{32}\right) \\ &= \frac{1}{2}\,\left((+1)\,(-cB_1) + (-1)\,cB_1\right) = -cB_1 \end{aligned}$$

Finally, one can form explicitly the two invariants involving the Faraday tensor:

1. Full contraction:
$$F^{\mu\nu}\,F_{\mu\nu} = -2\,(|\underline{E}|^2 - c^2|\underline{B}|^2)$$
which we had derived earlier, see above, and

2. Cross contraction, Hodge dual with Faraday
$$*F^{\mu\nu}\,F_{\mu\nu} = -4\,c\,\underline{EB}$$

There is also the possibility to get to the final Maxwell equation in its compact form as quoted at the start:

$$\partial_\mu\,(*F^{\mu\nu}) = 0$$

This can be tested quickly according to

$$\begin{aligned} \partial_\mu\,(*F^{\mu\nu}) &= \frac{1}{2}\,\epsilon^{\mu\nu\sigma\tau}\,\partial_\mu F_{\sigma\tau} \\ &= \frac{1}{2}\,\epsilon^{\mu\nu\sigma\tau}\,\partial_\mu\,(\partial_\sigma A_\tau - \partial_\tau A_\sigma) = 0 \end{aligned}$$

since differentiation is symmetric and the epsilon symbol is totally antisymmetric. The correspondence to the previous representation of the homogeneous Maxwell equations can also be established straightforwardly by simply counting through the indices and perform the summation for all possible ν-values:

$$\frac{1}{2}\,\epsilon^{\mu\nu\sigma\tau}\,\partial_\mu F_{\sigma\tau} = \partial_\mu F_{\sigma\tau} + \partial_\tau F_{\mu\sigma} + \partial_\sigma F_{\tau\mu} = 0$$

Next on the list are explicit field transformations.

3.5.2 Exercises

(Collection sheet [3,7-10])

1. A particular electromagnetic field has its $\underline{E}$ field at an angle θ to its $\underline{B}$ field and θ is invariant to all observers. Calculate the value of θ using the invariants of the electromagnetic field.

2. Given a plane wave in Cartesian coordinates

$$E_y = E_0 \sin(\omega\, t - k\, x)\,; \quad B_z = B_0 \sin(\omega\, t - k\, x),$$

where $E_0 = c\, B_0$, calculate the energy-momentum tensor. Test your result by requiring the trace of the tensor to vanish identically, $T^{\mu}_{\mu} = 0$. Use the expression

$$T^{\mu\nu} = F^{\mu}_{\alpha}\, F^{\alpha\nu} - \eta^{\mu\nu}\, \left(|\underline{E}^2| - c^2|\underline{B}|^2\right)$$

for the energy-momentum tensor.

3. Assume that for an electromagnetic wave the following relations are satisfied in all reference frames: $\underline{E}\,\underline{B} = 0$; $E^2 = c^2\, B^2$. If $\underline{K}$ is a three-vector in the direction of propagation of the wave, then also the following relations are true: $\underline{K}\,\underline{E} = \underline{K}\,\underline{B} = 0$. Show that these relations are also Lorentz invariant like the first set of relations above by showing their equivalence to the statement $n^{\mu}\, F_{\mu\nu} = 0$, where n^{μ} is the wave four-vector oriented in the direction of propagation of the wave and $F_{\mu\nu}$ is the Faraday tensor.

4. Calculate the value of the Lorentz invariant expression $*F^{\mu\nu}\, F_{\mu\nu}$.

3.6 Field transformations

This section specialises now to pure Lorentz transformations of the Faraday tensor. The idea is to demonstrate explicit effects due to such transformations on electric and magnetic fields. The prescription is always the same and rather straightforward:

$$F'^{\mu\nu}(\boldsymbol{x}') = L^{\mu}_{\alpha}\, L^{\nu}_{\beta}\, F^{\alpha\beta}(\boldsymbol{x})$$

Note the explicit display of the coordinate dependence of the fields (on the coordinate four-vector). A transformation of a field also always requires a transformation of the coordinate dependence!

Let's do this explicitly for a given Lorentz-transformation:

$$L^{\mu}_{\nu} = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Get for the transformed Faraday tensor, **F**':

$$\boldsymbol{L}^T \begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & -cB_3 & cB_2 \\ E_2 & cB_3 & 0 & -cB_1 \\ E_3 & -cB_2 & cB_1 & 0 \end{pmatrix} \boldsymbol{L}$$

$$= \boldsymbol{L} \begin{pmatrix} \gamma\beta E_1 & -\gamma E_1 & -E_2 & -E_3 \\ \gamma E_1 & -E_1\gamma\beta & -cB_3 & cB_2 \\ \gamma E_2 - c\gamma\beta B_3 & \gamma\beta E_2 + c\gamma B_3 & 0 & -cB_1 \\ \gamma E_3 + c\gamma\beta B_2 & -\gamma\beta E_3 - c\gamma B_2 & cB_1 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} \gamma^2\beta E_1 - \gamma^2\beta E_1 & -\gamma^2 E_1 + \gamma^2\beta^2 E_1 & -\gamma E_2 + c\gamma\beta B_3 & -\gamma E_3 - c\gamma\beta B_2 \\ \gamma^2 E_1 - \gamma^2\beta^2 E_1 & \gamma^2\beta E_1 - \gamma^2\beta E_1 & \gamma\beta E_2 - c\gamma B_3 & \gamma\beta E_3 + c\gamma B_2 \\ \gamma E_2 - c\gamma\beta B_3 & \gamma\beta E_2 + c\gamma B_3 & 0 & -cB_1 \\ \gamma E_3 + c\gamma\beta B_2 & -\gamma\beta E_3 - c\gamma B_2 & cB_1 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & -E_1 & -\gamma(E_2 - c\beta B_3) & -\gamma(E_3 + c\beta B_2) \\ E_1 & 0 & -\gamma(cB_3 - \beta E_2) & \gamma(cB_2 + \beta E_3) \\ \gamma(E_2 - c\beta B_3) & \gamma(cB_3 - \beta E_2) & 0 & -cB_1 \\ \gamma(E_3 + c\beta B_2) & -\gamma(cB_2 + \beta E_3) & cB_1 & 0 \end{pmatrix}$$

Therefore the $\underline{E}$ and $\underline{B}$ fields become mixed for a moving observer. However, this is merely what happens to the field components. Once you have explicit expressions, formulae, for the fields, also the coordinates need transforming. One exercise to demonstrate this procedure would be the humble point charge, simply because you know the expression for the fields, i.e. the Coulomb field:

$$\underline{B} = 0$$
$$\underline{E} = \frac{e}{r^2}\frac{\underline{r}}{|\underline{r}|}$$

Therefore, let's do the entire exercise, transform the fields **and** the field arguments, i.e. coordinates.

Assume a point charge at rest in a system S and transform to another system S' which moves parallel to the x-axis at some speed v. Get started by evaluating the fields at a point P with S and S' synchronized at $t = t' = 0$. Then label P in S with x^μ and in S' with x'^μ.

Spare yourself some clutter in expressions by abbreviating the perpendicular coordinates, y and z, by setting $b^2 = y^2 + z^2 = y'^2 + z'^2$ which is nothing else but the distance of P to the x-axis at the synchronization moment (x and x' axis coincide). Then the distance r from the origin of S to P is $r^2 = b^2 + x^2$. Hence expressed in primed coordinates, get

$$r^2 = b^2 + \gamma^2\,(x' + \beta x'^0)^2$$

where the x' above is the x'-coordinate and not the vector and $\beta x'^0 = vt'$.

Having clarified the coordinates, one can write immediately the explicit field descriptions for a point charge using the Coulomb field formula. Starting with the x-component:

$$E_1' = E_1 = \frac{e\,x}{r^3} = e\,\frac{\gamma(x' + vt')}{[\gamma^2(x' + vt')^2 + b^2]^{3/2}}$$

Note, the equality at the start of the above is the result of the Lorentz transformation of the Faraday tensor shown earlier. It's the identical set of assumptions, i.e. reference frames moving parallel along x with speed v. Additionally, the point charge in S does not move, hence there is no magnetic field in the original Faraday tensor. Nevertheless, as seen above, there are non-zero magnetic field values in the transformed Faraday tensor, all comprised of E-field components. Therefore all you need to do (and can do) is calculate the E-field components like above and in the following. The E_2 component transforms as follows:

$$E_2' = \gamma E_2 = \gamma\,\frac{e\,y}{r^3} = e\,\frac{\gamma y'}{[\gamma^2(x' + vt')^2 + b^2]^{3/2}}$$

and similarly for the E_3' component. Summarising the three expression for the three components of $\underline{E}'$, say at $t' = 0$, get

$$\underline{E}'(\underline{r}') = e\,\frac{\gamma\underline{r}'}{[\gamma^2 x'^2 + b^2]^{3/2}}$$

which when simplifying the denominator according to

$$\gamma^2 x'^2 + b^2 = \gamma^2 \left(x'^2 + (1-\beta^2)\, b^2\right) = \gamma^2 \left(r'^2 - \beta^2 b^2\right)$$

gives

$$\underline{E}'(\underline{r}') = e\, \frac{\underline{r}'\,(1-\beta^2)}{\left[r'^2 - \beta^2 b^2\right]^{3/2}}$$

This derivation is useful in order to see explicitly that the character, the topology, of the field does not change after a Lorentz transformation. The original field is purely radial and also the transformed field is purely radial. However, the geometrical behaviour (spatial dependence) of the field magnitude has changed. Explicitly calculated it is

$$|\underline{E}'| = e\, \frac{1-\beta^2}{r'^2\, \left(1 - \beta^2 \sin^2\theta'\right)^{3/2}}$$

using $\sin\theta' = b/r'$. Therefore what you get as the transformed electric field orthogonal to the x'-axis ($\sin\theta' = 1$) would be

$$|\underline{E}'|_\perp = \frac{e}{r'^2 (1-\beta^2)^{1/2}} = \frac{e\gamma}{r'^2}$$

and parallel to the x'-axis ($\sin\theta' = 0$):

$$|\underline{E}'|_{||} = \frac{e}{r'^2}\,(1-\beta^2) = \frac{e}{\gamma^2 r'^2}$$

This concludes the field transformation exercise. A brief look at these equations already reveals the general behaviour of fields of a moving point charge relative to a resting observer: As the speed increases, the gamma-factor increases, boosting the field magnitude in the transversal direction (transversal to the direction of relative motion) and diminishing quadratically the strength in the longitudinal direction. The initially spherical Coulomb field of a point charge turns elliptical as relative speed increases.

Clearly, field transformation exercises are quite attractive and useful to train all elements of special relativity. It should also be clear that such calculations are indeed quite simple on a conceptual level. If not, here is a recipe: (a) construct the initial Faraday tensor according to the exercise assumptions. Quite often it will not be the fully general Faraday tensor which needs being transformed. More often than not, the exercise would specify that only very few of the 6 independent components are actually non-zero. Then (b) transform the Faraday tensor as demonstrated above with the appropriate Lorentz transformation. Finally, make sure that you also transform the coordinates if you have to make calculations (give expressions) involving the transformed system (all the primed quantities in the example above). That's it, simple as that.

The next topic, time-dependent fields, is a little less trivial but as compensation offers far fewer opportunities to construct questions, so take it as an educational exercise, dipping your toes again into a more advanced topic.

3.6.1 Exercises

(Collection sheet [3,11-13])

1. By using the electromagnetic field transformations below, show that the quantity $\underline{E}\,\underline{B}$ is Lorentz invariant. Assume the field transformation for a frame S' that moves with velocity v in x-direction in the frame S:

$$E'_x = E_x\,; \quad E'_y = \gamma\,(E_y - v\,B_z)\ ; \quad E'_z = \gamma\,(E_z + v\,B_y)$$

$$B'_x = B_x\,; \quad B'_y = \gamma\left(B_y + \frac{v\,E_z}{c^2}\right)\ ; \quad B'_z = \gamma\left(B_z - \frac{v\,E_y}{c^2}\right)$$

2. A large parallel plate capacitor with plates parallel to the x-y plane and plate distance d moves in x-direction with velocity $\underline{v}$ relative to the laboratory frame (un-primed coordinates). The plates are biased with a constant voltage difference U in the rest frame of the plates (primed coordinates). Find the electric and magnetic field components in the laboratory frame, neglecting edge effects.

3. Field transformation:

   - Calculate the electric field components of a particle with charge e moving with a constant velocity v in the x-direction. This involves calculating the Lorentz transformed electric field as well as the explicitly transformed Coulomb field expression for a point charge.
   - Let the magnitude of the electric field at a distance D_1 in front of the moving charge be equal to E_0. The magnitude of the field is identical, E_0, perpendicular to the direction of motion at some distance D_2. Calculate D_2 in terms of D_1.

3.7 Time-dependent fields

This section serves to give you access to more advanced topics in relativistic electromagnetism and a few still controversial topics in the field. It is not suited to an assessment, hence this material is non-examinable.

The attraction, however, of time-dependent fields in electromagnetism is the close connection to quantisation and field theories. Most conceptual problems in those advanced topics can find a classical counterpart in time-dependent electromagnetism, enabling you to gain a deeper understanding of puzzling concepts in field theory or quantum mechanics with more accessible physical models in electromagnetism. One of the most important tools to make any ground on that journey is a proper, conceptually clear definition of radiation. Nothing else will be shown in this final chapter of the lecture but all self-study topics lead on from this base. Nevertheless, this chapter, being more a showcase than a proper lecture, will skip quite a few lengthy subjects and simply quote results as a shortcut to get to the important points.

Keywords for further studies could be: radiation and the equivalence principle, the radiation reaction and its fundamental challenges due to the assumed point-like nature of elementary charges, radiation at event horizons - information (energy) losses or not etc.

3.7.1 Lienard-Wiechert potential

The Lienard-Wiechert potential is the result of highly non-trivial work seeking to solve the inhomogeneous wave equation:

$$\Box A^{\mu} = \sqrt{\frac{\mu_0}{\epsilon_0}}\, j^{\mu}$$

for an arbitrarily moving point charge. For those interested in the process, the most straightforward way is the Green function method but we will not pursue this path any further.

The more physically important concept introduced in the derivation of the Lienard-Wiechert potential is the retardation condition between source and receiver. This is a crucial concept, closely related to the heart of special relativity, i.e. the finite speed of light. First of all, it's important to realise that radiation enables the study of **non-local** physics by relating **two** points, source and receiver in a hopefully unique and well-defined way. Likewise, it forces theory to confront the surprisingly uncomfortable concept of a **backreaction** on the source.

In order to be able to discuss any of the above, a clear directional concept must be introduced and one such possibility is the retardation condition. This condition requires for any point P in space-time to physically interact, i.e. be influenced by, only points such as Q on P's **past** light-cone. In particular P shall interact only with electromagnetic fields originating from its past light cone, i.e. points such as Q

Figure 3.1: Space-time diagram showing the retardation requirements on two points on the light-cone, P and Q. Retardation requires P to receive signals from Q only from the past, i.e. Q being on the past light-cone of P.

from an arbitrary world-line intersected by P's past light-cone. Q would be in this case the source of radiation and P the receiver.

This geometrically defined condition can also be put into algebra. Consider the wold-line of a point charge $z^\mu(\tau)$ in a fixed reference frame. Then, as we have seen earlier in the lecture, we have a velocity and acceleration defined at any point along the world-line according to:

$$v^\mu(\tau) = \frac{d}{d\tau}\, z^\mu(\tau)$$

$$a^\mu(\tau) = \frac{d}{d\tau}\, v^\mu(\tau)$$

Now assume that the coordinates for the point P are given by x^μ. Connect a point Q on the world-line with P by defining:

$$R^\mu \equiv x^\mu - z^\mu(\tau_0)$$

or for any Q

$$R^\mu \equiv x^\mu - z^\mu(\tau)$$

with the important constraint

$$R^\mu R_\mu = 0$$

i.e. the connection, R^μ, must be a light-like four-vector.

The task is to identify an invariant distance measure between P and Q. Many of the kinematic concepts from earlier in the lecture will be a great help at this point. As you know, at any point along a world-line, a momentary system of rest can be defined which in turn enables the definition of a local set of basis vectors.

Figure 3.2: Space-time diagram showing the connection vector between P and Q and its decomposition into invariant projections onto the momentary basis vectors along the arbitrary world-line $z^\mu(\tau)$.

The time-like basis vector was identified as the normalised four-velocity at that point and all remaining three space-like vectors are perpendicular to it.

A decomposition of the light-like connection four-vector into its projections on the basis vectors is hence always possible and unique. For instance, R^μ projected onto the time-like basis vector will define a scaling factor ρ by which the unit vector must be multiplied to give the projected length. The unique property of light–like four-vectors enables now to conclude, most conveniently, that it is the very same scaling factor ρ by which all projection can be calculated on all unit basis vectors, i.e. also the space-like basis vectors times ρ give the projected lengths of R^μ respectively. The reason being that the light-cone is uniquely invariant in special relativity.

Algebraically, we can write:

$$\rho\, \underline{e}_0 = \rho\, \frac{1}{c}\, v^\mu$$

and for the space-like unit vectors, call them n^μ here for convenience

$$n^\mu n_\mu = -1; \quad n^\mu v_\mu = 0$$

get the decomposition

$$R^\mu = \rho\, \left(n^\mu(\tau_0) + \frac{1}{c}\, v^\mu(\tau_0) \right)$$

or

$$\rho = -n_\mu R^\mu = \frac{1}{c}\, v_\mu R^\mu > 0$$

which is the precise definition of the retardation condition.

Once this is settled, the Lienard-Wiechert potential takes an especially simple form

$$A^\mu = \frac{e}{c}\frac{v^\mu}{\rho}$$

which is in fact the fully invariant solution of the inhomogeneous wave equation for any charge world-line $z^\mu(\tau)$.

From this starting point, we can attempt to get the electromagnetic fields for such a potential using the field tensor definition from above

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$$

However, how do you differentiate in this scenario? Any change in P, our observation point, the point where to evaluate the potential (and hence the fields), necessarily changes Q! Hence the instance τ_0 (associated with Q) depends on x^μ (coordinates of P). These are the tricky bits of dealing with non-local physics. Differentiation is always a local process but the retardation condition requires to transport the local differentiation along the connection to a second point, i.e. introducing additional dependencies. It is still all quite elementary, see below, but needs proper consideration when calculating anything. It is at this point, for example, where quantum field theories have to introduce challenging recipes in order to stay well defined. Here, all is needed now can be summarised as

$$\partial^\mu A^\nu = -\frac{e}{c}\frac{v^\nu}{\rho^2}\partial^\mu \rho + \frac{e}{c}\frac{1}{\rho}\, a^\nu\, \partial^\mu \tau$$

where the task is to calculate $\partial^\mu \rho$ and $\partial^\mu \tau$. Start with the known condition $R^\mu R_\mu = 0$:

$$R^\mu \left(\frac{dx_\mu}{d\tau} - v_\mu\right) = 0 = \left(n^\mu + \frac{v^\mu}{c}\right)\left(\frac{dx_\mu}{d\tau} - v_\mu\right)$$

from which follows

$$\left(n^\mu + \frac{v^\mu}{c}\right)\frac{dx_\mu}{d\tau} - c = 0$$

hence

$$\left(n^\mu + \frac{v^\mu}{c}\right)\frac{dx_\mu}{cd\tau} = 1$$

hence

$$\frac{1}{c}\left(n^\mu + \frac{v^\mu}{c}\right) = \frac{\partial \tau}{\partial x_\mu} = \partial^\mu \tau$$

simply the inverse of the more conventional $dx_\mu/d\tau$. Likewise for the second missing derivative, we get:

$$\partial^\mu \left(-n_\mu R^\mu\right) = -n_\mu - R^\mu(\partial^\nu n_\nu)$$

$$= -n_\mu - R^\mu \left[\frac{dn_\nu}{d\tau} \, \partial^\nu \tau \right]$$

where focussing on the square bracket term only gives:

$$\left[((v^\sigma n_\sigma) a_\nu - (a^\sigma n_\sigma) v_\nu) - \frac{1}{c} \left(n^\nu + \frac{v^\nu}{c} \right) \right] = -(a^\sigma n_\sigma) \equiv \, - a_n$$

where most terms disappear due to the orthogonality between $\boldsymbol{v}$ and $\boldsymbol{n}$ and you might also recognise the Fermi-Walker transport relation (first two terms). The abbreviation a_n is really only a convenience. The subscript 'n' is not an index. Finally, put together these results in the following derivative of the Lienard-Wiechert potential:

$$\partial^\mu \, A^\nu = -\frac{e}{c} \frac{v^\nu}{\rho^2} \, (n^\mu + a_n R^\mu) + \frac{e}{c^2} \frac{1}{\rho} \, a^\nu \left(n^\mu + \frac{v^\mu}{c} \right)$$

So, collecting terms results in

$$\begin{aligned} F^{\mu\nu} &= \frac{e}{c} \frac{1}{\rho^2} \, (v^\mu n^\nu - v^\nu n^\mu) \\ &+ \frac{e}{c^2} \frac{1}{\rho} \left[a^\nu \frac{v^\mu}{c} - a^\mu \frac{v^\nu}{c} + n^\mu \, (a^\nu - a_n c v^\nu) + n^\nu \, (a_n c v^\mu - a^\mu) \right] \end{aligned}$$

Two main terms become obvious: one Coulomb-like term $\propto \frac{1}{\rho^2}$ and one so called radiation-like term $\propto \frac{1}{\rho}$. These two parts can be written in a little more compact notation:

$$F_I^{\mu\nu} = \frac{e}{c} \frac{1}{\rho^2} \, v^{[\mu} \frac{R^{\nu]}}{\rho}$$

using $a^{[\mu} b^{\nu]} = a^\mu b^\nu - a^\nu b^\mu$ and

$$F_{II}^{\mu\nu} = \frac{e}{c^2} \frac{1}{\rho} \left(a_n v^{[\mu} \frac{R^{\nu]}}{\rho} + a^{[\nu} \frac{R^{\mu]}}{\rho} \right)$$

All that remains then is to plug $F^{\mu\nu}$ into the energy-momentum tensor and define radiation according to either

1. the asymptotic definition (near-field and far-field concepts) or
2. operational, i.e. quasi-local, hence at any finite distance from the charge.

Assume you do get the energy-momentum tensor (fairly involved calculation but brilliant practice of index notation, see reference below) then you might wish to split the tensor into three parts; two describe Coulomb-like distance behaviour of energy-momentum whereas the third part yields a directly conserved part, i.e. a tensor with vanishing divergence, i.e. source-free. It is this third part, $T^{\mu\nu}_{III}$ which describes radiation since it describes a purely light-like flux of energy-momentum according to $T^{\mu\nu}_{III} R_\mu = 0$. This constitutes the operational definition of radiation which is the preferred one conceptually. Surprisingly, this definition has been published only in 1970 (C. Teitelboim, Phys. Rev. D1 (1970) 1572). A full discussion of challenges and solutions for classical electromagnetic radiation can be picked up on in F. Rohrlich, Classical charged particles, 1965.

This concludes the lecture.

Appendix A

More on index notation

Let's collect a few basic relations first[1]: Components of a vector (one-form) from the vector (one-form) in an arbitrary basis. I'll leave out the (one-form) extension from now on and use index notation, naturally, i.e. upper indices represent vectors, lower indices represent one-forms.

$$v^i = \boldsymbol{v} \cdot \boldsymbol{g}^i \quad ; \quad v_i = \boldsymbol{v} \cdot \boldsymbol{g}_i$$

Likewise a representation of a vector in an arbitrary basis

$$\boldsymbol{v} = v^i \, \boldsymbol{g}_i = v_i \, \boldsymbol{g}^i$$

where (as a reminder) the index 'i' is a contracted index, sometimes called **dummy index** or loose index. Contraction means the operation of summation over that very index. It's a matter of training to deal with contracted indices confidently, hence these exercises might help. The notation requires indices over which to sum to be positioned diagonally with respect to each other, always an upper with its lower counterpart or vice versa. Each contraction therefore must have its own index, distinct from any other contraction operation in order not to confuse them.

Rank 2 tensors have similar relations to vectors:

$$T^{ij} = \boldsymbol{g}^i \cdot \mathcal{T} \, \boldsymbol{g}^j \quad ; \quad T_{ij} = \boldsymbol{g}_i \cdot \mathcal{T} \, \boldsymbol{g}_j$$

using

$$\boldsymbol{T}_j = \mathcal{T} \, \boldsymbol{g}_j = T_{ij} \boldsymbol{g}^i$$

this last equation derives from multiplying the basis vector $\boldsymbol{g}^i$ to the relation above for T_{ij}. It doesn't matter here whether you multiply from the left or right since all

[1]Most of this material comes from [3]

multiplications in the relations so far commute, i.e. multiplication and dot product. If you are uncomfortable commuting over the tensor $\mathcal{T}$ then you are right to do so, in principle. Just imagine it to be a matrix and they generally do not commute with anything other than numbers. However, here the tensor is always contracted to a vector (or one-form) at the least, hence commutation is fine.

So, multiplying $\boldsymbol{g}^i$ above gets you there but only if you remember that $\boldsymbol{g}^i \boldsymbol{g}_i = 1$ by definition. The basis vectors and basis one-forms are the reverse basis of each other respectively.

$$\boldsymbol{g}^i \boldsymbol{g}_j = \delta^i_j$$

A.1 Einstein summation convention, again

Vanderlinde [6] gives two rules to remember on index notation and Einstein summation convention:

1. Every index appearing **once** in an expression can take on values according to dimension. Notation: latin letters 1,2,3 typically; greek letters 0,1,2,3 for special relativity. Thus A_i denotes any member of the set $\{A_1, A_2, A_3\}$ and A_{ik} of the set $\{A_{11}, A_{12}, A_{13}, A_{21}, A_{22}, A_{23}, A_{31}, A_{32}, A_{33}\}$.

2. If a free index appears **twice** in a term, once as superscript and once as subscript or vice versa (the diagonal relation) summation over that index is implied.

$$\begin{aligned} A^i_i &\equiv A^1_1 + A^2_2 + A^3_3 \\ A_i B^i &\equiv A_1 B^1 + A_2 B^2 + A_3 B^3 \\ A_i B^k C^i &\equiv B^k \left(A_1 C^1 + A_2 C^2 + A_3 C^3 \right) \end{aligned}$$

An expression such as $A_i B^i$ is independent of the letter 'i', hence the letter is called a **dummy** index.

A.2 The metric tensor is special

A.2.1 The metric components and the identity tensor

Well, apart from the huge importance for physics ascribed to the metric tensor in general relativity, it's a little special also generally, for tensors and manipulations of them. It all follows from considering the components of the **identity tensor**. Strange as that might sound, there is more to it than merely

$$1 = \boldsymbol{g}_i \cdot \boldsymbol{g}^i = \boldsymbol{g}^i \cdot \boldsymbol{g}_i$$

In fact, mixed tensor δ^i_j is naturally considered to be the representation of the identity tensor 1^i_j components. But what about the $(0,2)$ or $(2,0)$ representations?

Let's define, rather provocatively, the components of the identity tensor as

$$1_{ij} = g_{ij}, \quad 1^i_j = \delta^i_j, \quad 1^i_j = \delta^i_j, \quad 1^{ij} = g^{ij}$$

with (and this is the key point):

$$g_{ij} = \boldsymbol{g}_i\, \boldsymbol{g}_j, \quad g^{ij} = \boldsymbol{g}^i\, \boldsymbol{g}^j$$

That makes sense when considering the representation of any tensor $\mathcal{T}$ from above in terms of arbitrary basis vectors $\boldsymbol{g}$. Here, all we've done is suppress the $\mathcal{I}$ symbol for the identity tensor since it's redundant.

Right, that needs some justification. Starting from the identity $\mathcal{I}\boldsymbol{v} = \boldsymbol{v}$ let's get the components of the identity tensor:

$$\begin{aligned} &\mathcal{I}(v^j\, \boldsymbol{g}_j) \\ &= \mathcal{I}\boldsymbol{g}_j\, v^j = \mathbf{1}_j\, v^j \\ &= 1_{ij}\boldsymbol{g}^i v^j = 1_{ij}\boldsymbol{g}^i(\boldsymbol{g}^j \cdot \boldsymbol{v}) \\ &= \left(1_{ij}\boldsymbol{g}^i\boldsymbol{g}^j\right)\, \boldsymbol{v} = \boldsymbol{v} \end{aligned}$$

hence the identity tensor can be represented by $1_{ij}\boldsymbol{g}^i\boldsymbol{g}^j$, no surprise since that is valid for any tensor, but now we also have

$$g_{ij} \equiv \boldsymbol{g}_i\boldsymbol{g}_j = \boldsymbol{g}_i\mathcal{I}\boldsymbol{g}_j$$

by simply inserting a redundant 1 in between the basis vectors, i.e. the identity tensor and now get, inserting the component representation of it

$$= \boldsymbol{g}_i 1_{ij}\boldsymbol{g}^i\boldsymbol{g}^j\boldsymbol{g}_j = 1_{ij}$$

since the basis vector and one-forms were defined as such to be inverse and 1_{ij} are simply numbers, the components of the identity tensor.

Now we still don't know the explicit form of the $g_{ij} = 1_{ij}$ but that is fine since this will heavily depend on the basis vectors. What we have, however, is a way to understand what makes the metric tensor special, in particular its next property singles it out from all other same-structure tensors - an operation called index raising or lowering.

A.2.2 Raising and lowering indices with the metric tensor

Typically this operation of raising and lowering indices is simply given as something the metric tensor is used for, nothing else. Here is how that comes about. First of all, changing positions of indices only looks trivial using index notation. What really happens are fundamental structural changes to the objects you work with. A $(3, 1)$ tensor is something other than a $(4, 0)$ tensor, both of rank 4, granted but for calculations they can be very different indeed.

We have seen the index lowering operation before, acting out on the dot-product between two 'vectors', i.e. when deriving that this rather corresponds to a product between a vector and a one-form, loosely speaking. Let's revisit this operation:

$$s^2 = g_{\mu\nu}\, x^\mu\, x^\nu = x_\nu\, x^\nu$$

The metric tensor 'pulls' the dummy index μ down by contraction and replacing it with the index ν. That's one graphical description of what happens. Of course, nothing of that kind actually happens at all.

Let's write this in some more detail, assuming a basis $\boldsymbol{e}_\mu$:

$$s^2 = g_{\mu\nu}\, x^\mu\, x^\nu = \boldsymbol{e}_\mu \boldsymbol{e}_\nu\, \boldsymbol{x}\, \boldsymbol{e}^\mu\, \boldsymbol{x}\, \boldsymbol{e}^\nu$$

using the expressions above to express the metric tensor coefficients in a given basis and the vector components. Now commute the inner vector $\boldsymbol{x}$ with the basis one-form $\boldsymbol{e}_\nu$ to get

$$= \boldsymbol{e}_\mu\, \boldsymbol{x}\, \boldsymbol{e}_\nu\, \boldsymbol{e}^\mu\, \boldsymbol{x}\, \boldsymbol{e}^\nu$$

$$= \boldsymbol{x}\, \boldsymbol{e}_\mu\, \delta^\mu_\nu\, \boldsymbol{x}\, \boldsymbol{e}^\nu = x_\nu\, x^\nu$$

and it's the mixed identity tensor components which actually perform the contraction (sum over μ here, trivially due to the Kronecker delta properties) and replacement $\mu \to \nu$, not the metric tensor as such. Raising indices works exactly the same. Also, from these elementary calculations it should be clear that this process works on any rank tensor, except numbers (no index, nothing to manipulate), of any structure.

Bibliography

[1] W.S.C. Williams, **Introducing Special Relativity**, Taylor and Francis, London 2002.

[2] N.M.J. Woodhouse, **Special Relativity**, Springer, London 2003.

[3] J.G. Simmonds, **A Brief on Tensor Analysis**, Springer, NY, 1982.

[4] C.W. Misner, K.S. Thorne and J.A, Wheeler, **Gravitation**, Freeman, 1973.

[5] D. Fleisch, **A Student Guide to Vectors and Tensors**, Cambridge Univ. Press, Cambridge, 2012.

[6] J. Vanderlinde, **Classical Electromagnetic Theory**, Kluwer Academic Publ., London, 2004, 2nd ed.

Index