Introduction to

Relativity and Cosmology I

J. Mark Heinzle
Gravitational Physics, Faculty of Physics
University of Vienna

Version 20/01/2010

CHAPTER 1

AETHER

1.1 Galilean space (...and time)

What is space? Which mathematical concept provides the best description for what we experience as our spatial world?

This question is neither naïve nor trivial. There exists a multitude of concepts in mathematics that could apply *a priori*. Space could be merely a set, it could be a topological space, maybe Hausdorff, a semi-group, a field, a manifold, a complex manifold, a vector space, ...

Our daily experience tells us that space is some kind of 'continuum' of 'points'. It is a matter of course for us to be able to form 'arrows' ('vectors') $\overrightarrow{pq}$ connecting points p, q in space. Most importantly, we can add vectors and multiply vectors with scalars ($\in \mathbb{R}$). This suggests that our spatial world has a structure intimately connected with the structure of a vector space. However, it is obvious that a vector space is *not* a good model for our spatial world. This is simply because there doesn't seem to exist a distinguished zero vector; quite the contrary, all spatial points seem to be on an equal footing. Collecting these intuitive observations results in the following statement:

Space is an *affine space* over a three-dimensional real vector space.

The affine space we live in is endowed with a geometric structure that is so fundamental that it is easily overlooked. We can measure the length of vectors (e.g., in meters) as well as angles between vectors. The mere existence of the concepts of length and angles entails that the vector space underlying our affine space is a Euclidean vector space, i.e., it carries a *scalar product*.

Space is an *affine space* over a three-dimensional Euclidean vector space.

Definition. *We call this model of our spatial reality* Galilean space.

Remark. All these "experiments" that we perform every day—we form vectors connecting points, compute lengths and angles— indicate that Galilean space is the correct mathematical description of the space of our daily experience. Note, however, that we would come to a radically different conclusion if the scale of our intuitive perception were different. But we think in meters and not in megaparsecs...

What is time? This is a somewhat more difficult question. Let us quote Sir Isaac Newton:

> "Absolute, true and mathematical time, of itself, and from its own nature flows equably without regard to anything external, and by another name is called duration: relative, apparent and common time, is some sensible and external (whether accurate or unequable) measure of duration by the means of motion, which is commonly used instead of true time ..."

Summarizing, we don't quite know what time is (at least not yet), but we have a lot of first hand experience with it. For our present purposes this is sufficient.

In order to formulate the laws of physics it is essential to choose a *coordinate system*.[1] A coordinate system enables us to make measurements; we can quantify position by real numbers, and we are able to perform a mathematical analysis of our physical fields: we form derivatives, gradients, divergences and formulate (differential) equations for our physical unknowns.

We are free to choose any curvilinear coordinate system we can come up with.[2]

[1] Strictly speaking, a coordinate system—or chart—is simply a bijective map of $\mathbb{R}^3$ into Galilean space. In the present context, however, dependence on time is permitted (so that a coordinate system is a one-parameter family of maps of $\mathbb{R}^3$ into Galilean space).

[2] And frequently, curvilinear coordinates are very useful. For instance, when one aims at studying rotating bodies, spherical coordinates are often advantageous.

However, among the vast number[3] of possible coordinate systems, there exists a special subfamily: the *inertial frames* (inertial coordinates). The existence of these inertial frames is guaranteed by the *principle of relativity*:

Special principle of relativity

There exists a family of coordinate systems, which we call inertial frames of reference, w.r.t. which the laws of nature take one and the same form. In other words, the mathematical formulation of the laws of physics is identical for all 'inertial observers'.

In the Galilean context, an inertial frame is thus a coordinate system of Galilean space, for which the laws of (Newtonian) physics assume their conventional form.

Consider, for example, Newton's first law which describes the motion of a point particle in the absence of exterior forces. Newton's first law states that these point particles move along straight lines in Galilean space. In an inertial frame of reference, Newton's first law is represented by the equation[4]

$$\ddot{x}^i(t) = 0 \,.$$

Another law of (Newtonian) physics is the Poisson equation describing the gravitational field of a massive body. In an inertial frame of reference we have

$$\Delta\phi = \rho \,,$$

where ϕ is the potential and ρ the matter density. The Laplacian Δ is given by

$$\Delta = \Big(\frac{\partial}{\partial x^1}\Big)^2 + \Big(\frac{\partial}{\partial x^2}\Big)^2 + \Big(\frac{\partial}{\partial x^3}\Big)^2 = \partial_1^2 + \partial_2^2 + \partial_3^2 = \delta^{ij}\,\partial_i\partial_j \,.$$

In a general curvilinear coordinate system, e.g., in spherical coordinates, a straight line does not appear as straight; consequently, the differential equation representing Newton's first law is different. (Of course, the same is true for the Poisson equation.) The price we pay by not choosing an inertial coordinate system is that the laws of physics take a more complicated form (that includes terms that are interpreted ficticious forces).

[3]The vast number is of course ∞.

[4]A point particle is represented by a curve $t \mapsto x(t)$; the components of the vector x are denoted by x^i, $i = 1, 2, 3$. This curve is a straight line if and only if $\ddot{x}(t) = 0$ (or, in component notation, $\ddot{x}^i(t) = 0$).

1.2 Galilean transformations

Definition 1.1. *A Galilean transformation is a coordinate transformation*[5] *of Galilean space that maps one inertial frame into another.*

Taking account of the definition of an inertial frame we find an alternative definition.

Definition 1.1′. *A Galilean transformation is a coordinate transformation*[6] *of Galilean space that leaves invariant the laws of (Newtonian) physics.*

Remark. We emphasize that the principle of relativity implies that *all* laws of (Newtonian) physics are invariant under Galilean transformations.

On the basis of definition 1.1′ we are able to derive the Galilean transformations; we choose to make use of Newton's first law.

Suppose that (t, x), where $x = (x^1, x^2, x^3)$, and $(\bar{t}, \bar{x})$, where $\bar{x} = (\bar{x}^1, \bar{x}^2, \bar{x}^3)$, are inertial coordinates. By definition, the Galilean transformation mapping the one inertial coordinate system into the other is of the form[7]

$$\bar{t} = t\,, \qquad \bar{x} = \bar{\mathsf{x}}(t, x)\,; \tag{1.1}$$

expressed in component form, the latter reads $\bar{x}^i = \bar{\mathsf{x}}^i(t, x^1, x^2, x^3)$. In (1.1) we assume that the origin of space and time remain unchanged; if this is not the case, (1.1) may contain additional constants representing a spatial translation and a temporal translation, i.e.,

$$\bar{t} = t + \bar{t}_0\,, \qquad \bar{x} = \bar{\mathsf{x}}(t, x) + \bar{x}_0\,. \tag{1.1′}$$

Let $x(t)$ represent the freely moving point particle w.r.t. the first inertial frame, i.e., $\ddot{x}^i(t) = 0$. W.r.t. the second inertial frame (1.1), this curve reads

[5] Strictly speaking, a Galilean transformation is a one-parameter family of coordinate transformations of Galilean space; this is because there might be a dependence on time, see (1.1).

[6] See footnote 5.

[7] It is implicit in definition 1.1′ that time remains unchanged under a Galilean transformation (because it is a coordinate transformation 'of Galilean space').

$\bar{x}(t) = \bar{\mathsf{x}}\big(t, x(t)\big)$; we may replace t by $\bar{t}$. We obtain

$$\begin{aligned}
\dot{\bar{x}}^i &= \frac{\partial \bar{\mathsf{x}}^i}{\partial t} + \frac{\partial \bar{\mathsf{x}}^i}{\partial x^j}\, \dot{x}^j\,, \\
\ddot{\bar{x}}^i &= \frac{\partial^2 \bar{\mathsf{x}}^i}{\partial t^2} + 2\frac{\partial^2 \bar{\mathsf{x}}^i}{\partial t \partial x^j}\, \dot{x}^j + \frac{\partial^2 \bar{\mathsf{x}}^i}{\partial x^j \partial x^k}\, \dot{x}^j \dot{x}^k + \frac{\partial \bar{\mathsf{x}}^i}{\partial x^j}\, \underbrace{\ddot{x}^j}_{=0}\,.
\end{aligned}$$

Since Newton's law must be invariant under the transformation (because $(\bar{t}, \bar{x})$ is assumed to be an inertial frame) we find that

$$0 = \frac{\partial^2 \bar{\mathsf{x}}^i}{\partial t^2} + 2\frac{\partial^2 \bar{\mathsf{x}}^i}{\partial t \partial x^j}\, \dot{x}^j + \frac{\partial^2 \bar{\mathsf{x}}^i}{\partial x^j \partial x^k}\, \dot{x}^j \dot{x}^k\,.$$

Since $\dot{x}^i$ is arbitrary, this results in

$$\frac{\partial}{\partial t}\frac{\partial \bar{\mathsf{x}}^i}{\partial t} = 0\,, \quad \frac{\partial}{\partial x^j}\frac{\partial \bar{\mathsf{x}}^i}{\partial t} = 0 \quad \text{and} \quad \frac{\partial}{\partial t}\frac{\partial \bar{\mathsf{x}}^i}{\partial x^k} = 0\,, \quad \frac{\partial}{\partial x^j}\frac{\partial \bar{\mathsf{x}}^i}{\partial x^k} = 0\,,$$

from which we conclude that

$$\frac{\partial \bar{\mathsf{x}}^i}{\partial t} \equiv -v^i = \text{const} \quad \text{and} \quad \frac{\partial \bar{\mathsf{x}}^i}{\partial x^k} \equiv R^i{}_k = \text{const}$$

Accordingly, (1.1) becomes

$$\bar{t} = t\,, \qquad \bar{x}^i = R^i{}_k x^k - v^i\, t\,; \tag{1.2}$$

in matrix notation the latter reads $\bar{x} = R\,x - v\,t$.

The matrix R is not arbitrary. It is straightforward to prove that $R \in \mathsf{O}(3)$, i.e., R is either a rotation ($R \in \mathsf{SO}(3)$) or a reflection ($R \in \mathsf{O}(3)$, $\det R = -1$). (Note, however, that this does not follow directly from the invariance of Newton's first law—it is necessary to invoke another law of physics.)

Exercise. Use the invariance of the Poisson equation under (1.2) to prove that $R \in \mathsf{O}(3)$.

In (1.2) the origin of space and time remain unchanged; if this is not the case, (1.2) may contain additional constants representing a spatial translation and a temporal translation, i.e.,

$$\bar{t} = t + \bar{t}_0\,, \qquad \bar{x}^i = R^i{}_k x^k - v^i\, t + \bar{x}_0\,. \tag{1.2$'$}$$

Of particular interest among the Galilean transformations (1.2′) are the so-called *Galilean boosts*

$$\bar{t} = t\,, \qquad \bar{x}^i = x^i - v^i\,t\,. \tag{1.3}$$

Galilean boosts are coordinate transformations between inertial frames that are in *uniform relative motion.* Let us summarize the statement of (1.2′) as a corollary.

Corollary 1.2. *An (orientation-preserving) Galilean transformation is a spatial translation by a constant vector, a temporal translation by a constant parameter, a rotation by a constant angle, a Galilean boost, or a combination of these.*

Remark. Also reflections are Galilean transformations; these do not, however, preserve the orientation of the coordinates.

Remark. We see that inertial frames are those coordinate systems that are adapted to the geometric structure of (Galilean) space (and time). Inertial frames essentially correspond to orthonormal frames or orthonormal frames in uniform motion.

Let us conclude this section with another example illustrating Galilean invariance. Consider the equations of motion of two gravitating point particles with masses m_1, m_2,

$$m_1 \ddot{x}_1 = -m_1 m_2 \,\frac{x_1 - x_2}{|x_1 - x_2|^3}\,, \tag{1.4a}$$

$$m_2 \ddot{x}_2 = -m_1 m_2 \,\frac{x_2 - x_1}{|x_1 - x_2|^3}\,. \tag{1.4b}$$

A change of inertial frame is a Galilean transformation (1.2′), i.e., $\bar{t} = t + \bar{t}_0$ and

$$\bar{x}_1{}^i = R^i{}_k x_1{}^k - v^i\,t + \bar{x}_0\,, \qquad \bar{x}_2{}^i = R^i{}_k x_2{}^k - v^i\,t + \bar{x}_0\,.$$

Let us restrict ourselves to a Galilean boost, i.e., $\bar{t} = t$ and

$$\bar{x}_1{}^i = x_1{}^k - v^i\,t\,, \qquad \bar{x}_2{}^i = x_2{}^k - v^i\,t\,.$$

It follows that

$$\dot{\bar{x}}_1{}^i = \dot{x}_1{}^i - v^i\,, \quad \dot{\bar{x}}_2{}^i = \dot{x}_2{}^i - v^i\,, \quad \ddot{\bar{x}}_1{}^i = \ddot{x}_1{}^i\,, \quad \ddot{\bar{x}}_2{}^i = \ddot{x}_1{}^i\,, \quad \bar{x}_1{}^i - \bar{x}_2{}^i = x_1{}^i - x_2{}^i\,,$$

and we infer that

$$m_1 \ddot{\bar{x}}_1 = -m_1 m_2 \,\frac{\bar{x}_1 - \bar{x}_2}{|\bar{x}_1 - \bar{x}_2|^3}\,, \tag{1.4a′}$$

$$m_2 \ddot{\bar{x}}_2 = -m_1 m_2 \,\frac{\bar{x}_2 - \bar{x}_1}{|\bar{x}_1 - \bar{x}_2|^3}\,, \tag{1.4b′}$$

hence the equations are identical w.r.t. the inertial frame $(\bar{t}, \bar{x})$.

Exercise. Show the invariance of (1.4) under a general Galilean transformation.

1.3 The Euler equations

The Euler equations describe the dynamics of inviscid fluids. The equations read

$$\partial_t \rho + \partial_k \big(\rho u^k\big) = 0 \,, \tag{1.5a}$$

$$\partial_t \big(\rho u^i\big) + \partial_k \big(\rho u^i u^k + p\delta^{ik}\big) = 0 \,, \tag{1.5b}$$

where $\rho = \rho(t, x)$ is the density and $p = p(t, x)$ the pressure of the fluid; $u = u(t, x)$ is the velocity field.

The first equation is the continuity equation and represents conservation of mass; in index-free notation it reads

$$\partial_t \rho + \nabla(\rho u) = 0 \,.$$

Introducing the Lagrangian derivative (material derivative, convective derivative)

$$\frac{d}{dt} = \frac{\partial}{\partial t} + u\nabla$$

we may write

$$\frac{d\rho}{dt} = -\rho \nabla u \,. \tag{1.5a$'$}$$

The second equation, i.e., (1.5b), corresponds to Newton's second law and encodes the conservation of momentum. Using

$$(\partial_t \rho) u^i + \rho \partial_t u^i + \partial_k (\rho u^k) u^i + \rho u^k \partial_k u^i + \partial^i p = \rho\big(\partial_t + u^k \partial_k\big) u^i + \partial^i p$$

we may rewrite it as

$$\rho\big(\partial_t + u\nabla\big) u = -\nabla p$$

or

$$\rho\, \frac{du}{dt} = -\nabla p \,. \tag{1.5b$'$}$$

The equations (1.5a) and (1.5b) do not form a closed system. It is required that we prescribe an equation of state,

$$p = \mathsf{p}(\rho) \,, \tag{1.5c}$$

relating the density and the pressure. It is common to also consider an equation representing the conservation of energy

$$\partial_t \epsilon + \partial_k \big[(\epsilon + p) u^k \big] = 0 \, ;$$

here, ϵ denotes the energy density.

The variables of the system of equation (1.5) are

$$\rho(t,x)\,, \qquad p(t,x) \qquad u(t,x)\,.$$

Let us consider a boosted inertial frame, i.e.,

$$\bar t = \bar{\mathsf{t}}(t,x) = t \qquad \bar x = \bar{\mathsf{x}}(t,x) = x - vt\,. \tag{1.6}$$

W.r.t. the second frame the functions are

$$\bar\rho(\bar t,\bar x)\,, \qquad \bar p(\bar t,\bar x)\,, \qquad \bar u(\bar t,\bar x)\,.$$

The transformation of these functions under (1.6) is straightforward;

$$\bar\rho(\bar t,\bar x) = \rho(t,x)\,, \qquad \bar p(\bar t,\bar x) = p(t,x)\,, \qquad \bar u^i(\bar t,\bar x) = u^j(t,x) - v^i\,; \tag{1.7}$$

the functions ρ and p transform as scalar fields, u as a vector field (when regarded as fields on Galilean spacetime).

The Euler equations are invariant under Galilean transformations. Consider, first, equation (1.5a):

$$\begin{aligned}
&\frac{\partial}{\partial t}\,\rho(t,x) + \frac{\partial}{\partial x^k}\Big(\rho(t,x) u^k(t,x)\Big) \\
&\quad = \left[\frac{\partial \bar{\mathsf{t}}}{\partial t}\,\frac{\partial}{\partial \bar t} + \frac{\partial \bar{\mathsf{x}}^i}{\partial t}\,\frac{\partial}{\partial \bar x^i}\right] \bar\rho(\bar t,\bar x) + \left[\frac{\partial \bar{\mathsf{t}}}{\partial x^k}\,\frac{\partial}{\partial \bar t} + \frac{\partial \bar{\mathsf{x}}^i}{\partial x^k}\,\frac{\partial}{\partial \bar x^i}\right]\Big(\bar\rho(\bar t,\bar x)\big(\bar u^k(\bar t,\bar x) + v^k\big)\Big) \\
&\quad = \frac{\partial}{\partial \bar t}\,\bar\rho(\bar t,\bar x) - v^i \frac{\partial}{\partial \bar x^i}\,\bar\rho(\bar t,\bar x) + \frac{\partial}{\partial \bar x^k}\Big(\bar\rho(\bar t,\bar x)\bar u^k(\bar t,\bar x) + \bar\rho(\bar t,\bar x) v^k\Big) \\
&\quad = \frac{\partial}{\partial \bar t}\,\bar\rho(\bar t,\bar x) + \frac{\partial}{\partial \bar x^k}\Big(\bar\rho(\bar t,\bar x)\bar u^k(\bar t,\bar x)\Big)\,.
\end{aligned} \tag{1.8}$$

Hence, if (1.5a) holds w.r.t. the first system of coordinates, it holds w.r.t. second system of coordinates as well.

To prove (1.8) we have made use of the relations

$$\frac{\partial}{\partial t} = \frac{\partial \bar{\mathsf{t}}}{\partial t}\,\frac{\partial}{\partial \bar t} + \frac{\partial \bar{\mathsf{x}}^i}{\partial t}\,\frac{\partial}{\partial \bar x^i} = \frac{\partial}{\partial \bar t} - v^i \frac{\partial}{\partial \bar x^i}\,, \tag{1.9a}$$

$$\frac{\partial}{\partial x^k} = \frac{\partial \bar{\mathsf{t}}}{\partial x^k}\,\frac{\partial}{\partial \bar t} + \frac{\partial \bar{\mathsf{x}}^i}{\partial x^k}\,\frac{\partial}{\partial \bar x^i} = \frac{\partial}{\partial \bar x^k}\,. \tag{1.9b}$$

These relations can be rewritten as

$$\frac{\partial}{\partial t} = \frac{\partial}{\partial \bar{t}} - v\bar{\nabla}\,, \qquad \bar{\nabla} = \nabla\,,$$

and used to show invariance of the Lagrangian derivative:

$$\frac{d}{dt} = \frac{\partial}{\partial t} + u\nabla = \frac{\partial}{\partial \bar{t}} - v\bar{\nabla} + (\bar{u}+v)\bar{\nabla} = \frac{\partial}{\partial \bar{t}} + \bar{u}\bar{\nabla} = \frac{d}{d\bar{t}}\,.$$

This invariance leads directly to the invariance of equation (1.5b); we use the form (1.5b′).

$$\begin{aligned}&\rho(t,x)\,\frac{d}{dt}\,u(t,x) + \nabla p(t,x)\\ &\quad = \bar{\rho}(\bar{t},\bar{x})\,\frac{d}{d\bar{t}}\big(\bar{u}(\bar{t},\bar{x}) + v\big) + \bar{\nabla}\bar{p}(\bar{t},\bar{x}) = \bar{\rho}(\bar{t},\bar{x})\,\frac{d}{d\bar{t}}\,\bar{u}(\bar{t},\bar{x}) + \bar{\nabla}\bar{p}(\bar{t},\bar{x})\,.\end{aligned}$$

Finally, to complete the proof that (1.5) is invariant under a Galilean boost, we note that

$$\bar{p}(\bar{t},\bar{x}) = p(t,x) = \mathsf{p}\big(\rho(t,x)\big) = \mathsf{p}\big(\bar{\rho}(\bar{t},\bar{x})\big)\,.$$

In summary, like every good law of (Newtonian) physics, the Euler equations (1.5) are invariant under Galilean transformations.

1.4 Propagation of perturbations

Let us consider a simple ('equilibrium') solution of the Euler equations (1.5):

$$\rho(t,x) \equiv \mathring{\rho} = \text{const}\,, \qquad u(t,x) \equiv 0\,, \qquad p(t,x) \equiv \mathring{p} = \mathsf{p}(\mathring{\rho})\,. \tag{1.10}$$

A small perturbation of this solution is given by

$$\rho(t,x) = \mathring{\rho}\big(1+\phi(t,x)\big)\,, \qquad |\phi(t,x)| \ll 1\,, \tag{1.11a}$$

$$|u(t,x)| \ll 1\,. \tag{1.11b}$$

Using this ansatz we find

$$\frac{d}{dt}\,\rho = \frac{\partial}{\partial t}\big(\mathring{\rho}(1+\phi)\big) + u\nabla\big(\mathring{\rho}(1+\phi)\big) = \mathring{\rho}\Big[\frac{\partial\phi}{\partial t} + u\nabla\phi\Big]\,,$$

which we may insert into the continuity equation (1.5a′) to obtain

$$\mathring{\rho}\Big[\frac{\partial\phi}{\partial t}+u\nabla\phi\Big]=-\mathring{\rho}(1+\phi)\,\nabla u$$

and thus

$$\frac{\partial\phi}{\partial t}=-(1+\phi)\nabla u-u\nabla\phi\,. \tag{1.12a}$$

For (1.5b′) we get

$$\mathring{\rho}(1+\phi)\left[\frac{\partial u}{\partial t}+u\nabla u\right]=-\nabla p=-\mathsf{p}'(\rho)\,\nabla\rho=\left[-\mathsf{p}'(\mathring{\rho})-\mathsf{p}''(\mathring{\rho})\mathring{\rho}\,\phi-\cdots\right]\mathring{\rho}\nabla\phi$$

and therefore

$$\frac{\partial u}{\partial t}=-u\nabla u-\left[c_{\mathrm{s}}^2+\mathsf{p}''(\mathring{\rho})\mathring{\rho}\,\phi+\cdots\right](1+\phi)^{-1}\,\nabla\phi\,. \tag{1.12b}$$

In this equation we have introduced the quantity c_{s} which has the dimension of a velocity,

$$\mathsf{p}'(\mathring{\rho})=\frac{d\mathsf{p}}{d\rho}\Big|_{\rho=\mathring{\rho}}=c_{\mathrm{s}}^2\,; \tag{1.13}$$

we will identify it with the speed of sound.

Provided that the spatial gradients of ϕ and u are of the same small order as ϕ and u themselves, we may neglect the terms of higher order and thereby obtain a simple linear system of equations,

$$\frac{\partial\phi}{\partial t}=-\nabla u\,, \tag{1.14a}$$

$$\frac{\partial u}{\partial t}=-c_{\mathrm{s}}^2\,\nabla\phi\,. \tag{1.14b}$$

From (1.14) we infer that

$$\Box\phi=-\frac{1}{c_{\mathrm{s}}^2}\,\frac{\partial^2}{\partial t^2}\,\phi+\Delta\phi=0\,, \tag{1.15}$$

i.e., $\phi=\phi(t,x)$ satisfies a wave equation. The solutions of (1.15) are compression and rarefaction waves that propagate with the speed c_{s}. Similarly,

$$\Box u=-\frac{1}{c_{\mathrm{s}}^2}\,\frac{\partial^2}{\partial t^2}\,u+\Delta u=0\,.$$

The main observation for us is the following: The Euler equations (1.5) are invariant under Galilean transformations. From these equations we derive an equation describing the propagation of perturbations of an equilibrium state of the medium; this is the wave equation (1.15). But this equation is *not* invariant under Galilean transformations. To see this we consider a Galilean boost and use (1.9) to find

$$\begin{aligned}
\Box\,\phi(t,x) &= -\frac{1}{c_{\mathrm{s}}^2}\,\partial_t^2\phi(t,x) + \delta^{ij}\,\partial_i\partial_j\phi(t,x) \\
&= -\frac{1}{c_{\mathrm{s}}^2}\left(\bar\partial_t - v^i\bar\partial_i\right)\left(\bar\partial_t - v^j\bar\partial_j\right)\bar\phi(\bar t,\bar x) + \delta^{ij}\,\bar\partial_i\bar\partial_j\bar\phi(\bar t,\bar x) \\
&= -\frac{1}{c_{\mathrm{s}}^2}\,\bar\partial_t^2\bar\phi(\bar t,\bar x) + \frac{2}{c_{\mathrm{s}}^2}\,v^i\bar\partial_i\,\bar\partial_t\bar\phi(\bar t,\bar x) - \frac{v^iv^j}{c_{\mathrm{s}}^2}\,\bar\partial_i\bar\partial_j\bar\phi(\bar t,\bar x) + \delta^{ij}\,\bar\partial_i\bar\partial_j\bar\phi(\bar t,\bar x) \\
&= \bar\Box\,\bar\phi(\bar t,\bar x) + \frac{2}{c_{\mathrm{s}}^2}\,v^i\bar\partial_i\,\bar\partial_t\bar\phi(\bar t,\bar x) - \frac{v^iv^j}{c_{\mathrm{s}}^2}\,\bar\partial_i\bar\partial_j\bar\phi(\bar t,\bar x) \;\neq\; \bar\Box\,\bar\phi(\bar t,\bar x)\,.
\end{aligned}$$

As a simple an explicit example let us consider a one-dimensional problem. In one (spatial) dimension, the wave equation

$$\Box\phi(t,x) = 0 \tag{1.16}$$

reduces to

$$\left[-\frac{1}{c_{\mathrm{s}}^2}\,\partial_t^2 + \partial_x^2\right]\phi(t,x) = \left[-\frac{1}{c_{\mathrm{s}}}\,\partial_t + \partial_x\right]\left[\frac{1}{c_{\mathrm{s}}}\,\partial_t + \partial_x\right]\phi(t,x) = 0\,. \tag{1.17}$$

Its general solution can be represented as the linear combination of a wave traveling to the right and a wave traveling to the left, i.e.,

$$\phi(t,x) = \lambda_1\phi_1(x - c_{\mathrm{s}}t) + \lambda_2\phi_2(x + c_{\mathrm{s}}t)\,, \tag{1.18}$$

where ϕ_1 and ϕ_2 are arbitrary functions and $\lambda_1,\lambda_2 \in \mathbb{R}$.

In a boosted inertial frame the equation describing the perturbation is given by

$$\bar\Box\bar\phi(\bar t,\bar x) + \frac{2}{c_{\mathrm{s}}^2}\,v^i\bar\partial_i\,\bar\partial_t\bar\phi(\bar t,\bar x) - \frac{v^iv^j}{c_{\mathrm{s}}^2}\,\bar\partial_i\bar\partial_j\bar\phi(\bar t,\bar x) = 0\,. \tag{$\overline{1.16}$}$$

In one dimension this reduces to

$$\begin{aligned}
&\left[-\frac{1}{c_{\mathrm{s}}^2}\,\bar\partial_t^2 + \bar\partial_x^2\right]\bar\phi(\bar t,\bar x) + \frac{2}{c_{\mathrm{s}}^2}\,v\,\bar\partial_x\bar\partial_t\bar\phi(\bar t,\bar x) - \frac{v^2}{c_{\mathrm{s}}^2}\,\bar\partial_x^2\bar\phi(\bar t,\bar x) \\
&\quad = \left(1 - \frac{v^2}{c_{\mathrm{s}}^2}\right)\left[-\frac{1}{c_{\mathrm{s}}+v}\,\bar\partial_t + \bar\partial_x\right]\left[\frac{1}{c_{\mathrm{s}}-v}\,\bar\partial_t + \bar\partial_x\right]\bar\phi(\bar t,\bar x) = 0\,,
\end{aligned} \tag{$\overline{1.17}$}$$

where v is merely a number instead of a vector. The general solution of ($\overline{1.17}$) is

$$\bar{\phi}(\bar{t},\bar{x}) = \bar{\lambda}_1\bar{\phi}_1\big(\bar{x}-(c_\mathrm{s}-v)\bar{t}\big) + \bar{\lambda}_2\bar{\phi}_2\big(\bar{x}+(c_\mathrm{s}+v)\bar{t}\big)\ , \tag{$\overline{1.18}$}$$

where again $\bar{\phi}_1$ and $\bar{\phi}_2$ are arbitrary functions and $\bar{\lambda}_1, \bar{\lambda}_2 \in \mathbb{R}$. The solution is thus a linear combination of a wave traveling at the speed $c_\mathrm{s} - v$ and a wave traveling at the speed $c_\mathrm{s} + v$.

As a simple special case consider an inertial frame whose relative velocity to the original inertial frame is $v = c_\mathrm{s}$. Then ($\overline{1.18}$) reads

$$\bar{\phi}(\bar{t},\bar{x}) = \bar{\lambda}_1\bar{\phi}_1(\bar{x}) + \bar{\lambda}_2\bar{\phi}_2\big(\bar{x}+2c_\mathrm{s}\bar{t}\big)\ .$$

This makes perfect sense. An observer who is traveling at the speed c_s observes a class of waves that do not propagate at all (i.e., “standing waves”), and and a class of waves that propagate in the direction opposite to the direction of motion, which are waves with velocity $2c_\mathrm{s}$. These two classes are represented by $\bar{\phi}_1$ and $\bar{\phi}_2$, respectively.

The example shows that inertial frames need not necessarily be equivalent. Clearly, the *fundamental* laws of (Newtonian) physics must be the same under a change of inertial frame (Galilean transformation). But if there exists a medium (represented by a ‘background’ solution), then this medium automatically distinguishes a frame of reference (“absolute space”); *derived* equations (which rely on the existence of a background solution) might take a distinguished (simple) form with respect to the distinguished frame of reference. (In the inertial frame (t, x), where the fluid is at rest, the equation for the perturbation is (1.16); in a boosted inertial frame $(\bar{t}, \bar{x})$ we obtain ($\overline{1.16}$); the rest frame of the fluid is obviously distinguished.)

Let us conclude this section by analyzing the properties of the wave equation in some more detail. As has been demonstrated above, the wave equation

$$\Box\phi(t,x) = -\frac{1}{c_\mathrm{s}^2}\,\partial_t^2\phi(t,x) + \delta^{ij}\,\partial_i\partial_j\phi(t,x) = 0 \tag{1.19}$$

is not invariant under Galilean transformations. However, there exists a class of transformations under which (1.19) is in fact invariant. To see this let us first define

$$\eta_\mathrm{s} = \begin{pmatrix} -\frac{1}{c_\mathrm{s}^2} & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix}\ . \tag{1.20}$$

In addition we set $x^0 = t$ and thus $\partial_0 = \partial_t$. Using this notation the wave equation (1.19) becomes

$$\Box\,\phi\big(x^0,x^1,x^2,x^3\big) = \eta_{\rm s}^{\mu\nu}\,\partial_\mu\partial_\nu\,\phi\big(x^0,x^1,x^2,x^3\big) = 0\,. \qquad (1.19')$$

It is common to specify the arguments of the functions collectively; e.g., instead of $\phi\big(x^0,x^1,x^2,x^3\big)$ we will write $\phi(x^\sigma)$.

Consider a (linear) change of coordinates

$$\bar{x}^\mu = \bar{\mathsf{x}}^\mu(x^\sigma) = [L_{\rm s}]^\mu{}_\nu\,x^\nu\,. \qquad (1.21)$$

This implies

$$\partial_\mu = \frac{\partial}{\partial x^\mu} = \frac{\partial\bar{\mathsf{x}}^\nu}{\partial x^\mu}\,\frac{\partial}{\partial\bar{x}^\nu} = \frac{\partial\bar{\mathsf{x}}^\nu}{\partial x^\mu}\,\bar{\partial}_\nu = [L_{\rm s}]^\nu{}_\mu\,\bar{\partial}_\nu\,.$$

Furthermore, the function $\phi(x^\sigma)$ is assumed to transform like a scalar function, i.e., $\bar{\phi}(\bar{x}^\sigma) = \phi(x^\sigma)$.

A straightforward computation now leads to

$$\begin{aligned}\Box\,\phi(x^\sigma) &= \eta_{\rm s}^{\mu\nu}\,\partial_\mu\partial_\nu\,\phi(x^\sigma) = \eta_{\rm s}^{\mu\nu}\,\partial_\mu\left([L_{\rm s}]^\beta{}_\nu\,\bar{\partial}_\beta\,\bar{\phi}(\bar{x}^\sigma)\right)\\ &= \eta_{\rm s}^{\mu\nu}\,[L_{\rm s}]^\beta{}_\nu\,\partial_\mu\bar{\partial}_\beta\,\bar{\phi}(\bar{x}^\sigma) = \eta_{\rm s}^{\mu\nu}\,[L_{\rm s}]^\alpha{}_\mu\,[L_{\rm s}]^\beta{}_\nu\,\bar{\partial}_\alpha\bar{\partial}_\beta\,\bar{\phi}(\bar{x}^\sigma)\\ &\overset{!}{=} \bar{\Box}\,\bar{\phi}(\bar{x}^\sigma) = \eta_{\rm s}^{\alpha\beta}\,\bar{\partial}_\alpha\bar{\partial}_\beta\,\bar{\phi}(\bar{x}^\sigma)\,,\end{aligned}$$

where the symbol $\overset{!}{=}$ means 'is required to be equal to'. Consequently, invariance of the d'Alembertian is equivalent to requiring, that

$$[L_{\rm s}]^\alpha{}_\mu\,[L_{\rm s}]^\beta{}_\nu\,\eta_{\rm s}^{\mu\nu} \;=\; \eta_{\rm s}^{\alpha\beta}\,, \qquad (1.22)$$

which means that $\eta_{\rm s}$ remains invariant under (1.21).

An example for a transformation (1.21) satisfying (1.22) is

$$\begin{pmatrix}\bar{t}\\ \bar{x}^1\\ \bar{x}^2\\ \bar{x}^3\end{pmatrix} = \begin{pmatrix}\gamma_{\rm s} & -\gamma_{\rm s}\,v/c_{\rm s}^2 & 0 & 0\\ -\gamma_{\rm s}\,v & \gamma_{\rm s} & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix}\begin{pmatrix}t\\ x^1\\ x^2\\ x^3\end{pmatrix}, \qquad (1.23)$$

where $\gamma_{\rm s} = (1-v^2/c_{\rm s}^2)^{-1}$. This is shown by a direct calculation. It is somewhat strange that the transformation (1.23) involves a non-trivial transformation of time, $t \mapsto \bar{t} = \gamma_{\mathrm{s}}\big(t - v\, c_{\mathrm{s}}^{-2}\, x^1\big)$. However, we don't have to wrack our brains about this curiosity now. We know that the equation (1.19) is not a fundamental equation of physics (but arising from a background solution); therefore, the transformation properties of (1.19) cannot tell us anything about fundamental physics.

1.5 Maxwell's equations and the wave equation

In standard notation, Maxwell's equations in vacuum read

$$\vec{\nabla}\vec{E} = 0 \qquad \vec{\nabla}\vec{B} = 0 \tag{1.24a}$$

$$\vec{\nabla} \times \vec{E} = -\partial_t \vec{B} \qquad \vec{\nabla} \times \vec{B} = \frac{1}{c^2}\,\partial_t \vec{E} \tag{1.24b}$$

where, of course, c is the speed of light,

$$c = 299\,792\,458\, m/s\ .$$

When we form $\vec{\nabla} \times \vec{\nabla} \times \vec{E}$, then we obtain

$$\vec{\nabla} \times \vec{\nabla} \times \vec{E} = \vec{\nabla}\,\underbrace{\vec{\nabla}\vec{E}}_{=0} - \Delta\vec{E}$$

on the one hand, and

$$\vec{\nabla} \times \vec{\nabla} \times \vec{E} = \vec{\nabla} \times \left(-\partial_t \vec{B}\right) = -\partial_t \left(\vec{\nabla} \times \vec{B}\right) = -\partial_t \left(\frac{1}{c^2}\,\partial_t \vec{E}\right)$$

on the other hand. Therefore,

$$\Box\vec{E} = -\frac{1}{c^2}\,\partial_t^2 \vec{E} + \Delta\vec{E} = 0\,. \tag{1.25}$$

Analogously, $\Box\vec{B} = 0$. Therefore, the components of the electric and the magnetic field, $\vec{E}$ and $\vec{B}$, satisfy the free wave equation.

1.6 The luminiferous and electromagnetic ether

So, this is the idea: Light and electromagnetic waves in general require a propagation medium; this medium is called luminiferous and electromagnetic *ether* (alternative spelling: aether).

Electromagnetic waves are like perturbations of the ether (which is the background solution). There exists some (non-linear) theory of the ether (a theory that has not been discovered yet); the linearized equations that describe the perturbations manifest themselves as the Maxwell equations (and thus as the wave equation for the electric and magnetic field). The (unknown) ether equations are certainly Galilean invariant; the Maxwell equations on the other hand are not; the validity of the Maxwell equations is restricted to a distinguished inertial frame, the rest frame of the ether.

Remark. A good model to have in mind is the one discussed in sections 1.3 and 1.4. There is one difference, however. In the case of a perfect fluid and the Euler equations, the perturbations are longitudinal waves (compression and rarefaction); in the electromagnetic case, perturbations of the ether are transversal waves (which is immediate from the polarization of light). As a consequence, a fluid cannot be a viable model for the ether; the ether must have more complicated properties.

Since light and electromagnetic waves are omnipresent, so is ether. In particular, (outer) space is filled with ether, and the earth travels through it on its course around the sun. What, then, is the movement of the earth relative to the ether? Does earth interact with ether? Does ether interact with matter?

To answer these questions several experiments were conceived and carried out in the nineteenth century. We focus on Hoek's experiment and the Michelson–Morley experiment.

1.7 Hoek

Based on an earlier experiment by Fizeau, in 1868, Martinus Hoek performed an experiment to measure the motion of the earth through the ether and/or the effect of matter on ether.

The basic idea is the detection of interference of light beams taking different paths. Suppose that the earth moves with velocity v through the ether. For a light beam that is parallel to the direction of motion of the earth, the velocity is $c - v$, when measured in the rest frame of the laboratory (earth); for a light beam that is antiparallel to the direction of motion of the earth, this velocity is $c + v$. Therefore, if a light ray makes a round-trip through a device of length

l, the travel time is

$$t = \frac{l}{c+v} + \frac{l}{c-v} = \frac{2l}{c} + \frac{2l}{c}\frac{v^2}{c^2} + O\big(v^4/c^4\big)\ . \tag{1.26}$$

The velocity of the earth in its course around the sun is approximately 30km/s; hence

$$\frac{v}{c} \simeq 10^{-4}\ , \qquad \frac{v^2}{c^2} \simeq 10^{-8}\ .$$

Therefore, since the travel time (1.26) differs from $2l/c$ at second order in v/c only, the effect to be measured is very small.

Hoek's experiment was designed in such a way that the expected effect was of first order in v/c, because, at that time, effects of second order in v/c were beyond experimental reach. Two light rays pass through vacuum (or air) in one direction and through a medium (water) in the opposite direction. Recall that the speed of light in a medium is

$$\frac{c}{n}\ ,$$

where n is the refractive index of the medium. In the simplest approach to the problem, we expect the travel times

$$\begin{aligned} t_1 &= \frac{l}{\frac{c}{n}+v} + \frac{l}{c-v} = \frac{(1+n)l}{c} + \frac{(1-n^2)l}{c}\frac{v}{c} + O\big(v^2/c^2\big) \\ t_2 &= \frac{l}{\frac{c}{n}-v} + \frac{l}{c+v} = \frac{(1+n)l}{c} - \frac{(1-n^2)l}{c}\frac{v}{c} + O\big(v^2/c^2\big) \end{aligned}$$

for the two light rays going through the apparatus in opposite directions. The difference in travel times (at first order in v/c) should manifest itself in interference fringes.

Remark. Since the direction of motion of the earth through ether is unknown, the experimental device could be rotated. (In addition, the experiment was performed several times several months apart—the earth could be at rest relative to ether at a certain time; since the earth revolves around the sun, some months later, it should be moving with 60km/s.

The experiment gave a null result. The motion of the earth through ether could not be detected. However, the explanation of this result was straightforward (and completely consistent with the earlier results by Fizeau). Ether interacts with matter which causes the phenomenon of 'ether drag':

Suppose that some matter moves with velocity v (w.r.t. the universal ether). Then the ether within the material is (partially) dragged along, so that the velocity of the 'local' ether is

$$v_{\text{ether in matter}} = d\,v\,; \tag{1.27}$$

d is the dragging coefficient. If $d = 0$ there is no dragging; if $0 < d < 1$ there is partial dragging; if $d = 1$ the dragging is complete.

Let us consider Hoek's experiment in the light of 'ether drag': If the velocity of the medium (water)[8] w.r.t. the universal ether is v, then its velocity w.r.t. the 'local' ether contained within it is $(1-d)v$. We thus obtain the travel times

$$t_1 = \frac{l}{\frac{c}{n} + (1-d)v} + \frac{l}{c-v} = \frac{(1+n)l}{c} + \frac{\big(1-(1-d)n^2\big)l}{c}\,\frac{v}{c} + O\big(v^2/c^2\big)$$
$$t_1 = \frac{l}{\frac{c}{n} - (1-d)v} + \frac{l}{c+v} = \frac{(1+n)l}{c} - \frac{\big(1-(1-d)n^2\big)l}{c}\,\frac{v}{c} + O\big(v^2/c^2\big)$$

for the two light rays going through the apparatus in opposite directions.

The null result of the experiment is explained by partial ether drag, if and only if $1-(1-d)n^2 = 0$, which corresponds to a dragging coefficient of

$$d = 1 - \frac{1}{n^2}\,. \tag{1.28}$$

This was exactly the value suggested by Fresnel some decade earlier and confirmed earlier measurements and results by Fizeau.

Due to the effect of partial ether drag, Hook's experiment could not detect the movement of the earth w.r.t. the (universal) ether. The terms of first order in v/c disappear; the remaining effect would be of second order in v/c and could not be detected with Hoek's experimental means.

1.8 Michelson and Morley

Toward the end of the nineteenth century the possibility to construct a high precision interferometer to measure effects of quadratic order in v/c came into reach.

[8] In Hoek's experiment, the velocity of the medium coincides with the velocity of earth through the (universal) ether, since the medium is at rest w.r.t. the laboratory.

In the Michelson interferometer a light beam is divided into two parts that travel along orthogonal paths to meet again where interference is observed. Suppose that one of the two light rays moves in the direction of the movement of earth through ether; we call this ray the parallel ray, the other ray is called the perpendicular ray. In the laboratory frame, the velocity vectors are

$$u_\parallel = \begin{pmatrix} c \pm v \\ 0 \end{pmatrix} \tag{$\parallel$}$$

for the parallel ray and

$$u_\perp = \begin{pmatrix} 0 \\ \pm w \end{pmatrix}$$

for the perpendicular ray, where w is to be determined. A simple Galilean transformation show that $u_\perp$ reads

$$u'_\perp = \begin{pmatrix} v \\ \pm w \end{pmatrix}$$

in the rest frame of the ether. Since the absolute value of this velocity must be c, i.e.,

$$u'_\perp u'_\perp = v^2 + w^2 = c^2 ,$$

we find

$$w = \sqrt{c^2 - v^2} = c\sqrt{1 - \tfrac{v^2}{c^2}} .$$

Hence,

$$u_\perp = \begin{pmatrix} 0 \\ \pm\sqrt{c^2 - v^2} \end{pmatrix} . \tag{$\perp$}$$

Let us compute the travel times of the two light rays. We obtain

$$\begin{aligned} t_\parallel &= \frac{l}{c+v} + \frac{l}{c-v} = \frac{2l}{c} + \frac{2l}{c}\frac{v^2}{c^2} + O\big(v^4/c^4\big) , \\ t_\perp &= \frac{l}{\sqrt{c^2-v^2}} + \frac{l}{\sqrt{c^2-v^2}} = \frac{2l}{c} + \frac{l}{c}\frac{v^2}{c^2} + O\big(v^4/c^4\big) . \end{aligned}$$

The difference causes a phase shift between the two beams that produces interference fringes. These interference pattern should change if the interferometer is rotated; e.g.,

$$\Delta t_{0^\circ} = t_\parallel - t_\perp = +\frac{l}{c}\frac{v^2}{c^2} + O\big(v^4/c^4\big) , \tag{1.29a}$$

$$\Delta t_{90^\circ} = t_\perp - t_\parallel = -\frac{l}{c}\frac{v^2}{c^2} + O\big(v^4/c^4\big) ; \tag{1.29b}$$

therefore, detection of the movement of the earth through ether should be observed straightforwardly.

The first experiment was performed by Michelson in 1881. However, there was a fundamental error in Michelson's computations; he used the value $w = c$ to compute the travel time of the perpendicular ray. This error leads to an additional factor of 2 in (1.29), hence Michelson expected an effect twice the actual size. The error of his experimental design would have been sufficiently small to measure this larger effect, but the actual effect is merely half the size and Michelson's experiment was not precise enough to measure (1.29).

A reduction of experimental error was necessary to obtain reliable results. The experiment was repeated (with a better design) by Michelson and Morley in 1887 and again in the eighteen nineties. The experimental error was reassuringly small.

The experiment gave a *null result*.

The conclusion drawn by Michelson and Morley was that the earth drags the entire ether along, possibly gravitatively. (This view was consistent with an earlier suggestion by Stokes.) However, as had been known long before, the observation of stellar aberration is proof that this kind of ether drag is impossible. (The phenomenon of stellar aberration would be very different, if the light reaching us from the stars passed through an ether whose velocity changes with position.) Moreover, the connection of that kind of complete ether drag with the partial dragging established by Fizeau and Hoek would remain mysterious.

The attempts to explain the null result of the interferometry experiment became rather desperate. Since the laboratory where Michelson and Morley performed their experiment was situated in the basement of a building, it was suggested that there could be complete ether in such a surrounding. But experiments on mountains and in balloons (!) confirmed the null results.[9] As a last resort, Fitzgerald and Lorentz proposed the contraction of lengths in the direction of motion of earth through ether. However, the mechanism that caused such a contraction remained unclear.

In 1905, Einstein made a clear cut. The main idea is simple: There is no ether.

[9] The occasional non-null results were all shown to be due to bad experimental design.

CHAPTER 2

EINSTEIN

2.1 Lorentz transformations

This is the line of reasoning: There is no ether. The Maxwell equations are not 'derived' equations that describe the perturbations of a medium; the Maxwell equations are 'fundamental'. In other words: 'Laws of nature'. But then, obviously, the principle of relativity comes into play. If the principle of relativity holds, then the set of coordinate systems w.r.t. which the Maxwell equations take their standard form are the inertial frames of reference.

The coordinate transformations that leave invariant the Maxwell equations must therefore be the transformations that map one inertial frame into another. (These transformation will be given the name Lorentz/Poincaré transformations.)

So, then, what are the coordinate transformations that leave invariant the Maxwell equations?

Let us use the results of section 1.5. The Maxwell equations encompass the wave equation. It thus makes sense to begin by studying the coordinate transformations that leave invariant the wave equation, i.e.,

$$\Box\phi = -\frac{1}{c^2}\,\partial_t^2\phi + \Delta\phi = 0\,. \tag{2.1}$$

Like at the end of section 1.4 we begin by writing the wave equation in a convenient form. Using the Kronecker symbol (and the Einstein summation convention) the wave equation (2.1) becomes

$$\Box\phi(t,x) = -\frac{1}{c^2}\,\partial_t^2\phi(t,x) + \delta^{ij}\,\partial_i\partial_j\phi(t,x) = 0\,. \tag{2.2}$$

We define a coordinate x^0 that encodes time by

$$x^0 = ct\,.$$

Accordingly, instead of coordinates (t, x^1, x^2, x^2) we use a system of coordinates (x^0, x^1, x^2, x^3). It is standard in relativity to have indices that run from $0, \ldots, 3$ (instead of from $1, \ldots, 4$).

Remark. One can interpret the coordinate x^0 as measuring time in 'light meters' instead of in seconds (where a 'light meter' is the time that light takes to travel through one meter). More or less equivalently, one can measure time in seconds and length in light seconds.

Using (x^0, x^1, x^2, x^3), the wave equation reads

$$\Box\phi(x^0,x^1,x^2,x^3) = -\partial_0^2\,\phi(x^0,x^1,x^2,x^3) + \delta^{ij}\,\partial_i\partial_j\,\phi(x^0,x^1,x^2,x^3) = 0\,. \tag{2.3}$$

Let us define η to be the matrix

$$\eta = \begin{pmatrix} -1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix}\,. \tag{2.4}$$

Using this notation the wave equation (2.3) becomes

$$\Box\,\phi\big(x^0,x^1,x^2,x^3\big) = \eta^{\mu\nu}\,\partial_\mu\partial_\nu\,\phi\big(x^0,x^1,x^2,x^3\big) = 0\,. \tag{2.5}$$

It is common to specify the arguments of the functions collectively; e.g., instead of $\phi\big(x^0,x^1,x^2,x^3\big)$ we will write $\phi(x^\sigma)$.

Having completed the preparatory steps, let us now consider an arbitrary change of coordinates,

$$\bar{x}^\mu = \bar{\mathsf{x}}^\mu(x^\sigma)\,. \tag{2.6}$$

This implies

$$\partial_\mu = \frac{\partial}{\partial x^\mu} = \frac{\partial\bar{\mathsf{x}}^\nu}{\partial x^\mu}\,\frac{\partial}{\partial\bar{x}^\nu} = \frac{\partial\bar{\mathsf{x}}^\nu}{\partial x^\mu}\,\bar{\partial}_\nu\,.$$

Furthermore, the function $\phi(x^\sigma)$ is assumed to transform like a scalar function, i.e., $\bar\phi(\bar x^\sigma) = \phi(x^\sigma)$.

A straightforward computation now leads to

$$\begin{aligned}
\Box\,\phi(x^\sigma) &= \eta^{\mu\nu}\,\partial_\mu\partial_\nu\,\phi(x^\sigma) = \eta^{\mu\nu}\,\partial_\mu\left(\frac{\partial\bar{\mathsf{x}}^\beta}{\partial x^\nu}\,\bar\partial_\beta\,\bar\phi(\bar x^\sigma)\right) \\
&= \eta^{\mu\nu}\,\frac{\partial\bar{\mathsf{x}}^\beta}{\partial x^\nu}\,\partial_\mu\bar\partial_\beta\,\bar\phi(\bar x^\sigma) + \eta^{\mu\nu}\,\frac{\partial^2\bar{\mathsf{x}}^\beta}{\partial x^\mu\partial x^\nu}\,\bar\partial_\beta\,\bar\phi(\bar x^\sigma) \\
&= \eta^{\mu\nu}\,\frac{\partial\bar{\mathsf{x}}^\alpha}{\partial x^\mu}\,\frac{\partial\bar{\mathsf{x}}^\beta}{\partial x^\nu}\,\bar\partial_\alpha\bar\partial_\beta\,\bar\phi(\bar x^\sigma) + \eta^{\mu\nu}\,\frac{\partial^2\bar{\mathsf{x}}^\beta}{\partial x^\mu\partial x^\nu}\,\bar\partial_\beta\,\bar\phi(\bar x^\sigma)\,. \qquad (2.7)
\end{aligned}$$

If and only if the coordinate system $(\bar x^0, \bar x^1, \bar x^2, \bar x^3)$ is an inertial frame of reference, the wave equation is invariant under the change of coordinates (2.6), i.e.,

$$\Box\,\phi(x^\sigma) \stackrel{!}{=} \bar\Box\,\bar\phi(\bar x^\sigma) = \eta^{\alpha\beta}\bar\partial_\alpha\bar\partial_\beta\,\bar\phi(\bar x^\sigma)\,. \qquad (2.8)$$

Using (2.7) we conclude that invariance of the wave equation is equivalent to requiring that

$$\eta^{\mu\nu}\,\frac{\partial\bar{\mathsf{x}}^\alpha}{\partial x^\mu}\,\frac{\partial\bar{\mathsf{x}}^\beta}{\partial x^\nu} = \eta^{\alpha\beta}\,, \qquad \eta^{\mu\nu}\,\frac{\partial^2\bar{\mathsf{x}}^\beta}{\partial x^\mu\partial x^\nu} = 0\,. \qquad (2.9)$$

From these equations it is possible (though not simple) to conclude that the transformation (2.6) must satisfy

$$\frac{\partial\bar{\mathsf{x}}^\mu}{\partial x^\nu} = \text{const}\,,$$

i.e., the transformation must be linear. Let us define

$$L^\mu{}_\nu := \frac{\partial\bar{\mathsf{x}}^\mu}{\partial x^\nu} = \text{const}\,.$$

Then the coordinate transformation (2.6) reads

$$\bar x^\mu = L^\mu{}_\nu x^\nu\,. \qquad (2.10)$$

Remark. Strictly speaking, we conclude from $L^\mu{}_\nu = \partial\bar{\mathsf{x}}^\mu/\partial x^\nu = \text{const}$ that

$$\bar x^\mu = L^\mu{}_\nu x^\nu + \bar a^\mu\,, \qquad (2.10')$$

for some constant vector $\bar a^\mu$. However, this merely corresponds to an additional translation (in time and/or space). For simplicity we suppress this translational freedom for the moment.

In addition to (2.10), we conclude from (2.9) that invariance of the wave equation requires

$$L^{\alpha}{}_{\mu}\, L^{\beta}{}_{\nu}\, \eta^{\mu\nu} \;=\; \eta^{\alpha\beta}\,, \tag{2.11}$$

which means that η remains invariant under (2.10).

Let us finally denote the coordinate transformation we consider here by their well-known name:

Definition 2.1. *A* **Lorentz transformation** *is a coordinate transformation of spacetime that maps one inertial frame into another.*

Taking account of the definition of an inertial frame we find an alternative definition.

Definition 2.1′. *A* **Lorentz transformation** *is a coordinate transformation of spacetime that leaves invariant the laws of physics*[1] *.*

Remark. So far, the only (relativistic) law of physics we know are the Maxwell equations and the wave equation (which derives from the Maxwell equations). The wave equation is the one we use above.

Remark. Strictly speaking, a Lorentz transformation is coordinate transformation of spacetime that maps one inertial frame into another (or: leaves invariant the laws of physics) *and keeps the origin (of spacetime) fixed.* If this is not the case we call the transformation a **Poincaré transformation**; compare (2.10) and (2.10′).

Our considerations on the wave equation that lead us to (2.10) and (2.11) thus prove the following theorem:

Theorem 2.2. *The Lorentz transformations are uniquely characterized as the transformations*

$$\bar{x}^{\mu} = L^{\mu}{}_{\nu} x^{\nu}$$

that leave η invariant, i.e.,

$$L^{\mu}{}_{\alpha}\, L^{\nu}{}_{\beta}\, \eta^{\alpha\beta} \;=\; \eta^{\mu\nu}\,. \tag{2.12}$$

More commonly, equation (2.12) is written as

$$L^{\mu}{}_{\alpha}\, L^{\nu}{}_{\beta}\, \eta_{\mu\nu} \;=\; \eta_{\alpha\beta}\,, \tag{2.12′}$$

[1] By laws of physics we mean the actual (relativistic) laws and not the Newtonian approximations.

where $\eta_{\mu\nu}$ is again given as the matrix

$$\eta = \begin{pmatrix} -1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix} ,$$

see (2.4).

Remark. In matrix notation, equation (2.12′) reads

$$L^{\mathrm{T}}\,\eta\,L \;=\; \eta\,.$$

It is simple to see that the two representations (2.12) and (2.12′) are equivalent: We multiply (2.12) with $[L^{-1}]^{\sigma}{}_{\mu}$ and $[L^{-1}]^{\lambda}{}_{\nu}$ and obtain

$$[L^{-1}]^{\sigma}{}_{\mu}\,[L^{-1}]^{\lambda}{}_{\nu}\,L^{\mu}{}_{\alpha}\,L^{\nu}{}_{\beta}\,\eta^{\alpha\beta} \;=\; [L^{-1}]^{\sigma}{}_{\mu}\,[L^{-1}]^{\lambda}{}_{\nu}\,\eta^{\mu\nu}\,.$$

Since $[L^{-1}]^{\sigma}{}_{\mu}\,L^{\mu}{}_{\alpha} = \delta^{\sigma}{}_{\alpha}$ and $[L^{-1}]^{\lambda}{}_{\nu}\,L^{\nu}{}_{\beta} = \delta^{\lambda}{}_{\beta}$ we get

$$\eta^{\sigma\lambda} \;=\; [L^{-1}]^{\sigma}{}_{\mu}\,[L^{-1}]^{\lambda}{}_{\nu}\,\eta^{\mu\nu}\,.$$

A simple change of indices leads to

$$[L^{-1}]^{\mu}{}_{\alpha}\,[L^{-1}]^{\nu}{}_{\beta}\,\eta^{\alpha\beta} \;=\; \eta^{\mu\nu}\,. \tag{2.12^{-1}}$$

The interpretation of (2.12^{-1}) is obvious: If L is a Lorentz transformation, i.e., if L satisfies (2.12), then L^{-1} is a Lorentz transformation as well, i.e., L^{-1} satisfies (2.12) as well. (Note that (2.12) and (2.12^{-1}) are identical.) In matrix notation, equation (2.12^{-1}) reads

$$L^{-1}\,\eta^{-1}\,(L^{-1})^{\mathrm{T}} = \eta^{-1}\,.$$

Note that η and η^{-1} are identical as matrices; the reason for the choosing η^{-1} here will become clear later (when we discuss metrics).[2] Therefore,

$$\Big(L^{-1}\,\eta^{-1}\,(L^{-1})^{\mathrm{T}}\Big)^{-1} = \eta\,,$$

which implies

$$L^{\mathrm{T}}\,\eta\,L = \eta\,.$$

[2]The index structure of the Minkowski metric η is $\eta_{\mu\nu}$; the index structure of its inverse η^{-1} is $\eta^{\mu\nu}$. (Regarded as matrices the two are equal.) The reason is that the matrix identity $\eta^{-1}\eta = \mathbb{1}$ corresponds to $\eta^{\mu\nu}\eta_{\nu\sigma} = \delta^{\mu}{}_{\sigma}$ in index notation.

In index notation, we thus arrive at

$$L^{\mu}{}_{\alpha}\, L^{\nu}{}_{\beta}\, \eta_{\mu\nu} \;=\; \eta_{\alpha\beta}\;, \tag{2.13}$$

which is (2.12′).

Remark. The equivalence of (2.12) and (2.12′) is in fact rather obvious and does not need any particular work (however, we may consider the above a useful exercise). One simply takes into account the action of L on contravariant and covariant tensors; see appendix.

2.2 Examples of Lorentz transformations

In section 2.1 we have seen that the Lorentz transformations

$$\bar{x}^{\mu} = L^{\mu}{}_{\nu} x^{\nu} \tag{2.14}$$

are those transformations that leave η invariant, i.e.,

$$L^{\mu}{}_{\alpha}\, L^{\nu}{}_{\beta}\, \eta_{\mu\nu} \;=\; \eta_{\alpha\beta}\;. \tag{2.15}$$

Recall that $\eta_{\mu\nu}$ is given as the matrix

$$\eta = \begin{pmatrix} -1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix}. \tag{2.16}$$

In matrix notation, equation (2.15) reads

$$L^{\mathrm{T}}\, \eta\, L = \eta\;. \tag{2.15′}$$

In this section we will see how Lorentz transformations actually look like.

The first question we ask is: How many Lorentz transformations are there? Every Lorentz transformation is represented by a (4×4) matrix, which makes 16 unknowns. Looking at equation (2.15) we see that there are 10 equations that L must satisfy (since η is a symmetric matrix, 6 of the 16 equations are redundant). $16 - 10 = 6$. In other words, we expect that *the Lorentz transformations form a 6-parameter family.*

More specifically, the Lorentz transformations form a **6-parameter group**.[3] The group property simply amounts to the statement that the composition of Lorentz transformations is again a Lorentz transformation. To see this, let L_1 and L_2 be Lorentz transformations, i.e., $L_1^{\mathrm{T}}\,\eta\,L_1 = \eta$ and $L_2^{\mathrm{T}}\,\eta\,L_2 = \eta$. Then $(L_1L_2)^{\mathrm{T}}\,\eta\,L_1L_2 = L_2^{\mathrm{T}}L_1^{\mathrm{T}}\,\eta\,L_1L_2 = \eta$; in other words, L_1L_2 is a Lorentz transformation.

To derive the Lorentz transformations from (2.15) we make an ansatz. We choose a matrix L such that is trivial apart from a two-by-two matrix block. These are six[4] possibilities, namely

$$\begin{pmatrix} L^0{}_0 & L^0{}_1 & 0 & 0 \\ L^1{}_0 & L^1{}_1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} L^0{}_0 & 0 & L^0{}_2 & 0 \\ 0 & 1 & 0 & 0 \\ L^2{}_0 & 0 & L^2{}_2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} L^0{}_0 & 0 & 0 & L^0{}_3 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ L^3{}_0 & 0 & 0 & L^3{}_3 \end{pmatrix},$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & L^1{}_1 & L^1{}_2 & 0 \\ 0 & L^2{}_1 & L^2{}_2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & L^1{}_1 & 0 & L^1{}_3 \\ 0 & 0 & 1 & 0 \\ 0 & L^3{}_1 & 0 & L^3{}_3 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & L^2{}_2 & L^2{}_3 \\ 0 & 0 & L^3{}_2 & L^3{}_3 \end{pmatrix}.$$

Provided that the ansatz is successful (i.e., provided that there exist Lorentz transformations with this structure), we automatically obtain the complete six-parameter family of Lorentz transformations.

Let us denote the non-trivial (2×2) block by

$$K = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

irrespective of where it sits within the matrix L. Using the ansatz for L in (2.15), i.e., in $L^{\mathrm{T}}\,\eta\,L = \eta$, we find that a matrix L of the (2×2) block form is a Lorentz transformation if and only if

$$K^{\mathrm{T}} \begin{pmatrix} \mp 1 & 0 \\ 0 & 1 \end{pmatrix} K = \begin{pmatrix} \mp 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{2.17}$$

There are two cases:

[3] This six-parameter group is the so-called Lorentz group, which is either denoted by $\mathcal{L}$ or by $\mathsf{O}(3,1)$.

[4] Oh, I see.

+ The sign is a plus sign, if the (2×2) block K is a part of the spatial components of the matrix L; this corresponds to the latter three of the six possibilities above.
− The sign is a minus sign, if the (2×2) block K involves time and space, i.e., the 0-component and one spatial component; this corresponds to the first three of the six possibilities above.

The + **case** is particularly simple. In this case, equation (2.17) reduces to

$$K^{\mathrm{T}} K = \mathbb{1} \,, \tag{2.18}$$

which means that K must be an orthogonal matrix, i.e., $K \in \mathsf{O}(2)$. If we require the transformation to preserve the orientation of the coordinate, then we restrict ourselves to $K \in \mathsf{SO}(2)$, i.e., to the **rotations**, i.e.,

$$K = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \,, \tag{2.19}$$

with $\varphi \in [0, 2\pi)$.

We have reached the conclusion that rotations are Lorentz transformations; this agrees with our intuition: A rotation (by a constant angle) changes one inertial frame into another.[5] By using our ansatz we have derived three different rotations: A rotation around the x^1-axis, a rotation around the x^2-axis, and a rotation around the x^3, i.e.,

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi & 0 \\ 0 & \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} , \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\varphi & 0 & -\sin\varphi \\ 0 & 0 & 1 & 0 \\ 0 & \sin\varphi & 0 & \cos\varphi \end{pmatrix} , \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\varphi & -\sin\varphi \\ 0 & 0 & \sin\varphi & \cos\varphi \end{pmatrix} .$$

Clearly, we can compose these three elementary rotations to a general rotation in space; therefore,

$$L = \begin{pmatrix} 1 & \\ & \boxed{R} \end{pmatrix} \,, \tag{2.20}$$

where $R = (R^i{}_j)_{i,j}$ is a (3×3) rotation matrix, i.e., $R \in \mathsf{SO}(3)$, is a Lorentz transformation. For these rotations, (2.14) becomes

$$\begin{aligned} \bar{x}^0 &= x^0 \,, \\ \bar{x}^i &= R^i{}_j x^j \,. \end{aligned}$$

[5] Evidently, rotations are Galilean transformations and Lorentz transformations at the same time.

Note again that there exist three rotational degrees of freedom, since the group $\mathsf{SO}(3)$ is a three-parameter group (characterized, e.g., by the Euler angles).

Let us now consider the **− case**, i.e., (2.17) with the minus sign. We obtain

$$K^{\mathrm{T}}\begin{pmatrix}-1 & 0\\ 0 & 1\end{pmatrix}K = \begin{pmatrix}a & c\\ b & d\end{pmatrix}\begin{pmatrix}-1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}a & b\\ c & d\end{pmatrix} = \begin{pmatrix}-a^2+c^2 & -ab+cd\\ -ab+cd & -b^2+d^2\end{pmatrix},$$

which is required to satisfy

$$\begin{pmatrix}-a^2+c^2 & -ab+cd\\ -ab+cd & -b^2+d^2\end{pmatrix} = \begin{pmatrix}-1 & 0\\ 0 & 1\end{pmatrix}. \tag{2.21}$$

From the equation $a^2 - c^2 = 1$ in (2.21) we conclude that there exists $u \in \mathbb{R}$ such that

$$a = \cosh u\,, \qquad c = -\sinh u\,.$$

It is not necessary to restrict oneself to the case $a > 0$ ($a = -\cosh u$ would be admissible as well); however, we choose to study Lorentz transformation that keep the direction of time fixed.; it is not difficult to see that this requires $a > 0$. (The minus in front of $\sinh u$ is merely a matter of convenience and not a restriction.) Analogously, the equation $-b^2 + d^2 = 1$ in (2.21) implies

$$b = -\sinh w\,, \qquad d = \cosh w$$

for some $w \in \mathbb{R}$. We choose not to consider spatial reflections, so that we assume $d > 0$. Inserting these relations into the remaining equation in (2.21) yields

$$\sinh u \cosh w - \cosh u \sinh w = \sinh(u-w) \overset{!}{=} 0\,,$$

from which we easily conclude that

$$u = w\,.$$

Therefore,

$$K = \begin{pmatrix}a & b\\ c & d\end{pmatrix} = \begin{pmatrix}\cosh u & -\sinh u\\ -\sinh u & \cosh u\end{pmatrix}. \tag{2.22}$$

Using K in the ansatzes we obtain the three remaining Lorentz transformations, i.e.,

$$L = \begin{pmatrix}\cosh u & -\sinh u & 0 & 0\\ -\sinh u & \cosh u & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix}, \tag{2.23}$$

and the two other ones where the (2×2) block involves x^2 or x^3 instead of x^1, i.e.,

$$\begin{pmatrix} \cosh u & 0 & -\sinh u & 0 \\ 0 & 1 & 0 & 0 \\ -\sinh u & 0 & \cosh u & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} \cosh u & 0 & 0 & -\sinh u \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\sinh u & 0 & 0 & \cosh u \end{pmatrix}.$$

Thereby we have successfully completed our search for the Lorentz transformations. We have found six elementary Lorentz transformations; because of the group property, every Lorentz transformation can be represented as a combination (composition) of these six elementary transformations.

Remark. The parameter u is called *rapidity*.

2.3 Lorentz boosts

Before we proceed, let us reintroduce dimensions into our considerations. The coordinate x^0 is defined as $x^0 = ct$; therefore, the Lorentz transformation (2.14), i.e., $\bar{x}^\mu = L^\mu{}_\nu x^\nu$, which reads

$$\begin{pmatrix} \bar{x}^0 \\ \bar{x}^1 \\ \bar{x}^2 \\ \bar{x}^3 \end{pmatrix} = \begin{pmatrix} \cosh u & -\sinh u & 0 & 0 \\ -\sinh u & \cosh u & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix}$$

can be rewritten as

$$\begin{pmatrix} \bar{t} \\ \bar{x}^1 \\ \bar{x}^2 \\ \bar{x}^3 \end{pmatrix} = \begin{pmatrix} \cosh u & -c^{-1}\sinh u & 0 & 0 \\ -c\sinh u & \cosh u & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} t \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} \tag{2.24}$$

when we use t (and $\bar{t}$) instead of x^0 (and $\bar{x}^0$).

It is not difficult to interpret this Lorentz transformation. Let us denote the inertial frame with coordinates (t, x) by X and the inertial frame with coordinates $(\bar{t}, \bar{x})$ by $\bar{X}$. Take a family of particles that are at rest w.r.t. the inertial observer (inertial frame of reference) $\bar{X}$; these particles are described by $\bar{x}^1 = \text{const}$, $\bar{x}^2 = \text{const}$, $\bar{x}^3 = \text{const}$ in the coordinate system $\bar{X}$. According to (2.24), since $\bar{x}^1 = -c(\sinh u)\,t + (\cosh u)\,x^1$, the coordinates the inertial observer X ascribes to these particles are

$$x^1 = \text{const} + \underbrace{c \tanh u}_{v}\, t\,, \qquad x^2 = \text{const} \qquad x^2 = \text{const}\,, \tag{2.25}$$

i.e., the particles are in uniform motion in the direction of the x^1-axis with some constant velocity v which is given by

$$v = c \tanh u\,. \tag{2.26}$$

Consequently, v is the constant velocity of the inertial observer $\bar{X}$ as seen by the observer X. (Note that a motion in the direction of the positive x^1-axis is described by positive values of v, while a motion in the direction of the negative x^1-axis is described by negative values of v.) We draw an interesting conclusion from the formula (2.26): Since the modulus of $\tanh u$ is always less than 1, we find

$$|v| < c\,. \tag{2.27}$$

This means that if X and $\bar{X}$ are inertial observers, then the relative velocity $|v|$ describing their relative motion must necessarily be less than the speed of light.

Definition 2.3. *The* ***γ-factor*** *associated with the relative velocity $|v|$ is defined as*

$$\gamma = \gamma(v) = \frac{1}{\sqrt{1 - v^2/c^2}}\,. \tag{2.28}$$

Rewriting (2.24) in terms of v and $\gamma(v)$ yields the standard form for the **Lorentz boost** in x^1-direction:

$$\begin{pmatrix} \bar{t} \\ \bar{x}^1 \\ \bar{x}^2 \\ \bar{x}^3 \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma v/c^2 & 0 & 0 \\ -\gamma v & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} t \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} \tag{2.29}$$

A **Lorentz boost** is the prototype of a Lorentz transformation. Often the terms 'Lorentz boost' and 'Lorentz transformation' are used synonymously and interchangeably.[6] Lorentz boosts involve a change of both the time and the spatial coordinates; they are thus the special relativistic analog of the Galilean boosts. This is because Lorentz boosts describe the change of coordinates between an observer X and an observer $\bar{X}$ that moves with constant velocity v w.r.t. X.

[6] Admittedly, this is a bit misleading since also rotations are Lorentz transformations. But rotations are boring transformations, so it is easy to turn a blind eye on them.

Remark. For all inertial observers the speed of light is equal to c. This is already implicit in our assumptions, since we have required the wave equation (which contains c) to hold w.r.t. all inertial frames. As an exercise we can check consistency with (2.29). To that end consider a photon which moves according to $x^1 = ct$ w.r.t. the frame X. In coordinates $\bar{X}$ we obtain $\bar{t} = \gamma(1-v/c)t$ and $\bar{x}^1 = \gamma(1-v/c)ct$ from (2.29). Consequently, $\bar{x}^1 = c\bar{t}$, i.e., also for the observer $\bar{X}$ the photon moves with velocity c.

Let us summarize. A Lorentz boost in x^1-direction is the coordinate transformation between inertial observers in uniform relative motion; the inertial observer $\bar{X}$ is seen by X to move with velocity v in the direction of x^1.

In particular, the origin of $\bar{X}$, i.e., $\bar{x} = \bar{o}$ is seen by X to move according to

$$\bar{o} : x(t) = \begin{pmatrix} vt \\ 0 \\ 0 \end{pmatrix} .$$

The Lorentz boosts associated with the other axes are completely analogous. Here is the complete list:

Lorentz boost in direction of x^1:

$$\bar{o} : x(t) = \begin{pmatrix} vt \\ 0 \\ 0 \end{pmatrix} \qquad \begin{pmatrix} \bar{t} \\ \bar{x}^1 \\ \bar{x}^2 \\ \bar{x}^3 \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma v/c^2 & 0 & 0 \\ -\gamma v & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} t \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} \tag{2.30a}$$

Lorentz boost in direction of x^2:

$$\bar{o} : x(t) = \begin{pmatrix} 0 \\ vt \\ 0 \end{pmatrix} \qquad \begin{pmatrix} \bar{t} \\ \bar{x}^1 \\ \bar{x}^2 \\ \bar{x}^3 \end{pmatrix} = \begin{pmatrix} \gamma & 0 & -\gamma v/c^2 & 0 \\ 0 & 1 & 0 & 0 \\ -\gamma v & 0 & \gamma & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} t \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} \tag{2.30b}$$

Lorentz boost in direction of x^3:

$$\bar{o} : x(t) = \begin{pmatrix} 0 \\ 0 \\ vt \end{pmatrix} \qquad \begin{pmatrix} \bar{t} \\ \bar{x}^1 \\ \bar{x}^2 \\ \bar{x}^3 \end{pmatrix} = \begin{pmatrix} \gamma & 0 & 0 & -\gamma v/c^2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\gamma v & 0 & 0 & \gamma \end{pmatrix} \begin{pmatrix} t \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} \tag{2.30c}$$

Naturally there exist Lorentz boosts associated with an arbitrary direction. It is given by

$$\bar{o} : x(t) = \vec{v}t = \begin{pmatrix} v^1 \\ v^2 \\ v^3 \end{pmatrix} t \qquad \begin{pmatrix} \bar{t} \\ \bar{x}^1 \\ \bar{x}^2 \\ \bar{x}^3 \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma \vec{v}^T/c^2 \\ -\gamma\vec{v} & \boxed{\mathbb{1} + \dfrac{\gamma - 1}{v^2}\vec{v}\vec{v}^T} \end{pmatrix} \begin{pmatrix} t \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} ,$$

i.e., the transformation matrix is

$$L = L_{\text{general}}(\vec{v}) = \begin{pmatrix} \gamma & -\gamma \vec{v}^T/c^2 \\ -\gamma\vec{v} & \boxed{\mathbb{1} + \dfrac{\gamma - 1}{v^2}\vec{v}\vec{v}^T} \end{pmatrix} , \tag{2.31}$$

where $\vec{v}$ is an arbitrary vector (with $|\vec{v}| < c$ of course). This *general Lorentz boost* can be obtained by applying suitable rotations to (2.30a),

$$L_{\text{general}}(\vec{v}) = \begin{pmatrix} 1 & \\ & \boxed{R^{\mathrm{T}}} \end{pmatrix} \begin{pmatrix} \gamma & -\gamma|v|/c^2 & & \\ -\gamma|v| & \gamma & & \\ & & 1 & \\ & & & 1 \end{pmatrix} \begin{pmatrix} 1 & \\ & \boxed{R} \end{pmatrix} ,$$

i.e., R is a rotation matrix that rotates the vector $\vec{v}$ into the vector $(|v|, 0, 0)^{\mathrm{T}}$. We see that the family of Lorentz boosts is a three-parameter family, where the parameters are the three components of $\vec{v}$ in (2.31).

To obtain the inverse of Lorentz boost in x^1-direction (2.29) we perform a straightforward calculation—we simply compute the inverse matrix. We find that the inverse Lorentz boost is

$$\begin{pmatrix} t \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} = \begin{pmatrix} \gamma & \gamma v/c^2 & 0 & 0 \\ \gamma v & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \bar{t} \\ \bar{x}^1 \\ \bar{x}^2 \\ \bar{x}^3 \end{pmatrix} . \tag{2.32}$$

The result (2.32) is intuitively clear. The transformation (2.29) represents the fact that the observer $\bar{X}$ is in uniform motion with velocity v w.r.t. the observer X. Then the inverse transformation must represent the fact that X is in uniform motion with velocity $(-v)$ w.r.t. the observer $\bar{X}$. Accordingly, the inverse transformation is again described by the Lorentz transformation (2.29), where v is replaced by $(-v)$, and we obtain (2.32).

Finally, let us discuss the **Newtonian limit** of a Lorentz boost. Suppose we have two inertial frames

$$X : \{t, x^1, x^2, x^3\} \qquad\qquad \bar{X} : \{\bar{t}, \bar{x}^1, \bar{x}^2, \bar{x}^3\} \tag{2.33}$$

related by a Lorentz boost (2.29) with a relative velocity $v \ll c$. Expanding γ we obtain

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}} = \frac{1}{1 - \frac{1}{2}\frac{v^2}{c^2} + O(\frac{v^4}{c^4})} = 1 + \frac{1}{2}\frac{v^2}{c^2} + O\Big(\frac{v^4}{c^4}\Big) , \tag{2.34}$$

and therefore

$$\begin{aligned} \bar{t} &= \left[1 + \frac{1}{2}\frac{v^2}{c^2} + O\Big(\frac{v^4}{c^4}\Big)\right] \left[t - \frac{v}{c^2}x^1\right] = t\left(1 + O\Big(\frac{v^2}{c^2}\Big)\right) - \frac{x^1}{c}\,O\Big(\frac{v}{c}\Big) \\ \bar{x}^1 &= \left[1 + \frac{1}{2}\frac{v^2}{c^2} + O\Big(\frac{v^4}{c^4}\Big)\right] \left[-vt + x^1\right] = \left[-vt + x^1\right]\left(1 + O\Big(\frac{v^2}{c^2}\Big)\right) . \end{aligned}$$

When we keep only the highest order terms we get

$$\bar{t} \simeq t \tag{2.35a}$$

$$\bar{x} \simeq x^1 - vt \tag{2.35b}$$

i.e., we recover the *Galilean boost* in x^1-direction.

2.4 The Minkowski metric

As demonstrated in section 2.1 it is convenient to use

$$(x^0, x^1, x^2, x^3)$$

instead of coordinates (t, x^1, x^2, x^3). It is understood that the zero component x^0 encodes time. To account for units, one sets

$$x^0 = ct .$$

Remark. Recall that one can interpret the coordinate x^0 as measuring time in 'light meters' instead of in seconds (where a 'light meter' is the time that light takes to travel through one meter). More or less equivalently, one can measure time in seconds and length in light seconds.

The components of a four-vector v will be denoted by Greek indices; by v^μ we typically mean the collection of the four components (abstract index notation) and not one particular component. The spatial components will be denoted by Latin indices, which run from 1 to 3 (and are thus 'spatial indices'); by v^i we typically mean the collection of the three spatial components; alternatively we write $\vec{v}$.

$$\begin{aligned} \mu, \nu, \ldots &= 0, 1, 2, 3 \\ i, j, \ldots &= 1, 2, 3 \end{aligned}$$

Using these conventions a Poincaré transformation relating an inertial frame of reference X with another inertial frame X' can be written as

$$x^{\mu\prime} = L^\mu{}_\nu x^\nu + a^{\mu\prime} .$$

A Lorentz transformation reads

$$x^{\mu\prime} = L^\mu{}_\nu x^\nu .$$

For example, a boost in direction of x^1 takes the form

$$\begin{pmatrix} x^{0\prime} \\ x^{1\prime} \\ x^{2\prime} \\ x^{3\prime} \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma v & & \\ -\gamma v & \gamma & & \\ & & 1 & \\ & & & 1 \end{pmatrix} \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} , \tag{2.36}$$

i.e., using the coordinate x^0 corresponds to setting

$$c = 1 .$$

in (2.29) and other formulas. In particular, the γ-factor is

$$\gamma = \gamma(v) = \frac{1}{\sqrt{1 - v^2}} . \tag{2.37}$$

Remark. Since $x^0 = ct$, all velocities are measured w.r.t. c. In particular, $|v| < c$ now becomes

$$|v| < 1 . \tag{2.38}$$

W.r.t. some inertial frame of reference we make the following definition:

Definition 2.4. *The Minkowski pseudo-scalar product (Minkowski metric) is the non-degenerate bilinear form*

$$\eta(v,w) = -v^0w^0 + v^1w^1 + v^2w^2 + v^3w^3 = -v^0w^0 + \vec{v}\vec{w}\,, \tag{2.39a}$$

where v and w are arbitrary vectors. Equivalently, we can write

$$\eta(v,w) = (v^0, v^1, v^2, v^3)\begin{pmatrix}-1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1\end{pmatrix}\begin{pmatrix}w^0\\ w^1\\ w^2\\ w^3\end{pmatrix} = v^{\mathrm{T}}\eta\, w\,, \tag{2.39b}$$

or

$$\eta(v,w) = \eta_{\mu\nu}v^\mu w^\nu\,, \qquad \text{where} \quad \left(\eta_{\mu\nu}\right)_{\mu,\nu} = \begin{pmatrix}-1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1\end{pmatrix}. \tag{2.39c}$$

Theorem 2.5. *The Lorentz transformations are uniquely characterized as the transformations that leave $\eta(\cdot,\cdot)$ invariant, i.e.,*

$$\eta(Lv, Lw) = \eta(v,w) \tag{2.40}$$

for all v, w and L.

Proof. Suppose that L is a Lorentz transformation. By theorem 2.2 of section 2.1 this means that

$$L^{\mathrm{T}}\,\eta\,L \;=\; \eta$$

(or $L^\mu{}_\nu L^\sigma{}_\lambda \eta_{\mu\sigma} = \eta_{\nu\lambda}$ in index notation). We obtain

$$\eta(Lv, Lw) = (Lv)^{\mathrm{T}}\eta(Lw) = v^{\mathrm{T}}\big(L^{\mathrm{T}}\eta L\big)w = v^{\mathrm{T}}\eta w = \eta(v,w)\,,$$

i.e., $\eta(\cdot,\cdot)$ is left invariant. Conversely, suppose that $\eta(\cdot,\cdot)$ is left invariant by a transformation L, i.e.,

$$\eta(Lv, Lw) = \eta(v,w)$$

for all v, w. Then

$$0 = \eta(Lv,Lw) - \eta(v,w) = (Lv)^{\mathrm{T}}\eta(Lw) - v^{\mathrm{T}}\eta w = v^{\mathrm{T}}\big(L^{\mathrm{T}}\eta L - \eta\big)w\,.$$

Since this holds for all v, w, the term in brackets must vanish, i.e.,

$$L^{\mathrm{T}}\eta L - \eta = 0\,.$$

We conclude that L is a Lorentz transformation. □

Theorem 2.5 is central to our following considerations. The Minkowski metric (2.39a), although defined w.r.t. some inertial observer, is the same for every other inertial observer. This shows that *spacetime is endowed with a natural geometric structure*, a *pseudo-scalar product*: the (*coordinate-independent*, *frame-independent*, *observer-independent*) *Minkowski metric*.

The geometric structure of spacetime has been unveiled. In the following sections, this geometric structure will be the cornerstone for everything.

CHAPTER 3

MINKOWSKI

3.1 Minkowski spacetime

We collect our previous considerations and condense our findings into what is in fact the actual postulate of special relativity: *The model of our spatiotemporal reality is* **Minkowski spacetime**.

Spacetime is an affine space modeled over a four-dimensional Lorentzian vector space.

Remark. Recall that (Galilean) space is an affine space modeled over a three-dimensional Euclidean vector space, see section 1.1. Note the analogy.

In connection with this definition some explanations are in order.

Definition 3.1. *A* Lorentzian vector space *is a four-dimensional vector space that is endowed with a pseudo-scalar product of signature* $(-+++)$*. We denote this pseudo-scalar product by* $\eta = \eta(\cdot,\cdot)$ *and call it* Minkowski metric*.*

The existence of a pseudo-scalar product $\eta(\cdot,\cdot)$ of signature $(-+++)$ means that there exists bases of the vector space, $\{e_0, e_1, e_2, e_3\}$, such that

$$\eta(e_\mu, e_\nu) = \eta_{\mu\nu} = \begin{pmatrix} -1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix} , \tag{3.1}$$

i.e., $\eta(e_0, e_0) = -1$, $\eta(e_1, e_1) = 1$, $\eta(e_2, e_2) = 1$, $\eta(e_3, e_3) = 1$, and $\eta(e_\mu, e_\nu) = 0$ for $\mu \neq \nu$. Bases of this type are called *pseudo-orthonormal.*

Let v and w be four-vectors, i.e.,

$$v^\mu = \begin{pmatrix} v^0 \\ \vec{v} \end{pmatrix} = \begin{pmatrix} v^0 \\ v^1 \\ v^2 \\ v^3 \end{pmatrix} , \qquad w^\mu = \begin{pmatrix} w^0 \\ \vec{w} \end{pmatrix} = \begin{pmatrix} w^0 \\ w^1 \\ w^2 \\ w^3 \end{pmatrix} , \tag{3.2}$$

w.r.t. some (pseudo-)orthonormal basis $\{e_0, e_1, e_2, e_3\}$. W.r.t. this basis, the (pseudo-)scalar product of these vectors reads

$$\eta(v, w) = \eta_{\mu\nu} v^\mu w^\nu = v^{\mathrm{T}} \eta\, w , \tag{3.3}$$

or, alternatively,

$$\eta(v, w) = v^\mu w_\mu \tag{3.4}$$

where $w_\mu = \eta_{\mu\nu} w^\nu$, or, explicitly,

$$\eta(v, w) = -v^0 w^0 + \vec{v}\vec{w} = -v^0 w^0 + v^1 w^1 + v^2 w^2 + v^3 w^3 . \tag{3.5}$$

By definition, *Minkowski spacetime* is an affine space modeled over a four-dimensional Lorentzian vector space. (In fact, in special relativity, spacetime *is* Minkowski spacetime.) By choosing an origin in Minkowski spacetime, it becomes a vector space (i.e., identified with the Lorentzian vector space underlying it).

Definition 3.2. *A point in Minkowski space is called* event. *An event contains the information of when and where. An element of a Lorentzian vector space we call* four-vector.

Remark. Frequently, for brevity, Minkowski spacetime is called Minkowski space.

3.2 The geometry of Minkowski spacetime

Since the pseudo-scalar product $\eta(u,u)$ is by definition not positive definite, the square $\eta(u,u) = u^\mu u_\mu$ of a four-vector u can have any sign.

Definition 3.3. *A four-vector $u \neq 0$ is called*

- *spacelike, if $\eta(u,u) > 0$,*
- *null, if $\eta(u,u) = 0$,*
- *timelike, if $\eta(u,u) < 0$.*

The basis vector e_0 satisfies $\eta(e_0, e_0) = \eta_{00} = -1$ and is therefore timelike. The basis vectors e_i ($i = 1, 2, 3$) satisfy $\eta(e_i, e_i) = \eta_{ii} = 1$ and are therefore spacelike.

Remark. We will see later that timelike four-vectors are used to describe the motion of massive particles, i.e., velocities less than the speed of light; null vectors are used to describe the motion of massless particles (photons). Finally, spacelike four-vectors are used to describe the motion of superluminous particles a.k.a. tachyons; since such particles probably do not exist, the main role of spacelike vectors is to measure lengths.

☞ A note of advice. The reader is *strongly* recommended to prepare their own lecture notes and to include the figures/Minkowski diagrams drawn on the blackboard during the lecture course.

A special case we will consider frequently is *two-dimensional Minkowski space.* It is given by suppressing the 2- and the 3-component of vectors; in other words: it is the subspace spanned by the vectors $\{e_0, e_1\}$. In two-dimensional Minkowski space, w.r.t. the (pseudo-)orthonormal basis $\{e_0, e_1\}$, we have

$$\eta(v,w) = -v^0 w^0 + v^1 w^1 = (v^0, v^1) \begin{pmatrix} -1 & \\ & 1 \end{pmatrix} \begin{pmatrix} w^0 \\ w^1 \end{pmatrix} . \tag{3.6}$$

Consider a vector

$$u = r \begin{pmatrix} \sin\varphi \\ \cos\varphi \end{pmatrix}$$

in two-dimensional Minkowski space; $r \neq 0$, $\varphi \in (-\pi, \pi]$. Let us compute $\eta(u,u)$:

$$\eta(u,u) = r^2(-\sin^2\varphi + \cos^2\varphi) = r^2 \cos(2\varphi)$$

It follows that

- u is spacelike, if $\varphi \in (-45°, 45°)$,
- u is null, if $\varphi = -45°$ or $\varphi = 45°$,
- u is timelike, if $\varphi < -45°$ or $\varphi > 45°$.

The most conspicuous structure is the set of the two *null lines*, which are the straight lines consisting of null vectors (at $\varphi = \pm 45°$).

Now consider two vectors

$$u = r_u \begin{pmatrix} \sin\varphi \\ \cos\varphi \end{pmatrix} \qquad \text{and} \qquad v = r_v \begin{pmatrix} \sin\psi \\ \cos\psi \end{pmatrix}$$

in two-dimensional Minkowski space. We ask the question when these two vectors are orthogonal. Let us therefore compute $\eta(u, v)$:

$$\eta(u, v) = r_u r_v(-\sin\varphi\sin\psi + \cos\varphi\cos\psi) = r_u r_v \cos(\varphi + \psi)$$

It follows that

$$u \text{ and } v \text{ are orthogonal if} \qquad \psi = 90° - \varphi\,, \tag{3.7}$$

which is the case if u and v are related by a reflection at the null line with angle 45°.

In two-dimensional Minkowski space we thus arrive at the simple consequence: when u is timelike, then its orthogonal vector v is spacelike, and conversely, when u is spacelike, then its orthogonal vector v is timelike. Note that a null vector k is self-orthogonal, since $\eta(k, k) = 0$.

Finally, let us investigate the set of 'unit vectors', i.e., vectors

$$u = \begin{pmatrix} u^0 \\ u^1 \end{pmatrix}$$

such that $\eta(u, u) = 1$ (for spacelike vectors) or $\eta(u, u) = -1$ (for timelike vectors); note that null vectors k cannot be normalized since $\eta(k, k) = 0$.

If u is spacelike and normalized, i.e., $\eta(u, u) = 1$, this means that

$$\eta(u, u) = -(u^0)^2 + (u^1)^2 = 1\,.$$

This is the equation of a *unit hyperbola* (situated in the spacelike part of Minkowski space), whose asymptotes are the two null lines. Likewise, if u is timelike and normalized, i.e., $\eta(u, u) = -1$, this means that

$$\eta(u, u) = -(u^0)^2 + (u^1)^2 = -1\,,$$

which is again the equation of a *unit hyperbola*, whose asymptotes are the two null lines.

Let us now consider four-dimensional Minkowski space. Consider a four-vector

$$u = \begin{pmatrix} u^0 \\ u^1 \\ u^2 \\ u^3 \end{pmatrix} \tag{3.8}$$

(w.r.t. some inertial basis $\{e_0, e_1, e_2, e_3\}$).

- Suppose that u is a *null vector*. Then

  $$\eta(u,u) = -(u^0)^2 + \vec{u}^2 = 0 \qquad \Leftrightarrow \qquad |u^0| = |\vec{u}|\,.$$

  Accordingly, the null vectors lie on a cone in Minkowski space, which we call the ***light cone***. The set $u^0 > 0$ forms the *future light cone*, the set $u^0 < 0$ is the *past light cone*.
- Suppose that u is a *timelike vector* that is *normalized*, i.e., $\eta(u,u) = -1$. Then

  $$\eta(u,u) = -(u^0)^2 + \vec{u}^2 = -1 \qquad \Leftrightarrow \qquad (u^0)^2 = 1 + \vec{u}^2\,.$$

  This is a unit hyperboloid in Minkowski space, which consists of two disconnected parts. The part with $u^0 > 0$ we call the *unit mass shell*. (The nomenclature will become clear later.) If u is not normalized, i.e., $\eta(u,u) = -\text{const}$, then $(u^0)^2 = \text{const} + \vec{u}^2$, which is also a hyperboloid.
- Suppose that u is a *spacelike vector* that is *normalized*, i.e., $\eta(u,u) = 1$. Then

  $$\eta(u,u) = -(u^0)^2 + \vec{u}^2 = 1 \qquad \Leftrightarrow \qquad (u^0)^2 = -1 + \vec{u}^2\,,$$

  which is again a unit hyperboloid in Minkowski space.

In two-dimensional Minkowski space the light cone reduces to the two null lines with a slope of $\pm 45°$, and the hyperboloids reduce to hyperbolas.

We see that the family of timelike vectors falls into two disconnected classes: Future-directed timelike vectors and past-directed timelike vectors. The analog is true for null vectors.

Definition 3.4. *A time-like four-vector is called* **future-directed** *if* $u^0 > 0$ *and past-directed if* $u^0 < 0$. *The analog holds for null vectors.*

Concerning orthogonality let us consider the following generalization of spacelike versus timelike for four-dimensional Minkowski space.

Proposition 3.5. *Suppose that u is a timelike four-vector. Then its orthogonal complement (i.e., all four-vectors orthogonal to u) is spacelike.*

Proof. Since $\eta(u,u) < 0$, w.l.o.g. we can assume $\eta(u,u) = -1$ (otherwise we perform a rescaling). Set $e_0 = u$ and complete e_0 to an orthonormal basis $\{e_0, e_1, e_2, e_3\}$. W.r.t. this basis we have

$$\eta(v,v) = -(v^0)^2 + (v^1)^2 + (v^2)^2 + (v^3)^2\,, \tag{3.9}$$

where $v = v^\mu e_\mu$. By construction, the orthogonal complement of $u = e_0$ is the subspace spanned by $\{e_1, e_2, e_3\}$, i.e., the subspace of all four-vectors v with $v^0 = 0$. Therefore, if we assume that v is orthogonal to u (where we assume $v \neq 0$, of course), then equation (3.9) yields

$$\eta(v,v) = (v^1)^2 + (v^2)^2 + (v^3)^2 > 0\,,$$

i.e., v is spacelike. □

Remark. The orthogonal complement of a null vector k is the tangential hyperplane of the light cone through k; it contains spacelike vectors and the line $\langle k \rangle$ of null vectors. The orthogonal complement of a spacelike vector contains spacelike vectors, null vectors, and timelike vectors.

Definition 3.6. *Two events p, q in Minkowski space are*

- *spacelike separated, if $u = \overrightarrow{pq}$ is spacelike,*
- *null separated, if $u = \overrightarrow{pq}$ is null,*
- *timelike separated, if $u = \overrightarrow{pq}$ is timelike.*

It is important to emphasize that everything we did in this section is completely *observer-independent* (frame-independent, coordinate-independent). The definition of timelikeness (and the other concepts) is only based on the scalar product $\eta(\cdot,\cdot)$, which exists independently of any choice of observer.

3.3 Inertial observers and the relativity of simulaneity

By an **inertial frame of reference** (**inertial observer**) we simply mean a coordinate system that is adapted to the geometric structure; this involves the choice of an arbitrary event (point) as the origin o and a *pseud-orthonormal basis* $\{e_0, e_1, e_2, e_3\}$, i.e.,

$$\eta(e_\mu, e_\nu) = \eta_{\mu\nu} ,$$

which we can also write as

$$\eta(e_0, e_0) = -1 \qquad \eta(e_0, e_i) = 0 \qquad \eta(e_i, e_j) = \delta_{ij} ,$$

where $i, j = 1, 2, 3$. The vector e_0 thus lies on the unit mass shell (since it is timelike, future-oriented, and normalized), while the vectors e_i are normalized spacelike vectors orthogonal to it.

Remark. Recall that in two-dimensional Minkowski space we consider only e_0 and e_1, where the two vectors are related by a reflection at the straight line with slope 45°, see (3.7).

An inertial observer X assigns coordinates to each event (point) in Minkowski space; these coordinates are denoted by

$$(x^\mu)_{\mu=0,\dots,3} = (x^0, \vec{x}) = (t, \vec{x}) .$$

Henceforth we suppress the distinction between x^0 and t and use the two interchangeably. In other words, when we write t, then we refer to time measured in units such that

$$c = 1 .$$

These coordinates rely on the decomposition of four-vectors v w.r.t. the basis $\{e_0, e_1, e_2, e_3\}$, i.e.,

$$v = v^\mu e_\mu = v^0 e_0 + v^i e_i .$$

We write

$$v = \begin{pmatrix} v^0 \\ \vec{v} \end{pmatrix}$$

w.r.t. the observer X.

The time-lines of an observer X are the straight lines given by $\vec{x} = \text{const}$. The special time-line $\vec{x} = 0$ is the time-axis of the observer. Clearly, the time-lines are associated with the vector e_0 of the basis, which is simply because the direction of the time-lines is given by e_0. Since the vector e_0 is distinguished from the other basis vectors by the fact that $\eta(e_0, e_0) = -1$ (instead of $+1$), and since the time-lines are constructed from it, the vector e_0 is of central importance.

Frequently, this vector is thus denoted by a separate symbol (which is typically u, v, or w). It is called the **four-velocity** of the inertial observer X; it is X's 'arrow of time'. When we write

"Consider an inertial observer X with four-velocity **u** . . . " ,

it is understood that u is a four-vector satisfying $u^2 = \eta(u, u) = -1$ (i.e., a timelike vector on the unit mass shell), and we mean that the observer's timelines are determined by u. Accordingly, the observer is associated with a coordinate system where $e_0 = u$; this vector is simply supplemented by three orthonormal basis vectors $\{e_1, e_2, e_3\}$, so that $\{u = e_0, e_1, e_2, e_3\}$ form an orthonormal basis.

Remark. The four-velocity u of an observer determines his coordinate system up to the freedom of choosing $\{e_1, e_2, e_3\}$ in the orthogonal complement of u. This freedom corresponds to rotations of $\{e_1, e_2, e_3\}$.

Remark. These considerations are intimately connected with proposition 3.5, which guarantees that the orthogonal complement of u consists of spacelike vectors (since u is timelike).

Suppose X is an inertial observer with four-velocity u. It is trivial to remark that

$$u = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \qquad \text{w.r.t.} \quad \{u = e_0, e_1, e_2, e_3\} \,, \tag{3.10}$$

in other words, X's own four-velocity w.r.t. X's own coordinate system is the zeroth unit vector (3.10).

Consider an inertial observer X with four-velocity u, i.e., an inertial frame $\{u = e_0, e_1, e_2, e_3\}$. The time-lines of the observer are the straight lines of constant $\vec{x}$, e.g., the time-axis $\vec{x} = 0$ is given as the line $\{tu \,|\, t \in \mathbb{R}\}$. In contrast, the **planes of simultaneity** are the planes of constant time t, i.e., $t = \text{const}$.

Let x be an event in Minkowski spacetime; X ascribes to x the coordinates

$$x = \begin{pmatrix} t \\ \vec{x} \end{pmatrix} .$$

Multiplication (in the sense of the scalar product) with the four-velocity u of X yields

$$\eta(u, x) = -t \,.$$

Therefore, the plane of simultaneity $t = \tau$ (where $\tau = \text{const}$) is given by the set of all events x such that $\eta(u, x) = tau$, i.e.,

$$\text{Plane } t = \tau \;\Leftrightarrow\; \big\{x \,|\, \eta(u, x) = tau\big\} \,. \tag{3.11}$$

This is reminiscent of the standard Hessian normal form[1] for the representation of a plane in a Euclidean space. In (3.11) the four-vector u is the normal vector (in the sense of the Minkowski metric, not in the standard Euclidean sense) of the plane.

The plane $t = 0$ is $\big\{x \,|\, \eta(u, x) = 0\big\}$, which is simply the plane orthogonal to X's four-velocity u (and thus the plane spanned by the spatial basis vectors $\{e_1, e_2, e_3\}$); the planes $t = \tau$ are the paralllel planes.

Corollary 3.7. *For an observer X with four-velocity u (where $\eta(u, u) = -1$), the plane of simultaneity $t = 0$ is given by the set of all events x that are orthogonal [in the sense of the Minkowski metric $\eta(\cdot, \cdot)$] to u, i.e.,*

$$\eta(u, x) = \eta_{\mu\nu} u^\mu x^\nu = 0 \,. \tag{3.12}$$

The planes of simultaneity $t = \tau$ (where τ is arbitrary) are the planes parallel to it, i.e.,

$$\eta(u, x) = \eta_{\mu\nu} u^\mu x^\nu = -\tau \,. \tag{3.13}$$

The corollary provides a *coordinate-independent* (*frame-independent*, *observer-independent*) way of defining simultaneity. This is because (3.13) is written only in terms of the Minkowski metric $\eta(\cdot, \cdot)$ and does therefore not make use of coordinates. We will come back to the issue of coordinate-independence in the next section.

Simultaneity is a relative concept, i.e., observer-dependent. To see that consider two arbitrary events p and q that are spacelike separated (i.e., the four-vector $\overrightarrow{pq}$ is spacelike). There exist observers such that

- the event p comes before the event q;
- the event p and the event q are simultaneous;
- the event p comes after the event q;

[1] A plane (in $\mathbb{R}^3$) is the set of all $\vec{x}$ such that $\vec{n}\vec{x} = \text{const}$, where $\vec{n}$ is the Euclidean normal vector of the plane. We could also write $\delta(\vec{n}, \vec{x}) = \text{const}$, where $\delta(\cdot, \cdot)$ is the standard scalar product.

in other words: two events that are spacelike separated do not have a unique chronological order—their chronological order is observer-dependent. In the following we prove this claim.

W.l.o.g. we assume that p coincides with the origin (otherwise we make a translation). Since q is spacelike separated from p we are able to choose an inertial observer X whose (timelike) four-velocity u is orthogonal to (the spacelike vector) q. W.r.t. X we have

$$p = \begin{pmatrix} 0 \\ \vec{o} \end{pmatrix} , \qquad q = \begin{pmatrix} 0 \\ \vec{q} \end{pmatrix} .$$

By construction, for the observer X, p and q lie in the same plane of simultaneity (namely $t = 0$) and are thus simultaneous; note that $\eta(u,p) = 0$ and $\eta(u,q) = 0$.

Let X' be another observer with four-velocity w, where we assume that w has the form

$$w = \begin{pmatrix} w^0 \\ \vec{w} \end{pmatrix}$$

for the observer X, i.e., w.r.t. $\{u = e_0, e_1, e_2, e_3\}$. The planes of simultaneity $t' = \tau$ of the observer X' are given by the planes $\{x \,|\, \eta(w,x) = -\tau\}$, see (3.13). Clearly, since $\eta(w,p) = 0$, p lies on the plane $t' = 0$; in contrast, by computing $\eta(w,q) = \vec{w}\vec{q}$ we find that q lies on a plane $t' = \tau > 0$ when $\vec{w}\vec{q} < 0$ and on a plane $t' = \tau < 0$ when $\vec{w}\vec{q} < 0$. In other words, for an observer X' with four-velocity w and $\vec{w}$ such that $\vec{w}\vec{q} < 0$, the event q comes at a time $t' = \tau$ after $t' = 0$ (which is the time of p); for an observer X' with four-velocity w and $\vec{w}$ such that $\vec{w}\vec{q} > 0$, the event q comes at a time $t' = \tau$ before $t' = 0$. This proves the claim.

☞ For a better understanding of these results, Minkowski diagrams are essential.

Finally, consider two arbitrary events p and q that are timelike/null separated (where the timelike/null four-vector $\overrightarrow{pq}$ is assumed to be future-oriented). *For all observers the event p comes before the event q.* The proof is left as an exercise.

Definition 3.8. *The* **future** *of an event o is the set of all events that are timelike/null separated from o by a future-oriented timelike/null vector. The* **past** *of an event o is the set of all events that are timelike/null separated from*

o by a past-oriented timelike/null vector. The **present** *of an event is the set of all events that are spacelike separated from o.*

Remark. In other words, the future light cone and its interior are the future, the past light cone and its interior are the past, and the exterior of the light cone is the present of an event o.

3.4 Four-velocities and three-velocities of observers

Consider an inertial observer X with four-velocity u, i.e., X has an inertial frame $\{u = e_0, e_1, e_2, e_3\}$. Let's denote by X' a second inertial observer, whose four-velocity is given by w and whose frame is $\{w = e'_0, e'_1, e'_2, e'_3\}$.

Consider the observer X', whose four-velocity is w. The time-lines of X' are the lines of constant values of $\vec{x}'$; in particular, the time-axis is $\vec{x}' = 0$, or $\langle w \rangle = \{t'w \,|\, t' \in \mathbb{R}\}$. Evidently, X' says that his four-velocity w is

$$\begin{pmatrix} 1 \\ \vec{o} \end{pmatrix} .$$

The observer X does not agree. W.r.t. the basis used by X, the four vector w is

$$\begin{pmatrix} w^0 \\ \vec{w} \end{pmatrix} ,$$

where $-(w^0)^2 + \vec{w}^2 = -1$ since w is a unit vector. Note that w is always the same vector, but once as seen by X' (i.e., decomposed w.r.t. the frame of X'), once as seen by X (i.e., decomposed w.r.t. the basis of X).

Accordingly, the time-axis of X' looks like

$$\langle w \rangle = \left\{ \lambda \begin{pmatrix} w^0 \\ \vec{w} \end{pmatrix} \,|\, \lambda \in \mathbb{R} \right\} = \left\{ \lambda w^0 \begin{pmatrix} 1 \\ \vec{w}/w^0 \end{pmatrix} \,|\, \lambda \in \mathbb{R} \right\} = \left\{ \tilde{\lambda} \begin{pmatrix} 1 \\ \vec{w}/w^0 \end{pmatrix} \,|\, \tilde{\lambda} \in \mathbb{R} \right\}.$$

Hence, $t = \tilde{\lambda}$ and $\vec{x} = \tilde{\lambda}\, \vec{w}/w^0$, which implies

$$\vec{x} = \frac{\vec{w}}{w^0}\, t \,.$$

In other words, for the inertial observer X, the observer X' is in uniform motion with velocity

$$\vec{v} = \frac{\vec{w}}{w^0} = \frac{\vec{w}}{\sqrt{1 + \vec{w}^2}} \,.$$

Accordingly,

$$\vec{w} = \frac{\vec{v}}{\sqrt{1-\vec{v}^2}} = \gamma\vec{v}$$

and $w^0 = \gamma$. This implies that, while w is the zeroth unit vector as seen by X', it looks like

$$w = \begin{pmatrix} w^0 \\ \vec{w} \end{pmatrix} = \gamma \begin{pmatrix} 1 \\ \vec{v} \end{pmatrix} \qquad \text{w.r.t.} \quad X\,, \tag{3.14}$$

i.e., w.r.t. the basis $\{e_0, e_1, e_2, e_3\}$ used by X.

Let us summarize. For the four-velocity w of the inertial observer X' we find

$$w = \begin{pmatrix} w^{0\prime} \\ \vec{w}' \end{pmatrix} = \begin{pmatrix} 1 \\ \vec{o} \end{pmatrix} \qquad \text{w.r.t.} \quad X' = \{e'_0, e'_1, e'_2, e'_3\}\,, \tag{3.15a}$$

$$w = \begin{pmatrix} w^0 \\ \vec{w} \end{pmatrix} = \gamma \begin{pmatrix} 1 \\ \vec{v} \end{pmatrix} \qquad \text{w.r.t.} \quad X = \{e_0, e_1, e_2, e_3\}\,, \tag{3.15b}$$

where $\vec{v}$ is the relative velocity of the observer X' as seen by X.

Obviously, the scenario (3.15) can also be described by interchanging the roles of X and X'.

$$u = \begin{pmatrix} u^0 \\ \vec{u} \end{pmatrix} = \begin{pmatrix} 1 \\ \vec{o} \end{pmatrix} \qquad \text{w.r.t.} \quad X = \{e_0, e_1, e_2, e_3\}\,, \tag{3.16a}$$

$$u = \begin{pmatrix} u^{0\prime} \\ \vec{u}' \end{pmatrix} = \gamma \begin{pmatrix} 1 \\ -\vec{v} \end{pmatrix} \qquad \text{w.r.t.} \quad X' = \{e'_0, e'_1, e'_2, e'_3\}\,. \tag{3.16b}$$

When X' moves with $\vec{v}$ w.r.t. X, then X' sees X move away with $(-\vec{v})$.

When we form the scalar product $\eta(u, w)$ from (3.15) we obtain

$$\eta(u, w) = -\gamma\,. \tag{3.17}$$

Since the expression $\eta(u, w)$ is *coordinate-independent* (*frame-independent*, *observer-independent*), the result is the same w.r.t. all coordinate systems. (In particular, one can easily check that the result is the same when computed w.r.t. $\{e_0, e_1, e_2, e_3\}$ and $\{e'_0, e'_1, e'_2, e'_3\}$.) Forming the product $\eta(u, w)$ thus leads to a *coordinate-independent* way of defining the absolute value of the *relative velocity* $\vec{v}$ between two observers.

Corollary 3.9. *Consider two observers X and X' represented by the four-velocities u and w, respectively. Then $|\vec{v}|$ is given by*

$$\gamma = \frac{1}{\sqrt{1-|\vec{v}|^2}} = -\eta(u,w)\ . \tag{3.18}$$

CHAPTER 4

PARTICLES

4.1 World lines

A point particle is represented in Minkowski space by a curve

$$\mathbb{R} \ni \lambda \mapsto x(\lambda) ,$$

where the parametrization is completely irrelevant. It is the image of the curve (i.e., the so-called 'geometric curve') that encodes all the information on the whereabouts and 'whenabouts' of the particle.[1]

Definition 4.1. *The curve in Minkowski space representing a point particle is called the* world line *of the particle.*

Remark. An inertial observer X is characterized by a *family* of world lines, namely by the straight time-lines $\vec{x} = \text{const}$.

Consider the world line of a particle, i.e., the curve

$$\mathbb{R} \ni \lambda \mapsto x(\lambda) . \tag{4.1}$$

We can form the (field of) tangent vectors along the world line; the tangent vector $w(\lambda) = dx(\lambda)/d\lambda$ is a four-vector,

$$\mathbb{R} \ni \lambda \mapsto w(\lambda) = \frac{d}{d\lambda} x(\lambda) . \tag{4.2}$$

[1] See, however, the treatment and interpretation of null lines in section 4.6.

W.r.t. some inertial coordinate system X we have

$$x(\lambda) = \begin{pmatrix} t(\lambda) \\ \vec{x}(\lambda) \end{pmatrix} ,$$

$$w(\lambda) = \frac{d}{d\lambda} x(\lambda) = \frac{d}{d\lambda} \begin{pmatrix} t(\lambda) \\ \vec{x}(\lambda) \end{pmatrix} = \begin{pmatrix} w^0(\lambda) \\ \vec{w}(\lambda) \end{pmatrix} .$$

The normal three-velocity $\vec{v}$ of the particle, as seen by X, is given by

$$\vec{v}(\lambda) = \frac{\vec{w}(\lambda)}{w^0(\lambda)} .$$

To see this we note that

$$\vec{v} = \frac{d\vec{x}}{dt} = \frac{d\vec{x}}{d\lambda}\frac{d\lambda}{dt} = \frac{d\vec{x}(\lambda)}{d\lambda}\Big(\frac{dt(\lambda)}{d\lambda}\Big)^{-1} = \frac{\vec{w}}{w^0} .$$

Therefore, we can write

$$\begin{aligned} x(\lambda) &= \begin{pmatrix} t(\lambda) \\ \vec{x}(\lambda) \end{pmatrix} , \\ w(\lambda) &= \frac{d}{d\lambda} x(\lambda) = \frac{d}{d\lambda} \begin{pmatrix} t(\lambda) \\ \vec{x}(\lambda) \end{pmatrix} = w^0(\lambda) \begin{pmatrix} 1 \\ \vec{v}(\lambda) \end{pmatrix} . \end{aligned} \tag{4.3}$$

Definition 4.2. *A curve $\lambda \mapsto x(\lambda)$ is called*

- *spacelike, if $w(\lambda) = \frac{d}{d\lambda}x(\lambda)$ is spacelike for all λ,*
- *null, if $w(\lambda) = \frac{d}{d\lambda}x(\lambda)$ is null for all λ,*
- *timelike, if $w(\lambda) = \frac{d}{d\lambda}x(\lambda)$ is timelike for all λ.*

Timelike world lines describe the motion of particles at velocities less than the speed of light, null world lines describe the motion of particles at the speed of light. This is a simple consequence of (4.3): If the velocity is smaller than the speed of light, i.e., $|\vec{v}| < 1$, then the vector $(1, \vec{v})^{\mathrm{T}}$ is timelike (and, incidentally, future-oriented), because $-1 + \vec{v}^2 < 0$; hence w is timelike. If the velocity equals the speed of light, then $(1, \vec{v})^{\mathrm{T}}$ is null, because $-1 + \vec{v}^2 = 0$; hence the tangent vector w is a null vector. Finally, if the velocity is superluminous, then

$(1, \vec{v})^{\mathrm{T}}$ and thus w is spacelike.

$$|\vec{v}| < 1 \quad \Leftrightarrow \quad w = w^0 \begin{pmatrix} 1 \\ \vec{v} \end{pmatrix} \text{ is timelike ,} \tag{4.4a}$$

$$|\vec{v}| = 1 \quad \Leftrightarrow \quad w = w^0 \begin{pmatrix} 1 \\ \vec{v} \end{pmatrix} \text{ is null ,} \tag{4.4b}$$

$$|\vec{v}| > 1 \quad \Leftrightarrow \quad w = w^0 \begin{pmatrix} 1 \\ \vec{v} \end{pmatrix} \text{ is spacelike .} \tag{4.4c}$$

It is important to stress that this classification is observer-independent (which is simply because the r.h. side does not make use of coordinates). While different observers might not agree on the issue of how fast (in m/s) a particle actually is (for one observer, the particle is at rest, for the other it might move fast), they always agree on the issue of whether the particle moves slower than the speed of light, at the speed of light, or faster.

A simple parametrization of the world line (4.3) is to use the coordinate time t of some observer X as the parameter λ. Then (4.3) becomes

$$x(t) = \begin{pmatrix} t \\ \vec{x}(t) \end{pmatrix} ,$$

$$w(t) = \frac{d}{dt} x(t) = \frac{d}{dt} \begin{pmatrix} t \\ \vec{x}(t) \end{pmatrix} = \begin{pmatrix} 1 \\ \vec{v}(t) \end{pmatrix} .$$

In general, reparametrizations of a curve change the length of the tangent vectors. Obviously, the coordinate time reparametrization corresponds to setting $w^0 = 1$ in (4.3).

4.2 Tachyons

Tachyons are supposed to be particles that are represented by spacelike curves and are thus associated with superluminous velocities. Probably, tachyons do not exist, since they would cause violation of causality. This will be shown in the following.

☞ Again it is recommended to use Minkowski diagrams to illustrate the text.

Assume that an observer X sends a tachyon from event p to event q. By definition, tachyons move along spacelike curves, i.e., in the 'present', which is the exterior of the light cone of p. In brief: p and q are spacelike separated. This has some counterintuitive consequences: For X, p comes before q (p is 'emission', q is 'reception' of the tachyon). However, there exist observers, for who p and q are simultaneous. There exist even observers, for who p comes after q. For such observers, the tachyon is seen to come out of the receiver at q, move in the direction to the emitter, and vanish in the emitter p. This is weird.

Violation of causality becomes inevitable, when we consider a scenario involving two observers X and X' who are both able to send tachyons and make the following agreement: X sends a tachyon to X' and X' send a tachyon back immediately upon reception. For simplicity, we assume that the tachyons move extremely fast, namely with infinite velocity. (Note, however, that the same argument still holds if we assume the velocity to be only slightly larger than the speed of light.) The observer X has installed a tachyon emitter/receiver at $\vec{x} = 0$, so that the tachyon emitter's world line is the time-axis of X. X sends a tachyon to X' at $t = 0$, i.e., from the event $p = (0, \vec{o})$, along the plane of simultaneity $t = 0$ to the event $q = (0, \vec{q})$. X' receives the tachyon at q and sends it back immediately; it reaches the emitter/receiver that X uses at the event $p' = (t, \vec{o})$. Since X' is an observer who is completely equivalent to X, X' is able to return tachyons at the same speed, i.e., again along a plane of simultaneity. However, for X', whose four-velocity is u', the planes of simultaneity are not $t = \text{const}$, but $t' = \tau' = \text{const}$, or, in coordinate-independent notation, the planes of all events x such that $\eta(u', x) = -\tau'$, see (3.13). Let the four-velocity u' of X' be given by

$$u' = \gamma \begin{pmatrix} 1 \\ \vec{v} \end{pmatrix} ,$$

where $\vec{v}$ is the velocity of X' as seen by X, see (3.15). The event q lies on the plane of simultaneity $t' = \tau'$ with

$$\eta(u', q) = \eta(\gamma \begin{pmatrix} 1 \\ \vec{v} \end{pmatrix}, \begin{pmatrix} 0 \\ \vec{q} \end{pmatrix}) = \gamma \vec{v}\vec{q} = -\tau' .$$

To compute the time t when X receives the tachyon that comes back we compute the intersection of that plane with the world line $(t, 0)$ of the emitter/receiver X uses:

$$\eta(u', p') = \underbrace{\eta(\gamma \begin{pmatrix} 1 \\ \vec{v} \end{pmatrix}, \begin{pmatrix} t \\ 0 \end{pmatrix})}_{-\gamma t} \stackrel{!}{=} \gamma \vec{v}\vec{q} .$$

It follows that the time of reception (corresponding to event p') is given by

$$t = -\vec{v}\vec{q}\,, \tag{4.5}$$

which is negative (when we have made the right choices, i.e., $\vec{v}\vec{q} > 0$). But this means that X receives the tachyon that X' has sent back *before* the original tachyon has been emitted (at $t = 0$). Then what is supposed to happen, if X decides, during the time-interval $(-\vec{v}\vec{q}, 0)$, "I don't feel like sending a tachyon any longer." Then X' couldn't have received anything, and consequently, wouldn't have sent back anything. But then where did the tachyon that X received at $t = -\vec{v}\vec{q}$ come from?

Evidently, the existence of tachyons leads to a *violation of causality*. Henceforth, we exclude the possibility that such particles could exist and postulate:

Particles move along world lines that are timelike or null.

4.3 Proper time

Consider a particle described by the world line

$$\mathbb{R} \ni \lambda \mapsto x(\lambda)\,, \tag{4.6a}$$

and let

$$w(\lambda) = \frac{d}{d\lambda}x(\lambda)\,. \tag{4.6b}$$

be the tangent vector, which we suppose to be *timelike* for all λ. We ask the question of how much time passes for the particle between two events $x_1 = x(\lambda_1)$ and $x_2 = x(\lambda_2)$.

Consider the particle at a fixed value of λ, i.e., at a fixed event $x(\lambda)$. To measure how much time passes in the infinitesimal interval $[\lambda, \lambda + d\lambda]$, the particle needs an observer. But not any observer:

Time is measured in an inertial coordinate system in which the particle is (momentarily) at rest, the *momentary rest frame*.

If the particle is not in uniform motion, there does not exist one global rest frame. However, there exists a *momentary* rest frame, i.e., an inertial observer for whom the particle is at rest at the instant of time λ.

This inertial frame is associated with the observer whose four-velocity u is parallel to the tangent vector w at λ, i.e., u is given by

$$u = \frac{w(\lambda)}{\sqrt{-\eta\big(w(\lambda), w(\lambda)\big)}} \,. \tag{4.7}$$

Note that $u^2 = \eta(u, u) = -1$ as it is required for an observer's four-velocity. W.r.t. the rest frame $\{u = e_0, e_1, e_2, e_3\}$ the world line's tangent vector at λ, i.e., the tangent at the event $x(\lambda)$, reads

$$w(\lambda) = \begin{pmatrix} w^0(\lambda) \\ \vec{w}(\lambda) \end{pmatrix} = \begin{pmatrix} \sqrt{-\eta\big(w(\lambda), w(\lambda)\big)} \\ \vec{o} \end{pmatrix} \qquad \text{w.r.t.} \quad \{u = e_0, e_1, e_2, e_3\} \,.$$

In the infinitesimal interval $[\lambda, \lambda + d\lambda]$ the world line connects the event $x(\lambda)$ with the event

$$x(\lambda + d\lambda) = x(\lambda) + w(\lambda)d\lambda = x(\lambda) + \begin{pmatrix} \sqrt{-\eta\big(w(\lambda), w(\lambda)\big)} \\ \vec{o} \end{pmatrix} d\lambda \,.$$

Accordingly,

$$x(\lambda + d\lambda) - x(\lambda) = \begin{pmatrix} t(\lambda + d\lambda) \\ \vec{x}(\lambda + d\lambda) \end{pmatrix} - \begin{pmatrix} t(\lambda) \\ \vec{x}(\lambda) \end{pmatrix} = \begin{pmatrix} \sqrt{-\eta\big(w(\lambda), w(\lambda)\big)}\, d\lambda \\ \vec{o} \end{pmatrix}$$

The time that has elapsed for the particle between the events $x(\lambda)$ and $x(\lambda{+}d\lambda)$ we call ds. It coincides with the element $dt = t(\lambda{+}d\lambda){-}t(\lambda)$ in the coordinates of the momentary rest frame and can thus be read off straightforwardly as the 0^{th} component of the four-vector $x(\lambda + d\lambda) - x(\lambda)$. Therefore,

$$ds = \sqrt{-\eta\big(w(\lambda), w(\lambda)\big)}\, d\lambda \,. \tag{4.8}$$

Remark. Since the particle is not in uniform motion in general, the momentary rest frame changes along the world line of the particle. Therefore, to avoid ambiguities in the notation we do not denote the time that elapses along a particle's world line by t (which is the coordinate time of some particular momentary rest frame), but by s. This is the *proper time* of the particle.

Definition 4.3. *The* proper time *(denoted by* **s***) along a world line of a particle is the flow of time as measured in the momentary rest frames of the particle.*

In order to obtain the proper time s it merely remains to integrate (4.8) along the particle's world line. The proper time s along a world line is given by

$$s = \int \sqrt{-\eta\big(w(\lambda), w(\lambda)\big)}\, d\lambda \,. \tag{4.9a}$$

Accordingly, the proper time Δs that elapses between two events $x_1 = x(\lambda_1)$ and $x_2 = x(\lambda_2)$ is

$$\Delta s = s|_{x_2} - s|_{x_1} = \int_{\lambda_1}^{\lambda_2} \sqrt{-\eta\big(w(\lambda), w(\lambda)\big)}\, d\lambda \,. \tag{4.9b}$$

Remark. It is not difficult to show that these formulas are in fact independent on the chosen parametrization of the world line: Let κ denote the parameter of an alternative parametrization, i.e., $\lambda = \lambda(\kappa)$; then the tangent vector w.r.t. the reparametrized curve is

$$\tilde{w}(\kappa) = \frac{d}{d\kappa}\, x(\lambda) = \Big(\frac{d}{d\lambda}\, x(\lambda)\Big) \Big(\frac{d\lambda}{d\kappa}\Big) = \frac{d\lambda}{d\kappa}\, w(\lambda)\,,$$

where $\lambda = \lambda(\kappa)$. Therefore,

$$\sqrt{-\eta\big(\tilde{w}(\kappa), \tilde{w}(\kappa)\big)}\, d\kappa = \sqrt{-\eta\big(w(\lambda), w(\lambda)\big)}\, \frac{d\lambda}{d\kappa}\, d\kappa = \sqrt{-\eta\big(w(\lambda), w(\lambda)\big)}\, d\lambda\,,$$

i.e., as expected, the concept of proper time depends only on the geometric curve and not on the actual parametrization of the world line.

Remark. The definition of proper time is coordinate-independent (frame-independent), since only the Minkowski metric $\eta(\cdot,\cdot)$ appears in (4.9).

Let us consider a (fixed) inertial observer X, whose coordinates are $(t, \vec{x})$. (This coordinate system is completely arbitrary; in particular, it need not be the momentary rest frame of the particle under consideration at any time.) W.r.t. this observer's coordinate time t we parametrize the world line of the particle. The world line then reads

$$\mathbb{R} \ni t \mapsto x(t) = \begin{pmatrix} t \\ \vec{x}(t) \end{pmatrix}\,, \tag{4.10}$$

and the tangent vector is

$$\frac{d}{dt} x(t) = \begin{pmatrix} 1 \\ d\vec{x}(t)/dt \end{pmatrix} = \begin{pmatrix} 1 \\ \vec{v}(t) \end{pmatrix}\,,$$

where $\vec{v}$ is the three-velocity of the particle at time t as seen by the observer X. Inserting this into (4.8) and (4.9) and replacing λ by our current choice of parameter t, we obtain

$$ds = \sqrt{1 - \vec{v}^2(t)}\, dt \tag{4.11a}$$

$$s = \int \sqrt{1 - \vec{v}^2(t)}\, dt \tag{4.11b}$$

The proper time that elapses for the particle during an interval $[t_\mathrm{i}, t_\mathrm{f}]$ of coordinate time t, i.e., between the events $x_\mathrm{i} = x(t_\mathrm{i})$ and $x_\mathrm{f} = x(t_\mathrm{f})$, is thus

$$\Delta s = s|_{x_\mathrm{f}} - s|_{x_\mathrm{i}} = \int_{t_\mathrm{i}}^{t_\mathrm{f}} \sqrt{1 - \vec{v}^2(t)}\, dt\,. \tag{4.11c}$$

The factor $\sqrt{1 - \vec{v}^2}$ is the *time dilation* factor. During the time dt, which passes for the observer X, the particle's proper time only increases by $ds < dt$. This gives rise to the so-called *twin paradox*.

4.4 The twin paradox

We consider two particles—let's call them twins—that are represented by the two (timelike) world lines

$$\mathbb{R} \ni t \mapsto x_A(t) = \begin{pmatrix} t \\ \vec{x}_A(t) \end{pmatrix} \quad \text{and} \quad \mathbb{R} \ni t \mapsto x_B(t) = \begin{pmatrix} t \\ \vec{x}_B(t) \end{pmatrix}\,.$$

Let us assume that twin A "stays at home", i.e., twin A is always at rest w.r.t. an inertial coordinate system. W.l.o.g. the spatial coordinates of twin A are zero, i.e.,

$$x_A(t) = \begin{pmatrix} t \\ \vec{o} \end{pmatrix}\,.$$

Twin B is supposed to lead a more adventurous life than her sibling. At time $t = t_\mathrm{i} = 0$, twin B leaves A for a life "somewhere in Minkowski space". Only at time $t = t_\mathrm{f}$, twin B returns and keeps company with A again. Accordingly,

$$x_B(t) = \begin{pmatrix} t \\ \vec{x}_B(t) \end{pmatrix}\,, \quad \text{where} \quad \vec{x}_B(0) = \vec{x}_B(t_\mathrm{f}) = 0\,,$$

and

$$\frac{d}{dt} x_B(t) = \begin{pmatrix} 1 \\ \vec{v}_B(t) \end{pmatrix}\,.$$

For twin A, his sister has been away for a time t_{f}. In other words, the proper time Δs_A for A between departure and arrival of B is $\Delta s_A = t_{\mathrm{f}}$. This is obvious since A is always at rest w.r.t. the inertial coordinate system. Alternatively, one can compute Δs_A explicitly by inserting $\vec{v} = \vec{v}_A = \vec{o}$ into (4.11).

The proper time Δs_B that passes for twin B is different. We obtain

$$\Delta s_B = \int_0^{t_{\mathrm{f}}} \sqrt{1 - \vec{v}_B^2(t)}\, dt$$

from (4.11). By assumption, $\vec{v}_B(t) \neq 0$ at least for some t; therefore the square root is less than one at least for some t. Consequently,

$$\Delta s_B = \int_0^{t_{\mathrm{f}}} \underbrace{\sqrt{1 - \vec{v}_B^2(t)}}_{<1}\, dt \;<\; \int_0^{t_{\mathrm{f}}} dt = t_{\mathrm{f}} = \Delta s_A$$

and we find that

$$\Delta s_B < \Delta s_A\,. \tag{4.12}$$

We conclude that, during her voyage, B has aged at a slower rate than A, so that, upon her return, twin B is younger than twin A.

Remark. Note that we have not made any assumptions about the specifics of B's journey. The fact that $\Delta s_B < \Delta s_A$ holds independently of these details.

Remark. A common point of confusion is the following. Doesn't relativity mean that I could also take twin B's point of view? Viewed by B, it is A who moves away and comes back. Then $\Delta s_B > \Delta s_A$, no? Indeed: no. The difference is clear: A is at rest, all the time, w.r.t. an inertial coordinate system; B isn't (at best, B is described by a sequence of momentary rest frames). A can take the point of view of an inertial observer, while B can't, because accelerations are involved in the world line of B.

4.5 Four-velocities

Consider again a particle represented by a timelike world line $\mathbb{R} \ni \lambda \mapsto x(\lambda)$. Let s denote the proper time along this world line. It suggests itself to use proper time s instead of the (arbitrary) parameter λ to parametrize the world line, i.e.,

$$\mathbb{R} \ni s \mapsto x(s)\,. \tag{4.13}$$

The tangent vector of this curve in this parametrization we call u; it is

$$u(s) = \frac{d}{ds} x(s) \,, \tag{4.14}$$

and it has an important property:

Proposition 4.4. *The tangent vector $u = u(s)$ of a world line that is parametrized w.r.t. proper time s is normalized, i.e., $\eta(u, u) = -1$.*

Proof. To prove the proposition we make a straightforward computation. Since

$$u = \frac{dx}{ds} = \frac{dx}{d\lambda}\frac{d\lambda}{ds} = \frac{dx}{d\lambda}\Big(\frac{ds}{d\lambda}\Big)^{-1} = \frac{w}{\sqrt{-\eta(w, w)}} \,,$$

we find

$$\eta(u, u) = \frac{1}{\big(-\eta(w, w)\big)}\, \eta(w, w) = -1 \,,$$

as claimed. □

Remark. The proposition is a direct consequence of our definition of proper time. Proper time is the coordinate time associated with the momentary rest frame; the four-velocity of the momentary rest frame and the tangent vector (4.14) must thus coincide; $u = (1, 0, 0, 0)^{\mathrm{T}}$ in the coordinates of the momentary rest frame (at some instant of time). Accordingly, $\eta(u, u) = -1$.

Definition 4.5. *The tangent vector $u = u(s)$ of a world line representing a particle (where the world line is parametrized w.r.t. proper time) is called the particle's* four-velocity; *$\eta(u, u) = -1$.*

Remark. The concepts of a particle's four-velocity and the four-velocity of an inertial observer are intimately related, the main difference being that the four-velocity of a particle is defined along the particle's world line, while the four-velocity of an inertial observer can be thought of as a (constant) vector field on the entire Minkowski space.

Consider a (fixed) inertial observer X, whose coordinates are $(t, \vec{x})$. As seen previously, w.r.t. this observer's coordinate time t, the world line of the particle reads

$$\mathbb{R} \ni t \mapsto x(t) = \begin{pmatrix} t \\ \vec{x}(t) \end{pmatrix} \,, \tag{4.15a}$$

and the tangent vector is

$$\frac{d}{dt}x(t) = \begin{pmatrix} 1 \\ d\vec{x}(t)/dt \end{pmatrix} = \begin{pmatrix} 1 \\ \vec{v}(t) \end{pmatrix} ,$$

where $\vec{v} = \vec{v}(t)$ is the three-velocity of the particle at time t as seen by the observer X. The four-velocity u associated with the particle corresponds to the normalized tangent vector; since its squared norm is $(-1 + \vec{v}^2) = \gamma^{-2}$ we again

$$u(t) = \gamma \begin{pmatrix} 1 \\ \vec{v}(t) \end{pmatrix} . \tag{4.15b}$$

Corollary 4.6. *W.r.t. a given observer X (whose basis is $\{e_0, e_1, e_2, e_3\}$), the four-velocity of a particle can be written as*

$$\boldsymbol{u} = \boldsymbol{\gamma} \begin{pmatrix} \mathbf{1} \\ \boldsymbol{\vec{v}} \end{pmatrix} \qquad \textit{w.r.t. } X , \tag{4.16}$$

where $\vec{v} = \vec{v}(t)$ is the standard three-velocity of the particle w.r.t. this observer.

Remark. Equation (3.15) and (4.16) resemble each other closely, the main difference being that $\vec{v} = \text{const}$ in (3.15) while $\vec{v} = \vec{v}(t)$ varies along the world line of the particle in (4.16).

As an exercise let us rederive (4.16) in a slightly different way. Let us reparametrize the world line (4.15a) by proper time s, i.e., we express t as a function of s and $\vec{x} = \vec{x}\big(t(s)\big)$,

$$\mathbb{R} \ni s \mapsto x\big(t(s)\big) = \begin{pmatrix} t(s) \\ \vec{x}\big(t(s)\big) \end{pmatrix} . \tag{4.17}$$

The derivative w.r.t. s yields

$$u\big(t(s)\big) = \frac{d}{ds}\begin{pmatrix} t(s) \\ \vec{x}\big(t(s)\big) \end{pmatrix} = \begin{pmatrix} \frac{dt}{ds}(s) \\ \frac{d\vec{x}}{dt}\big(t(s)\big)\,\frac{dt}{ds}(s) \end{pmatrix} = \frac{dt}{ds}(s)\begin{pmatrix} 1 \\ \vec{v}\big(t(s)\big) \end{pmatrix} . \tag{4.18}$$

From

$$\frac{ds}{dt} = \sqrt{1 - \vec{v}^2(t)} \,,$$

see (4.11), we conclude that

$$\frac{dt}{ds}(s) = \frac{1}{\sqrt{1 - \vec{v}^2\big(t(s)\big)}} = \gamma \,,$$

therefore (4.18) becomes

$$u\big(t(s)\big) = \gamma \begin{pmatrix} 1 \\ \vec{v}\big(t(s)\big) \end{pmatrix} ,$$

i.e., we reproduce (4.16).

Corollary 3.9 of section 3.4 states that the (absolute value of the) relative velocity between two inertial observers is obtained in a *coordinate-independent* (*observer-independent*) manner via the scalar product. The analog holds for the velocity of a particle w.r.t. a given observer:

Consider a particle described by a timelike world line $\mathbb{R} \ni s \mapsto x^\mu(s)$, where s denotes proper time; let $u^\mu = dx^\mu/ds$ be the four-velocity.[2] W.r.t. an observer X the four-velocity u^μ becomes

$$u^\mu = \gamma \begin{pmatrix} 1 \\ \vec{v} \end{pmatrix} , \tag{4.19}$$

see (4.16); canonically, u is parametrized w.r.t. proper time s, but, equivalently, we may parametrize it w.r.t. X's coordinate time t. Let u_X^μ denote the four-velocity of the observer X. W.r.t. the coordinates of X, we have

$$u_X^\mu = \begin{pmatrix} 1 \\ \vec{o} \end{pmatrix} . \tag{4.20}$$

Hence, $\eta_{\mu\nu} u_X^\mu u^\nu = -\gamma$, i.e.,

$$\eta(u_X, u) = -\gamma , \tag{4.21}$$

or, when we write out the dependence on t,

$$\gamma(t) = \frac{1}{\sqrt{1 - \vec{v}^2(t)}} = -\eta\big(u_X, u(t)\big) . \tag{4.21$'$}$$

This equation thus determines $|\vec{v}(t)|$ in a coordinate-independent way.[3]

[2] Here we are using abstract index notation for a change.

[3] Recall that coordinate-independence means that the objects involved (which are $\eta(\cdot,\cdot)$, u_X, and u in the present context) are given in an abstract way (i.e., without resorting to coordinates). To compute the product $\eta(u_X, u)$ we can choose any coordinate system that comes to mind.

Example. Consider the two world lines

$$x_A(t) = \begin{pmatrix} t \\ \vec{x}_A(t) \end{pmatrix} = \begin{pmatrix} t \\ -\frac{1}{2}\sin t \\ 0 \\ 0 \end{pmatrix}, \quad x_B(t) = \begin{pmatrix} t \\ \vec{x}_B(t) \end{pmatrix} = \begin{pmatrix} t \\ \frac{3}{4}\arctan t \\ 0 \\ 0 \end{pmatrix}, \tag{4.22}$$

which are given in some inertial coordinate system $\{t, \vec{x}\}$. (In principle we could reparametrize each of these world lines with proper time; however, since this involves elliptic integrals, we refrain from doing so in the present context.) The four-velocities associated with (4.22) are

$$u_A(t) = \gamma_A \begin{pmatrix} 1 \\ \vec{v}_A(t) \end{pmatrix} = \gamma_A \begin{pmatrix} 1 \\ -\frac{1}{2}\cos t \\ 0 \\ 0 \end{pmatrix}, \quad u_B(t) = \gamma_B \begin{pmatrix} 1 \\ \vec{v}_B(t) \end{pmatrix} = \gamma_B \begin{pmatrix} 1 \\ \frac{3}{4}\frac{1}{1+t^2} \\ 0 \\ 0 \end{pmatrix},$$

where $\gamma_A = \gamma(\vec{v}_A)$ and $\gamma_B = \gamma(\vec{v}_B)$. For instance, at $t = 0$, we have

$$u_A\big|_{t=0} = \gamma_A \begin{pmatrix} 1 \\ -\frac{1}{2} \\ 0 \\ 0 \end{pmatrix}, \qquad u_B\big|_{t=0} = \gamma_B \begin{pmatrix} 1 \\ \frac{3}{4} \\ 0 \\ 0 \end{pmatrix}, \tag{4.23}$$

where $\gamma_A = 2/\sqrt{3}$ and $\gamma_B = 4/\sqrt{7}$, since $|\vec{v}_A| = 1/2$ and $|\vec{v}_B| = 3/4$. To obtain the relative velocity $|\vec{v}| = |\vec{v}_{AB}|$ between the two particles (for $t = 0$) we may apply (4.21) in the straightforward way, i.e.,

$$\gamma = \frac{1}{\sqrt{1 - |\vec{v}|^2}} = -\eta\big(u_A, u_B\big),$$

which yields

$$|\vec{v}|\,\Big|_{t=0} = \frac{10}{11}. \tag{4.24}$$

Exercise. In the above example, when we proceed analogously for $t \neq 0$, the 'relative velocity' is found to be

$$|\vec{v}| = 2\,\frac{3 + 2(1+t^2)\cos t}{8(1+t^2) + 3\cos t}. \tag{4.25}$$

Argue that (4.25) is basically meaningless in general for $t \neq 0$. For instance, for $t = 1.90172$, the formula yields $|\vec{v}| = 0$; this suggests that we can conclude that the particles A and B represented by the world lines are at rest w.r.t. each other at that time. But this does not make sense; why? And why is the situation for $t = 0$ different?

Exercise. And now it becomes really challenging.... Define $t_A = 1.86274$ and $t_B = 2.05225$ ($\pm$ errors in the next digits). Argue that it makes sense to say that the particles are at rest w.r.t. each other when particle A is at the event $(t_A, x_A(t_A))$ and B at $(t_B, x_B(t_B))$. There exist other pairs (t_A, t_B) which exhibit the same property. Compute another pair (using Mathematica or Maple). Can one deduce something more general from this example?

4.6 Photons

The simplest solutions of Maxwell's (vacuum) equations are *plane waves*. Let X be an inertial observer, whose basis is $\{u = e_0, e_1, e_2, e_3\}$ and whose coordinates are $(t, \vec{x})$. In these coordinates, a plane wave is given by

$$\phi(x) = \phi(t, \vec{x}) = a e^{i(-\omega t + \vec{k}\vec{x})} , \tag{4.26}$$

where a is the amplitude, ω the angular frequency, and $\vec{k}$ the wave vector. The phase velocity $\omega/|\vec{k}|$ coincides with the speed of light (where $c = 1$), hence

$$\omega = |\vec{k}| . \tag{4.27}$$

A simple computation shows that (4.26) with (4.27) is indeed a solution of the free wave equation $\Box\phi = 0$.

Define the four-vector k according to

$$k = (k^\mu)_{\mu=0,\dots 3} = \begin{pmatrix} \omega \\ \vec{k} \end{pmatrix} . \tag{4.28}$$

(Note that this is w.r.t. $\{e_0, e_1, e_2, e_3\}$, i.e., w.r.t. the observer X under consideration.) By construction, k is a *null vector*, since $\eta(k, k) = -\omega^2 + \vec{k}^2 = 0$. Using the Minkowski metric, the plane wave (4.26) can be written as

$$\phi(x) = a e^{i\eta(k,x)} = a e^{ik_\mu x^\mu} , \tag{4.26'}$$

which is a coordinate-independent (observer-independent) expression.

A plane wave (of light) corresponds to free photons. As we will see in the following, the world looks different for these particles, which is due to the fact that k is null.

In the limit of geometric optics we deal with light rays. As implied by (4.26′), in Minkowski space, a light ray (a photon) is described by a *null line*, a straight world line whose tangent is a null vector.

$$\mathbb{R} \ni \lambda \mapsto x^\mu(\lambda) = a^\mu + \lambda k^\mu \,, \tag{4.29}$$

where a^μ and k^μ are constant four-vectors, and where

$$k^\mu k_\mu = \eta(k,k) = 0 \,. \tag{4.30}$$

For photons, the concept of proper time does not make sense. This is because we have used momentary rest frames in the derivation of proper time, which do not exist for photons—there do not exist observers who see photons at rest. (Formally, by using the formulas of section 4.3, time seems to be at a standstill for photons. However, it is beyond speculation how photons "perceive" time—in fact, we can be quite certain that photons do not "perceive" anything at all.) Accordingly, the concept of a four-velocity does not exist for photons either.

W.r.t. a chosen observer X (whose basis is $\{e_0, e_1, e_2, e_3\}$, the four-vector k naturally decomposes into

$$k = \begin{pmatrix} \omega \\ \vec{k} \end{pmatrix} \qquad \text{w.r.t. } X \,, \tag{4.31}$$

where $k^0 = \omega$ is the angular frequency and $\vec{k}$ (with $|\vec{k}| = \omega$) the wave vector as seen by the observer X. This is the direct analog of (4.16) for particles that are represented by timelike world-lines (massive particles).

Remark. In the context of timelike world-lines we have seen that the parametrization of the world-line is irrelevant. The parametrization w.r.t. proper time is beneficial, in particular when we aim at reading off directly the particle's three-velocity, but this parametrization need not be enforced. With photons, the case is different. A photon with a prescribed frequency and wave vector is described not by the null world-line alone, but by the null line plus the length of the null tangent vector k. (Reparametrizations of (4.29) would leave invariant the "path" of the photon in space and time, but they would change the frequency and the wave vector of the photon.)

Let us conclude by discussing how to obtain the frequency ω in a coordinate-independent way (i.e., we are looking for an analog of (4.21)). Consider a null line

$$\mathbb{R} \ni \lambda \mapsto x^\mu(\lambda) = a^\mu + \lambda k^\mu \,,$$

which describes a photon; k^μ is a null vector, which can be represented as in (4.31) w.r.t. X. The value of the (angular) frequency ω can be obtained in a *coordinate-independent* way. We simply note that

$$\omega = -\eta(u_X, k) = -\eta_{\mu\nu} u_X^\mu k^\nu \; . \tag{4.32}$$

CHAPTER 5

RELATIVISTIC EFFECTS

5.1 Addition of velocities and the Doppler effect

Let there be given two observers (or particles), which we call X and Y, and which are represented by the four-velocities u_X and u_Y, respectively. In particular, we have $u_X^2 = \eta(u_X, u_X) = -1$ and $u_Y^2 = \eta(u_Y, u_Y) = -1$.

In addition consider a third four-vector, which is assumed to be either another timelike normalized vector u_Z representing a third observer (or particle) Z, or a null vector k representing a light ray.

Case I. The third four-vector is u_Z, which is timelike and normalized and thus represents a third observer (or particle). Let v_{XY} denote the (absolute value of the) relative velocity between X and Y and v_{YZ} the relative velocity between Y and Z. The question we ask is the following: Given v_{XY} and v_{YZ}, what is the relative velocity v_{XZ} between X and Z?

Example. Obviously, in Galilean physics, in one (spatial) dimension, we have $v_{XZ} = |v_{XY} \pm v_{YZ}|$, where the $\pm$ sign depends on the directions of the velocities, i.e., on whether the motion is parallel or antiparallel. In relativity, this simple addition of velocities fails. For instance, in the example of section 4.5 we have seen that when the velocity of a particle A w.r.t. a given observer is $1/2$ (in the negative x^1-direction), and the velocity of another particle B is $3/4$ (in the positive x^1-direction), then the relative velocity between A and B is not $1/2 + 3/4 = 5/4$ (which would be larger than the speed of light) but $10/11$; see (4.23) and (4.24).

Case II. Alternatively, we consider the case when the third four-vector is a null vector k, i.e., $k^2 = \eta(k,k) = 0$. In this case, k describes a null line, i.e., a light ray. Let ω_Y denote the (angular) frequency as seen by Y and by v_{XY} the relative velocity between X and Y. The question we ask is the following: Given v_{XY} and ω_Y, what is the angular frequency ω_X of the photon as seen by X?

The two cases can be treated rather analogously. In the following we will take Y as our reference observer; hence, in case I,

$$u_X = \gamma_{XY} \begin{pmatrix} 1 \\ \vec{v}_{XY} \end{pmatrix}, \qquad u_Y = \begin{pmatrix} 1 \\ \vec{o} \end{pmatrix}, \qquad u_Z = \gamma_{YZ} \begin{pmatrix} 1 \\ \vec{v}_{YZ} \end{pmatrix}. \tag{5.1}$$

and, in case II,

$$u_X = \gamma_{XY} \begin{pmatrix} 1 \\ \vec{v}_{XY} \end{pmatrix}, \qquad u_Y = \begin{pmatrix} 1 \\ \vec{o} \end{pmatrix}, \qquad k = \begin{pmatrix} \omega_Y \\ \vec{k}_Y \end{pmatrix}. \tag{5.2}$$

Case I. Relativistic addition of velocities.
To compute the relative velocity v_{XZ} between X and Z we use (4.21). We obtain

$$\gamma_{XZ} = -\eta(u_X, u_Z) = -\big(-\gamma_{XY}\gamma_{YZ} + \gamma_{XY}\gamma_{YZ}\vec{v}_{XY}\vec{v}_{YZ} \big),$$

and thus

$$\gamma_{XZ}^2 = \gamma_{XY}^2\gamma_{YZ}^2(1 - \vec{v}_{XY}\vec{v}_{YZ})^2,$$

Let α denote the angle between $\vec{v}_{XY}$ and $\vec{v}_{YZ}$ and set $v_{XY} = |\vec{v}_{XY}|$, $v_{XZ} = |\vec{v}_{XZ}|$, and $v_{YZ} = |\vec{v}_{YZ}|$. A straightforward calculation then yields

$$v_{XZ} = \frac{\sqrt{v_{XY}^2 + v_{YZ}^2 - 2v_{XY}v_{YZ}\cos\alpha - v_{XY}^2 v_{YZ}^2(1-\cos^2\alpha)}}{1 - v_{XY}v_{YZ}\cos\alpha}. \tag{5.3}$$

Let us concentrate on some special cases. The case $\alpha = 0$ means that $\vec{v}_{XY}$ and $\vec{v}_{YZ}$ are parallel; $\alpha = \pi$ means that $\vec{v}_{XY}$ and $\vec{v}_{YZ}$ are antiparallel. In these cases we obtain

$$v_{XZ} = \left| \frac{v_{XY} \pm v_{YZ}}{1 \pm v_{XY}v_{YZ}} \right|. \tag{5.4}$$

The sign is a $+$ sign, if the velocities are antiparallel, and a $-$ sign, if the velocities are parallel (as seen by Y).

In the case $\alpha = \pi/2$ the velocities $\vec{v}_{XY}$ and $\vec{v}_{YZ}$ are orthogonal. It easily follows that

$$v_{XZ} = \sqrt{v_{XY}^2 + v_{YZ}^2 - v_{XY}^2 v_{YZ}^2}\,. \tag{5.5}$$

Case II. Relativistic Doppler effect.
To compute the frequency ω_X that is measured by the observer X we use (4.32). We obtain

$$\omega_X = -\eta(u_X, k) = -\big(-\gamma_{XY}\omega_Y + \gamma_{XY}\vec{v}_{XY}\vec{k}_Y\big)\,.$$

Let α denote the angle between $\vec{v}_{XY}$ and $\vec{k}_Y$ (and recall that $|\vec{k}_Y| = \omega_Y$). Then we get

$$\omega_X^2 = \gamma_{XY}^2 \omega_Y^2 \big(1 - v_{XY}\cos\alpha\big)^2$$

and, finally,

$$\omega_X = \omega_Y \frac{1 - v_{XY}\cos\alpha}{\sqrt{1 - v_{XY}^2}}\,. \tag{5.6}$$

The longitudinal Doppler effect corresponds to the case $\alpha = 0$ or $\alpha = \pi$, i.e., the direction of motion and the direction of the light ray coincide. In this case we arrive at

$$\omega_X = \omega_Y \frac{1 \mp v_{XY}}{\sqrt{1 - v_{XY}^2}} = \omega_Y \sqrt{\frac{1 \mp v_{XY}}{1 \pm v_{XY}}}\,. \tag{5.7}$$

The interpretation is straightforward: If X moves parallel to the direction of the photon ("X chases the photon"), where our reference is Y, then we have a $-$ sign in the numerator of (5.7) and

$$\omega_X = \omega_Y \sqrt{\frac{1 - v_{XY}}{1 + v_{XY}}} < \omega_Y\,; \tag{5.8a}$$

accordingly, there is a *redshift*. On the other hand, if X moves antiparallel to the direction of the photon ("X heads toward the photon"), where our reference is Y, then we have a $+$ sign in (5.7) and

$$\omega_X = \omega_Y \sqrt{\frac{1 + v_{XY}}{1 - v_{XY}}} > \omega_Y\,, \tag{5.8b}$$

i.e., the frequency increases and the wave length undergoes a *blueshift*.

The transversal Doppler effect corresponds to $\alpha = \pi/2$. We find

$$\omega_X = -\eta(u_X, k) = \gamma_{XY}\omega_Y \,. \tag{5.9}$$

Since $\gamma_{XY} = (1 - v_{XY}^2)^{-1/2}$ is larger than 1, we find $\omega_X > \omega_Y$, i.e., a blue-shift.

Remark. Note that for the transversal Doppler effect the 'symmetry' between the observers is broken. The fact that Y sees the observer X and the light ray k move in orthogonal directions does not imply that X observes the same for Y and k. On the contrary, the 'transversality property' is not 'relative'.

5.2 The aberration of light

Aberration is a well-known effect. It is best experienced on a rainy day. When it rains, the angle at which the rain drops are falling on my umbrella (provided I have one) depends on how fast I walk. (From the perspective of a car driver rain seems to fall almost horizontally.) The aberration of light underlies the same principle; however, since a consistent treatment of light is necessarily founded in relativity, we need the full machinery of special relativity to obtain a satisfying mathematical description.

To be able to study aberration we need a concept of **angles**. In Euclidean geometry (where the vector space is equipped with a positive definite scalar product $\langle\cdot|\cdot\rangle$), the angle between two vectors v and w is defined as

$$\cos\alpha_{vw} = \frac{\langle v|w\rangle}{\|v\|\,\|w\|} \,, \tag{5.10}$$

where $\|\cdot\| = \sqrt{\langle\cdot|\cdot\rangle}$. However, Minkowski space is not a Euclidean space—the Minkowski metric is not a scalar product, but a pseudo-scalar product. An attempt to generalize (5.10) would be

$$\cos\alpha_{vw} = \frac{\eta(v,w)}{\sqrt{|\eta(v,v)|}\,\sqrt{|\eta(w,w)|}} \,. \tag{5.11}$$

The absolute values of $\eta(\cdot,\cdot)$ in the denominator ensure that the square roots exist. However, if v (or w) is a null vector, then the r.h.s. of (5.11) would be infinite in general; hence, (5.11) does not make sense for null vectors. Suppose v is timelike and w is timelike or spacelike; then the inverse Cauchy-Schwarz inequality states that

$$\eta(v,w)^2 \geq \eta(v,v)\eta(w,w) \,,$$

where the equality sign refers to the case when v and w are collinear. Hence, in general, $\eta(v,w) < -\sqrt{|\eta(v,v)|}\,\sqrt{|\eta(w,w)|}$ or $\eta(v,w) > \sqrt{|\eta(v,v)|}\,\sqrt{|\eta(w,w)|}$. For (5.11) this means $\cos\alpha_{vw} < -1$ or $\cos\alpha_{vw} > 1$. This is impossible.

Exercise. Let v and w be spacelike four-vectors and define α_{vw} according to (5.11). What is the interpretation of α_{vw}?

Since an observer-independent concept of angles between four-vectors does not exist, we turn our attention to the case when an inertial coordinate system (inertial observer) X is given. (Let X's four-velocity be denoted by u_X.) Let v and w be two four-vectors; w.r.t. the coordinate system X we have the decomposition[1]

$$u_X = \begin{pmatrix} 1 \\ \vec{o} \end{pmatrix}\,, \qquad v = \begin{pmatrix} v^0 \\ \vec{v} \end{pmatrix}\,, \qquad w = \begin{pmatrix} w^0 \\ \vec{w} \end{pmatrix}\,. \tag{5.12}$$

Now, life is easy: We can consider the spatial parts of the four-vectors ('spatial' meaning 'spatial' w.r.t. X) and compute the conventional (Euclidean) angles between these three-vectors. The spatial parts are $\vec{v}$ and $\vec{w}$ and the angle α_{vw} between the two vectors is then simply given by

$$\cos\alpha_{vw} = \frac{\vec{v}\vec{w}}{|\vec{v}||\vec{w}|}\,, \tag{5.13}$$

cf. (5.10).

Let us find a coordinate-independent formula for (5.13). Define the **spatial projection** of a four-vector v w.r.t. the observer X as

$$P_X v = v + \eta(u_X, v)u_X\,. \tag{5.14}$$

W.r.t. X's coordinates we find

$$P_X v = \begin{pmatrix} 0 \\ \vec{v} \end{pmatrix}\,, \qquad P_X w = \begin{pmatrix} 0 \\ \vec{w} \end{pmatrix}\,.$$

Therefore, $P_X v$ and $P_X w$ are indeed the projections of v and w onto the (linear) plane of simultaneity $\{t = 0\}$ of X. Since $P_X v$ and $P_X w$ are purely spatial (w.r.t. X), the (Euclidean) scalar product coincides with the Minkowski product on such objects; e.g.,

$$\vec{v}\vec{w} = \eta(P_X v, P_X w)\,, \quad |\vec{v}| = \sqrt{\vec{v}\vec{v}} = \sqrt{\eta(P_X v, P_X v)}\,.$$

[1] Let us exclude the degenerate case where $\vec{v} = \vec{o}$ or $\vec{w} = \vec{o}$.

Consequently, equation (5.13) takes the form

$$\cos\alpha_{vw} = \frac{\eta(P_X v, P_X w)}{\sqrt{\eta(P_X v, P_X v)}\sqrt{\eta(P_X w, P_X w)}} \,. \tag{5.15}$$

In the following we restrict ourselves to light rays and angles in between, since this is the relevant case for our applications. Let k and l be null vectors representing two light rays. According to (5.15) the angle between k and l is

$$\cos\alpha_{kl} = \frac{\eta(P_X k, P_X l)}{\sqrt{\eta(P_X k, P_X k)}\sqrt{\eta(P_X l, P_X l)}} \,. \tag{5.16}$$

Let us insert (5.14) for $P_X k$ and $P_X l$. On the one hand,

$$\begin{aligned}\eta(P_X k, P_X l) &= \eta\big(k + \eta(u_X, k)u_X, l + \eta(u_X, l)u_X\big)\\ &= \eta(k,l) + 2\eta(u_X,k)\eta(u_X,l) + \eta(u_X,k)\eta(u_X,l)\underbrace{\eta(u_X,u_X)}_{-1}\\ &= \eta(k,l) + \eta(u_X,k)\eta(u_X,l)\,,\end{aligned}$$

and, on the other hand,

$$\begin{aligned}\eta(P_X k, P_X k) &= \eta\big(k + \eta(u_X, k)u_X, k + \eta(u_X, k)u_X\big)\\ &= \underbrace{\eta(k,k)}_{0} + 2\eta(u_X,k)\eta(u_X,k) + \eta(u_X,k)\eta(u_X,k)\underbrace{\eta(u_X,u_X)}_{-1}\\ &= \eta(u_X,k)^2\,,\end{aligned}$$

and the analog for $P_X l$.

Accordingly, the angle between (the spatial parts of) k and l is given by

$$\cos\alpha_{kl} = \frac{\eta(k,l) + \eta(u_X,k)\eta(u_X,l)}{\eta(u_X,k)\eta(u_X,l)} \,. \tag{5.17}$$

Equivalently, we have

$$\cos\alpha_{kl} - 1 = \frac{\eta(k,l)}{\eta(u_X,k)\eta(u_X,l)} \,. \tag{5.17$'$}$$

Remark. The formula (5.17′) is quite useful. Imagine an astronomer who observes two stars. The two stars correspond to points on the celestial sphere; by construction, the angle between the two corresponds to the angle α_{kl} between the (projections of the) null lines k and l, which represent the light rays emitted by the stars. Formula (5.17′) thus shows how the apparent angle between the stars changes depending on the state of motion of the observer (as described by u_X).

Remark. Rescalings of k and l (i.e., $k \mapsto \lambda_k k$ and $l \mapsto \lambda_l l$ where λ_k, λ_l are positive), leave α_{kl} invariant; this is intuitively clear, because angles are defined between the directions determined by the vectors and thus do not depend on the actual lengths of the vectors.

Equipped with (5.17′) we are now able to address the problem of the aberration of light: Consider two observers, X (with four-velocity u_X) and Y (with u_Y), and two (future-pointing) null vectors k and l (representing light rays). We ask the question of how the angle between the light rays is observed by X and Y changes depending on the relative velocity between X and Y.

Let us introduce coordinates associated with X; then u_X, u_Y, and the two (future-pointing) null vectors read

$$u_X = \begin{pmatrix} 1 \\ \vec{o} \end{pmatrix}, \qquad u_Y = \gamma \begin{pmatrix} 1 \\ \vec{v} \end{pmatrix}, \qquad k = \begin{pmatrix} k^0 \\ \vec{k} \end{pmatrix}, \qquad l = \begin{pmatrix} l^0 \\ \vec{l} \end{pmatrix}. \tag{5.18}$$

Since $\eta(k,k) = 0$ we have $k^0 = |\vec{k}|$ (and recall that k^0 is essentially the frequency of the light ray). Rescalings of k and l do not affect the angle between the two rays, hence, w.l.o.g., we may set

$$u_X = \begin{pmatrix} 1 \\ \vec{o} \end{pmatrix}, \qquad u_Y = \gamma \begin{pmatrix} 1 \\ \vec{v} \end{pmatrix}, \qquad k = \begin{pmatrix} 1 \\ \vec{n}_k \end{pmatrix}, \qquad l = \begin{pmatrix} 1 \\ \vec{n}_l \end{pmatrix}, \tag{5.18′}$$

where $\vec{n}_k$ and $\vec{n}_l$ are unit vectors, i.e., $|\vec{n}_k| = |\vec{n}_l| = 1$.

We denote the angle between (the spatial projections of) k and l w.r.t. X by θ_X; the angle the observer Y measures, we denote by θ_Y. According to (5.17′) we have

$$\cos\theta_X - 1 = \frac{\eta(k,l)}{\eta(u_X,k)\eta(u_X,l)} = \frac{-1+\vec{n}_k\vec{n}_l}{(-1)(-1)} = -1 + \vec{n}_k\vec{n}_l\,. \tag{5.19}$$

This does not come as a surprise, of course; $\cos\theta_X = \vec{n}_k\vec{n}_l$ is the standard formula. However, Y measures a different angle:

$$\begin{aligned} \cos\theta_Y - 1 &= \frac{\eta(k,l)}{\eta(u_Y,k)\eta(u_Y,l)} = \frac{-1+\vec{n}_k\vec{n}_l}{\gamma(-1+\vec{v}\vec{n}_k)\gamma(-1+\vec{v}\vec{n}_l)} \\ &= \frac{\cos\theta_X - 1}{\gamma^2(1-\vec{v}\vec{n}_k)(1-\vec{v}\vec{n}_l)}\,. \end{aligned} \tag{5.20}$$

Setting $\vec{v}\vec{n}_k = |\vec{v}||\vec{n}_k|\cos\alpha_k = v\cos\alpha_k$ and $\vec{v}\vec{n}_l = v\cos\alpha_l$, we have expressed θ_Y in terms of θ_X and the angles α_k, α_l between the light rays and the direction of motion of Y. Here and henceforth we use the abbreviation $v = |\vec{v}|$.

Let us specialize (5.20) to the important case where one of the light rays, say k, is aligned with the direction of relative motion between the observers, i.e., $\vec{n}_k \parallel \vec{v}$. In other words, we set $\vec{n}_k = \vec{v}/v$, so that

$$k = \begin{pmatrix} 1 \\ \frac{\vec{v}}{v} \end{pmatrix} , \qquad l = \begin{pmatrix} 1 \\ \vec{n}_l \end{pmatrix} . \tag{5.21}$$

In this case, we obtain

$$\cos\theta_X = \frac{\vec{v}\vec{n}_l}{v} ,$$

and (5.20) reduces to

$$\begin{aligned} \cos\theta_Y - 1 &= \frac{\cos\theta_X - 1}{\gamma^2(1 - \vec{v}\vec{v}/v)(1 - \vec{v}\vec{n}_l)} = \frac{\cos\theta_X - 1}{\gamma^2(1 - v)(1 - v\cos\theta_X)} \\ &= \frac{(1+v)(\cos\theta_X - 1)}{1 - v\cos\theta_X} . \end{aligned} \tag{5.22}$$

This formula describes the *relativistic aberration* (*aberration of light*). Applying standard algebraic manipulations we see that

$$\cos\theta' = \frac{\cos\theta - v}{1 - v\cos\theta} , \tag{5.23a}$$

$$\tan\frac{\theta'}{2} = \sqrt{\frac{1+v}{1-v}}\,\tan\frac{\theta}{2} , \tag{5.23b}$$

$$\sin\theta' = \sqrt{1 - v^2}\,\frac{\sin\theta}{1 - v\cos\theta} . \tag{5.23c}$$

Note that equivalence between the formulas (5.23a) and (5.23b) follows from the identity $\tan\alpha = (1 - \cos 2\alpha)/(\sin 2\alpha)$. Since $\sin\theta'$ is not bijective on $\theta' \in [0, \pi)$, to invert equation (5.23c) it must be completed by the additional requirement that $\theta' \in [0, \pi/2)$ if $\cos\theta > v$ and $\theta' \in (\pi/2, \pi)$ if $\cos\theta < v$.

From (5.23b) it is simple to conclude that

$$\theta' > \theta , \tag{5.24}$$

which means that the angle measured by the observer X' is larger than the angle measured by X (provided that $\theta > 0$).

If the velocity v is close to the velocity of light (which is 1 in our units), then $\theta' \gg \theta$. This has a strange effect on the field of vision. Objects that are actually located behind appear right in front of an observer if the observer moves sufficiently fast. For details we refer to the lecture course.

5.3 Lorentz contraction

So far we have only considered particles in Minkowski space (which were represented as world lines). An extended rod (in a state of uniform motion) is represented by a *world sheet*, which is a two-dimensional domain whose boundaries are two parallel world lines (which represent the two ends of the rod).

The two ends of the rod (which we call E1 and E2) are represented by world lines that are straight lines (since the rod is in uniform motion). W.l.o.g. we assume that the world line of E1 passes through the origin. We thus have

$$\mathrm{E1} : s \mapsto x_1^\mu(s) = s u^\mu \,, \qquad \mathrm{E2} : s \mapsto x_2^\mu(s) = \ell^\mu + s u^\mu \,, \tag{5.25}$$

where u^μ is the four-velocity of the rod, i.e., $\eta(u,u) = \eta_{\mu\nu} u^\mu u^\nu = -1$, and ℓ^μ is a constant four-vector representing the displacement of E1 and E2. Every vector in $\ell + \langle u \rangle$ can take the role of the displacement vector; hence, w.l.o.g. we may assume that ℓ^μ is orthogonal to u^μ, i.e., $\eta(\ell, u) = \eta_{\mu\nu} \ell^\mu u^\nu = 0$.

In the rest frame of the rod, i.e., for a comoving observer, we have $u = (1, \vec{o})^{\mathrm{T}}$, and hence

$$\mathrm{E1} : s \mapsto x_1^\mu(s) = s \begin{pmatrix} 1 \\ \vec{o} \end{pmatrix} \,, \qquad \mathrm{E2} : s \mapsto x_2^\mu(s) = \begin{pmatrix} 0 \\ \vec{\ell} \end{pmatrix} + s \begin{pmatrix} 1 \\ \vec{o} \end{pmatrix} \,. \tag{5.25$'$}$$

Definition 5.1. *The* **proper length** *of a rod is the length that is measured in the rest frame of the rod.*

The proper length L of (5.25′) is obviously given by $L^2 = |\vec{l}|^2$; equivalently, we write

$$L^2 = \eta(\ell, \ell) \,. \tag{5.26}$$

Let X denote an inertial observer with four-velocity u_X. What is the length of the rod according to X? The length of the rod is simply the distance between the two ends E1 and E2 at a given instant of time (where by time we obviously mean X's time t).

Let us therefore calculate two events on the world lines E1 and E2, respectively, that are simultaneous (for X); for simplicity, we determine the two events on the plane of simultaneity $t = 0$. From (3.13) we obtain

$$\begin{aligned} \text{E1}: \quad & \eta(x_1(s), u_X) = 0 \quad \Leftrightarrow \quad s\eta(u, u_X) = 0 && \Leftrightarrow \quad s = 0\,, \\ \text{E2}: \quad & \eta(x_2(s), u_X) = 0 \quad \Leftrightarrow \quad \eta(\ell, u_X) + s\eta(u, u_X) = 0 && \Leftrightarrow \quad s = -\frac{\eta(\ell, u_X)}{\eta(u, u_X)}\,. \end{aligned}$$

Therefore, the events

$$x_1 = 0 \text{ on E1} \quad \text{and} \quad x_2 = \ell - \frac{\eta(\ell, u_X)}{\eta(u, u_X)}\, u \text{ on E2} \tag{5.27}$$

are simultaneous for the inertial observer X. (In coordinates adapted to X, these events would read $(0, \vec{o})^{\mathrm{T}}$ and $(0, \vec{\ell}_X)^{\mathrm{T}}$ for some $\vec{\ell}_X$. In the coordinates (5.29) we have $(0, \vec{o})^{\mathrm{T}}$ and $(\vec{\ell}\vec{v}, \vec{\ell})^{\mathrm{T}}$. We could use the latter to derive (5.30).)

The length L_X of the rod that X measures is the distance between the two events; hence,

$$L_X^2 = \eta\Big(\ell - \frac{\eta(\ell, u_X)}{\eta(u, u_X)}\, u, \ell - \frac{\eta(\ell, u_X)}{\eta(u, u_X)}\, u\Big)\,.$$

Simple algebraic manipulations using $\eta(\ell, u) = 0$ and $\eta(u, u) = -1$ show that

$$L_X^2 = \eta(\ell, \ell) - \frac{\eta(\ell, u_X)^2}{\eta(u, u_X)^2} = L^2 - \frac{\eta(\ell, u_X)^2}{\eta(u, u_X)^2}\,. \tag{5.28}$$

An immediate consequence is that $L_X \leq L$; note that the comoving observer reproduces the result $L_X = L$, since the second term in (5.28) vanishes when $u_X = u$.

Let us use the coordinates of the comoving observer to compute L_X in terms of the relative velocity. (Since (5.28) is formulated in a coordinate-independent way we may choose any coordinate we like.) In these coordinates we have

$$u = \begin{pmatrix} 1 \\ \vec{o} \end{pmatrix}, \qquad \ell = \begin{pmatrix} 0 \\ \vec{\ell} \end{pmatrix}, \qquad u_X = \gamma \begin{pmatrix} 1 \\ \vec{v} \end{pmatrix}, \tag{5.29}$$

where $\vec{v}$ is the velocity of X as seen by the comoving observer. Then (5.28) yields

$$L_X^2 = L^2 - (\vec{v}\vec{\ell})^2 = L^2 - \vec{v}^2\vec{\ell}^2 \cos^2\alpha\,, \tag{5.30}$$

where α is the angle between $\vec{v}$ and $\vec{\ell}$ (as seen by the comoving observer). Therefore,

$$L_X = L\sqrt{1 - |\vec{v}|^2 \cos^2 \alpha}\ ; \tag{5.31}$$

in particular,

$$L \geq L_X\ . \tag{5.32}$$

Remark. We see that the proper length of the rod (which is the length in the rest frame) can also be viewed as the maximizer of the lengths of the rod as seen by inertial observers.

Two cases are of special interest. First, the transversal case: Suppose that the rod and the direction of motion are orthogonal, i.e., $\vec{\ell} \perp \vec{v}$. Then $\cos \alpha = 0$ (i.e., $\alpha = \pi/2$) and

$$L_X = L\ . \tag{5.33}$$

In other words, there does not exist a transversal contraction of lengths.

Second, the longitudinal case: Suppose that the direction of motion is parallel (or antiparallel) to the rod itself, i.e., $\vec{\ell} \parallel \vec{v}$. Then $\cos \alpha = 1$ (i.e., $\alpha = 0$) and

$$L_X = L\gamma^{-1} = L\sqrt{1 - |\vec{v}|^2}\ . \tag{5.34}$$

This effect is called the *Lorentz contraction*. For the observer X the rod appears contracted by a factor γ^{-1}.

A well-known "paradox" involving Lorentz contractions is the *ladder paradox*. If a ladder is traveling at high speed it may undergo a sufficient length contraction to fit into a much smaller garage. On the other hand, from the point of view of an observer who is comoving with the ladder, it is the garage that is Lorentz contracted to a length that is smaller than its proper length, which means that the garage will be unable to contain the ladder at all. For a discussion we refer to the lecture course.

CHAPTER 6

ENERGY AND MOMENTUM

6.1 Introduction

In Newtonian (Galilean) physics the momentum $\vec{p}$ of a point particle is defined as $\vec{p} = m\vec{v}$, where m is the particle's mass; the kinetic energy is $E = \frac{1}{2}m\vec{v}^2$; here, $\vec{v}$ is the velocity of the particle w.r.t. a (Galilean) inertial observer. Conservation of momentum and energy is crucial. Consider a system of particles m_i $(i = 1, 2, \ldots, n)$ that come together, interact, and move out again as a different system m'_j $(j = 1, 2, \ldots, n')$. Suppose for simplicity that we have only one spatial degree of freedom. Then

$$\sum_i m_i v_i = \sum_j m'_j v'_j \qquad \text{and} \qquad \sum_i m_i v_i^2 = \sum_j m'_j v'^2_j \,. \tag{6.1}$$

The (Galilean) principle of relativity implies that conservation of momentum and energy must hold in each inertial frame. For an observer that moves with velocity u w.r.t. the original observer, the Galilean transformation yields

$$\sum_i m_i(v_i + u) = \sum_j m'_j(v'_j + u)\,, \qquad \sum_i m_i(v_i + u)^2 = \sum_j m'_j(v'_j + u)^2 \,. \tag{6.2}$$

Expanding the first equation in the variable u results in

$$\sum_i m_i v_i + u \sum_i m_i = \sum_j m_j v'_j + u \sum_j m'_j \,,$$

and the second equation becomes

$$\sum_i m_i v_i^2 + 2u \sum_i m_i v_i + u^2 \sum_i m_i = \sum_j m'_j v'^2_j + 2u \sum_j m'_j v'_j + u^2 \sum_j m'_j \,,$$

which is consistent with (6.1) and in addition implies a conservation of the sum of the masses.

And in relativity? Suppose the definition of energy and momentum are the same in relativistic physics. Since (6.1) must hold in each (relativistic) inertial frame by the principle of relativity, we find that conservation of momentum looks like

$$\sum_i m_i \frac{v_i + u}{1 + v_i u} = \sum_j m'_j \frac{v'_j + u}{1 + v_j u} \tag{6.3}$$

for an observer who moves with velocity u (w.r.t. the nameless observer we had chosen initially); here we have used the relativistic addition of velocities (5.4). Expanding (6.3) in u yields

$$\sum_i m_i v_i + u\Big(\sum_i m_i - \sum_i m_i v_i^2\Big) + \\ + u^2\Big(-\sum_i m_i v_i + \sum_i m_i v_i^3\Big) + u^3\Big(\sum_i m_i v_i^2 - \sum_i m_i v_i^4\Big) + \ldots$$

and the analogous expression for the r.h. side. Equating the two sides then leads to

$$\sum_i m_i (v_i)^k = \sum_j m'_j (v'_j)^k \tag{6.4}$$

for all $k \in \mathbb{N}$. This is ridiculous.

We conclude that either conservation of momentum and energy fails in relativity or the concept of momentum and energy is different. It is the latter…

6.2 Four-momentum of massive particles

Definition 6.1 (Four-momentum)**.** *Consider a particle represented by a time-like world line $s \mapsto x(s)$. Let u denote the four-velocity of the particle. Then the particle's four-momentum is given by*

$$p = mu \,, \tag{6.5}$$

where m is the mass of the particle.

Since $u^2 = \eta(u,u) = -1$ we have

$$p^2 = \eta(p,p) = -m^2\,. \tag{6.6}$$

Let X be an arbitrary inertial observer. W.r.t. X, the four-momentum of the particle is

$$p = mu = m\gamma \begin{pmatrix} 1 \\ \vec{v} \end{pmatrix}\,. \tag{6.7}$$

Remark. In SI units (where $c \neq 1$) we obtain

$$p = m\gamma \begin{pmatrix} c \\ \vec{v} \end{pmatrix}\,. \tag{6.7$'$}$$

Note that this leaves (6.6) untouched, since $\eta(x,y) = -\frac{1}{c^2}x^0y^0 + \vec{x}\vec{y}$ in this case.

As usual we distinguish a temporal and spatial part of p, i.e., we have

$$p = \begin{pmatrix} p^0 \\ \vec{p} \end{pmatrix} = m\gamma \begin{pmatrix} c \\ \vec{v} \end{pmatrix}\,. \tag{6.8}$$

The vector $\vec{p}$ is the three-momentum of the particle w.r.t. the observer X,

$$\vec{p} = m\gamma\vec{v}\,. \tag{6.9}$$

In particular it is different from the Newtonian momentum. However, in the limit of small velocities we obtain

$$\vec{p} = m\vec{v}\left[1 + \frac{|\vec{v}|^2}{2c^2} + O\Big(\frac{|\vec{v}|^4}{c^4}\Big)\right]\,, \tag{6.10}$$

i.e., we recover the Newtonian momentum for small velocities.

Remark. Despite the claim in numerous textbooks, it is a fact that the mass of a particle does not (I repeat: *does not*) increase with the velocity of the particle.[1] The idea of a "relativistic mass" has its roots in the obsession that the momentum must be written as $\vec{p} = m\vec{v}$. This implies that one must tamper with the mass and make it depend on $|\vec{v}|$. However, the simple truth is that the three-momentum is not proportional to the three-velocity, but contains a γ-factor, see (6.9). This leaves the mass as it is and as we love it: as a constant.

[1] This sentence is scandalous anyway, since it suggests that velocity is an absolute quantity.

The temporal part of p is given by p^0. Evidently it has the dimension of a momentum, hence p^0c has the dimension of an *energy*:

$$E = p^0 c = m\gamma c^2 \,. \tag{6.11}$$

In the Newtonian approximation we obtain

$$E = mc^2 \left[1 + \frac{|\vec{v}|^2}{2c^2} + \frac{3|\vec{v}|^4}{8c^4} + \dots \right] = mc^2 + \frac{m|\vec{v}|^2}{2} + \frac{3m|\vec{v}|^4}{8c^2} + \dots \,. \tag{6.12}$$

The energy E is the *energy of the particle* (w.r.t. the observer X). W.r.t. a frame where the particle is at rest the energy reduce to the *rest energy*,

$$E_{\text{rest}} = mc^2 \,. \tag{6.13}$$

This relationship between mass and energy is the foundation of the *equivalence of mass and energy.*

W.r.t. a frame where the particle is in motion the energy consists of the rest energy term and a term that is easily identified as the generalization of the Newtonian kinetic energy. (In fact, for small velocities, the kinetic energy reduces to the standard Newtonian kinetic energy.) Hence,

$$E = E_{\text{rest}} + E_{\text{kin}} \,, \tag{6.14}$$

where

$$E_{\text{kin}} = E - E_{\text{rest}} = mc^2\big(\gamma - 1\big) = \frac{m|\vec{v}|^2}{2} + \frac{3m|\vec{v}|^4}{8c^2} + \dots \,. \tag{6.15}$$

In units where

$$c = 1 \,, \tag{6.16}$$

the formulas are even simpler: Since

$$p = m\gamma \begin{pmatrix} 1 \\ \vec{v} \end{pmatrix} \tag{6.17}$$

we have

$$E = p^0 = m\gamma \,, \tag{6.18a}$$
$$\vec{p} = m\gamma\vec{v} \,. \tag{6.18b}$$

Moreover,

$$p^2 = \eta(p,p) = -E^2 + \vec{p}^2 = -m^2 \,. \tag{6.19}$$

6.3 Four-momentum of photons

Definition 6.2 (Four-momentum)**.** *Consider a photon represented by a null line with null vector k. Then the photon's four-momentum is given by*

$$p = \hbar k \,, \tag{6.20}$$

where $\hbar$ is the reduced Planck constant.

This definition has its roots in quantum mechanics. When ν is the frequency and ω the angular frequency of a photon, then its energy is given by $E = h\nu = \hbar\omega$ Likewise the three-momentum is $\vec{p} = \hbar\vec{k}$. Since ω and $\vec{k}$ are the building blocks of k, the definition (6.20) ensues.

Note that

$$p^2 = \eta(p,p) = 0 \tag{6.21}$$

for photons.

6.4 Four-momentum conservation

Suppose we have a collision of particles. The n incoming particles have four-momenta p_i, $i = 1, 2, \ldots, n$. The n' outgoing particles have four-momenta p'_j, $j = 1, 2, \ldots, n'$. Then the four-momenta are conserved, i.e.,

$$\sum_i p_i = \sum_j p'_j \,. \tag{6.22}$$

Since this equation includes the energies and the spatial components, it formulates conservation of energy and (three-)momentum at the same time.

Example (Decay of a particle). Suppose we have one particle with mass m that splits into two particles, each of mass m'. Conservation of four-momentum reads

$$p = p'_1 + p'_2 \,, \tag{6.23}$$

where $p = mu$ and $p'_1 = m'u'_1$ and $p'_2 = m'u'_2$ and where u, u'_1, u'_2 are the four-velocities. W.l.o.g. we can do the computations in the rest frame of the first particle, hence

$$m \begin{pmatrix} 1 \\ \vec{o} \end{pmatrix} = m'\gamma'_1 \begin{pmatrix} 1 \\ \vec{v}'_1 \end{pmatrix} + m'\gamma'_2 \begin{pmatrix} 1 \\ \vec{v}'_2 \end{pmatrix} \,. \tag{6.24}$$

The equation for the (three-)momentum implies that $\vec{v}_1' = \vec{v}'$ and $\vec{v}_2' = -\vec{v}'$ and $\gamma_1' = \gamma_2' = \gamma'$. Therefore, the equation for the energy results in

$$m = 2m'\gamma' = 2m' \frac{1}{\sqrt{1 - |\vec{v}|^2}} . \tag{6.25}$$

In particular we find

$$m > 2m' . \tag{6.26}$$

The Newtonian conservation of mass states that the sum of the masses on the l.h. side equals the sum of the masses on the r.h. side. This type of *conservation of mass does not hold* in general in special relativity.

6.5 Relativistic billiards[2]

To analyze the physics of billiards we study an elastic collision of two particles of equal rest mass, one of which is originally at rest. An elastic collision is one in which the masses of the particles are unchanged.

Let the four-momenta of the particles before the collision be denoted by p and q, respectively, and by p' and q' after the collision. Four-momentum conservation implies

$$p + q = p' + q' . \tag{6.27}$$

W.l.o.g. we set the mass m of the particles to one, so that $p^2 = -1$, $q^2 = -1$, $p'^2 = -1$, $q'^2 = -1$ (where in our notation $p^2 = \eta(p,p)$,...). From (6.27) we are able to derive a number of simple identities,

$$\eta(p,p') = \eta(q,q') , \qquad \eta(p,q) = \eta(p',q') , \qquad \eta(p,q') = \eta(p',q) ; \tag{6.28}$$

for instance, the latter follows by computing and equating the (Minkowski) norms of $p - q'$ and $p' - q$.

By assumption, one of the particles is originally at rest; we choose this to be the second particle, which is described by q. Hence the inertial frame we use is represented by a four-velocity that is equal to q. This inertial frame is called the *laboratory frame*.

[2]The treatment of relativistic billiards in this section is inspired by [Beig:2001] R. Beig, Relativistic Billiards in the Lab Frame, Preprint UWTh-Ph-2001-9, 2001 (unpublished).

Let $\vartheta_{pp'}$ denote the angle between the (spatial) direction of the incoming particle p and the outgoing particle p'. Using (5.14) in (5.15) we get

$$\cos\vartheta_{pp'} = \frac{\eta(p,p') + \eta(q,p)\eta(q,p')}{\sqrt{-1+\eta(q,p)^2}\sqrt{-1+\eta(q,p')^2}} \,.$$

Since

$$\eta(p,p') = \eta(p,p+q-q') = -1 + \eta(p,q) - \eta(p,q') = -1 + \eta(p,q) - \eta(p',q) \,,$$

we further obtain

$$\cos\vartheta_{pp'} = \frac{(-1+\eta(p,q))(1+\eta(p',q))}{\sqrt{-1+\eta(q,p)^2}\sqrt{-1+\eta(q,p')^2}} = \sqrt{\frac{1-\eta(p,q)}{-1-\eta(p,q)}}\sqrt{\frac{-1-\eta(p',q)}{1-\eta(p',q)}} \,,$$

or, by straightforward manipulations,

$$\tan\vartheta_{pp'} = \frac{\sqrt{2\eta(p',q) - 2\eta(p,q)}}{\sqrt{1-\eta(p,q)}\sqrt{-1-\eta(p',q)}} \,.$$

In complete analogy we compute the angle $\vartheta_{pq'}$ between the (spatial) direction of the incoming particle p and the outgoing particle q':

$$\tan\vartheta_{pq'} = \frac{\sqrt{2\eta(q',q) - 2\eta(p,q)}}{\sqrt{1-\eta(p,q)}\sqrt{-1-\eta(q',q)}} \,.$$

We can view the product $\eta(p,q)$ as given (since this quantity is characteristic of the experimental set-up), while the quantities $\eta(p',q)$ and $\eta(q',q)$ appearing in the formulas are unknown. However,

$$\eta(q',q) = \eta(p+q-p',q) = -1 + \eta(p,q) - \eta(p',q) \,,$$

hence the only true variable is $\eta(p',q)$. Consequently, we can eliminate $\eta(p',q)$ from the equations and then express $\vartheta_{pq'}$ as a function of $\vartheta_{pp'}$ and the given quantity $\eta(p,q)$. The most clever way is to simply multiply $\tan\vartheta_{pp'}$ and $\tan\vartheta_{pq'}$; we obtain

$$\tan\vartheta_{pp'}\,\tan\vartheta_{pq'} = \frac{2}{1-\eta(p,q)} \,. \tag{6.29}$$

Since $\eta(p,q) = -\gamma_v = -(1-|v|^2)^{-1/2}$, where v is the relative velocity between the particles p and q (or, simply, the velocity of the particle p in the lab frame), we have

$$\tan\vartheta_{pp'}\,\tan\vartheta_{pq'} = \frac{2}{1+\gamma_v} \,. \tag{6.30}$$

Consequently,

$$\vartheta_{pp'} + \vartheta_{pq'} < 90^\circ \,, \tag{6.31}$$

since $\gamma_v > 1$.

Remark. In the Newtonian limit we deal with small velocities, hence $\gamma \simeq 1$. The formula thus reduces to

$$\tan\vartheta_{pp'}\,\tan\vartheta_{pq'} = 1 \,, \tag{6.32}$$

which implies that $\vartheta_{pp'} + \vartheta_{pq'} = 90^\circ$. For a proof we use that

$$\tan a\,\tan b = \frac{\cos(a-b) - \cos(a+b)}{\cos(a-b) + \cos(a+b)} \,.$$

If you've ever played billiard, you needn't be convinced that 90° is the correct result.

Let us conclude this section with an extended remark. Let us try to understand what four-momentum conservation is able to tell us. We assume that p and q are known from the experimental set-up of the problem; in particular, the (three-)velocity associated with p is given. Therefore, since the masses are known, the unknowns of the problem are the (three-)velocities of the two particles after the collision. (From their three-velocities we are able to construct their four-velocities.) In other words, there are 6 unknown variables. On the other hand, four-momentum conservation provides us with 4 equations. Consequently, we will not be able to solve the problem completely when we only use the conservation equations; to compeletly solve the problem we must analyze the equations of motion and the (field) theory that models the interaction between the particles. In the present case, however, and in a large number of other cases, we pose a type of question that can be answered by using four-momentum conservation alone.

Let us discuss the problem at hand in detail. Physical intuition suggests that the problem is effectively two-dimensional; the three-velocities involved will be coplanar (i.e., lie in a plane). In fact, this is the first result implied by four-momentum conservation. (We have thus used one of the four equations of (6.27); three more to go.) Since the three-velocity associated with p must lie in the plane (and p is known), the unknown position of the plane corresponds to one degree of freedom (which is, e.g., a variable representing the plane's rotational angle about the axis through p). The one equation of (6.27) that we have used so far thus reduces the 6 unknowns to $4 + 1$ (the latter being the plane's angle). There are 4 remaining unknowns which we take to be the following: The absolute values of the velocities $v_{p'}$ and $v_{q'}$ of the outgoing particles and the angles $\vartheta_{pp'}$ and $\vartheta_{pq'}$ between the (spatial) direction of the incoming particle p and the outgoing particles p' and q', respectively. Since four-momentum conservation provides three conditions that we have not used up yet, we will be able to compute all four unknown $v_{p'}$, $v_{q'}$, $\vartheta_{pp'}$, and $\vartheta_{pq'}$ except for one that remains unspecified. For example, we will be able to express three of the variables as functions of the remaining one. (A good choice for the remaining unknown is $v_{p'}$, which is equivalent to $\eta(p', q)$.) The reader is encouraged to reread the analysis of this section in the light of the above. Da capo al fine.

6.6 The Compton effect

Let us consider a set-up that is similar to the one for the billiards. We assume that p is the four-momentum of an ingoing photon that scatters at an electron with four-momentum q. The interaction leads to p' for the photon and q' for the electron,

$$p + q = p' + q' \,. \tag{6.33}$$

The vectors p and p' are null vectors, $p^2 = 0$ and $p'^2 = 0$, while $q^2 = -m_\mathrm{e}^2$ and $q'^2 = -m_\mathrm{e}^2$, where m_e is the electron mass.

The lab frame is the inertial frame where the electron is at rest initially, i.e., the inertial frame with four-velocity u such that

$$q = m_\mathrm{e} u \,. \tag{6.34}$$

Let us compute the angle $\vartheta_{pp'}$ between the direction of the incoming photon p and the outgoing photon p' w.r.t. the lab frame. Since p and p' are null we can directly apply (5.17′), i.e.,

$$\cos\vartheta_{pp'} - 1 = \frac{\eta(p, p')}{\eta(u, p)\eta(u, p')} \,. \tag{6.35}$$

By definition, $p = \hbar k$, where k is the null vector of the photon; hence

$$\eta(u, p) = \hbar\eta(u, k) = -\hbar\omega \,, \tag{6.36}$$

and analogously, $\eta(u, p') = -\hbar\omega'$; here, ω and ω' are the (angular) frequencies of the photons p and p', respectively, as seen in the lab frame defined by u.

Also the product $\eta(p,p')$ is related to the angular frequencies. When we multiply (6.33) with p we obtain

$$\eta(p,p)+\eta(p,q)=\eta(p,p')+\eta(p,q')=\eta(p,p')+\eta(p',q)\ ,$$

where we have used (6.28) in the last step. Since $\eta(p,p)=0$ we conclude that

$$\eta(p,p')=\eta(q,p)-\eta(q,p')=m_{\mathrm{e}}\hbar\left[\eta(u,k)-\eta(u,k')\right]=m_{\mathrm{e}}\hbar\left[-\omega+\omega'\right]\ .$$

Combining these results leads to

$$\cos\vartheta_{pp'}-1=\frac{m_{\mathrm{e}}}{\hbar}\,\frac{(-\omega+\omega')}{\omega\omega'} \tag{6.37}$$

or

$$\frac{1}{\omega'}-\frac{1}{\omega}=\frac{\hbar}{m_{\mathrm{e}}}\left[1-\cos\vartheta_{pp'}\right]\ , \tag{6.38}$$

which is in turn equivalent to

$$\lambda'-\lambda=\frac{h}{m_{\mathrm{e}}}\left[1-\cos\vartheta_{pp'}\right]\ , \tag{6.38$'$}$$

where λ and λ' denote the wave lengths of p and p'. In units where $c\neq 1$ the factor on the r.h. side reads

$$\lambda_{\mathrm{C}}=\frac{h}{m_{\mathrm{e}}c}\ ; \tag{6.39}$$

this is the Compton wave length.

Equation (6.38) shows the change of wave length of the scattered photon in dependence on the angle.

CHAPTER 7

ACCELERATED MOTION

7.1 Acceleration

Consider a particle represented by a timelike world line

$$\mathbb{R} \ni s \mapsto x^\mu(s) \,, \tag{7.1}$$

where s is proper time. The four-velocity is given by the derivative of $x^\mu(s)$ w.r.t. s, i.e.,

$$u^\mu(s) = \dot{x}^\mu(s) = \frac{d}{ds} x^\mu(s) \,. \tag{7.2}$$

Definition 7.1. *The* four-acceleration *is the second derivative of $x(s)$ w.r.t. proper time, i.e.,*

$$a^\mu(s) = \ddot{x}^\mu(s) = \frac{d^2}{ds^2} x^\mu(s) \,. \tag{7.3}$$

Differentiating the condition $\eta\big(\dot{x}(s), \dot{x}(s)\big) = -1$ we obtain

$$\eta\big(\dot{x}(s), \ddot{x}(s)\big) = 0 \,,$$

which is the same as

$$\eta\big(u, a\big) = 0 \,. \tag{7.4}$$

Hence, the four-acceleration a^μ is orthogonal (in the Minkowski sense) to the four-velocity u^μ. This makes the four-acceleration a *spacelike* vector, see proposition 3.5.

Suppose a particle with world line $x^\mu(s)$ has vanishing four-acceleration. From $\ddot{x}^\mu(s) = 0$ it follows that $\dot{x}^\mu(s)$ is a constant four-vector (whose norm must be -1 because s is supposed to be proper time). Accordingly, $x^\mu(s)$ describes a straight world line, i.e., a particle in uniform motion.

Evidently, the concept of four-acceleration is rather simple. However, to make contact with the concept of 'three-acceleration' we have to work hard. We begin by recalling that (w.r.t. a given observer) a four-velocity u^μ can be written as

$$u^\mu = \begin{pmatrix} u^0 \\ \vec{u} \end{pmatrix} = \gamma \begin{pmatrix} 1 \\ \vec{v} \end{pmatrix} . \tag{7.5}$$

Hence the three-velocity $\vec{v}$ (which is the relative velocity between the particle and the observer) has a simple relationship with the spatial components of the four-velocity; namely, $\vec{u} = \gamma\vec{v}$. For accelerations life is not so simple. Let us compute $a^\mu(s)$ from $u^\mu(s)$.

$$a^\mu = \frac{d}{ds}\left[\gamma \begin{pmatrix} 1 \\ \vec{v} \end{pmatrix}\right] = \frac{d\gamma}{ds}\begin{pmatrix} 1 \\ \vec{v} \end{pmatrix} + \gamma \begin{pmatrix} 0 \\ \frac{d\vec{v}}{ds} \end{pmatrix} . \tag{7.6}$$

Since $ds = \gamma^{-1}dt$, see (4.11), we obtain

$$a^\mu = \gamma\frac{d\gamma}{dt}\begin{pmatrix} 1 \\ \vec{v} \end{pmatrix} + \gamma^2 \begin{pmatrix} 0 \\ \frac{d\vec{v}}{dt} \end{pmatrix} = \gamma^4\left(\vec{v}\,\frac{d\vec{v}}{dt}\right)\begin{pmatrix} 1 \\ \vec{v} \end{pmatrix} + \gamma^2 \begin{pmatrix} 0 \\ \frac{d\vec{v}}{dt} \end{pmatrix} , \tag{7.7}$$

where we have used that $d\gamma/dt = \gamma^3\vec{v}(d\vec{v}/dt)$. We call

$$\vec{a}_N = \frac{d\vec{v}}{dt} \tag{7.8}$$

the Newtonian three-acceleration. Accordingly,

$$a^\mu = \gamma^4(\vec{v}\,\vec{a}_N)\begin{pmatrix} 1 \\ \vec{v} \end{pmatrix} + \gamma^2 \begin{pmatrix} 0 \\ \vec{a}_N \end{pmatrix} . \tag{7.9}$$

The Newtonian three-acceleration $\vec{a}_N$ is a poor measure of the acceleration. To see this we consider a particle whose velocity is close to the speed of light (w.r.t. the given observer). When this particle is further accelerated, its velocity increases only marginally; hence $\vec{a}_N$ is essential zero. However, the momentum still increases by large amounts because $\vec{p} = m\gamma\vec{v}$ (and not $\vec{p} = m\vec{v}$). A large gain in momentum should correspond to a large acceleration. This is not properly reflected by the quantity $\vec{a}_N$.

In search for a better measure for the three-acceleration we analyze the particle in a coordinate system in which the particle is instantaneously at rest. Let $\utilde{X}$ be a coordinate system (with coordinates $\{\utilde{t}, \utilde{x}^i\}$) such that the particle is at rest at time $\utilde{t} = \utilde{t}_r$. Hence the velocity $\vec{\utilde{v}}(\utilde{t})$ (which is the relative velocity of the particle w.r.t. $\utilde{X}$) satisfies $\vec{\utilde{v}}(\utilde{t}_r) = \vec{o}$, i.e.,

$$u^\mu = \utilde{\gamma}\begin{pmatrix}1\\ \vec{\utilde{v}}\end{pmatrix}, \qquad u^\mu\Big|_{\utilde{t}=\utilde{t}_r} = \begin{pmatrix}1\\ \vec{o}\end{pmatrix}. \tag{7.10}$$

Furthermore, inserting $\vec{\utilde{v}}(\utilde{t}_r) = \vec{o}$ in (7.7) we get

$$a^\mu\Big|_{\utilde{t}=\utilde{t}_r} = \begin{pmatrix}0\\ \frac{d\vec{\utilde{v}}}{d\utilde{t}}\Big|_{\utilde{t}=\utilde{t}_r}\end{pmatrix}. \tag{7.11}$$

We call

$$\vec{a}_p = \frac{d\vec{\utilde{v}}}{d\utilde{t}}\Big|_{\utilde{t}=\utilde{t}_r} \tag{7.12}$$

the *proper acceleration* of the particle—it is the (Newtonian) acceleration *measured in a momentary rest frame.*

Remark. It is not difficult to argue that the proper acceleration is the actual acceleration experienced by the particle. Namely, in the momentary rest frame the particle's velocities for times $\utilde{t}$ in a small interval $[\utilde{t}_r - \epsilon, \utilde{t}_r + \epsilon]$ are small. Consequently, Newtonian physics is a good approximation in $[\utilde{t}_r - \epsilon, \utilde{t}_r + \epsilon]$ (and, in fact, exact at $\utilde{t} = \utilde{t}_r$), so that the Newtonian concepts (like $d\vec{\utilde{v}}/d\utilde{t}$) reflect the physically reality in a correct way.

We write (7.11) as

$$a^\mu = \begin{pmatrix}0\\ \vec{a}_p\end{pmatrix}, \tag{7.13}$$

where it is understood that the coordinate system that is used is a momentary rest frame. From (7.13) it is easy to see that

$$a^2 = \eta(a, a) = \vec{a}_p^{\,2}. \tag{7.14}$$

Hence the magnitude of the proper acceleration coincides with the norm of the four-acceleration.

Our next aim is express the proper acceleration $\vec{a}_p$ in terms of the Newtonian acceleration $\vec{a}_N$ for arbitrary observers. Equation (7.9) results in

$$\begin{aligned} a^2 &= \gamma^8(\vec{v}\,\vec{a}_N)^2(-1+\vec{v}^2) + 2\gamma^6(\vec{v}\,\vec{a}_N)^2 + \gamma^4\vec{a}_N^2 = \gamma^6(\vec{v}\,\vec{a}_N)^2 + \gamma^4\vec{a}_N^2 \\ &= \gamma^6\left[(\vec{v}\,\vec{a}_N)^2 + (1-\vec{v}^2)\vec{a}_N^2\right] = \gamma^6\left[1-\vec{v}^2\sin^2\alpha\right]\vec{a}_N^2 \,, \end{aligned} \tag{7.15}$$

where α is the angle between $\vec{v}$ and $\vec{a}_N$.

If $\vec{v}$ and $\vec{a}_N$ are (anti)parallel, i.e., in the longitudinal case, we have $\alpha = 0$, so that

$$a^2 = \vec{a}_p^2 = \gamma^6\vec{a}_N^2$$

and thus

$$\vec{a}_p = \gamma^3\vec{a}_N \,. \tag{7.16}$$

Since the problem is effectively (spatially) one-dimensional, we can simply write

$$a_p = \gamma^3 a_N \,. \tag{7.16$'$}$$

where a_p and a_N denote the magnitude of the respective accelerations. This leads to the formula

$$a_p = \gamma^3\,\frac{dv}{dt} \tag{7.16$''$}$$

that describes the *linear (proper) acceleration* of a particle (where the word 'linear' is synonymous with 'longitudinal').

If $\vec{v}$ and $\vec{a}_N$ are orthogonal, i.e., in the transversal case, we have $\alpha = 90°$, so that

$$a^2 = \vec{a}_p^2 = \gamma^4\vec{a}_N^2$$

and thus

$$\vec{a}_p = \gamma^2\vec{a}_N \,, \tag{7.17}$$

which describes the orthogonal (proper) acceleration of a particle.

Combining the two formulas it follows that in the general case

$$\vec{a}_p = \gamma^3\vec{a}_N^{||} + \gamma^2\vec{a}_N^{\perp} \,, \tag{7.18}$$

where $\vec{a}_N^{||}$ is the component of $\vec{a}_N$ parallel to $\vec{v}$ and $\vec{a}_N^{\perp}$ the component orthogonal to $\vec{v}$, i.e.,

$$\vec{a}_N^{||} = \left(\vec{a}_N\vec{v}\right)\frac{\vec{v}}{|\vec{v}|^2} \,, \qquad \vec{a}_N^{\perp} = \vec{a}_N - \left(\vec{a}_N\vec{v}\right)\frac{\vec{v}}{|\vec{v}|^2} \,.$$

Remark. Finally, the four-acceleration can be expressed in terms of the proper acceleration as

$$a^\mu = \gamma(\vec{v}\vec{a}_p)\begin{pmatrix}1\\ \frac{\vec{v}}{\vec{v}^2}\end{pmatrix} + \begin{pmatrix}0\\ \vec{a}_p - (\vec{v}\vec{a}_p)\frac{\vec{v}}{\vec{v}^2}\end{pmatrix} = \gamma\begin{pmatrix}\vec{v}\vec{a}_p\\ \vec{a}_p^{\|}\end{pmatrix} + \begin{pmatrix}0\\ \vec{a}_p^{\perp}\end{pmatrix}\,.$$

Remark. Occasionally, equation (7.15) is expressed not in terms of $\vec{a}_N = d\vec{v}/dt$ but in $d\vec{v}/ds$. Since $d\vec{v}/ds = \gamma\, d\vec{v}/dt$, we have

$$a^2 = \gamma^4\left[1 - \vec{v}^2\sin^2\alpha\right]\left(\frac{d\vec{v}}{ds}\right)^2 \tag{7.15$'$}$$

and thus

$$a_p = \gamma^2\,\frac{dv}{ds} \tag{7.16$'''$}$$

in the case of linear acceleration.

7.2 Constant linear acceleration

Constant acceleration means constant proper acceleration. Since $a^2 = \vec{a}_p^2$, we require that the norm of the four-acceleration is constant, i.e.,

$$a^2 = \eta(a,a) = \text{const}\,. \tag{7.19}$$

In the case of linear acceleration, the motion of the particle is along a straight line (in space). In slight abuse of notation we denote by a also the constant value of the acceleration, i.e., $a = \sqrt{\eta(a,a)} = \text{const}$. Equation (7.16$''$) reads

$$\gamma^3\frac{dv}{dt} = a = \text{const}\,, \tag{7.20}$$

which is a differential equation can be integrated rather easily by noting that

$$\int (1-v^2)^{-3/2}dv = \frac{v}{\sqrt{1-v^2}}\,. \tag{7.21}$$

Exercise. (A time travel to the beginner's course in analysis). If I don't want to bother Mathematica or Maple, then a standard way to solve the integral is the following. We substitute v by $\sin w$ and get

$$\int (1-v^2)^{-3/2}dv = \int \frac{1}{\cos^2 w}dw = \tan w = \tan\arcsin v = \frac{v}{\sqrt{1-v^2}}\,;$$

that's it.

Solving

$$\frac{v}{\sqrt{1-v^2}} = at$$

for v yields

$$v(t) = \pm\frac{at}{\sqrt{1+a^2t^2}}\ ; \tag{7.22}$$

note that by neglecting the constant of integration we have set $v(0) = 0$. Another integration then leads to

$$x(t) = x_0 \pm \frac{1}{a}\left(\sqrt{1+a^2t^2} - 1\right)\ . \tag{7.23}$$

When we consider the case $x_0 = \pm 1/a$, then

$$x(t) = \pm\frac{1}{a}\ \sqrt{1+a^2t^2}\ . \tag{7.24}$$

Since

$$-t^2 + x(t)^2 = \frac{1}{a^2}\ , \tag{7.25}$$

the world line of the particle

$$t \mapsto \begin{pmatrix} 1 \\ x(t) \\ 0 \\ 0 \end{pmatrix} \tag{7.26}$$

represents an (equilateral) hyperbola. Therefore, constant linear acceleration leads to '*hyperbolic motion*'.

In the following we present an alternative derivation of this result that is better adapted to the four-vector formalism (and more elegant). Linear acceleration means that the motion of the particle is along a straight line (in space); hence, the world line of the particle must be in a two-dimensional plane (which we choose to be the (t, x)-plane). The four-acceleration

$$a^\mu = \begin{pmatrix} a^0 \\ a^1 \\ 0 \\ 0 \end{pmatrix} \tag{7.27}$$

satisfies $a^2 = \eta(a, a) = -(a^0)^2 + (a^1)^2 = \text{const}$, hence we can parametrize a^μ as

$$a^\mu(s) = a\begin{pmatrix} \sinh\alpha(s) \\ \cosh\alpha(s) \\ 0 \\ 0 \end{pmatrix}\ , \tag{7.28}$$

where $\alpha(s)$ is a function of s that is unspecified a priori. Since u^μ must be orthogonal to a^μ, see (7.4), and normalized, $\eta(u,u) = -1$, we automatically obtain

$$u^\mu(s) = \begin{pmatrix} \cosh\alpha(s) \\ \sinh\alpha(s) \\ 0 \\ 0 \end{pmatrix} . \tag{7.29}$$

On the other hand,

$$\frac{d}{ds} u^\mu = a^\mu \, , \tag{7.30}$$

from which we easily infer that $\alpha(s) = as$ (plus a constant which we set to zero). Integrating u^μ we arrive at x^μ,

$$x^\mu(s) = \frac{1}{a} \begin{pmatrix} \sinh as \\ \cosh as \\ 0 \\ 0 \end{pmatrix} , \tag{7.31}$$

where we have again set the constants of integration to zero. This equation describes a hyperbola in Minkowski space.

Exercise. In units where $c \neq 1$ we have

$$t(s) = \frac{c}{a} \sinh \frac{as}{c} \qquad \text{and} \qquad x(s) = \frac{c^2}{a} \cosh\left(\frac{as}{c}\right) .$$

Since $x(0)$ is c^2/a, the distance Δx a particle covers in proper time s is

$$\Delta x = \frac{c^2}{a} \left[\cosh\left(\frac{as}{c}\right) - 1\right] .$$

Suppose that the acceleration a equals $1g$, where g is the acceleration due to gravity. (This is the most pleasant acceleration for spacefarers.) Then

$$\Delta x[\text{in light years}] = 0.97 \left[\cosh\Big(1.03\, s[\text{in years}]\Big) - 1\right] .$$

Compute how long it takes to get to the center of our galaxy (which is about 25000 light years from us). (Be amazed: it's only 10 years 6 months.)

7.3 Circular motion

Consider a particle in uniform circular motion. W.r.t. some observer its world line is given by

$$x^\mu(t) = \begin{pmatrix} t \\ r\cos\omega t \\ r\sin\omega t \\ 0 \end{pmatrix} , \tag{7.32}$$

where r is the radius of the circular orbit and ω the constant angular velocity. Since $|\vec{v}| = \omega r$, the product ωr must be less than 1 to ensure that the circular velocity is less than the speed of light.

The four-velocity is given by

$$u^\mu(t) = \gamma \begin{pmatrix} 1 \\ -r\omega\sin\omega t \\ r\omega\cos\omega t \\ 0 \end{pmatrix} , \tag{7.33}$$

where

$$\gamma = \frac{1}{\sqrt{1-\omega^2 r^2}} = \text{const} .$$

This implies that proper time s is a linear function of t, namely

$$s = \int \gamma^{-1} dt = \gamma^{-1} t . \tag{7.34}$$

Consequently, when we parametrize x^μ and u^μ w.r.t. proper time we get

$$x^\mu(s) = \begin{pmatrix} \gamma s \\ r\cos(\omega\gamma s) \\ r\sin(\omega\gamma s) \\ 0 \end{pmatrix} ,$$

$$u^\mu(s) = \frac{d}{ds} x^\mu(s) = \gamma \begin{pmatrix} 1 \\ -r\omega\sin(\omega\gamma s) \\ r\omega\cos(\omega\gamma s) \\ 0 \end{pmatrix} .$$

The four-acceleration is the derivative of $u^\mu(s)$ w.r.t. s, i.e.,

$$a^\mu(s) = \frac{d}{ds} u^\mu(s) = \gamma^2 \begin{pmatrix} 0 \\ -r\omega^2\cos(\omega\gamma s) \\ -r\omega^2\sin(\omega\gamma s) \\ 0 \end{pmatrix} . \tag{7.35}$$

It is easy to see that the acceleration is constant,

$$a^2 = \left(\gamma^2 r\omega^2\right)^2 . \tag{7.36}$$

Furthermore, the spatial part of the acceleration $\vec{a}$ is orthogonal to $\vec{v}$.

Remark. An alternative derivation of (7.35) uses the general considerations of section 7.1. We have

$$\vec{v} = \begin{pmatrix} -r\omega \sin \omega t \\ r\omega \cos \omega t \\ 0 \end{pmatrix} \qquad \text{and therefore} \qquad \vec{a}_N = \begin{pmatrix} -r\omega^2 \cos \omega t \\ -r\omega^2 \sin \omega t \\ 0 \end{pmatrix} .$$

Inserting this result into (7.9) leads to (7.35). The proper acceleration is

$$\vec{a}_p = \gamma^2 \vec{a}_N$$

and the $\vec{a}$ (which is the spatial part of the four-acceleration) is equal to $\vec{a}_p$. Since the acceleration—either of $\vec{a}_N$, $\vec{a}_p$, $\vec{a}$—is orthogonal to $\vec{v}$ we are in the 'transversal case' of section 7.1.

The magnitude a of the acceleration (which coincides with $|\vec{a}_p|$) is given by

$$a = \gamma^2 r\omega^2 , \tag{7.37}$$

which can be written as

$$a = \frac{r\omega^2}{1 - r^2\omega^2} . \tag{7.38}$$

This is the formula for the *relativistic centripetal acceleration.* In the Newtonian case we simply have

$$a_N = r\omega^2 . \tag{7.39}$$

The relativistic centripetal acceleration is larger and diverges as $r\omega \to 1$.

7.4 The relativistic rocket

The rocket we consider is represented by a world line $x(s)$ in Minkowski spacetime, where we choose to parametrize the world line by proper time s; at lift-off we set $s = 0$. If we denote the rocket's four velocity by $u = \dot{x}(s)$ and by m its mass, its four momentum is given by

$$p = mu . \tag{7.40}$$

Since a rocket obtains thrust by ejecting its propellant, its mass m decreases with time. We are thus careful and write

$$p(s) = m(s)u(s)\ . \tag{7.40'}$$

The four momentum of the propellant (let's say 'gas' for simplicity) that is emitted in a small time interval $[s, s+ds]$ is given by

$$dp_{\text{gas}} = dm_{\text{gas}}u_{\text{gas}}\ , \tag{7.41}$$

where $dm_{\text{gas}} = \dot{m}_{\text{gas}}ds$ is the exhaust mass. (Note that each of the quantities depends on s in general.) Consequently, the four momentum of the gas that has been emitted up to time s is given by

$$p_{\text{gas}} = \int dm_{\text{gas}}u_{\text{gas}} = \int_0^s \dot{m}_{\text{gas}}(s')u_{\text{gas}}(s')ds'\ . \tag{7.42}$$

The overdot denotes differentiation w.r.t. proper time s.

Conservation of momentum means that the total momentum, i.e., the sum of the momentum of the rocket and the gas, remains constant, i.e.,

$$p + p_{\text{gas}} = \text{const}\ ; \tag{7.43}$$

recall that $p = p(s)$ is the momentum of the rocket and p_{gas} the momentum of the gas emitted up to time s. Differentiating (7.43) we obtain

$$\dot{p} + \dot{p}_{\text{gas}} = \dot{m}u + m\dot{u} + \dot{m}_{\text{gas}}u_{\text{gas}} = 0\ . \tag{7.44}$$

When we multiply this equation with u (in the Minkowski sense), then

$$\dot{m}\eta(u,u) + m\eta(u,\dot{u}) + \dot{m}_{\text{gas}}\eta(u,u_{\text{gas}}) = 0\ . \tag{7.45}$$

Since $\eta(u,u) = -1$, we have $\eta(u,\dot{u}) = 0$, see (7.4); moreover, $\eta(u,u_{\text{gas}}) = -\gamma_{\text{gas}}$ with $\gamma_{\text{gas}} = \left(1 - v_{\text{gas}}^2\right)^{-1/2}$, where v_{gas} is the relative velocity between the gas emitted at time s and the rocket at time s. Therefore, equation (7.45) results in

$$-\dot{m} - \dot{m}_{\text{gas}}\gamma_{\text{gas}} = 0\ . \tag{7.46}$$

We can thus replace $\dot{m}_{\text{gas}}$ in (7.44) to obtain

$$\dot{m}u + m\dot{u} - \dot{m}\gamma_{\text{gas}}^{-1}u_{\text{gas}} = 0\ . \tag{7.47}$$

Equivalently, we have

$$\dot{m}u + ma = \dot{m}\gamma_{\mathrm{gas}}^{-1}u_{\mathrm{gas}} , \tag{7.48}$$

where we have written a instead of $\dot{u}$. It is now straightforward to take the Minkowski norms of the l.h. side and the r.h. side,

$$-\dot{m}^2 + m^2a^2 = -\dot{m}^2\gamma_{\mathrm{gas}}^{-2} , \tag{7.49}$$

from which we conclude that

$$m^2a^2 = \dot{m}^2\left[-\gamma_{\mathrm{gas}}^{-2} + 1\right] = \dot{m}^2v_{\mathrm{gas}}^2 . \tag{7.50}$$

Finally, we have arrived at

$$ma = -\dot{m}v_{\mathrm{gas}} , \tag{7.51}$$

where (in slight abuse of notation) a $(= a_p)$ now denotes the magnitude of the (proper) acceleration. (Clearly, the acceleration is antiparallel to the velocity of the emitted gas, hence the minus sign.) Equation (7.51) is the *relativistic rocket equation.*

Exercise. The Newtonian rocket equation takes exactly the same form. But there are obvious differences. Which ones?

As an example, let us consider a very simple rocket: let us assume that the velocity v_{gas} is a constant over time, i.e., that the gas is always emitted at the same speed. From (7.16‴) we have

$$a = \gamma^2\,\frac{dv}{ds} = \gamma^2\,\dot{v} , \tag{7.52}$$

where v denotes the velocity of the rocket w.r.t. some given observer. (Since we have assumed that lift-off is at time $s = 0$, we have $v = 0$ at $s = 0$.) The rocket equation becomes

$$m\gamma^2\dot{v} = -\dot{m}v_{\mathrm{gas}} , \tag{7.53}$$

which we can write as

$$\frac{\dot{v}}{1 - v^2} = -v_{\mathrm{gas}}\,\frac{\dot{m}}{m} \tag{7.54}$$

and thus integrate:

$$\frac{1}{2}\left[\log(1 + v) - \log(1 - v)\right] = -v_{\mathrm{gas}}(\log m - \log m_0) . \tag{7.55}$$

Here, m_0 is an integration constant—it is the mass of the rocket at lift-off. Finally,

$$\sqrt{\frac{1 - v}{1 + v}} = \left(\frac{m_0}{m}\right)^{-v_{\mathrm{gas}}} \tag{7.56}$$

and thus

$$v = \frac{1 - \left(\frac{m}{m_0}\right)^{2v_{\rm gas}}}{1 + \left(\frac{m}{m_0}\right)^{2v_{\rm gas}}} \ . \tag{7.57}$$

We conclude that the rocket's velocity comes arbitrarily close to 1 (i.e., c) provided m/m_0 becomes arbitarily small.

7.5 Four-force

Consider a particle represented by a timelike world line $x^\mu(s)$, where s is proper time. Let p^μ denote the four-momentum of the particle.

Definition 7.2. *The* four-force *acting on the particle is given as the derivative of the four-momentum w.r.t. proper time, i.e.,*

$$F^\mu(s) = \dot{p}^\mu(s) = \frac{d}{ds} p^\mu(s) \ . \tag{7.58}$$

Since

$$p^\mu = \begin{pmatrix} E \\ \vec{p} \end{pmatrix} \tag{7.59}$$

w.r.t. some observer, we have

$$F^\mu = \begin{pmatrix} \frac{dE}{ds} \\ \frac{d\vec{p}}{ds} \end{pmatrix} = \gamma \begin{pmatrix} \frac{dE}{dt} \\ \frac{d\vec{p}}{dt} \end{pmatrix} = \gamma \begin{pmatrix} \frac{dE}{dt} \\ \vec{F} \end{pmatrix} \ , \tag{7.60}$$

where $\vec{F}$ is the (relativistic) three-force. Interestingly enough, the *power* dE/dt appears as the zero component of the four-force.

Written out explicitly, we have

$$\frac{dE}{dt} = \frac{d}{dt}\,(m\gamma) = \frac{dm}{dt}\,\gamma + m\gamma^3 \vec{v}\,\frac{d\vec{v}}{dt} \ , \tag{7.61a}$$

$$\vec{F} = \frac{d}{dt}\,\big(m\gamma\vec{v}\big) = \frac{dm}{dt}\,\gamma\vec{v} + m\left[\gamma^3 \left(\vec{v}\,\frac{d\vec{v}}{dt}\right)\vec{v} + \gamma\,\frac{d\vec{v}}{dt}\right] \ . \tag{7.61b}$$

When $\{\underset{\sim}{t}, \underset{\sim}{\vec{x}}\}$ is an inertial system in which the particle is at rest at some time $\underset{\sim}{t} = \underset{\sim}{t}_r$, then

$$\left.\frac{d\underset{\sim}{E}}{d\underset{\sim}{t}}\right|_{\underset{\sim}{t}=\underset{\sim}{t}_r} = \left.\frac{dm}{d\underset{\sim}{t}}\right|_{\underset{\sim}{t}=\underset{\sim}{t}_r} \tag{7.62}$$

and

$$\underset{\sim}{\vec{F}}\Big|_{\underset{\sim}{t}=\underset{\sim}{t_r}} = m\,\frac{d\underset{\sim}{\vec{v}}}{d\underset{\sim}{t}}\Big|_{\underset{\sim}{t}=\underset{\sim}{t_r}} =: \vec{F}_p\,. \tag{7.63}$$

The quantity $\vec{F}_p$ is the 'proper force', which is experienced by the particle in the momentary rest frame. Obviously,

$$\vec{F}_p = m\vec{a}_p\,. \tag{7.64}$$

Remark. We call a force *heatlike* if it does not change the particle's velocity. Then (7.61) becomes

$$F^\mu = \gamma\,\frac{dm}{dt}\,u^\mu\,, \tag{7.65}$$

where u^μ is the (constant) four-velocity of the particle. In the rest frame,

$$F^\mu = \begin{pmatrix} \frac{dm}{d\underset{\sim}{t}} \\ \vec{o} \end{pmatrix}\,. \tag{7.66}$$

By definition, a particle is not accelerated by a heatlike force, but there is an increase in mass.

We call a force *pure* if it does not change the particle's mass, i.e., if

$$m \equiv \text{const} \tag{7.67}$$

for all times. Then

$$F^\mu\Big|_{\underset{\sim}{t}=\underset{\sim}{t_r}} = \begin{pmatrix} 0 \\ \vec{F}_p \end{pmatrix} \tag{7.68}$$

in the momentary rest frame. Accordingly, the particle's four-velocity and the four-force are orthogonal, i.e.,

$$\eta(F,u) = 0\,. \tag{7.69}$$

This is consistent with our expectations, since for a pure force we have

$$F^\mu = \frac{d}{ds}p^\mu(s) = ma^\mu\,, \tag{7.70}$$

and a^μ is orthogonal to u^μ.

As a direct consequence of $\eta(F,u) = 0$ we have

$$\frac{dE}{dt} = \vec{F}\vec{v}\,. \tag{7.71}$$

Evidently, w.r.t. the momentary rest frame of the particle, the four-velocity of the observer—let's call it w^μ— is given by

$$\gamma \begin{pmatrix} 1 \\ -\vec{v} \end{pmatrix} . \tag{7.72}$$

Accordingly,

$$-\gamma \frac{dE}{dt} = \eta(F, w) = -\gamma \vec{F}_p \vec{v} \,, \tag{7.73}$$

where we have evaluated the Minkowski product once in the momentary rest frame of the particle and once in the observer's inertial frame. We conclude that

$$\frac{dE}{dt} = \vec{F}\vec{v} = \vec{F}_p \vec{v} \,. \tag{7.74}$$

Alternatively, we can write

$$dE = \vec{F}_p d\vec{x} \,, \tag{7.75}$$

which represents the (infinitesimal) work done by the force on the particle.

Remark. Finally, we note that the four-force can be expressed in terms of the proper force as

$$F^\mu = \gamma(\vec{v}\vec{F}_p) \begin{pmatrix} 1 \\ \frac{\vec{v}}{\vec{v}^2} \end{pmatrix} + \begin{pmatrix} 0 \\ \vec{F}_p - (\vec{v}\vec{F}_p)\frac{\vec{v}}{\vec{v}^2} \end{pmatrix} = \gamma \begin{pmatrix} \vec{v}\vec{F}_p \\ \vec{F}_p^{\parallel} \end{pmatrix} + \begin{pmatrix} 0 \\ \vec{F}_p^{\perp} \end{pmatrix} . \tag{7.76}$$

This is in complete analogy with the remark at the end of section 7.1.

Our analysis in this section so far can be summarized as follows: Given the motion of a particle and its associated four-velocity, four-momentum, and four-acceleration we are able to compute (what we call) the four-force (and the proper force) that is connected with the particle's motion. In particular, if the particle's mass m is constant along its world line, then

$$ma^\mu = F^\mu \,, \tag{7.77}$$

which is the analog of Newton's second law.

However, in Newtonian physics, it is straightforward to do the 'opposite'. If there is given a force field $\vec{F}(t,\vec{x})$, then Newton's second law $m\vec{a} = \vec{F}$ is a system of ordinary differential equations (ODEs) for $\vec{x}(t)$, i.e.,

$$\ddot{\vec{x}}(t) = \frac{d^2}{dt^2}\, \vec{x}(t) = \vec{F}\big(t, \vec{x}\big) \,,$$

which can be solved for arbitrary initial data $\vec{x}(0)$, $\vec{v}(0) = \dot{\vec{x}}(0)$, to yield $\vec{x}(t)$ and thus the motion of the particle in the force field.

The relativistic analog (7.77) can *not* be interpreted in the analogous way. Suppose that we are given a force field $F^\mu(x^\sigma)$ on Minkowski space and initial data for the particle, i.e., $x^\mu(0) = \mathring{x}^\mu$ and $u^\mu(0) = \dot{x}^\mu(0) = \mathring{u}^\mu$ (where we continue to use the dot as referring to proper time). Then

$$m\ddot{x}^\mu = F^\mu(x^\sigma)\,, \quad x^\mu(0) = \mathring{x}^\mu\,, \quad \dot{x}^\mu(0) = \mathring{u}^\mu$$

is not a well-posed IVP (initial value problem). The reason is that $u^\mu u_\mu = -1$ is required; hence the l.h.s. is orthogonal to $\mathring{u}^\mu$ at initial time, independently of the choice of $\mathring{u}^\mu$, but the r.h.s., i.e., $F^\mu(x^\sigma)$, cannot be orthogonal to all four-vectors $\mathring{u}^\mu$ simultaneously (unless $F^\mu = 0$).

We conclude that the concept of a four-force field (as a vector field on Minkowski space) does not makes sense. To resolve the problem we must realize that it is impossible to prescribe $F^\mu = F^\mu(x^\sigma)$; instead, we attempt to prescribe F^μ as a function of the spacetime position *and the four-velocity*, i.e., $F^\mu = F^\mu(x^\sigma, u^\lambda)$.

Let us test the simplest ansatz that comes to mind: We assume linearity in u^λ, i.e., we make the ansatz

$$F_\mu(x^\sigma, u^\lambda) = F_{\mu\nu}(x^\sigma)\, u^\nu\,. \tag{7.78}$$

Note that, equivalently, we are able to write $F^\mu(x^\sigma, u^\lambda) = F^\mu{}_\nu(x^\sigma)\, u^\nu$, when we raise indices. Let us suppress the dependence on the spacetime position and simply write $F_\mu = F_{\mu\nu}u^\nu$. The requirement is that F^μ be orthogonal to u^μ (for arbitrary u^μ), i.e., $F_\mu u^\mu = 0$ (for arbitrary u^μ). Let's see whether the ansatz (7.78) can guarantee that. We obtain

$$F_\mu u^\mu = F_{\mu\nu}u^\mu u^\nu = \big(F_{(\mu\nu)} + F_{[\mu\nu]}\big)u^\mu u^\nu = F_{(\mu\nu)}u^\mu u^\nu \overset{!}{=} 0\,,$$

which is zero for all u^μ if and only if $F_{(\mu\nu)} = 0$.

Exercise. Prove that, if

$$F_{(\mu\nu)}u^\mu u^\nu = 0$$

for arbitrary u^μ (where $u^\mu u_\mu = -1$), then $F_{(\mu\nu)} = 0$.

We thus find that if $F_{\mu\nu} = F_{[\mu\nu]}$, i.e., if $F_{\mu\nu}$ is assumed to be antisymmetric, then Newton's second law (7.77) together with the ansatz (7.78) lead to a well-posed IVP,

$$m\ddot{x}^\mu = F^\mu{}_\nu(x^\sigma)\, u^\nu\,, \quad x^\mu(0) = \mathring{x}^\mu\,, \quad \dot{x}^\mu(0) = \mathring{u}^\mu\,.$$

In this context, the prescribed field is not a vector field but an antisymmetric tensor field $F_{\mu\nu}(x^\sigma)$ on Minkowski space. (The force is then a derived quantity, see (7.78).) We will see in the next chapter that these considerations are part of an important theory, namely the theory of electromagnetism. It is remarkable that nature does the same thing as we did with the simplest possible ansatz (7.78).

Now we become overzealous, of course. Having been so successful with the simple ansatz (7.78) to obtain (one aspect of) electromagnetism, it suggests itself to try the next simple one to obtain relativistic gravity. We assume that F^μ does not depend linearly but quadratically on u^μ, i.e., we make the ansatz

$$F_\mu(x^\lambda, u^\rho) = G_{\mu\nu\sigma}(x^\lambda)\, u^\nu u^\sigma \,. \tag{7.79}$$

Since $u^\nu u^\sigma$ is symmetric in ν and σ, it is w.l.o.g. when we assume that $G_{\mu\nu\sigma}$ is symmetric in ν and σ as well. One tensor that is symmetric, we have come to know really well: $\eta_{\nu\sigma}$. Our intention is to build $G_{\mu\nu\sigma}$ from the Minkowski metric and one additional four-vector field that is regarded as the four-gradient of a function $V = V(x^\lambda)$. A first try is to put

$$G_{\mu\nu\sigma} = V_{,\mu}\eta_{\nu\sigma} \,,$$

where we use the comma notation for the derivative, e.g., $V_{,\mu} = \partial_\mu V = \partial V/\partial x^\mu$. However, the required orthogonality $F_\mu u^\mu = 0$ does not hold for this simple ansatz; to ensure $F_\mu u^\mu = 0$ we set

$$G_{\mu\nu\sigma} = V_{,\mu}\eta_{\nu\sigma} - \eta_{\mu(\nu}V_{,\sigma)} \,. \tag{7.80}$$

It is straightforward to check that

$$F_\mu u^\mu = G_{\mu\nu\sigma}u^\mu u^\nu u^\sigma = \big(V_{,\mu}\eta_{\nu\sigma} - \eta_{\mu(\nu}V_{,\sigma)}\big)u^\mu u^\nu u^\sigma = -V_{,\mu}u^\mu + V_{,\sigma}u^\sigma = 0 \,,$$

independently of u^μ, as required.

Let us investigate Newton's second law with (7.79), i.e.,

$$a^\mu = G^\mu{}_{\nu\sigma}(x^\lambda)\, u^\nu u^\sigma \,; \tag{7.81}$$

for simplicity we have set the particle's mass to 1. The r.h.s. is

$$G^\mu{}_{\nu\sigma}\, u^\nu u^\sigma = \big(V^{,\mu}\eta_{\nu\sigma} - \delta^\mu{}_{(\nu}V_{,\sigma)}\big)u^\nu u^\sigma = -V^{,\mu} - u^\sigma V_{,\sigma}\, u^\mu \,; \tag{7.82}$$

its temporal component (zero component) is

$$G^0{}_{\nu\sigma}\, u^\nu u^\sigma = -\eta^{00}V_{,t} - \gamma^2\big(V_{,t} + \vec{v}\vec{\nabla}V\big) \,, \tag{7.82^0}$$

since $u^\mu = \gamma(1, \vec{v})^{\mathrm{T}}$ and thus $u^\sigma V_{,\sigma} = \gamma V_{,t} + \gamma v^i V_{,i}$. The spatial components of (7.82) are

$$G^i{}_{\nu\sigma}\, u^\nu u^\sigma = -\delta^{ij} V_{,j} - \gamma^2 \big(V_{,t} + \vec{v}\vec{\nabla}V\big) v^i \,. \qquad (7.82^i)$$

The temporal and spatial components of the l.h.s. of (7.81) are given by

$$a^\mu = \gamma^4 \Big(\vec{v}\,\frac{d\vec{v}}{dt}\Big) \begin{pmatrix} 1 \\ \vec{v} \end{pmatrix} + \gamma^2 \begin{pmatrix} 0 \\ \frac{d\vec{v}}{dt} \end{pmatrix} ,$$

see (7.7).

Let us consider the Newtonian limit, i.e., the limit of small velocities. Then equation (7.82^0) becomes $\partial_t V - \partial_t V - \vec{v}\vec{\nabla}V = -\vec{v}\vec{\nabla}V$ and (7.82^i) becomes $-\vec{\nabla}V$. Therefore, from (7.81), in the Newtonian limit, we obtain

$$\vec{v}\frac{d\vec{v}}{dt} = -\vec{v}\vec{\nabla}V \,, \qquad \frac{d\vec{v}}{dt} = -\vec{\nabla}V \,.$$

If $V = V(\vec{x})$, then $\vec{v}\vec{\nabla}V\big(\vec{x}(t)\big) = \partial_t V\big(\vec{x}(t)\big)$; accordingly, the equations read

$$\frac{d}{dt}\Big(\frac{m\vec{v}^2}{2} + V(\vec{x})\Big) = 0 \,, \qquad \vec{a} = \vec{F} \,,$$

where $\vec{F} = -\vec{\nabla}V$ is the negative gradient of the potential, i.e., the force. (Since we did not explicitly introduce a coupling constant in (7.79), V should be regarded as mass times potential, which makes $-\vec{\nabla}V$ the force.) We recover the fundamental aspects of Newtonian gravity in the limit of small velocities: The first equation is conservation of energy, which is the sum of kinetic plus potential energy; the second equation is the law of motion.

The theory based on (7.79) and (7.80), which leads to (7.81), is a relativistic scalar theory of gravity. Unfortunately, despite its appeal, it is false (i.e., in contradiction with observations).

CHAPTER 8

MAXWELL

8.1 The electromagnetic field tensor

Suppose we have an (exterior) electromagnetic field represented by an electric field $\vec{E}(t,\vec{x})$ and a magnetic field $\vec{B}(t,\vec{x})$, where $\{t,\vec{x}\}$ are the coordinates of an inertial frame. A (test) particle that is moving in this electromagnetic field is subject to the *Lorentz force*, i.e.,

$$\frac{d\vec{p}}{dt} = e\left(\vec{E} + \vec{v} \times \vec{B}\right), \tag{8.1}$$

where $\vec{p}$ and $\vec{v}$ are the particle's three-momentum and three-velocity, respectively, and e denotes the particle's charge. This equation of motion is known in the case of a slowly moving particle; it is unclear a priori how to extend this equation to any $\vec{v}$ with $|\vec{v}| < 1$.

In a momentary rest frame, which is represented by $\{\utilde{t},\utilde{\vec{x}}\}$, the Lorentz force is a purely electric force,

$$\frac{d\utilde{\vec{p}}}{d\utilde{t}} = e\utilde{\vec{E}}\,, \tag{8.2}$$

since $\utilde{\vec{v}} = \vec{o}$ (at the considered time $\utilde{t} = \utilde{t}_r$ where the particle is instantaneously at rest). Using (7.68) we associate (8.2) with a four-force[1]

$$\mathcal{F}^\mu = \begin{pmatrix} 0 \\ \vec{F}_p \end{pmatrix} = \begin{pmatrix} 0 \\ e\vec{\underset{\sim}{E}} \end{pmatrix} , \tag{8.3}$$

where $\underset{\sim}{t} = \underset{\sim}{t}_r$ is understood. Equation (7.76) then yields

$$\mathcal{F}^\mu = e\gamma \begin{pmatrix} \vec{v}\vec{\underset{\sim}{E}} \\ \vec{\underset{\sim}{E}}^{\|} \end{pmatrix} + e \begin{pmatrix} 0 \\ \vec{\underset{\sim}{E}}^{\perp} \end{pmatrix} \tag{8.4}$$

for the four-force w.r.t. an arbitrary observer. However, from (8.1) we obtain

$$\mathcal{F}^\mu = e\gamma \begin{pmatrix} \vec{E}\vec{v} \\ \vec{E}^{\|} + \vec{E}^{\perp} + \vec{v} \times \vec{B} \end{pmatrix} , \tag{8.5}$$

where we have used (7.60) and (7.74). Comparing (8.4) and (8.5) we obtain

$$\vec{\underset{\sim}{E}}^{\|} = \vec{E}^{\|} , \qquad \vec{\underset{\sim}{E}}^{\perp} = \gamma\Big(\vec{E}^{\perp} + \vec{v} \times \vec{B}\Big) . \tag{8.6}$$

In particular, we note that $\vec{E}$ and $\vec{B}$ mix: There is no observer-independent distinction between the electric and the magnetic field. For instance, while one observer might see only an electric field (and no magnetic field), another observer sees both an electric and a magnetic field (according to the above transformation).

Strictly speaking. so far we have only considered the limit of small velocities, since we hesitated to automatically assume (8.1) for large velocities. However, it is a fact (which is amply supported by experiments) that (8.1) is indeed the correct representation of the Lorentz force for all velocities. In the following we present an additional plausibility argument that supplements the experimental evidence.

We have argued that the decomposition of the electromagnetic field into an electric part $\vec{E}$ and a magnetic part $\vec{B}$ is not observer-independent. So, our aim is to collect these fields into one quantity that represents the *electromagnetic field* as a whole. What could this quantity be? Obviously, the electromagnetic field cannot be a four-vector, which is simply because we have six independent quantities (namely, the three components of $\vec{E}$ and the three components of $\vec{B}$). It is suggestive to assume that the electromagnetic field is a tensor

$$F_{\mu\nu} . \tag{8.7}$$

[1] To avoid ambiguities we denote the four-force by a calligraphic letter in this chapter. The normal letter F is reserved for the electromagnetic field (8.7).

Clearly, such a tensor has 16 entries; hence, many of the components must be redundant to encode the electromagnetic field.

Given $F_{\mu\nu}$, there is not much choice to derive a four-force from it other than via the simple rule

$$\mathcal{F}^\mu = eF^\mu{}_\nu u^\nu \,, \tag{8.8}$$

where u^μ is the four-velocity of the particle. If we require the *Lorentz four-force* (8.8) to be a pure force, then

$$\mathcal{F}^\mu u_\mu = eF_{\mu\nu} u^\mu u^\nu = 0 \,, \tag{8.9}$$

see (7.69). Consequently, the field tensor $F_{\mu\nu}$ must be *antisymmetric*, i.e., $F_{\mu\nu} = -F_{\nu\mu}$, or

$$F_{\mu\nu} = F_{[\mu\nu]} \,, \tag{8.10}$$

cf. the considerations of section 7.5.

Since $F_{\mu\nu}$ is antisymmetric, there are six independent components. Let us choose an arbitrary observer; the temporal components are denoted by 0, the spatial components are denoted by Latin indices. Using the antisymmetry we see that, on the one hand, we have the three components F_{i0} (or, equivalently, their negative counterparts F_{0i}); we set $F_{i0} = \delta_{ij}\mathcal{E}^j$. On the other hand, we have the three components encoded by F_{ij}, which we can write as $F_{ij} = \epsilon_{ijk}\mathcal{B}^k$. W.r.t. the chosen observer we further get

$$u^\mu = \gamma \begin{pmatrix} 1 \\ v^i \end{pmatrix} \,, \tag{8.11}$$

hence

$$\mathcal{F}_\mu = eF_{\mu\nu}u^\nu = e\gamma\Big(F_{0i}v^i, F_{i0} + F_{ij}v^j\Big) = e\gamma\Big(-\delta_{ij}v^i\mathcal{E}^j, \delta_{ij}\mathcal{E}^j + \epsilon_{ijk}v^j\mathcal{B}^k\Big) \tag{8.12}$$

and the Lorentz force reads

$$\mathcal{F}^\mu = e\gamma \begin{pmatrix} \delta_{ij}v^i\mathcal{E}^j \\ \mathcal{E}^i + \epsilon^i{}_{jk}v^j\mathcal{B}^k \end{pmatrix} = e\gamma \begin{pmatrix} \vec{\mathcal{E}}\vec{v} \\ \vec{\mathcal{E}} + \vec{v} \times \vec{\mathcal{B}} \end{pmatrix} \,. \tag{8.13}$$

As a consequence we obtain

$$\vec{F} = \frac{d\vec{p}}{dt} = e\big(\vec{\mathcal{E}} + \vec{v} \times \vec{\mathcal{B}}\big) \tag{8.14}$$

for the Lorentz (three-)force, cf. (7.60). Comparison with (8.1) suggests

$$\vec{\mathcal{E}} = \vec{E} \,, \qquad \vec{\mathcal{B}} = \vec{B} \,. \tag{8.15}$$

We have thus found the four-tensor representation of the electromagnetic field and the (observer-dependent) decomposition into electric part $\vec{E}$ and magnetic part $\vec{B}$.

Let us summarize. The electromagnetic field is described by an antisymmetric tensor field $F_{\mu\nu}$ that encodes both the electric and magnetic degrees of freedom. This electromagnetic field tensor is observer-independent.[2] When an observer is chosen, the electromagnetic field tensor is represented by an antisymmetric matrix, where $F_{i0} = \delta_{ij} E^j$ and $F_{ij} = \epsilon_{ijk} B^k$, i.e.,

$$F_{\mu\nu} = \begin{pmatrix} 0 & -E^1 & -E^2 & -E^3 \\ E^1 & 0 & B^3 & -B^2 \\ E^2 & -B^3 & 0 & B^1 \\ E^3 & B^2 & -B^1 & 0 \end{pmatrix} \tag{8.16}$$

w.r.t. the chosen observer. The Lorentz force acting on a particle with four-velocity u^μ is

$$\mathcal{F}^\mu = eF^\mu{}_\nu u^\nu \,, \tag{8.17}$$

where e is the charge of the particle. In the chosen coordinates we have

$$\mathcal{F}^\mu = \gamma \begin{pmatrix} e\vec{E}\vec{v} \\ e\big(\vec{E} + \vec{v} \times \vec{B}\big) \end{pmatrix} \,, \tag{8.18}$$

so that the three-force takes its conventional form

$$\vec{F} = e\left(\vec{E} + \vec{v} \times \vec{B}\right) . \tag{8.19}$$

Let us conclude this section by defining another antisymmetric tensor that is closely connected with $F_{\mu\nu}$. The *dual of the field tensor* is defined as

$$*F_{\gamma\delta} = \tfrac{1}{2}\, \epsilon_{\alpha\beta\gamma\delta} F^{\alpha\beta} \tag{8.20}$$

where $F^{\alpha\beta} = \eta^{\alpha\mu}\eta^{\beta\nu} F_{\mu\nu}$ as usual. (In the theory of (differential) n-forms, $*$ is known as the Hodge star operator; $*F$ is the Hodge dual of F.) The tensor $*F_{\mu\nu}$ is antisymmetric.

Remark. Viewed as a map on antisymmetric tensors $T_{\mu\nu}$, duality is an *anti-involution*, i.e.,

$$** = -\mathrm{id}\,. \tag{8.21}$$

[2]The indices of $F_{\mu\nu}$ are abstract indices.

To show this we use the definition (8.20).

$$\begin{aligned} **F_{\alpha\beta} &= \tfrac{1}{2}\epsilon_{\alpha\beta\gamma\delta} *F^{\gamma\delta} = \tfrac{1}{4}\epsilon_{\alpha\beta\gamma\delta}\epsilon^{\gamma\delta\varepsilon\zeta}F_{\varepsilon\zeta} = \tfrac{1}{4}\epsilon_{\gamma\delta\alpha\beta}\epsilon^{\gamma\delta\varepsilon\zeta}F_{\varepsilon\zeta} \\ &= \tfrac{1}{4}(-4)\delta_\alpha^{[\varepsilon}\delta_\beta^{\zeta]}F_{\varepsilon\zeta} = -\delta_\alpha^\varepsilon\delta_\beta^\zeta F_{\varepsilon\zeta} = -F_{\alpha\beta}\,. \end{aligned}$$

Here we have used that

$$\epsilon_{\gamma\delta\alpha\beta}\epsilon^{\gamma\delta\varepsilon\zeta} = -4\delta^\varepsilon_{[\alpha}\delta^\zeta_{\beta]} = -4\delta_\alpha^{[\varepsilon}\delta_\beta^{\zeta]} = -4\delta_{[\alpha}^{[\varepsilon}\delta_{\beta]}^{\zeta]}\,,$$

which can be proved by elementary considerations about permutations, as done in (B.25).

Making use of either the definition (8.20) between $*F_{\mu\nu}$ and $F_{\mu\nu}$ we find that $*F_{\mu\nu}$ takes the form

$$*F_{\mu\nu} = \begin{pmatrix} 0 & B^1 & B^2 & B^3 \\ -B^1 & 0 & E^3 & -E^2 \\ -B^2 & -E^3 & 0 & E^1 \\ -B^3 & E^2 & -E^1 & 0 \end{pmatrix} \tag{8.22}$$

w.r.t. the chosen frame of reference. Note that (8.22) arises from (8.16) by replacing

$$B^i \mapsto E^i\,, \qquad E^i \mapsto -B^i\,.$$

8.2 Transformations of $\vec{E}$ and $\vec{B}$

Suppose we are given an electromagnetic field tensor $F_{\mu\nu}$ (where we reemphasize that this is an abstract object that exists independently of any choice of coordinates). W.r.t. a chosen coordinate frame X (= observer X), the tensor components take the form (8.16), where $\vec{E}$ and $\vec{B}$ are the electric and magnetic 3-field as seen by X. W.r.t. a different observer X', the field tensor is represented by a different matrix, whose entries $\vec{E}'$ and $\vec{B}'$ are the electric and magnetic 3-field as seen by X'.

Let L denote the Lorentz transformation connecting the inertial frames of reference X and X'; e.g., L is a boost (2.31). Then[3]

$$\begin{pmatrix} 0 & -E'^1 & -E'^2 & -E'^3 \\ E'^1 & 0 & B'^3 & -B'^2 \\ E'^2 & -B'^3 & 0 & B'^1 \\ E'^3 & B'^2 & -B'^1 & 0 \end{pmatrix} = (L^{-1})^{\mathrm{T}} \begin{pmatrix} 0 & -E^1 & -E^2 & -E^3 \\ E^1 & 0 & B^3 & -B^2 \\ E^2 & -B^3 & 0 & B^1 \\ E^3 & B^2 & -B^1 & 0 \end{pmatrix} L^{-1} \,.$$

We thereby obtain the transformation of the electric and the magnetic fields under a change of observer.

However, in the spirit of these lectures notes, we prefer to avoid computations of this kind; we favor coordinate-independent considerations.

Consider an arbitary inertial observer X whose four-velocity is u^μ. The *electric four-field* seen by the observer X is given as

$$E^\mu = F^\mu{}_\nu \, u^\nu \,, \tag{8.23a}$$

which takes the form

$$E^\mu = \begin{pmatrix} 0 \\ \vec{E} \end{pmatrix} \tag{8.23b}$$

in the observer's coordinates. Analogously, the *magnetic four-field* seen by X is given as

$$B^\mu = -\big(*F^\mu{}_\nu \, u^\nu \big) \,, \tag{8.24a}$$

which takes the form

$$B^\mu = \begin{pmatrix} 0 \\ \vec{B} \end{pmatrix} \tag{8.24b}$$

w.r.t. X. Although evident from the definitions (8.23a) and (8.24a), we emphatically stress that the electric/magnetic four-fields—the four-fields themselves, not merely their coordinate representations—depend on the choice of observer.[4]

Remark. Consider a particle with charge e and four-velocity u^μ. The electric four-field sensed by the particle is $E^\mu = F^\mu{}_\nu \, u^\nu$. (Strictly speaking, this is the four-field seen by the co-moving observer X, whose four-velocity coincides with that of the particle.) Therefore, the Lorentz force acting on the particle is simply $\mathcal{F}^\mu = eE^\mu$. Hence, from the particle's point of view, the Lorentz force is a purely electric force.

[3] We refer to (A.19) and (A.22), where we note that the transformation of tensors is a straightforward generalization; e.g., $\hat{T}_{ij} = (A^{-1})^k{}_i (A^{-1})^l{}_j \check{T}_{kl}$.

[4] If there existed four-vector fields E^μ and B^μ representing the electric and magnetic field in an observer-independent fashion, we would automatically have an observer-independent decomposition of the electromagnetic field into an electric and a magnetic part. But this contradicts the mixing of the electric and magnetic field, see section 8.1.

Remark for experts. Let E^μ and B^μ be the electric four-field and magnetic four-field associated with the four-velocity u^μ. It is not difficult to show that

$$F_{\mu\nu} = \quad u_\mu E_\nu - u_\nu E_\mu + \epsilon_{\mu\nu\sigma\tau} u^\sigma B^\tau \tag{8.25a}$$

$$*F_{\mu\nu} = -u_\mu B_\nu + u_\nu B_\mu + \epsilon_{\mu\nu\sigma\tau} u^\sigma E^\tau \; . \tag{8.25b}$$

Equipped with the definitions (8.23a) and (8.24a) it becomes straightforward to investigate the transformation of the electric and magnetic field. Obviously, the magnitude of the electric field $|\vec{E}|$ (as it is experienced by X) has the representation[5]

$$|\vec{E}|^2 = E^\mu E_\mu \; . \tag{8.26a}$$

Analogously, the magnitude of the magnetic field $|\vec{B}|$ is given by

$$|\vec{B}|^2 = B^\mu B_\mu \; . \tag{8.26b}$$

Furthermore, the angle between the fields $\vec{E}$ and $\vec{B}$ is obtained from

$$\vec{E}\vec{B} = E_\mu B^\mu \; . \tag{8.26c}$$

The formulas (8.26) extract the physically relevant information about the electric field and the magnetic field from a given electromagnetic field tensor $F_{\mu\nu}$.

Now suppose there are two observers, X and X', with four-velocities u and u', respectively. In the coordinates used by X we have

$$u^\mu = \begin{pmatrix} 1 \\ \vec{o} \end{pmatrix} \; , \qquad u'^\mu = \gamma \begin{pmatrix} 1 \\ \vec{v} \end{pmatrix} \; , \tag{8.27}$$

and

$$F^\mu{}_\nu = \begin{pmatrix} 0 & E^1 & E^2 & E^3 \\ E^1 & 0 & B^3 & -B^2 \\ E^2 & -B^3 & 0 & B^1 \\ E^3 & B^2 & -B^1 & 0 \end{pmatrix} \; , \tag{8.28}$$

where $\vec{E}$ and $\vec{B}$ are the electric and the magnetic field as seen by X. (Compare the index structure of (8.16) and (8.28).) Let us denote by $\vec{E}'$ and $\vec{B}'$ the fields as seen by X'. The associated four-fields are E'^μ and B'^μ, which are given by

$$E'^\mu = F^\mu{}_\nu u'^\nu = \gamma \begin{pmatrix} \vec{v}\vec{E} \\ \vec{E} + \vec{v} \times \vec{B} \end{pmatrix} = \gamma \begin{pmatrix} \vec{v}\vec{E}^\parallel \\ \vec{E}^\parallel + \vec{E}^\perp + \vec{v} \times \vec{B} \end{pmatrix} \tag{8.29a}$$

[5]These considerations are in complete the analogy with (7.13) and (7.14). A comment on the notation, though: Where we wrote $\eta(v, v)$ earlier, we write $v^\mu v_\mu$ here.

and[6]

$$B'^{\mu} = -\left(*F^{\mu}{}_{\nu}\, u'^{\nu} \right) = \gamma \begin{pmatrix} \vec{v}\vec{B} \\ \vec{B} - \vec{v} \times \vec{E} \end{pmatrix} = \gamma \begin{pmatrix} \vec{v}\vec{B}^{\parallel} \\ \vec{B}^{\parallel} + \vec{B}^{\perp} - \vec{v} \times \vec{E} \end{pmatrix} . \tag{8.29b}$$

Equation (8.26) then leads to

$$\begin{aligned} |\vec{E}'|^2 = E'^{\mu} E'_{\mu} &= \gamma^2 \left[-\vec{v}^2 (\vec{E}^{\parallel})^2 + (\vec{E}^{\parallel})^2 + (\vec{E}^{\perp} + \vec{v} \times \vec{B})^2 \right] \\ &= (\vec{E}^{\parallel})^2 + \gamma^2 (\vec{E}^{\perp} + \vec{v} \times \vec{B})^2 \end{aligned} \tag{8.30a}$$

and, in complete analogy,

$$\begin{aligned} |\vec{B}'|^2 = B'^{\mu} B'_{\mu} &= \gamma^2 \left[-\vec{v}^2 (\vec{B}^{\parallel})^2 + (\vec{B}^{\parallel})^2 + (\vec{B}^{\perp} - \vec{v} \times \vec{E})^2 \right] \\ &= (\vec{B}^{\parallel})^2 + \gamma^2 (\vec{B}^{\perp} - \vec{v} \times \vec{E})^2 . \end{aligned} \tag{8.30b}$$

In addition we have

$$\vec{E}'\vec{B}' = E'_{\mu} B'^{\mu} = \vec{E}\vec{B} . \tag{8.30c}$$

Hence, summarizing,

$$\begin{aligned} |\vec{E}'^{\parallel}|^2 + |\vec{E}'^{\perp}|^2 &= |\vec{E}^{\parallel}|^2 + \gamma^2 (\vec{E}^{\perp} + \vec{v} \times \vec{B})^2 , \\ |\vec{B}'^{\parallel}|^2 + |\vec{B}'^{\perp}|^2 &= |\vec{B}^{\parallel}|^2 + \gamma^2 (\vec{B}^{\perp} - \vec{v} \times \vec{E})^2 , \\ |\vec{E}'^{\parallel}|\,|\vec{B}'^{\parallel}| + |\vec{E}'^{\perp}|\,|\vec{B}'^{\perp}| &= |\vec{E}^{\parallel}|\,|\vec{B}^{\parallel}| + |\vec{E}^{\perp}|\,|\vec{B}^{\perp}| . \end{aligned}$$

From these formulas we infer that

$$\vec{E}'^{\parallel} = \vec{E}^{\parallel} \qquad \vec{E}'^{\perp} = \gamma \Big(\vec{E}^{\perp} + \vec{v} \times \vec{B} \Big) \tag{8.31a}$$

$$\vec{B}'^{\parallel} = \vec{B}^{\parallel} \qquad \vec{B}'^{\perp} = \gamma \Big(\vec{B}^{\perp} - \vec{v} \times \vec{E} \Big) , \tag{8.31b}$$

when the spatial axes of the observers are properly aligned. Equations (8.31) are the *transformation rules for the electric field and the magnetic field under a change of inertial frame of reference*. While X sees an electric field $\vec{E}$ and a magnetic field $\vec{B}$, the observer X' sees $\vec{E}'$ and $\vec{B}'$.

[6]The 'strange' form of E'^{μ} and B'^{μ} is due to the fact that we use the coordinates of the observer X to write down the fields experienced by X'. Clearly, in the own coordinates of the observer X', $E'^{\mu} = (0, \vec{E}')^{\mathrm{T}}$ and analogously for B'^{μ}.

We conclude this section by discussing the **invariants of the electromagnetic field $\boldsymbol{F_{\mu\nu}}$**. As we have already seen from (8.30c), the scalar product

$$\vec{E}\vec{B}$$

is an invariant—it is the same for all observers. A straightforward computation using (8.16) and (8.22) yields

$$\vec{E}\vec{B} = \tfrac{1}{4}\, F_{\mu\nu} *F^{\mu\nu} \,. \tag{8.32}$$

Since this is a manifestly coordinate-independent expression, invariance is a matter of course.

The squared magnitudes of the electric and magnetic field are not invariant under a change of observer, see (8.30). In terms of the electromagnetic field tensor we have

$$|\vec{E}|^2 = \eta^{\alpha\beta} F_{\alpha\gamma} F_{\beta\delta}\, u^\gamma u^\delta \,, \qquad |\vec{B}|^2 = \eta^{\alpha\beta} *F_{\alpha\gamma} *F_{\beta\delta}\, u^\gamma u^\delta \,.$$

However, the difference

$$\vec{E}^2 - \vec{B}^2$$

is an invariant quanitity, i.e., this quantity remains unchanged under a change of observer,

$$\vec{E}'^{\,2} - \vec{B}'^{\,2} = \vec{E}^2 - \vec{B}^2 \,.$$

To prove this claim, we can either perform a straightforward computation based on (8.31), or we note that

$$\vec{E}^2 - \vec{B}^2 = -\tfrac{1}{2}\, F_{\mu\nu} F^{\mu\nu} = \tfrac{1}{2} *F_{\mu\nu} *F^{\mu\nu} \,. \tag{8.33}$$

Since $\vec{E}^2 - \vec{B}^2$ has a coordinate-independent representation, it is automatically an invariant.

Remark. The invariants $\vec{E}\vec{B}$ and $\vec{E}^2 - \vec{B}^2$ are the only invariants of the electromagnetic field tensor. To show this we invoke a rather general argument. Let A_{jk} be an antisymmetric matrix. Antisymmetric matrices have imaginary eigenvalues. Moreover, eigenvalues come in complex conjugate pairs, i.e., if $\lambda \in \mathbb{C}$ is an eigenvalue, so is $\bar{\lambda}$. Therefore, if our space is $2n$-dimensional, the eigenvalues of A_{jk} must be $ia_1, -ia_1, ia_2, -ia_2, \ldots, ia_n, -ia_n$ for some $a_k \in \mathbb{R}$, $k = 1, \ldots, n$. The invariants of A_{jk} must be combinations of the eigenvalues; accordingly, A_{jk} can possess merely n independent invariants: a_i, $i = 1, \ldots, n$. Therefore, in our particular case, there can be up to two independent invariants (which we have already found). In fact, one easily finds that the characteristic polynomial of $F^{\mu}{}_{\nu}$ is given by

$$\lambda^4 - \big(\vec{E}^2 - \vec{B}^2\big)\lambda^2 + \big(\vec{E}\vec{B}\big) = 0\,,$$

which explicitly contains the two invariants.[7]

Remark for experts. Computations become simpler[8] in the language of differential forms. We refrain from discussing differential forms in these lecture notes; however, here's a teaser. Let E and B denote the 1-forms associated with the electric and magnetic four-field; the components are E_μ and B_μ, respectively. Then equation (8.25a) becomes

$$F = u \wedge E + *(u \wedge B)\,,$$

which implies

$$*F = *(u \wedge E) + *\,*\,(u \wedge B) = -u \wedge B + *(u \wedge E)\,,$$

i.e., (8.25b). Let A and B be arbitrary 2-forms; the definition of the Hodge star, $A \wedge (*B) = \eta(A,B)\mathrm{vol}$, leads to $A \wedge (*B) = (*A) \wedge B$. In particular, since $*\,* = -\mathrm{id}$, we find $(*A) \wedge (*A) = -A \wedge A$ for arbitrary A. Now, on the one hand,

$$F \wedge (*F) = \eta(F,F)\mathrm{vol} = \tfrac{1}{2}\, F_{\mu\nu}F^{\mu\nu}\,\mathrm{vol}\,;$$

on the other hand,

$$\begin{aligned}
F \wedge (*F) &= \Big(u \wedge E + *(u \wedge B)\Big) \wedge \Big(-u \wedge B + *(u \wedge E)\Big) \\
&= (u \wedge E) \wedge *(u \wedge E) - *(u \wedge B) \wedge (u \wedge B) \\
&= \eta(u \wedge E, u \wedge E)\,\mathrm{vol} - \eta(u \wedge B, u \wedge B)\,\mathrm{vol} \\
&= \left(\begin{vmatrix} \eta(u,u) & \eta(u,E) \\ \eta(E,u) & \eta(E,E) \end{vmatrix} - \begin{vmatrix} \eta(u,u) & \eta(u,B) \\ \eta(B,u) & \eta(B,B) \end{vmatrix} \right) \mathrm{vol} \\
&= \big[-\eta(E,E) + \eta(B,B)\big]\,\mathrm{vol}\,.
\end{aligned}$$

This yields (8.33) and thus invariance of $\vec{E}^2 - \vec{B}^2$.

[7] There is a small subtlety here that involves the index structure of the tensor. We refrain from giving any details, since the general picture remains unaffected.

[8] "Simpler".

Equation (8.32) is obtained by considering $F \wedge F$. On the one hand,

$$\begin{aligned} F \wedge F &= \Big(u \wedge E + *(u \wedge B)\Big) \wedge \Big(u \wedge E + *(u \wedge B)\Big) \\ &= (u \wedge E) \wedge *(u \wedge B) + *(u \wedge B) \wedge (u \wedge E) \\ &= 2\,\eta(u \wedge E, u \wedge B)\,\mathrm{vol} = 2\left(\begin{vmatrix} \eta(u,u) & \eta(u,E) \\ \eta(B,u) & \eta(E,B) \end{vmatrix}\right)\mathrm{vol} \\ &= -2\,\eta(E,B)\,\mathrm{vol}\,. \end{aligned}$$

On the other hand, $F \wedge F = -\eta(F, *F)\,\mathrm{vol}$, and $\eta(F, *F) = (1/2) F_{\mu\nu} *F^{\mu\nu}$; hence equation (8.32) follows.

8.3 The field of a uniformly moving charge

Consider a uniformly moving point charge. It is characterized by a straight world line with a four-velocity vector that we denote by w^μ; w.l.o.g. we assume that the world line passes through the origin. What is the electromagnetic field generated by this point charge?

Let's pretend that we haven't ever heard about a Coulomb field. But still we would like to know the electromagnetic field generated by the point charge. Is this possible? Yes. We can derive the electromagnetic field of a uniformly moving charge from basic mathematical and geometric considerations.

The field tensor of a uniformly moving charge

The electromagnetic field is represented by an antisymmetric tensor field $F_{\mu\nu}$; at event x^μ, the field is $F_{\mu\nu}(x^\sigma)$. This tensor must be built from coordinate-independent entities; however, there are but three available coordinate-independent entities: The event x^μ itself, the four-velocity w^μ of the point charge, and the distance r between the event and the world line of the charge.

By the **distance** $\boldsymbol{r}$ between an event x^μ and a timelike (straight) line

$$\{z^\mu + s w^\mu \,|\, w^\mu w_\mu = -1, s \in \mathbb{R}\}$$

we mean the normal distance, i.e.,

$$r^2 = \big(P_w(x-z)\big)^2 = \eta\big(P_w(x-z), P_w(x-z)\big) = \big(P_w x^\mu - P_w z^\mu\big)\big(P_w x_\mu - P_w z_\mu\big)\,,$$

where $P_w v^\mu = v^\mu + \eta(v,w) w^\mu$ is the spatial projection of an arbitrary four-vector v^μ onto the orthogonal complement of w^μ, see (5.14). We thus obtain

$$r^2 = \big(P_w(x-z)\big)^2 = \eta(x-z,x-z) + \eta(w,x-z)^2\,. \tag{8.34}$$

The notion of distance in Minkowski space is analogous to the notion of distance in a Euclidean space. However, note that the distance between an event and a timelike (straight) line is the maximal distance; this is in contrast to Euclidean geometry, where the normal distance is the minimal distance.

Remark. The reader is encouraged to go back to section 5.3 and reinvestigate the considerations on proper length and the Lorentz contraction by making use of the concept of distance.

Let us return to the problem at hand. The assumption that the world line of the charge passes through the origin amounts to setting $z^\mu = 0$ in (8.34); hence the distance between the event x^μ and the particle's world line is

$$r^2 = (P_w x)^2 = \eta(x,x) + \eta(w,x)^2\,. \tag{8.35}$$

The electromagnetic field tensor must be built from x^μ, w^μ, and r. There are only two possible ways: Either

$$F_{\mu\nu} = \frac{1}{f(r)}\, w_{[\mu} x_{\nu]}\,, \tag{8.36$_E$}$$

or

$$\bar{F}_{\mu\nu} = \frac{1}{f(r)}\,\frac{1}{2}\,\epsilon_{\mu\nu\sigma\tau}\, w^{[\sigma} x^{\tau]}\,, \tag{8.36$_B$}$$

where $f(r)$ is some function of r.

Let us deonote the rest frame of the charge by X' and the temporal and spatial coordinates of this comoving observer by t' and $\vec{x}'$, respectively. Let us compute $w_{[\mu} x_{\nu]}$ in these coordinates. Since, w.r.t. X', the four-velocity w^μ of the particle is

$$w^\mu = \begin{pmatrix} 1 \\ \vec{o} \end{pmatrix}\,,$$

and the event x^μ has coordinates

$$x^\mu = \begin{pmatrix} t' \\ \vec{x}' \end{pmatrix}\,,$$

it is easy to see that $w_{[\mu}x_{\nu]}$ is represented by a matrix of the form

$$w_{[\mu}x_{\nu]} = \frac{1}{2}\begin{pmatrix} 0 & -x'^1 & -x'^2 & -x'^3 \\ x'^1 & 0 & 0 & 0 \\ x'^2 & 0 & 0 & 0 \\ x'^3 & 0 & 0 & 0 \end{pmatrix}$$

in the coordinates of X'. (Recall that $w_\mu = \eta_{\mu\nu}w^\nu$; hence $w_0 = -1$ whenever $w^0 = 1$.)

Therefore, with (8.36_E),

$$F_{\mu\nu} = \frac{1}{2f(r)}\begin{pmatrix} 0 & -x'^1 & -x'^2 & -x'^3 \\ x'^1 & 0 & 0 & 0 \\ x'^2 & 0 & 0 & 0 \\ x'^3 & 0 & 0 & 0 \end{pmatrix}$$

in the coordinates of X', so that (8.16) leads to

$$\vec{E}' = \vec{E}'(t', \vec{x}') = \frac{\vec{x}'}{2f(r)}\,, \qquad \vec{B}' = \vec{B}'(t', \vec{x}') = \vec{o}\,. \tag{8.37}$$

The distance r is simply given by $r = |\vec{x}'|$ is these coordinates. (Note that $P_w x^\mu = (0, \vec{x}')^{\mathrm{T}}$.)

The function $f(r)$ can be determined (at least heuristically) by using geometric arguments involving surfaces of spheres. We obtain $f(r) \propto r^3$ and thus

$$\vec{E}' = \tilde{e}\,\frac{\vec{x}'}{r^3}\,, \qquad \vec{B}' = \vec{o}\,. \tag{8.38}$$

This is the *Coulomb field* of a point charge. In this context, $\tilde{e} = e/(4\pi\epsilon_0)$, where e is the charge, and ϵ_0 is the vacuum permittivity or dielectric constant.

Summarizing, the electromagnetic field tensor that corresponds to the Coulomb field is

$$F_{\mu\nu}(x^\rho) = \tilde{e}\,\frac{2w_{[\mu}x_{\nu]}}{r^3} = \tilde{e}\,\frac{w_\mu x_\nu - w_\nu x_\mu}{r^3}\,, \tag{8.39}$$

where w^μ is the four-velocity of the charge, r the distance between x^ρ and the particle's world line, and $\tilde{e} = e/(4\pi\epsilon_0)$. In the coordinates of the comoving observer X' we have

$$F_{\mu\nu} = \frac{\tilde{e}}{r^3} \begin{pmatrix} 0 & -x'^1 & -x'^2 & -x'^3 \\ x'^1 & 0 & 0 & 0 \\ x'^2 & 0 & 0 & 0 \\ x'^3 & 0 & 0 & 0 \end{pmatrix} \tag{8.39'}$$

and thus (8.38). Some comments are in order.

Remark. To derive (8.39) we have employed (8.36_E). What about (8.36_B)? Using (8.36_B) we obtain the dual of (8.39′), i.e.,

$$\bar{F}_{\mu\nu} = *F_{\mu\nu} = \frac{\tilde{e}}{r^3} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & x'^3 & -x'^2 \\ 0 & -x'^3 & 0 & x'^1 \\ 0 & x'^2 & -x'^1 & 0 \end{pmatrix} ,$$

when we use the coordinates of X'. This electromagnetic field tensor would correspond to

$$\vec{\vec{E}}' = \vec{o}\,, \qquad \vec{\vec{B}}' = \tilde{e}\,\frac{\vec{x}'}{r^3}\,, \tag{8.40}$$

which is the field of a (hypothetic) magnetic monopole. In this context, $\tilde{e}$ encodes the magnetic charge of the monopole. We reject this possibility on physical grounds.

Remark. An elegant derivation of (8.38) from (8.39) makes use of the concepts of section 8.2. The electric four-field E'^μ seen by X' is

$$E'^\mu = F^\mu{}_\nu\, w^\nu = \tilde{e}\,\frac{w^\mu x_\nu - w_\nu x^\mu}{r^3}\, w^\nu = \tilde{e}\,\frac{\eta(w,x)w^\mu + x^\mu}{r^3} = \tilde{e}\,\frac{P_w x^\mu}{r^3}\,.$$

Since $P_w x^\mu = (0, \vec{x}')^{\mathrm{T}}$ we obtain (8.38). The magnetic four-field B'^μ seen by X' is

$$B'^\mu = -\tfrac{1}{2}\,\epsilon^\mu{}_{\nu\sigma\tau} F^{\nu\sigma} w^\tau = -\frac{\tilde{e}}{r^3}\,\epsilon^\mu{}_{\nu\sigma\tau}\, w^{[\nu} x^{\sigma]}\, w^\tau = -\frac{\tilde{e}}{r^3}\,\epsilon^\mu{}_{\nu\sigma\tau}\, w^\nu x^\sigma w^\tau = 0\,.$$

Electric field for an arbitrary observer

Our goal is to compute the electric and magnetic field as seen by an arbitrary observer X, whose four-velocity is u^μ and thus different from the particles' four-velocity w^μ. W.r.t. the coordinates $\{t, \vec{x}\}$ used by X we have

$$u^\mu = \begin{pmatrix} 1 \\ \vec{o} \end{pmatrix}, \qquad w^\mu = \gamma \begin{pmatrix} 1 \\ \vec{v} \end{pmatrix} \qquad x^\mu = \begin{pmatrix} t \\ \vec{x} \end{pmatrix}. \tag{8.41}$$

The electric field $\vec{E}$ as seen by X is given via the electric four-field

$$E_\mu = F_{\mu\nu} u^\nu \,, \tag{8.42}$$

see (8.23a). We obtain

$$E^\mu = \frac{\tilde{e}}{r^3} \left(w^\mu x_\nu - w_\nu x^\mu\right) u^\nu = \frac{\tilde{e}}{r^3} \left(- t w^\mu + \gamma x^\mu\right) \,, \tag{8.43}$$

which results in

$$E^\mu(t, \vec{x}) = \frac{\tilde{e}}{r^3}\, \gamma \begin{pmatrix} 0 \\ \vec{x} - \vec{v} t \end{pmatrix} \tag{8.44}$$

in the coordinates of X. The vector on the r.h.s. of this equation represents the distance between the point at which the field is measured and the position of the charged particle at time t,

$$\varkappa^\mu = x^\mu - t\gamma^{-1} w^\mu = \begin{pmatrix} 0 \\ \vec{x} - \vec{v} t \end{pmatrix} \,, \qquad \vec{\varkappa} = \vec{x} - \vec{v} t \,. \tag{8.45}$$

We thus write

$$E^\mu(t, \vec{x}) = \frac{\tilde{e}}{r^3}\, \gamma \varkappa^\mu \qquad \text{and} \qquad \vec{E}(t, \vec{x}) = \frac{\tilde{e}}{r^3}\, \gamma \vec{\varkappa} \,. \tag{8.46}$$

It remains to express r^3 in terms of the coordinates $(t, \vec{x})$, ideally in terms of $\vec{\varkappa}$. To this end we first recall from (8.34) that

$$r^2 = \eta(P_w x, P_w x) = x^\mu x_\mu + \left(w^\nu x_\nu\right)^2 \,.$$

Second we note that

$$P_w x^\mu = P_w \varkappa^\mu \,,$$

therefore

$$r^2 = \eta(P_w \varkappa, P_w \varkappa) = \varkappa^\mu \varkappa_\mu + \left(w^\nu \varkappa_\nu\right)^2$$

which can be treated in a straightforward manner.[9] We obtain

$$r^2 = \vec{\varkappa}^2 + \left(\gamma \vec{\varkappa} \vec{v}\right)^2 = \vec{\varkappa}^2 + \gamma^2 \vec{v}^2 \vec{\varkappa}^2 \cos^2\alpha = \gamma^2 \vec{\varkappa}^2 \left(1 - \vec{v}^2 \sin^2\alpha\right) \,, \tag{8.47}$$

where α is the angle between $\vec{\varkappa}$ and $\vec{v}$. Inserting this result into (8.46) we finally arrive at

$$E^\mu(t, \vec{x}) = \frac{\tilde{e}}{\gamma^2 \left(1 - \vec{v}^2 \sin^2\alpha\right)^{3/2}} \, \frac{\varkappa^\mu}{|\vec{\varkappa}|^3} \,, \tag{8.48}$$

[9] Note the intimate connection of these considerations with the results of section 5.3.

so that

$$\vec{E}(t,\vec{x}) = \frac{\tilde{e}}{\gamma^2\big(1-\vec{v}^2\sin^2\alpha\big)^{3/2}}\,\frac{\vec{\varkappa}}{|\vec{\varkappa}|^3}\,. \tag{8.49}$$

We conclude that the electric field of a moving charge is compressed by a factor of γ^{-2} in the direction of motion. However, it is stretched by a factor of γ in the plane orthogonal to to the velocity. If the velocity is close to the speed of light, the electric field is almost concentrated in that plane.

Magnetic field for an arbitrary observer

The magnetic four-field seen by the observer X with four-velocity u^μ is

$$B_\mu = -\big(*F_{\mu\nu}u^\nu\big) = -\frac{1}{2}\,\epsilon_{\mu\nu\sigma\tau}\,\frac{\tilde{e}}{r^3}\,2\,w^{[\sigma}x^{\tau]}u^\nu = -\epsilon_{\mu\nu\sigma\tau}\,\frac{\tilde{e}}{r^3}\,w^\sigma x^\tau u^\nu$$

We anticipate that there is a simple relation between E^μ and B^μ. Hence we use (8.43) in the form $\tilde{e}x^\mu/r^3 = \gamma^{-1}(E^\mu + tw^\mu)$ and obtain

$$B_\mu = -\gamma^{-1}\epsilon_{\mu\nu\sigma\tau}w^\sigma(E^\tau + tw^\tau)u^\nu = \gamma^{-1}u^\nu\epsilon_{\nu\mu\sigma\tau}w^\sigma E^\tau\,.$$

W.r.t. the observer's coordinates we have (8.41), hence u^μ is represented by δ^μ_0, which leads to

$$B_i = \gamma^{-1}\epsilon_{0i\sigma\tau}w^\sigma E^\tau = \gamma^{-1}\epsilon_{0ijk}w^j E^k = \epsilon_{ijk}v^j E^k\,,$$

i.e.,

$$\vec{B} = \big(\vec{v}\times\vec{E}\big)\,, \tag{8.50}$$

where $\vec{E}$ is given by (8.49).

An alternative derivation

Naturally there exist alternative derivations of the electromagnetic field of a uniformly moving charge. The 'standard' derivation is based on the transformation formula (8.31) (which did not enter at all in our considerations above!) and application of the Lorentz transformation. Let us sketch this alternative derivation.

We presuppose equation (8.38), which is the electric (Coulomb) field in the rest frame of the particle. W.r.t. to an inertial frame X, in which the charged particle is moving with velocity $\vec{v}$, the field is $\vec{E}$ and $\vec{B}$. From (8.31) we infer

$$\vec{E}^{\|} = \vec{E}'^{\|} \qquad \vec{E}^{\perp} = \gamma\Big(\vec{E}'^{\perp} - \vec{v} \times \vec{B}'\Big) \tag{8.51a}$$

$$\vec{B}^{\|} = \vec{B}'^{\|} \qquad \vec{B}^{\perp} = \gamma\Big(\vec{B}'^{\perp} + \vec{v} \times \vec{E}'\Big)\,, \tag{8.51b}$$

hence

$$\vec{E}^{\|} = \vec{E}'^{\|} \qquad \vec{E}^{\perp} = \gamma\,\vec{E}'^{\perp} \tag{8.52a}$$

$$\vec{B}^{\|} = 0 \qquad \vec{B}^{\perp} = \gamma\left(\vec{v} \times \vec{E}'\right). \tag{8.52b}$$

It remains to express $\vec{x}'$ in terms of t and $\vec{x}$, which can be done by using the Lorentz transformation (2.31),

$$\begin{pmatrix} t' \\ x^{1\prime} \\ x^{2\prime} \\ x^{3\prime} \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma\vec{v}^T \\ -\gamma\vec{v} & \boxed{\mathbb{1} + \dfrac{\gamma - 1}{v^2}\vec{v}\vec{v}^T} \end{pmatrix} \begin{pmatrix} t \\ x^1 \\ x^2 \\ x^3 \end{pmatrix}. \tag{8.53}$$

Standard (but tiresome) algebraic manipulations show that the results can be expressed in terms of the vector $\vec{\varkappa} = \vec{x} - \vec{v}t$. Finally, we arrive again at (8.49).

The simplest way to compute the magnetic field $\vec{B}$ is to again make use of (8.52):

$$\vec{E}^{\|} = \vec{E}'^{\|} \qquad \vec{E}^{\perp} = \gamma\,\vec{E}'^{\perp}$$

$$\vec{B}^{\|} = 0 \qquad \vec{B}^{\perp} = \gamma\left(\vec{v} \times \vec{E}'\right) = \gamma\left(\vec{v} \times \vec{E}'^{\perp}\right).$$

Consequently,

$$\vec{B}^{\|} = 0 \qquad \vec{B}^{\perp} = \Big(\vec{v} \times \left(\gamma\vec{E}'^{\perp}\right)\Big) = \left(\vec{v} \times \vec{E}^{\perp}\right) = \left(\vec{v} \times \vec{E}\right),$$

and therefore

$$\vec{B} = \left(\vec{v} \times \vec{E}\right). \tag{8.54}$$

This completes our discussion of the electromagnetic field of a uniformly moving charge.

8.4 Distributions of particles

Previously we have studied in detail the motion of a single particle. In this section we analyze a continuous *distribution of particles*.

A distribution of particles is characterized by a function $\rho(t, \vec{x})$ that represents the *particle density* at time t and position $\vec{x}$. (Alternatively, the function $\rho(t, \vec{x})$ might represent the mass density or the charge density. It is the latter that is relevant in electromagnetism.) Accordingly,

$$N(t) = \int_V \rho(t, \vec{x})\, d^3x \tag{8.55a}$$

represents the number of particles in a given volume V at time t. It is customary to write

$$dN = \rho\, d^3x\,. \tag{8.55b}$$

In addition to $\rho(t, \vec{x})$ there exists a *velocity field* $\vec{v}(t, \vec{x})$, which represents the velocity of the particles at position $\vec{x}$, at time t. The *particle current* is simply given as $\vec{\jmath} = \rho\vec{v}$. If particles are conserved, the *continuity equation* holds,

$$\frac{\partial \rho}{\partial t} + \vec{\nabla}\vec{\jmath} = 0\,. \tag{8.56}$$

(If ρ represents the charge density, then the continuity equation reflects the conservation of charge.)

Obviously, the definitions and formulas given here make sense only once coordinates $\{t, \vec{x}\}$ have been chosen. However, while the velocity field $\vec{v}(t, \vec{x})$ undergoes the obvious transformation under a change of coordinates, the density $\rho(t, \vec{x})$ does not: ρ is not a scalar function. This is because its definition involves volumes, which change under a change of observer (Lorentz transformation), see section 5.3, where we discuss the Lorentz contraction. Using the results of that section we can find an easy fix by simply compensating for the change of volume with the right γ factor. However, to neatly embed the treatment of distributions of particles into the four-vector formalism of Minkowski spacetime, let us proceed a bit more systematically and geometrically.

A distribution of particles in Minkowski space is a collection of particles represented by a scalar field $\rho_{\text{p}}(x^\sigma)$ and a vector field $u^\mu(x^\sigma)$ on Minkowski space. The latter encodes the four-velocity of the particles at the event x^σ, the former is the particle density *measured in the rest frame* of the particles at the event x^σ. We call ρ_{p} the *proper density* of the distribution.

Remark. Consider a fixed event x^σ. For an observer with four-velocity $u^\mu(x^\sigma)$, we have

$$u^\mu = \begin{pmatrix} 1 \\ \vec{o} \end{pmatrix} \tag{8.57}$$

w.r.t. the observer's coordinates $\{t, \vec{x}\}$ and basis $\{u = e_0, e_1, e_2, e_3\}$. The proper density (at x^σ) is the density in these coordinates, i.e.,

$$dN = \rho_{\mathrm{p}}\, d^3x\,. \tag{8.58}$$

(Note that the volume element d^3x is derived from the rest frame coordinates we use.)

Definition 8.1. *The* particle four-current density *of a distribution of particles is*

$$j^\mu = \rho_{\mathrm{p}}\, u^\mu\,. \tag{8.59}$$

Since j^μ is constructed from a scalar and a four-vector, it is a four-vector.

W.r.t. an arbitrary observer X we have

$$j^\mu = \rho_{\mathrm{p}}\, u^\mu = \rho_{\mathrm{p}}\gamma \begin{pmatrix} 1 \\ \vec{v} \end{pmatrix}\,. \tag{8.60}$$

In order to interpret the quantity $\rho_{\mathrm{p}}\gamma$ in terms of the density ρ measured by X we must consider volumes.

By definition, the proper density ρ_{p} at the event x^μ represents the number of particles contained in a unit volume in the rest frame of the particles at x^μ. For an observer X that moves w.r.t. this rest frame with velocity $-\vec{v}$, 'longitudinal lengths' undergo a Lorentz contraction by a factor of γ^{-1}, while 'transversal lengths' are unaffected.[10] Consequently, volumes (like a cube) appear 'contracted' by a factor of γ^{-1} in the direction of motion of X. Under the change of observer we consider, we thus find that

$$\rho = \rho_{\mathrm{p}}\gamma \tag{8.61}$$

is the density seen by X.

Accordingly, we can write the four-current density (8.60) as

$$j^\mu = \rho \begin{pmatrix} 1 \\ \vec{v} \end{pmatrix}\,, \tag{8.60$'$}$$

[10]The observer X moves with velocity $-\vec{v}$ w.r.t. the local rest frame of the particles. Hence, in the observer's coordinates, the particles move with velocity $\vec{v}$, which is reflected in (8.60).

where ρ and $\vec{v}$ are the particle density and the velocity field of the particle distribution as seen by X.

The *continuity equation* takes a manifestly covariant form when we use the four-current j^μ. From (8.56) we directly obtain

$$\partial_\mu j^\mu = 0 \,, \tag{8.62}$$

i.e., the four-current density is *divergence-free*. Recall that $\partial_0 = \partial/\partial x^0 = \partial/\partial t$ and that $\partial_i = \partial/\partial x^i$.

A little more about volumes

In analogy to the rigid rod considered in section 5.3, consider an extended body in uniform motion that takes up a fixed volume; let the four-velocity of the body be u^μ. The rest frame of the body is given by the co-moving observer X, whose four-velocity coincides with that of the body.

In the simplest case, the volume is a cuboid that is spanned by three spacelike vectors a^μ, b^μ, c^μ. W.l.o.g. we could choose these vectors to be orthogonal to the four-velocity u^μ, so that they lie in the observer's plane of simultaneity; this is not necessary, however. W.r.t. the co-moving observer's coordinates we have

$$u^\mu = \begin{pmatrix} 1 \\ \vec{o} \end{pmatrix} \,, \qquad a^\mu = \begin{pmatrix} a^0 \\ \vec{a} \end{pmatrix} \,, \qquad b^\mu = \begin{pmatrix} b^0 \\ \vec{b} \end{pmatrix} \,, \qquad c^\mu = \begin{pmatrix} c^0 \\ \vec{c} \end{pmatrix} \,. \tag{8.63}$$

Each point p^μ in the volume taken up by the body propagates along a world line $p^\mu + su^\mu$. Now, the proper volume of the cuboid (which is the volume measured in the rest frame X) is simply

$$V = V_p = \det \begin{pmatrix} \vec{a} & \vec{b} & \vec{c} \end{pmatrix} = \epsilon_{ijk} a^i b^j c^k \,. \tag{8.64}$$

Using the coordinate representation (8.63) it is not difficult to convince oneself that this is equivalent to

$$V = V_p = \epsilon_{\mu\alpha\beta\gamma} u^\mu a^\alpha b^\beta c^\gamma \,. \tag{8.65}$$

Since the volume form $\epsilon_{\alpha\beta\gamma\delta}$ is invariant under a change of inertial coordinates, this formula holds independently of the chosen coordinates.

For a different observer X', whose four-velocity is u'^μ, the body moves with some velocity $\vec{v}$, i.e.,

$$u'^\mu = \begin{pmatrix} 1 \\ \vec{o} \end{pmatrix} , \qquad u^\mu = \gamma \begin{pmatrix} 1 \\ \vec{v} \end{pmatrix} \tag{8.66}$$

w.r.t. X'. W.l.o.g. the three vectors a^μ, b^μ, and c^μ can be arranged to be simultaneous in the coordinates of X', i.e.,

$$a^\mu = \begin{pmatrix} 0 \\ \vec{a}' \end{pmatrix} , \qquad b^\mu = \begin{pmatrix} 0 \\ \vec{b}' \end{pmatrix} , \qquad c^\mu = \begin{pmatrix} 0 \\ \vec{c}' \end{pmatrix} . \tag{8.67}$$

w.r.t. X'. The volume measured by X' is given by

$$V' = \det \begin{pmatrix} \vec{a}' & \vec{b}' & \vec{c}' \end{pmatrix} = \epsilon_{ijk} a'^i b'^j c'^k . \tag{8.68}$$

or, equivalently, by[11]

$$V' = \epsilon_{\mu\alpha\beta\gamma} u'^\mu a^\alpha b^\beta c^\gamma . \tag{8.69}$$

In contrast, the proper volume is given by (8.65). Inserting $u^\mu = \gamma(u'^\mu + v^i e'^{\,\mu}_i)$ into (8.65) we compute

$$V_p = \gamma \epsilon_{\mu\alpha\beta\gamma} u'^\mu a^\alpha b^\beta c^\gamma + \gamma \underbrace{\epsilon_{\mu\alpha\beta\gamma} v^i e'^{\,\mu}_i a^\alpha b^\beta c^\gamma}_{=0} = \gamma V' . \tag{8.70}$$

Consequently, we find

$$V' = \gamma^{-1} V_p , \tag{8.71}$$

i.e., there is a 'volume contraction'.

8.5 Maxwell's equations

W.r.t. an inertial coordinate system, Maxwell's equations read

$$\vec{\nabla}\vec{E} = 4\pi\rho \qquad \vec{\nabla}\vec{B} = 0 \tag{8.72a}$$

$$\vec{\nabla} \times \vec{E} = -\partial_t \vec{B} \qquad \vec{\nabla} \times \vec{B} = \partial_t \vec{E} + 4\pi\vec{\jmath} , \tag{8.72b}$$

where $\rho = \rho(t, \vec{x})$ is the charge density and $\vec{\jmath} = \rho\vec{v}$ is the charge current density. Our aim is to find a version of Maxwell's equations that is manifestly covariant.

[11] The fact that we have chosen a^μ, b^μ, c^μ to be orthogonal to u'^μ is important here.

Let us first consider the equations with source terms, i.e.,

$$\vec{\nabla}\vec{E} = 4\pi\rho\,, \qquad \vec{\nabla}\times\vec{B} - \partial_t\vec{E} = 4\pi\vec{\jmath}\,. \tag{8.73}$$

In the preceding section we have seen that ρ and $\vec{\jmath}$ can be collected into a four-vector, the *four-current density*

$$j^\mu = \rho\begin{pmatrix}1\\ \vec{v}\end{pmatrix}\,.$$

Recalling that $F_{i0} = \delta_{ij}E^j$ and $F_{ij} = \epsilon_{ijk}B^k$ it is straightforward to see that the equations (8.73) take the form

$$\partial_\nu F^{\mu\nu} = 4\pi j^\mu\,. \tag{8.73$'$}$$

This is the first of the Maxwell equations in their relativistic formulation. (Clearly, since $\mu = 0,\ldots,3$, we have four equations encoded in (8.73′).)

Example. For instance, let μ be a spatial index k. Then

$$\partial_\nu F^{k\nu} = \partial_0 F^{k0} + \partial_i F^{ki} = \partial_t\big(\eta^{00}E^k\big) + \partial_i\big(\epsilon^{kij}B_j\big) = -\partial_t E^k + \big(\vec{\nabla}\times\vec{B}\big)^k\,,$$

from which the second equation of (8.73) follows.

Remark. Often, partial derivatives are denoted by commas, e.g., for a function f we have $f_{,\mu} \equiv \partial_\mu f$. Using this notation, $\partial_\nu F^{\mu\nu}$ becomes ${F^{\mu\nu}}_{,\nu}$, so that (8.73′) is written as

$${F^{\mu\nu}}_{,\nu} = 4\pi j^\mu\,. \tag{8.73$''$}$$

Second we consider the remaining equations

$$\vec{\nabla}\vec{B} = 0\,, \qquad \vec{\nabla}\times\vec{E} + \partial_t\vec{B} = 0\,. \tag{8.74}$$

A straightforward calculation shows that these equations are equivalent to

$$\partial_{[\mu}F_{\nu\sigma]} = 0\,. \tag{8.74$'$}$$

Equivalently, we can write

$$F_{[\mu\nu,\sigma]} = 0\,. \tag{8.74$''$}$$

Example. For instance, consider $\partial_{[0}F_{ij]}$, which is

$$\begin{aligned}\partial_{[0}F_{ij]} &= \frac{1}{3}\left[\partial_0 F_{ij} + \partial_i F_{j0} + \partial_j F_{0i}\right] = \frac{1}{3}\left[\partial_t\left(\epsilon_{ijk}B^k\right) + \partial_i E_j - \partial_j E_i\right] \\ &= \frac{1}{3}\left[\epsilon_{ijk}\left(\partial_t B^k\right) + 2\partial_{[i}E_{j]}\right] .\end{aligned}$$

Then

$$\begin{aligned}3\epsilon^{lij}\partial_{[0}F_{ij]} &= \epsilon^{lij}\epsilon_{ijk}\left(\partial_t B^k\right) + 2\epsilon^{lij}\partial_i E_j = \left(\delta^j{}_j\delta^l{}_k - \delta^j{}_k\delta^l{}_j\right)\left(\partial_t B^k\right) + 2\epsilon^{lij}\partial_i E_j \\ &= \left(3\delta^l{}_k - \delta^l{}_k\right)\left(\partial_t B^k\right) + 2\epsilon^{lij}\partial_i E_j = 2\partial_t B^l + 2\left(\vec{\nabla}\times\vec{E}\right)^l\end{aligned}$$

and the second equation in (8.74) follows.

Collecting the results we obtain the Maxwell equations in their manifestly covariant relativistic version:

$$F_{[\mu\nu,\sigma]} = 0 , \qquad F^{\mu\nu}{}_{,\nu} = 4\pi j^\mu . \tag{8.75}$$

Remark. Alternatively, the homogeneous Maxwell equation can be expressed in terms of the dual field tensor $*F_{\mu\nu}$. The Maxwell equations then read

$$*F^{\mu\nu}{}_{,\nu} = 0 , \qquad F^{\mu\nu}{}_{,\nu} = 4\pi j^\mu , \tag{8.76}$$

which is a simple consequence of (8.20) and (8.75).

Remark for experts. Let us define the complex tensor[12]

$$W_{\mu\nu} = F_{\mu\nu} + i *F_{\mu\nu} . \tag{8.77}$$

This tensor is antisymmetric like $F_{\mu\nu}$ and $*F_{\mu\nu}$; moreover, $W_{\mu\nu}$ is anti-self-dual, i.e.,

$$*W_{\mu\nu} = *F_{\mu\nu} + i **F_{\mu\nu} \overset{(8.21)}{=} *F_{\mu\nu} - iF_{\mu\nu} = -iW_{\mu\nu} . \tag{8.78}$$

The complex conjugate tensor $\overline{W}_{\mu\nu}$ is self-dual, i.e., $*\overline{W}_{\mu\nu} = i\overline{W}_{\mu\nu}$. In terms of $W_{\mu\nu}$, the Maxwell equations read

$$W^{\mu\nu}{}_{,\nu} = 4\pi j^\mu . \tag{8.79}$$

[12]Some mathematical background: The duality map $*$ is a linear map on the linear space of antisymmetric tensors $F_{\mu\nu}$. Since $*$ is an anti-involution, i.e., $** = -\mathrm{id}$, the only possible eigenvalues of $*$ can be $\pm i$. If $W_{\mu\nu}$ is an eigenvector of $*$ associated with the eigenvalue $(-i)$, then the complex conjugate vector $\overline{W}_{\mu\nu}$ is automatically an eigenvector associated with the eigenvalue i.

Remark for experts. In the language of differential forms, the antisymmetric field $F_{\mu\nu}$ is a 2-form, which is simply denoted by F. The operation $F_{[\mu\nu,\sigma]}$ then corresponds to the exterior derivative of this 2-form and is written as dF. The operation $F_{\mu}{}^{\nu}{}_{,\nu}$ is the co-differential $\delta F = *d*F$. Hence, using differential forms, the Maxwell equations (8.75) look particularly simple:

$$dF = 0\,, \qquad \delta F = 4\pi j\,. \tag{8.80}$$

Here, j is the current one-form, whose components are j_μ. Alternatively, one writes

$$dF = 0\,, \qquad d*F = 4\pi J\,, \tag{8.81}$$

where $J = *j$ is the current three-form.

A simple consequence of Maxwell's equations is the continuity equation. Indeed,

$$4\pi\partial_\mu j^\mu = \partial_\mu\partial_\nu F^{\mu\nu} = 0\,, \tag{8.82}$$

which follows directly from the antisymmetry of $F_{\mu\nu}$ (where we note that $\partial_\mu\partial_\nu$ is symmetric). The continuity equation implies the *conservation of charge*.

8.6 Four-potential

A fundamental result in analysis concerns the existence of potentials of vector fields. For instance, if $\vec{v}(\vec{x})$ is a vector field on $\mathbb{R}^3$ (or some other simply connected space), then there exists a (scalar) potential $\phi(\vec{x})$ if and only if the curl of the vector field vanishes, i.e.,

$$\vec{v} = \vec{\nabla}\phi \quad \Leftrightarrow \quad \vec{\nabla}\times\vec{v} = 0\,. \tag{8.83}$$

Since $(\vec{\nabla}\times\vec{v})^i = \epsilon^{ijk}\partial_j v_k$, this is in turn equivalent to requiring that $\partial_{[i}v_{j]} = 0$. In general, if $v^i(x)$ is a vector field on $x \in \mathbb{R}^n$, then there exists a (scalar) potential $\phi(x)$, if and only if $v_{[i,j]} = 0$, i.e.,

$$v_i = \partial_i\phi \quad \Leftrightarrow \quad v_{[i,j]} = 0\,. \tag{8.84}$$

This potential ϕ can be determined by a path integral,

$$\phi(x) = \int_0^x v_i\, ds^i\,. \tag{8.85}$$

The statement (8.84) can be generalized to antisymmetric tensors (n-forms): In particular, if $F_{\mu\nu}$ is an antisymmetric tensor, then there exists a *four-potential* A^μ if and only if $F_{[\mu\nu,\sigma]} = 0$. Therefore, for the electromagnetic field, the Maxwell equations automatically guarantee the existence of a potential A^μ, whose antisymmetric derivatives yield $F_{\mu\nu}$. More specifically,

$$F_{\mu\nu} = 2A_{[\mu,\nu]} \quad \Leftrightarrow \quad F_{[\mu\nu,\sigma]} = 0\,. \tag{8.86}$$

Remark. The first guess for the relation between $F_{\mu\nu}$ and the potential A^μ is probably $F_{\mu\nu} = A_{\mu,\nu}$. However, to ensure that the l.h. side is antisymmetric, we must perform an antisymmetrization. The factor of 2 in (8.86) is introduced for aesthetic reasons; alternatively, we can write (8.86) as

$$F_{\mu\nu} = A_{\mu,\nu} - A_{\nu,\mu} = \partial_\nu A_\mu - \partial_\mu A_\nu\,. \tag{8.86$'$}$$

It is not necessary to invoke a general theorem to prove (8.86). Instead, we can explicitly construct a four-potential in analogy to (8.85) (by means of what is known as the Poincaré-Lemma). Let us define A^μ according to

$$A_\mu(x) = \int_0^1 F_{\mu\nu}(\lambda x)\,\lambda x^\nu d\lambda\,. \tag{8.87}$$

To prove (the nontrivial direction of) the statement (8.86), we must show that $2A_{[\mu,\nu]} = F_{\mu\nu}$. We proceed step by step:

$$A_{\mu,\nu} = \int_0^1 \big(F_{\mu\sigma}(\lambda x)\,x^\sigma\big)_{,\nu}\lambda d\lambda = -\int_0^1 \big(F_{\sigma\mu}(\lambda x)\,x^\sigma\big)_{,\nu}\lambda d\lambda\,,$$

hence

$$\begin{aligned}
A_{[\mu,\nu]} &= \int_0^1 \big(F_{\sigma[\nu}(\lambda x)\,x^\sigma\big)_{,\mu]}\lambda d\lambda = \int_0^1 \Big(F_{\sigma[\nu,\mu]}(\lambda x)\,\lambda x^\sigma + F_{\sigma[\nu}\delta^\sigma_{\mu]}\Big)\lambda d\lambda \\
&= \int_0^1 \Big(F_{\sigma[\nu,\mu]}(\lambda x)\,\lambda x^\sigma + F_{[\mu\nu]}\Big)\lambda d\lambda \\
&= \int_0^1 F_{\sigma[\nu,\mu]}(\lambda x)\,\lambda^2 x^\sigma d\lambda + \int_0^1 F_{\mu\nu}\lambda d\lambda\,.
\end{aligned}$$

In the next step we use equation $(*)$ on this page.

$$\begin{aligned} A_{[\mu,\nu]} &\overset{(*)}{=} \int_0^1 \frac{1}{2} F_{\mu\nu,\sigma}(\lambda x)\, \lambda^2 x^\sigma d\lambda + \int_0^1 F_{\mu\nu} \lambda d\lambda \\ &= \int_0^1 \frac{1}{2} \left(\frac{d}{d\lambda} F_{\mu\nu}(\lambda x) \right) \lambda^2 d\lambda + \int_0^1 F_{\mu\nu} \lambda d\lambda \end{aligned}$$

Finally, by an integration by parts, we obtain

$$A_{[\mu,\nu]} = \frac{1}{2} F_{\mu\nu}(\lambda x)\, \lambda^2 \Big|_0^1 - \int_0^1 F_{\mu\nu} \lambda d\lambda + \int_0^1 F_{\mu\nu} \lambda d\lambda = \frac{1}{2} F_{\mu\nu}(x)$$

Here we have used $F_{[\mu\nu,\sigma]} = 0$ to derive

$$F_{\sigma[\nu,\mu]} = \frac{1}{2}\, F_{\mu\nu,\sigma}\,. \tag{$*$}$$

Summing up, we have explicitly constructed a four-potential A^μ from $F_{\mu\nu}$ via (8.86),

$$F_{\mu\nu} = A_{\mu,\nu} - A_{\nu,\mu}\,. \tag{8.86$'$}$$

It is important to note that there does *not* exist a unique four-potential. Let $\hat{A}^\mu$ and $\breve{A}^\mu$ be four-potentials of $F_{\mu\nu}$, i.e.,

$$F_{\mu\nu} = 2\hat{A}_{[\mu,\nu]}\,, \qquad F_{\mu\nu} = 2\breve{A}_{[\mu,\nu]}\,.$$

Then $\mathcal{A}_\mu = \breve{A}_\mu - \hat{A}_\mu$ satisfies

$$\mathcal{A}_{[\mu,\nu]} = 0\,,$$

which implies that there exists a scalar function Λ so that $\mathcal{A}_\mu = \partial_\mu \Lambda$, see equation (8.84). Accordingly,

$$\breve{A}_\mu = \hat{A}_\mu + \Lambda_{,\mu}\,. \tag{8.88}$$

Conversely, if $\hat{A}_\mu$ is a four-potential of $F_{\mu\nu}$, then $\breve{A}_\mu$ defined by (8.88) is a four-potential as well. The non-uniqueness of the four-potential expressed by (8.88) is called *gauge-freedom*.

In the following we formulate Maxwell's equations in terms of a four-potential A^μ. By construction, $F_{[\mu\nu,\sigma]} = 0$ holds automatically, since $F_{\mu\nu}$ is derived from a four-potential A^μ via (8.86′). The second of Maxwell's equations becomes

$$4\pi j^\mu = \partial_\nu F^{\mu\nu} = \partial_\nu \big(\partial^\nu A^\mu - \partial^\mu A^\nu\big) = \Box A^\mu - \partial^\mu \big(\partial_\nu A^\nu\big) \,. \tag{8.89}$$

To get rid of the second term, we require the four-potential to satisfy the *Lorenz gauge* condition[13]

$$\partial_\nu A^\nu = 0 \,. \tag{8.90}$$

Clearly, a given four-potential A^μ will in general not satisfy the Lorenz gauge condition (8.90); however, we can always make use of the gauge freedom (8.88) to achieve (8.90). To see this, assume that $\partial_\nu A^\nu \neq 0$; then there exists a scalar function Λ and a modified four-potential $\bar{A}^\mu = A^\mu + \partial^\mu \Lambda$, such that

$$\partial_\nu \bar{A}^\nu = \partial_\nu \big(A^\nu + \partial^\nu \Lambda\big) = \partial_\nu A^\nu + \Box \Lambda = 0 \,,$$

since the inhomogeneous wave equation

$$\Box \Lambda = -\partial_\nu A^\nu$$

possesses a solution Λ.

Exercise. Show that the four-potential A^μ defined by (8.87) automatically satisfies the Lorenz gauge condition, if $j^\mu = 0$.

Using a four-potential A^μ in Lorenz gauge, i.e., $\partial_\mu A^\mu = 0$, the second Maxwell equation thus reads

$$\Box A^\mu = 4\pi j^\mu \,. \tag{8.91}$$

Hence, each component of the four-potential satisfies a wave equation.

Remark. W.r.t. some inertial coordinate system we have

$$F_{\mu\nu} = \begin{pmatrix} 0 & -E^1 & -E^2 & -E^3 \\ E^1 & 0 & B^3 & -B^2 \\ E^2 & -B^3 & 0 & B^1 \\ E^3 & B^2 & -B^1 & 0 \end{pmatrix} \,,$$

[13] The Lorenz gauge condition was formulated by Ludvig Valentin Lorenz (⋆1829 in Helsingør, †1891) and not by Hendrik Antoon Lorentz (⋆1853 in Arnhem, †1928 in Haarlem).

see (8.16). Likewise, the four-potential A^μ is given by

$$A^\mu = \begin{pmatrix} \phi \\ \vec{A} \end{pmatrix} , \tag{8.92}$$

where we have set $A^0 = \phi$ (so that $A_0 = -\phi$). Since $F_{\mu\nu} = 2A_{[\mu,\nu]}$ we have

$$\begin{aligned} E_i &= F_{i0} = A_{i,0} - A_{0,i} = \partial_t A_i + \partial_i \phi \,, \\ B_i &= \tfrac{1}{2}\,\epsilon_{ijk} F^{jk} = \epsilon_{ijk} A^{j,k} = -\epsilon_{ikj} \partial^k A^j \end{aligned}$$

and hence

$$\vec{E} = \partial_t \vec{A} + \vec{\nabla}\phi \,, \qquad \vec{B} = -\vec{\nabla} \times \vec{A} \,. \tag{8.93}$$

Therefore, the components ϕ and $\vec{A}$ of the four-potential coincide (up to a minus sign[14]) with the electric scalar potential and the magnetic vector potential, which are known from the standard formulation of Maxwell's equations in terms of potentials.

The Maxwell equation (8.91) for the four-potential A^μ in Lorenz gauge can be solved if boundary conditions are prescribed. In particular, we obtain the well-known advanced and retarded solutions discussed in every textbook on electromagnetism.

8.7 Energy-momentum tensor

The energy-momentum tensor (stress-energy tensor) of the electromagnetic field is defined as

$$T_{\mu\nu} = \frac{1}{4\pi} \left[F_{\mu\sigma} F_\nu{}^\sigma - \frac{1}{4} F_{\sigma\lambda} F^{\sigma\lambda} \eta_{\mu\nu} \right] . \tag{8.94}$$

Remark for experts. Using the tensor $W_{\mu\nu} = F_{\mu\nu} + i * F_{\mu\nu}$, see (8.77), and its complex conjugate, the energy-momentum tensor $T_{\mu\nu}$ can be constructed in a simpler way,

$$T_{\mu\nu} = \frac{1}{4\pi} \frac{1}{2}\, W_{\mu\sigma} \overline{W}_\nu{}^\sigma \,. \tag{8.94$'$}$$

[14] If had chosen the signature of the Minkowski metric to be $\operatorname{sign}\eta = (+ - - -)$ instead of $(- + + +)$, there would not appear a minus sign here. However, our choice of signature is well adapted to other situations; in particular, when we go over from special relativity to general relativity.

Proof. To prove that the two expressions for $T_{\mu\nu}$ coincide we perform a straightforward calculation.[15]

$$\begin{aligned} W_{\mu\sigma}\overline{W}_{\nu}{}^{\sigma} &= \left(F_{\mu\sigma} + i\,{}^*F_{\mu\sigma}\right)\left(F_{\nu}{}^{\sigma} - i\,{}^*F_{\nu}{}^{\sigma}\right) \\ &= F_{\mu\sigma}F_{\nu}{}^{\sigma} \;+\; 2\,i\, \underbrace{{}^*F_{[\mu}{}^{\sigma}F_{\nu]\sigma}}_{\eta_{\tau[\mu}{}^*F^{\tau\sigma}F_{\nu]\sigma}} \;+\; {}^*F_{\mu\sigma} \underbrace{{}^*F_{\nu}{}^{\sigma}}_{\eta_{\nu\pi}{}^*F^{\pi\sigma}} \\ &= F_{\mu\sigma}F_{\nu}{}^{\sigma} + i\,\eta_{\tau[\mu}F_{\nu]\sigma}\epsilon^{\tau\sigma\rho\pi}F_{\rho\pi} + \tfrac{1}{4}\,\epsilon_{\mu\sigma\tau\rho}F^{\tau\rho}\eta_{\nu\pi}\epsilon^{\pi\sigma\lambda\xi}F_{\lambda\xi}\,. \end{aligned} \tag{8.95}$$

The second term can be manipulated by making use of the results of Appendix B. We have

$$\epsilon^{\tau\sigma\rho\pi}F_{\nu\sigma}F_{\rho\pi} = \epsilon^{\tau\sigma\rho\pi}F_{\nu[\sigma}F_{\rho\pi]}$$

because of (B.10) and further

$$\epsilon^{\tau\sigma\rho\pi}F_{\nu\sigma}F_{\rho\pi} = \epsilon^{\tau\sigma\rho\pi}F_{[\nu\sigma}F_{\rho\pi]}$$

because of (B.28a). In the case of a four-dimensional vector space (which is our case), modulo constants there exists only one totally antisymmetric tensor of order four, the ϵ-tensor, cf. (B.12); hence

$$F_{[\nu\sigma}F_{\rho\pi]} \propto \epsilon_{\nu\sigma\rho\pi}$$

and thus

$$\epsilon^{\tau\sigma\rho\pi}F_{\nu\sigma}F_{\rho\pi} \propto \epsilon^{\tau\sigma\rho\pi}\epsilon_{\nu\sigma\rho\pi} = \epsilon^{\sigma\rho\pi\tau}\epsilon_{\sigma\rho\pi\nu} = -6\delta^{\tau}_{\nu}\,,$$

where we have used (B.25). Therefore,

$$i\,\eta_{\tau[\mu}F_{\nu]\sigma}\epsilon^{\tau\sigma\rho\pi}F_{\rho\pi} \propto \eta_{\tau[\mu}\delta^{\tau}_{\nu]} = \eta_{[\nu\mu]} = 0\,,$$

i.e., the second term in (8.95) vanishes. The third term in (8.95) can be manipulated along the following lines:

$$\epsilon_{\mu\sigma\tau\rho}\epsilon^{\pi\sigma\lambda\xi}F^{\tau\rho}F_{\lambda\xi}\eta_{\nu\pi} = \epsilon_{\sigma\mu\tau\rho}\epsilon^{\sigma\pi\lambda\xi}F^{\tau\rho}F_{\lambda\xi}\eta_{\nu\pi} = -\delta^{\pi\lambda\xi}_{\mu\tau\rho}F^{\tau\rho}F_{\lambda\xi}\eta_{\nu\pi}$$

Here we have used (B.25). In the next step we apply (B.16), i.e.,

$$\epsilon_{\mu\sigma\tau\rho}\epsilon^{\pi\sigma\lambda\xi}F^{\tau\rho}F_{\lambda\xi}\eta_{\nu\pi} = \Big(-2\delta^{\pi}_{\mu}\delta^{[\lambda}_{\tau}\delta^{\xi]}_{\rho} - 4\delta^{[\lambda}_{\mu}\delta^{\xi]}_{[\tau}\delta^{\pi}_{\rho]}\Big)F^{\tau\rho}F_{\lambda\xi}\eta_{\nu\pi}\,.$$

[15] The calculation is straightforward, but this doesn't mean that it's easy. Actually, it isn't. In fact, it would be easier to define $F_{\mu\nu}$, $*F_{\mu\nu}$ and $W_{\mu\nu}$ as matrices in Mathematica or Maple and let the computer do the rest. However, we are here on a training ground for our later lives as (theoretical) physicists. (Yes, this might be part of what you'll be doing. . .)

The antisymmetrizations can be dropped because of (B.10) and we obtain

$$\epsilon_{\mu\sigma\tau\rho}\epsilon^{\pi\sigma\lambda\xi}F^{\tau\rho}F_{\lambda\xi}\eta_{\nu\pi} = -2\eta_{\mu\nu}\delta^{\lambda}_{\tau}\delta^{\xi}_{\rho}F^{\tau\rho}F_{\lambda\xi} - 4\delta^{\lambda}_{\mu}\delta^{\xi}_{\tau}\delta^{\pi}_{\rho}F^{\tau\rho}F_{\lambda\xi}\eta_{\nu\pi}\,.$$

Continuing in the obvious way we get

$$\begin{aligned}\tfrac{1}{4}\,\epsilon_{\mu\sigma\tau\rho}\epsilon^{\pi\sigma\lambda\xi}F^{\tau\rho}F_{\lambda\xi}\eta_{\nu\pi} &= -\tfrac{1}{2}\,\eta_{\mu\nu}F^{\lambda\xi}F_{\lambda\xi} - F^{\tau\rho}F_{\mu\tau}\eta_{\nu\rho}\\ &= -\tfrac{1}{2}\eta_{\mu\nu}F^{\lambda\xi}F_{\lambda\xi} + F_{\mu\tau}F_{\nu}{}^{\tau}\,.\end{aligned}$$

Inserting this result into (8.95) we finally arrive at

$$\tfrac{1}{2}W_{\mu\sigma}\overline{W}_{\nu}{}^{\sigma} = \tfrac{1}{2}\left[F_{\mu\sigma}F_{\nu}{}^{\sigma} - \tfrac{1}{2}\eta_{\mu\nu}F^{\lambda\xi}F_{\lambda\xi} + F_{\mu\tau}F_{\nu}{}^{\tau}\right] = F_{\mu\sigma}F_{\nu}{}^{\sigma} - \tfrac{1}{4}\eta_{\mu\nu}F^{\lambda\xi}F_{\lambda\xi},$$

which proves the claim (8.94′). □

The energy-momentum tensor (8.94) is a *symmetric* tensor, i.e.,

$$T_{\mu\nu} = T_{(\mu\nu)}\,. \tag{8.96a}$$

To see this we simply note that $F_{\nu\sigma}F_{\mu}{}^{\sigma} = F_{\nu}{}^{\sigma}F_{\mu\sigma} = F_{\mu\sigma}F_{\nu}{}^{\sigma}$, hence the first term is symmetric; for the second term, symmetry is evident.

Another important property of the electromagnetic energy-momentum tensor is its *tracelessness*, i.e.,

$$T^{\mu}{}_{\mu} = 0\,. \tag{8.96b}$$

The proof is straightforward:

$$T^{\mu}{}_{\mu} = \frac{1}{4\pi}\left[F^{\mu}{}_{\sigma}F_{\mu}{}^{\sigma} - \frac{1}{4}F_{\sigma\lambda}F^{\sigma\lambda}\underbrace{\delta^{\mu}{}_{\mu}}_{=4}\right] = \frac{1}{4\pi}\left[F_{\mu\sigma}F^{\mu\sigma} - F_{\sigma\lambda}F^{\sigma\lambda}\right] = 0\,.$$

Next we compute the divergence $T^{\mu\nu}{}_{,\nu}$ of the energy-momentum tensor:

$$\begin{aligned}4\pi T_{\mu}{}^{\nu}{}_{,\nu} &= F_{\mu\sigma,\nu}F^{\nu\sigma} + F_{\mu\sigma}F^{\nu\sigma}{}_{,\nu} - \tfrac{1}{4}\big(F_{\sigma\lambda,\nu}F^{\sigma\lambda} + F_{\sigma\lambda}F^{\sigma\lambda}{}_{,\nu}\big)\delta_{\mu}{}^{\nu}\\ &\overset{(8.75)}{=} F_{\mu\sigma,\nu}F^{\nu\sigma} - 4\pi F_{\mu\sigma}j^{\sigma} - \tfrac{1}{2}F_{\sigma\lambda,\mu}F^{\sigma\lambda}\\ &\overset{(*)}{=} F_{\mu\sigma,\nu}F^{\nu\sigma} - 4\pi F_{\mu\sigma}j^{\sigma} - F_{\mu[\lambda,\sigma]}F^{\sigma\lambda}\\ &= F_{\mu\sigma,\nu}F^{\nu\sigma} - 4\pi F_{\mu\sigma}j^{\sigma} - F_{\mu\lambda,\sigma}F^{\sigma\lambda}\\ &= -4\pi F_{\mu\sigma}j^{\sigma}\end{aligned}$$

where we have used equation (∗) of page 136. We note that $T_{\mu\nu}$ is divergence-free in the absence of sources, i.e.,

$$T^{\mu\nu}{}_{,\nu} = 0 \qquad \text{if } j^\mu = 0\,; \tag{8.97}$$

however, in general we obtain

$$T^{\mu\nu}{}_{,\nu} = -F^\mu{}_\nu\, j^\nu\,. \tag{8.98}$$

Let us investigate the r.h.s. of this equation. The charge four-current density j^μ is given by

$$j^\mu = \rho_e u^\mu\,,$$

see (8.59), where in the present context ρ_e is the proper charge density (i.e., charge per proper volume) of the charge distribution. Hence,

$$F^\mu{}_\nu\, j^\nu = \rho_e F^\mu{}_\nu\, u^\nu\,. \tag{8.99}$$

The Lorentz force acting on a point particle (with charge e) is

$$\mathcal{F}^\mu = eF^\mu{}_\nu\, u^\nu\,.$$

Since ρ_e is charge per proper volume, we conclude that (8.99) represents a *force density* $\mathfrak{F}^\mu$,

$$\mathfrak{F}^\mu = F^\mu{}_\nu\, j^\nu\,; \tag{8.100}$$

it is the Lorentz force (per proper volume) acting on the charge distribution of particles.

Now let $\{t, \vec{x}\}$ be inertial coordinates of an inertial observer X. The total four-force acting on an infinitesimal volume d^3x is

$$\mathcal{F}^\mu_{\text{tot}} = \mathfrak{F}^\mu\, \gamma\, d^3x\,, \tag{8.101}$$

where the factor γ enters because the proper volume of d^3x is $\gamma\, d^3x$, see (8.71). The (total) four-force is connected with the derivative (w.r.t. proper time[16]) of the (total) four-momentum in d^3x, which is

$$\frac{d}{ds}p^\mu_{\text{tot}} = \gamma\, \frac{d}{dt}p^\mu_{\text{tot}}\,. \tag{8.102}$$

[16]Proper time is defined along the integration curves of the four-vector field u^μ.

The four-momentum p^μ_{tot} is given by the four-momentum density $\mathcal{P}^\mu$ times the volume d^3x, i.e.,

$$p^\mu_{\text{tot}} = \mathcal{P}^\mu \, d^3x \, . \tag{8.103}$$

Hence, the expression $\mathcal{F}^\mu d^3x$ is connected with the time-derivative of $\mathcal{P}^\mu \, d^3x$. (Equality holds under appropriate conditions on the distribution of particles or when we consider not small volumes but the entire space $\mathbb{R}^3$.)

Remark. Note that the four-momentum density $\mathcal{P}^\mu$ is specific to the chosen observer—a different observer would define a different four-momentum density $\mathcal{P}^{\mu\,\prime}$.[17] This is because $\mathcal{P}^\mu$ is a density w.r.t. coordinate volume (and not proper volume). In the case of simple matter we have $\mathcal{P}^\mu = \rho u^\mu$, where in this context ρ is the mass density (w.r.t. coordinate volume). This will be discussed further in the next section.

Making use of these considerations we proceed with equation (8.98). The integral version is

$$\int T^{\mu\nu}{}_{,\nu} \, d^3x + \int F^\mu{}_\nu \, j^\nu d^3x = 0 \, . \tag{8.104}$$

When we use that $T^{\mu\nu}{}_{,\nu} = \partial_t T^{\mu 0} + \partial_i T^{\mu i}$ we obtain

$$\frac{d}{dt} \int T^{\mu 0} d^3x + \int \partial_i T^{\mu i} d^3x \tag{8.105}$$

for the first term of (8.104). The integral over the spatial divergence $\partial_i T^{\mu i}$ can be transformed into a boundary integral:

$$\int_V \partial_i T^{\mu i} d^3x = \int_{\partial V} T^{\mu i} d\sigma_i \, , \tag{8.106}$$

where $d\sigma_i$ is the surface element of the boundary ∂V of the volume V. If the fall-off as $|\vec{x}| \to \infty$ of the involved fields is sufficiently fast, then the boundary integral vanishes in the limit of infinitely large spheres. Therefore,

$$\int_{\mathbb{R}^3} \partial_i T^{\mu i} d^3x = 0 \, .$$

The second term in equation (8.104) becomes

$$\int_{\mathbb{R}^3} F^\mu{}_\nu \, j^\nu d^3x = \int_{\mathbb{R}^3} \mathcal{F}^\mu d^3x = \frac{d}{dt} \int_{\mathbb{R}^3} \mathcal{P}^\mu d^3x \, . \tag{8.107}$$

Collecting the terms we thus obtain

[17] This is in contrast to the four-current j^μ which is observer-independent.

$$\frac{d}{dt}\int_{\mathbb{R}^3}\Big(T^{\mu 0}+\mathcal{P}^{\mu}\Big)d^3x = 0\,. \tag{8.108}$$

The first conclusion we draw is that the quantity

$$T^{\mu 0} \tag{8.109}$$

is the *four-momentum density of the electromagnetic field* (as seen by the observer X). Consequently, the *energy density* is

$$\begin{aligned} T^{00} &= \frac{1}{4\pi}\left[F^{0\sigma}F^{0}{}_{\sigma} - \frac{1}{4}F_{\sigma\lambda}F^{\sigma\lambda}\eta^{00}\right] = \frac{1}{4\pi}\left[\vec{E}^2 - \frac{1}{2}(\vec{E}^2 - \vec{B}^2)\right] \\ &= \frac{1}{4\pi}\,\frac{1}{2}\Big(\vec{E}^2 + \vec{B}^2\Big)\,, \end{aligned} \tag{8.110}$$

where we have used (8.33). Analogously, the *three-momentum density* is

$$\begin{aligned} T^{i0} &= \frac{1}{4\pi}\left[F^{i\sigma}F^{0}{}_{\sigma} - \frac{1}{4}F_{\sigma\lambda}F^{\sigma\lambda}\eta^{i0}\right] = \frac{1}{4\pi}F^{ij}F^{0}{}_{j} = \frac{1}{4\pi}\,\epsilon^{ijk}B_kE_j \\ &= \frac{1}{4\pi}\Big(\vec{E}\times\vec{B}\Big)^i \end{aligned} \tag{8.111}$$

The vector $\vec{E}\times\vec{B}$ is the well-known *Poynting vector*.

The second conclusion concerns the balance equation (8.108). The sum of the total momentum (in $\mathbb{R}^3$) of the electromagnetic field and the total momentum of the charged matter is constant, i.e., we have conservation of energy and momentum of the system (matter + fields).

The energy density T^{00} and the momentum density T^{i0} are w.r.t. the chosen observer X. Let w^μ denote the four-velocity of this observer, i.e.,

$$w^\mu = \begin{pmatrix} 1 \\ \vec{o} \end{pmatrix}$$

w.r.t. the observer's coordinates. Then $T^{\mu 0}$ can be written in the coordinate-independent manner

$$T^{\mu 0} = -T^{\mu\nu}w_\nu = -T^{\mu}{}_{\nu}\,w^\nu\,. \tag{8.112}$$

In other words, $T^{\mu}{}_{\nu}\,w^\nu$ is the four-momentum density of the electromagnetic field as seen by an observer with four-velocity w^μ; likewise,

$$T^{00} = T^{\mu\nu}w_\mu w_\nu = T_{\mu\nu}w^\mu w^\nu \tag{8.113}$$

is the energy-density seen by this observer.

Finally, let us analyze the balance equation for general volumes in the vacuum case (which corresponds to $j^\mu = 0$, $\mathcal{P}^\mu = 0$). We obtain

$$\frac{d}{dt}\int_V T^{00} d^3x + \int_{\partial V} T^{0i} d\sigma_i = 0 \quad \text{and} \tag{8.114a}$$

$$\frac{d}{dt}\int_V T^{k0} d^3x + \int_{\partial V} T^{ki} d\sigma_i = 0\,. \tag{8.114b}$$

The first equation states that the energy of the electromagnetic field in a given volume is transported through the boundary via the Poynting vector T^{i0}. The second equation formulates the loss of momentum in a volume through its boundary in terms of the 'stress-tensor' T^{ij}.

CHAPTER 9

BEYOND MINKOWSKI

9.1 The equivalence principle

Newtonian gravity

The basis of Newtonian gravity is a gravitational potential Φ. The potential generated by a point particle located at $\vec{z}$ is

$$\Phi(\vec{x}) = -G\,\frac{m_{\rm ag}}{|\vec{x}-\vec{z}|}\,.$$

In this context, $m_{\rm ag}$ is a property of the point particle, which we call its active gravitational mass; a priori it might be different from its (inertial) mass. More generally, the gravitational potential is determined by the Poisson equation

$$\Delta\Phi = 4\pi G\rho_{\rm ag}\,.$$

In this equation, $\rho_{\rm ag}$ is the density of the active gravitational mass of the configuration that generates the gravitational field, and G is the gravitational constant (which takes the value $G = (6.6743 \pm 0.0007) \times 10^{-11}{\rm m}^3{\rm kg}^{-1}{\rm s}^{-2}$). The gravitational field is the vector field

$$\vec{\phi} = -\vec{\nabla}\Phi\,.$$

The force exerted on a (point) particle is proportional to the gravitational field

$$\vec{F} = m_{\rm pg}\vec{\phi}\,,$$

where $m_{\rm pg}$ is a property of the point particle that we call its passive gravitational mass; a priori it might be different from the active gravitational mass and the (inertial) mass. The passive gravitational mass is a measure of the coupling to a gravitational field.

When we invoke Newton's third law of motion we see that the concepts of active and passive gravitational mass coincide. Consider two point particles with gravitational masses $m_{\rm ag}$, $m_{\rm pg}$ and $M_{\rm ag}$, $M_{\rm pg}$, respectively. Then the force exerted by the first particle on the second must equal the force exerted by the second particle on the first, i.e.,

$$-G\,\frac{m_{\rm ag}M_{\rm pg}}{r^2} = -G\,\frac{M_{\rm ag}m_{\rm pg}}{r^2}\,,$$

where r is the distance between the particles. We conclude that

$$\frac{m_{\rm ag}}{m_{\rm pg}} = \frac{M_{\rm ag}}{M_{\rm pg}}\,,$$

which implies that we can set

$$m_{\rm ag} = m_{\rm pg}$$

by adjusting units. In other words, there is but one *gravitational mass* $m_{\rm g}$ that enters the equations, i.e.,

$$\Delta\Phi = 4\pi G\rho_{\rm g}\,, \qquad \vec{F} = m_{\rm g}\vec{\phi}\,,$$

universally for every type of matter.

The motion of a point particle in a gravitational field is described by Newton's second law, i.e., $m_{\rm i}\,\vec{a} = \vec{F}$, or

$$m_{\rm i}\,\ddot{\vec{x}} = m_{\rm g}\vec{\phi}\,. \tag{9.1}$$

Here, $m_{\rm i}$ is the *inertial mass* of the particle. The inertial mass is the mass that determines the particle's inertia, its 'resistance to accelerations'; it appears in the energy and momentum formulas—it is the 'kinematical mass'.

The Galilean *equivalence principle* is a postulate in Newtonian gravity.[1] 'The motion of a test particle in a gravitational field depends solely on its initial

[1] The equivalence principle in Einstein's general theory of relativity is not a postulate but a consequence of the field equations.

conditions (i.e., position and velocity).' Using this principle in combination with equation (9.1) we conclude that the ratio of inertial over gravitational mass is a universal constant, i.e.,

$$\frac{m_{\mathrm{i}}}{m_{\mathrm{g}}} = \mathrm{const}\,.$$

Adjusting units we obtain *equality of inertial and gravitational mass*,

$$m_{\mathrm{i}} = m_{\mathrm{g}}\,. \tag{9.2}$$

There is just one concept of mass m that enters the equations, i.e.,

$$\Delta\Phi = 4\pi G\rho\,, \qquad \vec{F} = m\vec{\phi}\,,$$

and (9.1) becomes

$$\ddot{\vec{x}} = \vec{\phi}\,. \tag{9.3}$$

It is evident that the equivalence principle is in fact equivalent to the equality of inertial and gravitational mass.

The equivalence principle is under constant experimental tests. Loránd Eötvös' experiments that were conducted at the beginning of the last century showed that the relative difference between inertial mass and gravitational mass must be less than 10^{-9}. Modern experiments have lowered that bound to approximately 10^{-12}; a result of 2008 (PRL 100, 041101) shows that

$$\left.\frac{m_{\mathrm{i}}}{m_{\mathrm{g}}}\right|_{\mathrm{Ti}} = \left(1 + (0.3 \pm 1.8) \times 10^{-13}\right) \left.\frac{m_{\mathrm{i}}}{m_{\mathrm{g}}}\right|_{\mathrm{Be}}$$

for titanium and beryllium test masses.

To study the implications of the equivalence principle let us consider a homogeneous gravitational field, i.e.,

$$\vec{\phi} = \vec{g}\,, \tag{9.4}$$

where $\vec{g}$ is a constant vector; we use the abbreviation $g = |\vec{g}|$. This gravitational field corresponds to a (time-independent) potential $\Phi(\vec{x}) = -\vec{g}\vec{x}$; it is a vacuum solution of the Poisson equation.[2] The motion of point particles in the homogeneous gravitational field (9.4) is determined by

$$\ddot{\vec{x}} = \vec{g}\,. \tag{9.5}$$

[2] But $\Phi(\vec{x}) \not\to 0$ as $|\vec{x}| \to \infty$. It is thus a valid viewpoint to say that $\vec{\phi} = \vec{g}$ isn't a real gravitational field at all, because it is not the field of an isolated body.

However, equation (9.5) can be interpreted in a completely equivalent manner as follows. Consider a (Galilean) spacetime in the absence of gravitational fields. Instead of inertial coordinates $\vec{x}$ consider an accelerated frame of reference, i.e.,

$$\vec{x}' = \vec{x} + \frac{1}{2}\, t^2\, \vec{g}\,.$$

W.r.t. these coordinates, Newton's second law (in the absence of gravitational fields), $m\ddot{\vec{x}} = 0$, becomes

$$\ddot{\vec{x}}' = \vec{g}\,. \tag{9.5'}$$

We conclude that the effect of a homogeneous gravitational field corresponds exactly to the effect of an apparent force in a uniformly accelerated frame of reference (where gravitational fields are absent). In fact, it is impossible to determine from experiments in a closed laboratory whether the lab is (at a 'fixed position') in a homogeneous gravitational field or in a state of uniform acceleration.

Remark. Using accelerated coordinates

$$\vec{x}' = \vec{x} - \frac{1}{2}\, t^2\, \vec{g}\,.$$

in a homogeneous gravitational field we find that the gravitational field can be compensated by the acceleration, i.e., $\ddot{\vec{x}}' = 0$. Therefore, it is impossible to determine whether a laboratory is an inertial frame (in the absence of gravitational fields) or freely falling in a homogeneous gravitational field.

This strict 'equivalence of gravitation and acceleration' is restricted to the case of homogeneous gravitational fields. If there are tidal forces (i.e., non-vanishing second derivatives of the potential), then, by considering separate points, it becomes possible to distinguish between effects of the gravitational field and effects from apparent forces in an accelerated frame.

Remark. Particularly, it becomes possible to distinguish between inertial frames (in the absence of gravitational fields) and freely falling frames of reference in a gravitational field.

Example. Consider a sphere consisting of (thousands of) particles. If these particles are at rest w.r.t. some inertial frame in the absence of gravitational fields, then this spherical configuration of particles remains unchanged forever. Now consider the same configuration in free fall in the gravitational field of the Earth; we assume that the particles are initially at rest. The trajectories of the particles are radial, i.e., straight lines meeting at the center of gravity of the Earth. Furthermore, the acceleration of a particle that is closer to the center of the Earth is larger than that of a particle farther away from the center. It is straightforward to see that what is initially a sphere of particles becomes distorted after some time—the sphere becomes a prolate spheroid. Therefore, by considering separate points, it is possible to detect the presence of a gravitational field despite the fact that the configuration is in free fall.

However, ‘equivalence of gravitation and acceleration’ is still approximately true, at least *locally*, i.e., in regions where (and as long as) the gravitational field is approximately constant. Let us make this statement more specific.

Consider

$$\ddot{\vec{x}}(t) = \vec{\phi}\big(t, \vec{x}(t)\big) + \vec{K}\ , \tag{9.6}$$

which models the motion of a particle in a gravitational field $\phi(t, \vec{x})$, where we include the possibility of an additional force.[3] For the majority of thought experiments, the gravitational field is assumed to be time-independent, i.e., $\vec{\phi} = \vec{\phi}(\cancel{t}, \vec{x})$; this is convenient but not necessary. Let $\underline{\vec{x}}(t)$ be a particular solution (obtained, e.g., by prescribing initial conditions) and define, in slight abuse of notation, $\vec{\phi}(t) = \vec{\phi}\big(t, \underline{\vec{x}}(t)\big)$. When we take a trajectory $\vec{x}(t)$ of (9.6) that is sufficiently close to $\underline{\vec{x}}(t)$ at $t = \underline{t}$, then $\vec{\phi}\big(t, \vec{x}(t)\big) \approx \vec{\phi}\big(t, \underline{\vec{x}}(t)\big) = \vec{\phi}(t)$ in the time interval containing $\underline{t}$ where $\vec{x}(t)$ is sufficiently close to $\underline{\vec{x}}(t)$. Accordingly, (9.6) reads

$$\ddot{\vec{x}}(t) \approx \vec{\phi}(t) + \vec{K} \tag{9.7}$$

in this time interval. Considering particles with initial data increasingly close to the initial data of $\underline{\vec{x}}(t)$ at $\underline{t}$, which is $\underline{\vec{x}}(\underline{t})$ and $\dot{\underline{\vec{x}}}(\underline{t})$, equation (9.7) holds for increasingly long times.

Let us define an accelerated frame of reference by

$$\vec{x}' = \vec{x} + \int dt \int dt\ \vec{\phi}(t)\ .$$

We then obtain

$$\ddot{\vec{x}}'(t) = \ddot{\vec{x}}(t) + \vec{\phi}(t)\ .$$

Now consider Newton's second law $\ddot{\vec{x}}(t) = \vec{K}$ in the absence of gravitational fields. When expressed in the accelerated frame we get

$$\ddot{\vec{x}}'(t) = \vec{\phi}(t) + \vec{K}\ . \tag{9.7$'$}$$

[3]For instance, imagine a (small) ball sitting on a table; the table provides the force $\vec{K}$ so that the ball doesn't move. In the general case, $\vec{K}$ depends on time and on the spatial position; we choose to suppress this dependence. Free fall corresponds to $\vec{K} \equiv \vec{o}$.

Comparing (9.7) and (9.7′) we conclude that in a sufficiently small neighborhood of $\underline{\vec{x}}(\underline{t})$, for some time, the effects of a gravitational field are virtually indistinguishable from the effects of an apparent force in an accelerated frame (where gravitation is absent).

Summing up, we observe 'equivalence of gravitation and acceleration' as a local phenomenon; in a small neighborhood of each trajectory, for short times, the gravitational effects can be reinterpreted, at least up to a certain accuracy, as effects of apparent forces in an accelerated frame. The closer the trajectory is followed, the better the 'equivalence'.

Remark. Using accelerated frames we are able to compensate the effects of gravity; these are the freely falling frames of reference. Along each separate trajectory, the gravitational field can be transformed to zero in these coordinates, i.e., a freely falling frame is equivalent to an inertial frame along the trajectory under consideration (and along this trajectory alone). In a (small) neighborhood of this trajectory, this strict equivalence is broken (when tidal forces are present); however, the statement is still true in an approximative sense: Define the freely falling frame of reference by

$$\vec{x}' = \vec{x} - \underline{\vec{x}}(t) ,$$

where $\underline{\vec{x}}(t)$ is a solution of (9.6) with $\vec{K} = \vec{o}$. Then (9.6) yields

$$\ddot{\vec{x}}'(t) = \ddot{\vec{x}}(t) - \ddot{\underline{\vec{x}}}(t) = \vec{\phi}\big(t, \vec{x}(t)\big) - \vec{\phi}\big(t, \underline{\vec{x}}(t)\big) = \vec{\phi}\big(t, \underline{\vec{x}}(t) + \vec{x}'(t)\big) - \vec{\phi}\big(t, \underline{\vec{x}}(t)\big) .$$

For small $\vec{x}'(t)$ the r.h.s. is approximately zero, i.e.,

$$\ddot{\vec{x}}'(t) \approx 0 ,$$

which implies that the freely falling frame in a gravitational field is, in a neighborhood of the trajectory under consideration, equivalent, to zeroth order, to an inertial frame (in the absence of gravitational fields). More specifically we obtain

$$\ddot{x}'_i(t) = \phi_{i,j}\big(t, \underline{\vec{x}}(t)\big)\, x'^j(t) + O(\vec{x}'^2) = -\Phi_{,ij}\big(t, \underline{\vec{x}}(t)\big)\, x'^j(t) + O(\vec{x}'^2) .$$

We conclude this section by addressing an additional aspect of the 'entanglement' between gravitation and acceleration: It is impossible from local observations to separate the effects of a gravitational field from those of an apparent force in an accelerated frame. Let us elaborate. Suppose the existence of a gravitational field

$$\vec{\phi}(\vec{x}) , \tag{9.8a}$$

which we assume to be time-independent for simplicity. The motion of point particles in this gravitational field is determined by

$$\ddot{\vec{x}} = \vec{\phi}(\vec{x}) \,. \tag{9.8b}$$

Consider a different gravitational field

$$\vec{\psi}(t,\vec{x}) = \vec{\phi}\big(\vec{x} + \tfrac{1}{2}\,t^2\,\vec{g}\big) - \vec{g}\,, \tag{9.9a}$$

where $\vec{g}$ is constant. The equation of motion of point particles in this gravitational field is $\ddot{\vec{x}} = \vec{\psi}(t,\vec{x})$.

Instead of inertial coordinates $\vec{x}$ consider an accelerated frame of reference, i.e.,

$$\vec{x}' = \vec{x} + \frac{1}{2}\,t^2\,\vec{g}\,.$$

W.r.t. this accelerated frame of reference, the equation of motion of point particles in the gravitational field $\vec{\psi}$ becomes

$$\ddot{\vec{x}}' = \ddot{\vec{x}} + \vec{g} = \vec{\psi}(t,\vec{x}) + \vec{g} = \vec{\phi}\big(\vec{x} + \tfrac{1}{2}\,t^2\,\vec{g}\big) - \vec{g} + \vec{g}\,,$$

i.e.,

$$\ddot{\vec{x}}' = \vec{\phi}(\vec{x}')\,. \tag{9.9b}$$

This equation is identical to (9.8b).

Let us summarize. The gravitational field (9.8a) yields the equation of motion (9.8b); the gravitational field (9.9a), which corresponds to

$$\vec{\psi}'(\vec{x}') = \vec{\psi}(t,\vec{x}) = \vec{\phi}(\vec{x}') - \vec{g} \tag{9.9a$'$}$$

in accelerated coordinates, leads to the same equation of motion, i.e., equation (9.9b), w.r.t. an accelerated frame.

We conclude that it is impossible to distinguish through local experiments whether scenario (9.8) or (9.9) is the ‘true one’: Do we perform our experiments in an inertial frame of reference, where the gravitational field is $\vec{\phi}$, or in a uniformly accelerated frame (with acceleration $\vec{g}$) where the gravitational field is $\vec{\psi} = \vec{\phi} - \vec{g}$?

Remark. Recall that if and only if $\vec{\phi}$ is constant, i.e., in the case of a homogeneous gravitational field, it is possible to get rid of the gravitational field entirely, i.e., $\vec{\psi} = \vec{o}$.[4] (But note that, in general, $\vec{\phi}$ is not homogeneous.)

[4] It is justified to say that a homogeneous gravitational field is not a real gravitational field, since it disappears through a coordinate transformation.

There is an important exception, however. The gravitational field of an isolated body satisfies 'boundary conditions at infinity', i.e., there are decay conditions; $\Phi \to 0$ as $|\vec{x}| \to \infty$. If the gravitational field is that of an isolated body, then this global piece of information removes the ambiguity between (9.8) or (9.9). The decay condition fixes the arbitrary constant in (9.9a′).

Beyond Newtonian gravity

In the (special) relativistic context, the equivalence principle is rather meaningless as long as we lack a relativistic theory of gravity. 'The motion of a test particle in a gravitational field depends solely on its initial conditions (i.e., position and velocity).' What is a gravitational field in relativity[5]?

Let us postpone this question. Instead, let us simply assume that the relativistic theory of gravity is a theory in which the equivalence principle is implemented in a well-defined manner. Let us further assume that the equivalence principle amounts to 'local equivalence of gravitation and acceleration', which implies that results on accelerated frames of reference will shed light on effects in a relativistic theory of gravity.

The world line of a uniformly accelerated observer[6] is a hyperbola in Minkowski space, cf. (7.31). W.l.o.g. we assume the motion to be in the x-direction of a (fixed) inertial frame of reference (whose coordinates we denote by $\{t, x, y, z\}$). Then

$$\begin{pmatrix} 0 \\ \mathring{x} - \frac{1}{a} \\ \mathring{y} \\ \mathring{z} \end{pmatrix} + \frac{1}{a} \begin{pmatrix} \sinh as \\ \cosh as \\ 0 \\ 0 \end{pmatrix} \quad \text{with } s \in \mathbb{R} \text{ and } \mathring{x}, \mathring{y}, \mathring{z} \text{ arbitrary} \tag{9.10}$$

is a family of uniformly accelerated observers (with proper acceleration a). As seen in (9.4) et seq., a uniformly accelerated frame of reference in Galilean spacetime is equivalent to a time-independent homogeneous gravitational field of Newtonian gravity. It thus seems natural to guess that the family (9.10) of observers, which represents a uniformly accelerated frame of reference in

[5] And what is a test particle? It turns out that the latter question is even trickier than the former.

[6] In special relativity we used to use the word 'observer' as a shorthand for 'inertial observer'; we break with this convention. Henceforth, by 'observer' we simply mean an 'idealized physicist' characterized by a timelike world line; a 'family of observers' is a congruence of timelike world lines that gives rise to local coordinates.

Minkowski spacetime, is equivalent to a time-independent homogeneous gravitational field in a relativistic theory of gravity. In other words, is it possible to reinterpret the world lines (9.10) as the world lines of observers 'sitting at fixed positions' in a gravitational field? Interestingly enough, the answer is *no*.

Consider world lines of the form (9.10) that are infinitesimally separated, i.e.,

$$\frac{1}{a}\begin{pmatrix}\sinh as\\ \cosh as\end{pmatrix} \quad\text{and}\quad \begin{pmatrix}0\\ dx\end{pmatrix}+\frac{1}{a}\begin{pmatrix}\sinh as'\\ \cosh as'\end{pmatrix}\ ; \tag{9.11}$$

we suppress the trivial y- and z-components. The proper distance $d\ell$ is obtained by solving

$$\frac{1}{a}\begin{pmatrix}\sinh as\\ \cosh as\end{pmatrix}+d\ell\begin{pmatrix}\sinh as\\ \cosh as\end{pmatrix}=\begin{pmatrix}0\\ dx\end{pmatrix}+\frac{1}{a}\begin{pmatrix}\sinh as'\\ \cosh as'\end{pmatrix}\ ,$$

because $(\sinh as, \cosh as)^{\mathrm{T}}$ is the spatial frame vector (i.e., the normalized spacelike vector orthogonal to the four-velocity). The result is

$$d\ell = (\cosh as)\, dx\ . \tag{9.12}$$

We infer that the (proper) *distance* between neighboring world lines *is growing* with time. This is despite the fact that the acceleration along the two world lines is identical.

Remark. Imagine two observers on two world lines (9.11) that are separated by some (finite) initial distance. To get a measure of their distance, observer one sends light signals to observer two, who signals back instantaneously upon receipt. Observer one measures the times that pass between emission and reception. Interestingly enough, these times increase over all bounds; in fact, after some time, the signals sent by observer two cannot even reach observer one any longer (which is a simple consequence of the fact that the asymptotes of the two world lines are parallel null lines). Note that this is despite the fact that the two observers experience the same acceleration.

It is clear that the property of increasing distances between observers is inconsistent with an interpretation of the world lines (9.11) as the world lines of observers 'sitting at fixed positions' in a gravitational field. We see that the strict equivalence of acceleration and gravitation (of uniformly accelerated frames and time-independent homogeneous fields) that is present in Newtonian theory is absent (or modified) in relativity.

For our present purposes this does not matter. The core of the equivalence principle, which is the *local* equivalence of gravitation and acceleration, is untouched by these considerations (which are of course of a global nature). Let us be persistent nonetheless; let us find a different family of accelerated observers.

We require a family of uniformly accelerated observers whose pairwise distances remain constant. In analogy with (9.11) consider one particular world line

$$x^\mu(s) = \frac{1}{a}\begin{pmatrix}\sinh as\\ \cosh as\end{pmatrix} \tag{9.13}$$

that corresponds to a uniformly accelerated observer. Note that the four-velocity is $(\cosh as, \sinh as)^{\mathrm{T}}$ while the spatial frame vector is $ax^\mu(s)$. As an alternative to (9.13) we may characterize the world line by the condition

$$x^\mu x_\mu = \frac{1}{a^2} \tag{9.13$'$}$$

(and the vanishing of the y- and z-component). Let $d \in \mathbb{R}$ and consider the set of events whose proper distance to the world line (9.13) is $|d|$, i.e.,

$$\Big\{\bar{x}^\mu(s) = x^\mu(s) + d\,a\,x^\mu(s)\,,\ s\in\mathbb{R}\Big\}\,. \tag{9.14}$$

Then

$$\bar{x}^\mu\bar{x}_\mu = (1+a\,d)^2\,x^\mu x_\mu = \Big(\frac{1+a\,d}{a}\Big)^2\,,$$

hence (9.14) is again a hyperbola and thus represents the motion of an observer with constant acceleration.[7] However, the acceleration differs from the original acceleration a; it is

$$\bar{a} = \frac{a}{1+a\,d}\,. \tag{9.15}$$

Conversely, the pair of world lines

$$x^\mu x_\mu = \frac{1}{a^2}\,, \qquad \bar{x}^\mu\bar{x}_\mu = \frac{1}{\bar{a}^2} \tag{9.16}$$

is equidistant (with $d = 1 - \bar{a}/a$). For two uniformly accelerated observers to retain a constant distance, the accelerations must be different.

Summarizing we see that the family of hyperbolas

$$\big\{x^\mu\,\big|\,x^\mu x_\mu = \text{const} > 0\big\} \tag{9.17}$$

[7] A posteriori we make the restriction $d > -a^{-1}$.

in some timelike 2-plane (e.g., the plane spanned by $\{t, x\}$) represents a family of uniformly accelerated observers (where, however, each observer experiences a different acceleration) such that the pairwise distances between the observers remain constant.

Take a particular world line (9.13). Consider, for $|\sigma| < a^{-2}$, the set

$$\left\{ \bar{x}^\mu(s) = x^\mu(s) + \sigma\, \dot{x}^\mu(s)\,,\ s \in \mathbb{R} \right\}. \tag{9.18}$$

Each event $\bar{x}^\mu(s)$ is separated from $x^\mu(s)$ by a constant amount of proper time σ along $\dot{x}^\mu(s)$. We find

$$\bar{x}^\mu \bar{x}_\mu = x^\mu x_\mu + \sigma^2\, \dot{x}^\mu \dot{x}_\mu = \frac{1}{a^2} - \sigma^2 = \text{const}\,,$$

i.e., the world line $\bar{x}^\mu(s)$ is again a hyperbola. Therefore, two world lines of (9.17) are spatially *and temporally* equidistant.

It is tempting to take a leap of faith and interpret the family of accelerated observers (9.17) as a family of observers, each at a fixed point, in a time-independent gravitational field. A straight line, on the other hand, would then correspond to an inertial observer; we would interpret this observer as being in free fall in the gravitational field. But there is a 'but': The 'gravitational field' so constructed is trivial. In the Newtonian case, a time-independent homogeneous gravitational field uniquely corresponds to a family of accelerated observers (in Galilean space without gravity). It is thus valid to say that this gravitational field is not a proper gravitational field at all. Likewise, in the relativistic case, we cannot construct a real gravitational field by simply considering a family of accelerated observers in Minkowski space. Minkowski space is Minkowski space, irrespective of whether we use inertial coordinates or accelerated coordinates. And Minkowski space is a spacetime where gravity is absent. We will come back to this issue in section 9.4.

However, we may repeat that, for our present purposes, this does not matter. The core of the equivalence principle, which is the *local* equivalence of gravitation and acceleration, is untouched by these considerations (which are of course of a global nature). We have good reasons to believe that a *single* accelerated observer, represented by a particular world line in Minkowski space, is completely equivalent to an observer experiencing accelerations caused by a gravitational field. Furthermore, locally, in a neighborhood of that observer, and for some finite time, the effects stemming from the acceleration of the frame of reference, are *approximately equivalent* to the effects stemming from gravity.

In the subsequent section we use the equivalence principle to gain a number of insights on gravitational effects. These considerations are *local* in nature, so that the equivalence principle is expected to hold (at least approximately).

9.2 Clocks and light in a gravitational field

We do not yet have a relativistic theory of gravity at hand. There is only one property of that theory we expect to hold (and which we thus assume): The equivalence principle, i.e., the local (approximate) equivalence between gravitation and acceleration discussed in section 9.1. Despite this severe restriction we are able to derive a number of results.

Imagine the Piazza del Duomo (Piazza dei Miracoli) in Pisa and Galileo Galilei standing on top of the Leaning Tower and performing free fall experiments. But let us twist history and imagine that Galileo intends to drop . . . clocks. On the lawn at the base of the Leaning Tower, Galileo's assistant, let's call him Albert, who is like Galileo equipped with precise chronometers, is eager and ready to begin with the measurements.

Suppose that the clocks Galileo drops are perfectly synchronized with his own chronometer—a millisecond on the clocks is a millisecond on Galileo's chronometer. The gravitational field of the Earth accelerates a falling clock until it reaches the velocity v at the base of the tower. Shortly before Albert breaks the fall of the clock, he makes time measurements by comparing the clock's time with his chronometer's time. The result is fascinating: The clock and Albert's chronometer are asynchronous and, really and truly, the clock runs faster than the chronometer by a factor of

$$1 + \frac{1}{2}\frac{v^2}{c^2} \,.$$

Galileo is sceptical of the findings his assistant relates to him upon his return at the base of the tower; he does not comprehend the results. But surprisingly, his assistant understands. “This is how it works,” says Albert, “we use the equivalence principle to explain the results.”

Galileo is a stationary observer in a gravitational field; the acceleration experienced by Galileo is g. By the equivalence principle, Galileo can be equivalently regarded as a uniformly accelerated observer (with acceleration g); in Minkowski space this corresponds to a hyperbolic world line with acceleration

g, see section 9.1. Albert and the falling clock are spatially close to Galileo and the experiment is rather short; the condition of locality in the equivalence principle is thus satisfied and we may regard Albert as another uniformly accelerating observer, represented by another hyperbola in Minkowski space. The freely falling clock, on the other hand, is to be identified with an inertial clock by the equivalence principle; free fall in a gravitational field is equivalent to inertial motion, which corresponds to a straight line in Minkowski space. Since the clock is inertial we are permitted to apply our collective special relativistic reasoning to measurements made w.r.t. this clock. At the moment Galileo drops the clock, the relative velocity between Galileo's chronometer and the clock is zero. Hence, obviously, there is no time dilation between the two. However, shortly before Albert catches the clock, the relative velocity between Albert's chronometer and the clock is v. Hence, as seen from the clock's perspective, the chronometer undergoes a time dilation by the factor of

$$\sqrt{1-\frac{v^2}{c^2}} \approx 1-\frac{1}{2}\frac{v^2}{c^2}\ ,$$

see the considerations of section 4.3. Note that the clock is inertial, while the chronometer is not (cf. the twin paradox in section 4.4).

We conclude that gravity influences the course of (proper) time; 'deeper down' in the gravitational field, time runs slower than 'higher up'.

Albert recommends to Galileo to solve an *exercise* to better understand the above: "Make a Minkowski diagram that represents the experiment. And note that the clock's world line is tangent to the hyperbola that represents you."

Galileo is not convinced but suggests a second experiment. He proposes to simply take a clock, which is initially synchronized with the chronometers, throw it up into the air, as high as the tower, and catch it again when it falls back to the ground. No sooner said then done. Galileo throws the clock up into the air and Albert manages to catch it again some seconds later. The readings corroborate Albert's ideas: The clock is ahead of the chronometers by some tiny fractions of a second. Galileo is astounded; he had thought the results of the first experiment to be due to some virtual effect resulting from the fact that the measurements are taken at different locations, on the base and on the top of the tower, and that these measurements should not be directly compared. Now, however, Galileo holds in his hands two clocks that do no longer show the same time despite the fact that had been perfectly synchronized initially. "Piece of cake," says Albert, "just remember the twin paradox and the equivalence principle."

Galileo and Albert correspond to a stationary observer in a gravitational field; the acceleration is g. By the equivalence principle, they can be equivalently regarded as a uniformly accelerated observer that is represented in Minkowski space by a hyperbolic world line. The clock, on the other hand, is freely falling on its entire trajectory; both on its way up and on its way down there are no forces except gravity that act on the clock. Free fall in a gravitational field is equivalent to inertial motion; therefore, the clock corresponds to a straight line, which represents inertial motion, in Minkowski spacetime. Albert correctly concludes that the experimental set-up can be represented, with good accuracy, since the condition of locality is satisfied, by a hyperbola, representing Galileo and Albert, and a secant of this hyperbola, representing the clock. The result then follows straightforwardly from the standard considerations on the twin paradox. The inertial twin, i.e., the clock, is 'older' than the accelerated twin, i.e., Galileo and Albert. The time that has passed on the clock during its round trip is greater than the time that has passed on Galileo's and Albert's chronometers: The clock is ahead of the chronometers by approximately

$$\int_{t_\mathrm{i}}^{t_\mathrm{f}} \left(1 - \sqrt{1 - \frac{v^2(t)}{c^2}}\right) dt \approx \int_{t_\mathrm{i}}^{t_\mathrm{f}} \frac{1}{2} \frac{v^2(t)}{c^2}\, dt\ ,$$

where t_i and t_f are the initial and final time, i.e., throw and catch, respectively; $v(t)$ is the relative velocity; $|v(t_\mathrm{i})| = |v(t_\mathrm{f})|$ and $v = 0$ at the turning point; see section 4.4.

Still hesitant to accept Albert's explanations, Galileo takes matters into his own hands. While he mounts the tower again, he orders Albert to stay at the foot of the tower until he will return. After some hour, Galileo descends again; with some despair he compares his chronometer with Albert's, which had remained at the ground for the entire time. And there it is again, the time difference. While on Galileo's chronometer the time Δt has passed, Albert's chronometer shows that only

$$\left(1 - \frac{gh}{c^2}\right) \Delta t$$

has passed for Albert; h is of course the height of the Leaning Tower. "Did you tamper with your chronometer?" Galileo cries accusingly. But Albert is a physicist of impeccable character. "Let me explain," he says.

Galileo and Albert are observers who take fixed positions in a stationary gravitational field. The equivalence principle tells us that, equivalently, Galileo and Albert are represented by uniformly accelerated observers in Minkowski space, i.e., by two hyperbolic world lines. The condition of locally in the equivalence principle is satisfied, since the spatial separation, i.e., h is small. We take the pair of hyperbolas (9.16) to be Galileo's and Albert's world lines, respectively, where a is replaced by g and the distance d corresponds to the height h. Therefore, we have

$$\begin{aligned} \text{Albert:} \quad & \frac{1}{g}\begin{pmatrix}\sinh g s_{\mathrm{A}} \\ \cosh g s_{\mathrm{A}}\end{pmatrix} \\ \text{Galileo:} \quad & \frac{1}{\bar{g}}\begin{pmatrix}\sinh \bar{g} s_{\mathrm{G}} \\ \cosh \bar{g} s_{\mathrm{G}}\end{pmatrix} , \end{aligned}$$

where s_{A} and s_{G} are Albert's and Galileo's proper time, respectively; furthermore, from (9.15), where we use SI units, we find

$$\bar{g} = \frac{g}{1 + gh/c^2} \approx g\left(1 - \frac{gh}{c^2}\right) .$$

Suppose that Galileo leaves Albert when $s_{\mathrm{A}} = 0$ and $s_{\mathrm{G}} = 0$. To compute the proper times s_{A} and s_{G} that have passed until Galileo's return, we equate the two time components, i.e.,

$$\frac{1}{g}\sinh(g s_{\mathrm{A}}) = \frac{1}{\bar{g}}\sinh(\bar{g} s_{\mathrm{G}}) ;$$

in this context we have neglected the time Galileo needs to ascend the tower and related subtleties. We thus need to solve

$$\sinh(\bar{g} s_{\mathrm{G}}) \approx \left(1 - \frac{gh}{c^2}\right)\sinh(g s_{\mathrm{A}}) ,$$

which yields

$$\bar{g}\, s_{\mathrm{G}} \approx g\, s_{\mathrm{A}} - \big(\tanh(g s_{\mathrm{A}})\big)\frac{gh}{c^2} \approx g\, s_{\mathrm{A}}$$

and thus

$$\begin{aligned} s_{\mathrm{G}} &\approx s_{\mathrm{A}}\left(1 + \frac{gh}{c^2}\right) , \\ s_{\mathrm{A}} &\approx s_{\mathrm{G}}\left(1 - \frac{gh}{c^2}\right) . \end{aligned}$$

This is in perfect accordance with the result of the experiment.

"You seem sure of yourself, Albert," Galileo observes, "but, according to the first experiment, should not the relation

$$s_{\mathrm{A}} \approx s_{\mathrm{G}}\sqrt{1 - \frac{v^2}{c^2}}$$

hold? Don't you remember the time dilation factor you computed with the clock's terminal velocity v?" "Sure, I do. Obviously, we have

$$v = gt \quad \text{and} \quad h = \tfrac{1}{2}\, gt^2 \,.$$

It follows that

$$v = \sqrt{2gh}$$

and we obtain

$$\sqrt{1 - \frac{v^2}{c^2}} \approx 1 - \frac{1}{2}\frac{v^2}{c^2} = 1 - \frac{1}{2}\frac{2gh}{c^2} = 1 - \frac{gh}{c^2} \,.$$

Quod erat demonstrandum."

Finally, Galileo is convinced: "Time in a gravitational field is relative! But I'm beginning to wonder whether we couldn't have obtained similar results by using light. Come Albert, let us try." Galileo procures a light source emitting light at a precisely specified frequency and mounts the tower again. From the top of the tower Galileo signals downward to Albert, who measures the frequency of the light beam he receives. Interestingly enough, the frequency Albert measures is larger by a factor of

$$1 + \frac{gh}{c^2}$$

than Galileo's frequency. A short debate between Albert and Galileo ensues, whereupon Galileo announces: "I've come to terms with the equivalence principle. So this time, I'll give the explanation myself."

By the equivalence principle, Galileo is represented by a hyperbolic world line in Minkowski space and Albert by a nearby one. The beam of light, on the other hand, is in 'free fall' and thus corresponds to a null line in Minkowski space intersecting Galileo's world line (at the time of emission) and Albert's world line (at the time of receipt). The time the light takes to reach the ground is h/c. Albert experiences the acceleration g during this time; hence, in the Minkowski picture, he has reached the velocity

$$v = \frac{gh}{c}$$

when the beam of light hits him. In this picture we may simply apply formula (5.7) of section 5.1, i.e., we regard the frequency shift as being due to the (longitudinal) Doppler effect. The frequency $\nu_{\rm A}$ that Albert measures is

$$\nu_{\rm A} = \sqrt{\frac{1 + v/c}{1 - v/c}}\; \nu_{\rm G} \approx \sqrt{\Big(1 + \frac{v}{c}\Big)\Big(1 + \frac{v}{c}\Big)}\; \nu_{\rm G} \approx \Big(1 + \frac{v}{c}\Big)\, \nu_{\rm G} \,.$$

Inserting v we obtain

$$\nu_{\mathrm{A}} \approx \left(1 + \frac{gh}{c^2}\right) \nu_{\mathrm{G}} \,,$$

which reproduces the measurement perfectly. Furthermore, this result is in perfect accord with the previous considerations on time. Since

$$s_{\mathrm{A}} \approx \left(1 - \frac{gh}{c^2}\right) s_{\mathrm{G}}$$

we expect that frequencies behave like the reciprocals of s_{A} and s_{G}, i.e.,

$$\nu_{\mathrm{A}} \approx \left(1 - \frac{gh}{c^2}\right)^{-1} \nu_{\mathrm{G}} \approx \left(1 + \frac{gh}{c^2}\right) \nu_{\mathrm{G}} \,.$$

A consistent picture of the *gravitational redshift* effect emerges. Light traveling from ‘deeper down’ to ‘higher up’ in a gravitational field is redshifted; on the opposite path there is a blueshift.

Galileo and Albert decide to pack their gear and return to their humble abode in the poorer quarters of Pisa. On their way home, Albert is deeply immersed in thought. Finally, he asks: “Master Galilei, may I confront you with a Gedanken-experiment of mine? It concerns another property of light, its bending in the gravitational field of the Earth.”

Suppose that the Earth is flat, which implies that the gravitational field is exactly the same (in magnitude and direction) along the surface. On this flat Earth Galileo and Albert stand some hundred meters apart from each other. Galileo sends a beam of light in the direction toward Albert; the initial height of the beam is precisely specified, and, at the point of emission, the beam is exactly parallel to the Earth’s surface. Interestingly enough, the height of the beam, when Albert receives it, is less than the original height by

$$\frac{1}{2}\,\frac{g\,d^2}{c^2}\,,$$

where d is the distance between Galileo and Albert.

It is again the equivalence principle that provides an explanation. Instead of the Earth and its gravitational field, imagine a planar surface that is uniformly accelerating through space, with Galileo and Albert on it. If the distance between the two is d, then the time the beam of light takes on its voyage from Galileo to Albert is

$$t = \frac{d}{c}\,.$$

During this time, Albert is uniformly accelerating with an acceleration g. Therefore, his position w.r.t. the original position at the time of emission of the light ray has changed by

$$\Delta h = \frac{1}{2}\, gt^2 = \frac{1}{2}\, \frac{g\, d^2}{c^2}\ ,$$

as claimed.

It is a little more complicated but instructive to consider the same Gedanken-experiment in the Minkowski space picture. Galileo's world line, the emitter's word line, to be exact, is represented by a hyperbola in Minkowski space; since the set-up requires two spatial dimensions it becomes necessary to picture a three-dimensional Minkowski space spanned by $\langle t, x, y\rangle$. W.l.o.g. we choose the emitter's hyperbolic world line to lie in the $\langle t, x\rangle$ plane and its vertex to coincide with the origin; furthermore, the origin is assumed to be the time of emission. Albert cannot be treated as a point particle; indeed, to measure the beam of light he requires a screen; the beam of light creates a spot on this screen whose height can then easily be measured. The screen corresponds to a family of hyperbolas which are parallel to Galileo's world line; the vertices of these hyperbolas are given by $t = 0$ and $y = d$, while x varies. To see this we simply note that x is the direction of acceleration in the Minkowski picture, which means that x is connected with the height in the gravitational field. (Of course, it is not the height itself; the family of accelerated observers define equal heights; the height is thus the x-coordinate of a world line at its point of intersection at $t = 0$.) The beam of light emitted from the origin is represented by a null line; it is the null line in the $\langle t, y\rangle$ plane (which is tangent to the emitter's world line); this requirement corresponds to the condition that the beam be parallel to the ground initially. The light ray intersects Albert's screen at the point

$$t = \frac{d}{c}\ , \quad x = 0\ , \quad y = d\ .$$

Tracing back this point along its hyperbolic world line to $t = 0$ yields

$$t = 0\ , \quad x = -\frac{g}{2}\frac{d^2}{c^2}\ , \quad y = d\ .$$

Since, at $t = 0$, the Minkowski coordinate x measures the height in the gravitational field, the height of the spot that the beam of light produces on the screen is less, by

$$\frac{g}{2}\frac{d^2}{c^2}\ ,$$

than the original height (since emission was at $t = 0$, $x = 0$). The Minkowski picture thus reproduces, in a more formal way, the previous result.

Galileo approves of Albert's ideas and entrusts him with an *exercise*. "Turn the Gedankenexperiment into an actual experiment; allow for the fact that the surface of the Earth is curved."

A day of exciting experiments comes to a close. Bidding Albert a goodnight, Galileo retires into his chamber. But his sleep is restless. Fragments of formulas appear and vanish in front of his inner eye. The term gh of

$$1 \pm \frac{gh}{c^2}$$

is particularly insistent and irritating; it dances in his mind and cries: "I'm the gravitational potential, haven't you noticed?" And the nightmare continues. Galileo rides on a beam of light; he leaves the Earth's surface and travels higher and higher until he sees the Earth as a blue ball floating in space. "Compute my redshift", the light ray commands. Galileo is willing to obey but he fails. And Albert steps forward from a dark corner of space and says: "Never, never ever, try to push the boundaries of the equivalence principle. Be aware of its limitations. Equivalence of gravitation and acceleration is merely local. Global considerations of this kind are bound to fail." And of out nothing a signpost appears that says: "Realm of General Relativity. Border crossing."

9.3 Metrics

Definition 9.1. *A metric is a sufficiently smooth non-degenerate symmetric tensor field of rank* $(0, 2)$ *on a differentiable manifold. A spacetime is a four-dimensional manifold equipped with a Lorentzian metric.*

In the following we discuss the concepts this definition is based on without going into too much detail.[8] The notion of a *manifold* is a fundamental concept in differential geometry. We avoid a formal definition and simply say that a manifold of dimension n is a 'space' that locally looks like (an open set of) $\mathbb{R}^n$, i.e., the defining property of a manifold of dimension n is to admit

[8] These lecture notes are not the right place to discuss the mathematics underlying general relativity in detail. The reader is referred to the lecture course "Einführung in die Relativitätstheorie und Kosmologie II" instead.

coordinate systems of n real coordinates, at least locally, at every point.[9] A simple example is the space $\mathbb{R}^2$ (or $\mathbb{R}^n$), which is automatically equipped with a (global) coordinate system. Similarly, every vector space and affine space of dimension n is a manifold of dimension n. Another two-dimensional manifold is the sphere S^2; it is not difficult to equip S^2 with coordinates.[10] A similar example is the torus T^2. In fact, every sufficiently well-behaved surface in $\mathbb{R}^3$ (or $\mathbb{R}^4$, like the Klein bottle) is a two-dimensional manifold. More generally, hypersurfaces of dimension n in $\mathbb{R}^{n+1}$ (or $\mathbb{R}^N$, $N > n+1$) are n-dimensional manifolds. It is possible, although probably not advisable, to visualize every n-dimensional manifold as a hypersurface in $\mathbb{R}^N$.

To (intuitively) understand the concept of tensor fields let us begin by considering vector fields. Consider a (fixed) point x of an manifold M and a coordinate system with coordinates $\{x^1, \dots, x^n\}$. Picture the coordinate system as a coordinate grid covering a neighborhood of the point. For a function f defined on M the expression

$$\partial_{x^i} f \,\Big|_x \quad \text{or} \quad \frac{\partial f}{\partial x^i}\Big|_x \tag{9.19}$$

is well-defined; it measures the change of the function f along the grid line denoted by x^i at the point x; note that the coordinate line x^i is characterized by $x^j = \text{const}$ for all j different from i. The expression (9.19) is the directional derivative of the function f along the coordinate line x^i. We interpret the derivative

$$\partial_{x^i}\Big|_x \quad \text{or} \quad \frac{\partial}{\partial x^i}\Big|_x \tag{9.20}$$

as a vector at the point $x \in M$. A coordinate grid defined by n coordinates gives rise to n linearly independent vectors

$$\frac{\partial}{\partial x^1}\Big|_x, \frac{\partial}{\partial x^2}\Big|_x, \frac{\partial}{\partial x^3}\Big|_x, \dots, \frac{\partial}{\partial x^n}\Big|_x .$$

Using these vectors as the basis we find that an arbitrary vector at $x \in M$ reads

$$v(x) = v^i(x)\,\frac{\partial}{\partial x^i}\Big|_x .$$

Using the language of differential geometry we define vectors as derivations on smooth functions.

[9] Consider a set M. A coordinate system—or chart— is simply an injective map φ of an open subset of $\mathbb{R}^n$ into M. The existence of a family of compatible charts ('atlas'), where compatibility of two charts φ_1, φ_2 means that $\varphi_2^{-1} \circ \varphi_1$ is a diffeomorphism on an open set, makes M a manifold of dimension n.

[10] Note, however, that at least two coordinate patches are necessary to cover the entire sphere S^2.

Example. A simple example is the two-dimensional plane. Choosing an origin and a basis $\{e_1, e_2\}$ we obtain a coordinate system; the coordinate grid consists of straight lines parallel to e_1 and e_2. Therefore,

$$\frac{\partial}{\partial x^1}\Big|_x = e_1 \qquad \frac{\partial}{\partial x^2}\Big|_x = e_2$$

for every point x.

Example. Consider the manifold S^2 as embedded into $\mathbb{R}^3$. A vector $v(x)$ at a point $x \in S^2$ is a tangent vector to the sphere; it is a vector in $\mathbb{R}^3$ but not, in any way, an element of S^2. However, by considering coordinate grids and directional derivatives we can give these tangent vectors an intrinsic characterization that does not make use of the ambient vector space $\mathbb{R}^3$.

Remark. The use of 'directional derivatives' as (basis) vectors is not merely a mathematical quirk. In general there does not exist any ambient vector space that simplifies matters, which in turn compels us to use concepts that are defined intrinsically. Note that the universe is a four-dimensional manifold that is not embedded in a five- or higher dimensional vector space.

When x is not fixed but regarded as varying on the manifold we obtain a *vector field.* Hence a vector field on a manifold M corresponds to a vector at each point, i.e.,

$$v = v^i \frac{\partial}{\partial x^i}\,, \tag{9.21}$$

where v^i depends on the position on M.

The basis vector fields

$$\frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, \frac{\partial}{\partial x^3}, \ldots, \frac{\partial}{\partial x^n}\,. \tag{9.22}$$

form a so-called *coordinate frame.*

Remark. It is important to note that a particular coordinate frame is in general not defined globally on the manifold—it is clear that its domain is the domain of the coordinate system (i.e., the chart).[11]

The representation of a vector field depends on the chosen coordinate system. A *change of coordinates* induces a change of basis. Let $\{x^1, \ldots, x^n\}$ be a coordinate system and $\{\bar{x}^1, \ldots, \bar{x}^n\}$ be a different coordinate system. Then

$$\frac{\partial}{\partial x^i} = \frac{\partial \bar{x}^j}{\partial x^i}\frac{\partial}{\partial \bar{x}^j} \tag{9.23}$$

[11] A good example is the manifold S^2. Since at least two charts are needed to cover S^2, there are at least two different coordinate frames needed to represent vector fields globally.

by the chain rule. The vector field v reads

$$v^i(x^k)\frac{\partial}{\partial x^i} \quad \text{and} \quad \bar{v}^i(\bar{x}^k)\frac{\partial}{\partial \bar{x}^i}$$

w.r.t. the first and the second coordinate system, respectively. It is straightforward to see that

$$\bar{v}^i(\bar{x}^k) = \frac{\partial \bar{x}^i}{\partial x^j}\, v^j(x^k)\,. \tag{9.24}$$

Compare (9.23) and (9.24) with (A.18) and (A.19) by setting $A^i{}_j = \partial\bar{x}^i/\partial x^j$. The transformation behavior (9.24) of vector fields under a change of coordinates generalizes the transformation behavior of scalar field (A.61).

To define *covector fields* (a.k.a. 1-forms) on a manifold M we use the coordinate coframe

$$dx^1,\, dx^2,\, dx^3,\, \ldots, dx^n\,. \tag{9.25}$$

It generalizes the concept of a dual basis, cf. (A.8′), since we require

$$dx^i\big(\frac{\partial}{\partial x^j}\big) = \delta^i{}_j\,.$$

Hence, at each point, a covector maps vectors to the real numbers. A covector field corresponds to a covector at each point, i.e.,

$$a = a_i dx^i\,, \tag{9.26}$$

where a_i depends on the position on M. A covector field a takes a vector field v and generates a function on M via

$$a(v) = a_i\, dx^i\big(v^j\frac{\partial}{\partial x^j}\big) = a_i v^i\,. \tag{9.27}$$

The transformation of covector fields under a change of coordinates is as intuitive as the transformation of vector fields. We have

$$dx^i = \frac{\partial x^i}{\partial \bar{x}^j}\, d\bar{x}^j \tag{9.28}$$

for the coframe and, from $a_i dx^i = \bar{a}_i d\bar{x}^i$,

$$\bar{a}_i(\bar{x}^k) = \frac{\partial x^j}{\partial \bar{x}^i}\, a_j(x^k)\,. \tag{9.29}$$

This is in complete analogy with (A.20) and (A.22), when we note that

$$\frac{\partial x^i}{\partial \bar{x}^j} = \Big(\frac{\partial \bar{x}^i}{\partial x^j}\Big)^{-1} .$$

The relation (9.27) is independent of the choice of coordinates. To see this explicitly we argue that

$$\bar{a}_i \bar{v}^i = \frac{\partial x^j}{\partial \bar{x}^i}\, a_j\, \frac{\partial \bar{x}^i}{\partial x^k}\, v^k = \frac{\partial x^j}{\partial \bar{x}^i}\, \frac{\partial \bar{x}^i}{\partial x^k}\, a_j v^k = \delta^j{}_k\, a_j v^k = a_j v^j .$$

Tensor fields are obtained via the tensor product of vector and covector fields. For example,

$$T = T^i{}_j{}^{kl}{}_m\, \frac{\partial}{\partial x^i} \otimes dx^j \otimes \frac{\partial}{\partial x^k} \otimes \frac{\partial}{\partial x^l} \otimes dx^m \tag{9.30}$$

is a tensor of rank $(3,2)$. The transformation of tensor fields under a change of coordinates generalizes the transformation of vector and covector fields in the obvious way: Each contravariant (i.e., upper) index follows (9.24), each covariant (i.e., lower) index behaves like (9.29), i.e.,

$$\bar{T}^i{}_j{}^{kl}{}_m\,(\bar{x}^n) = \frac{\partial \bar{x}^i}{\partial x^{i'}}\, \frac{\partial x^{j'}}{\partial \bar{x}^j}\, \frac{\partial \bar{x}^k}{\partial x^{k'}}\, \frac{\partial \bar{x}^l}{\partial x^{l'}}\, \frac{\partial x^{m'}}{\partial \bar{x}^m}\; T^{i'}{}_{j'}{}^{k'l'}{}_{m'}\,(x^n) . \tag{9.31}$$

The **metric** is a tensor field of rank $(0,2)$, i.e.,[12]

$$g = g_{ij}\, dx^i \otimes dx^j ; \tag{9.32}$$

the components are functions. *Symmetry* means that

$$g_{ij} = g_{ji} . \tag{9.33}$$

Non-degeneracy means that the inverse metric, which we denote by g^{ij}, exists, i.e.,

$$g^{ij} g_{jk} = \delta^i{}_k . \tag{9.34}$$

It is common to call a metric a *line element* ds^2. An equivalent notation for (9.32) is

$$ds^2 = g_{ij}\, dx^i \otimes dx^j . \tag{9.32$'$}$$

[12] When we go over to four-dimensional manifolds we typically use Greek indices, i.e.,

$$g = g_{\mu\nu}\, dx^\mu \otimes dx^\nu .$$

It is customary to omit the tensor product signs, i.e., to write

$$ds^2 = g_{ij}\, dx^i\, dx^j \; . \tag{9.32$''$}$$

A metric generalizes the concept of a non-degenerate bilinear form, see appendix A.3; it is a field of non-degenerate bilinear forms (scalar products or pseudo-scalar products), i.e., a non-degenerate bilinear form at each point of the manifold. Therefore, a metric functions exactly like a (pseudo-)scalar product, at each point of the manifold:

Metrics measure 'lengths' of vectors. The 'squared norm function' of a vector field v, i.e.,

$$g(v, v) = g_{ij} v^i v^j \; ,$$

is a function on $x \in M$. At the point $x \in M$, this expression defines the squared norm of the vector $v(x)$. Writing out the coordinate dependence explicitly, where x is represented by the coordinates $(x^1, x^2, \dots, x^n)$, we have

$$g(v, v)|_x = g_{ij}(x^k)\, v^i(x^k)\, v^j(x^k) \; .$$

If $g(v, v)|_x$ is positive for all $v(x) \neq 0$ and at each point x, which corresponds to positive definiteness at each point, then the metric is called *Riemannian*; below we discuss a prominent example. Likewise, the angle between the two vectors $v(x)$, $w(x)$ at the point $x \in M$ is defined by using $g(v, w)|_x$ in the obvious way.

Metrics are not necessarily Riemannian. If the signature of $g|_x$, at each point x, is $(-+++)$, then the metric is called *Lorentzian*. This means that it is possible to choose, at each point $x \in M$, a basis such that the components g_{ij} (or rather $g_{\mu\nu}$) of the metric form the diagonal matrix $\mathrm{diag}(-1, 1, 1, 1)$. Let us reiterate: It is possible to bring the metric to the standard Minkowski form, but merely at each point $x \in M$ *separately*. There does not exist a coordinate system and an associated coordinate frame such that $g = \eta_{ij} dx^i dx^j$ globally unless the metric is the Minkowski metric itself.[13]

Exercise. Show that the non-degeneracy of a metric, see (9.34), can be characterized alternatively as in (A.25c).

Metrics raise and lower indices; e.g., a vector field $v = v^i \partial_i$ becomes a covector field with components

$$v_i = g_{ij} v^j \; .$$

[13] If we achieve $g = g_{ij} dx^i dx^j$ to fulfill $g_{ij}|_x = \eta_{ij}$ at a point x, then $g_{ij}|_y \neq \eta_{ij}$ for almost all points in a neighborhood of x. The diagonal form can not be achieved on open sets.

Exercise. Show that the inverse metric g^{ij} is obtained by raising the two indices of the metric g_{ij}.

Let us discuss a prominent Riemannian metric, the *standard metric on the unit sphere* S^2. In a first step we take the standard metric[14]

$$dx^2 + dy^2$$

on $\mathbb{R}^2$ and express it in polar coordinates, i.e., $x = \rho\cos\varphi$, $y = \rho\sin\varphi$. Strictly speaking, the polar coordinates form a coordinate system (chart) on $\mathbb{R}^2\backslash\{0\}$ and not on the entire space $\mathbb{R}^2$. Therefore, the coordinate basis $\{\partial_\rho, \partial_\varphi\}$ and the coframe $\{d\rho, d\varphi\}$ are well-defined[15] on $\mathbb{R}^2\backslash\{0\}$ but not on the entire space $\mathbb{R}^2$. Since

$$dx = (d\rho)\cos\varphi - \rho\,(\sin\varphi)d\varphi\,, \quad dy = (d\rho)\sin\varphi + \rho\,(\cos\varphi)d\varphi\,,$$

we obtain

$$\begin{aligned} dx^2 + dy^2 &= \Big((\cos\varphi)d\rho - \rho\,(\sin\varphi)d\varphi\Big)^2 + \Big((\sin\varphi)d\rho + \rho\,(\cos\varphi)d\varphi\Big)^2 \\ &= d\rho^2 + \rho^2 d\varphi^2\,, \end{aligned} \tag{9.35}$$

which thus is the standard Euclidean metric in polar coordinates.

Exercise. Use $d\rho^2 + \rho^2 d\varphi^2$ to compute the length of the vector field ∂_φ in dependence on the position. What is the angle between ∂_ρ and ∂_φ? What is the length of the vector field $\rho^{-1}\partial_\varphi$? And what about the vector field $\rho\partial_\rho + \partial_\varphi$?

In the second step we take the standard metric

$$dx^2 + dy^2 + dz^2$$

on $\mathbb{R}^3$ and express it in spherical coordinates, i.e.,

$$x = r\sin\vartheta\,\cos\varphi\,, \quad y = r\sin\vartheta\,\sin\varphi\,, \quad z = r\cos\vartheta\,.$$

We first note that

$$x = \rho\cos\varphi\,, \quad y = \rho\sin\varphi\,, \quad z = r\cos\vartheta$$

[14] Note that $dx^2 + dy^2 = \delta_{ij}dx^i dx^j$ with $i, j = 1, 2$.

[15] There is one more subtlety connected with the periodic nature of φ; recall that $\varphi = 0$ and $\varphi = 2\pi$ are identified. We ignore this issue here.

with $\rho = r\sin\vartheta$. Using $d\rho = \sin\vartheta\, dr + r\cos\vartheta\, d\vartheta$ we obtain, from (9.35),

$$dx^2 + dy^2 = d\rho^2 + \rho^2 d\varphi^2 = \Big(\sin\vartheta\, dr + r\cos\vartheta\, d\vartheta\Big)^2 + r^2\sin^2\vartheta\, d\varphi^2\ ,$$

and thus

$$\begin{aligned} dx^2 + dy^2 + dz^2 &= \Big(\sin\vartheta\, dr + r\cos\vartheta\, d\vartheta\Big)^2 \\ &\qquad + r^2\sin^2\vartheta\, d\varphi^2 + \Big(\cos\vartheta\, dr - r\sin\vartheta\, d\vartheta\Big)^2 \\ &= dr^2 + r^2 d\vartheta^2 + r^2\sin^2\vartheta\, d\varphi^2\ . \end{aligned} \tag{9.36}$$

Setting $r =$ const in (9.36) yields the standard metric on a sphere of radius r, i.e., $r^2 d\vartheta^2 + r^2\sin^2\vartheta\, d\varphi^2$; setting $r = 1$ yields the standard metric on the unit sphere S^2, i.e.,

$$g_{S^2} = d\vartheta^2 + \sin^2\vartheta\, d\varphi^2\ . \tag{9.37}$$

The metric (9.37) allows the computation of lengths and angles of vectors on the sphere; these computations are purely intrinsic because the quantities employed are intrinsic to the sphere (like the metric (9.37) and vectors); the ambient vector space $\mathbb{R}^3$ does not enter these computations.

In the next section we turn our attention to the most basic example of Lorentzian metrics: The Minkowski metric itself.

9.4 The Minkowski metric in accelerated coordinates

The simplest Lorentzian metric is the Minkowski metric

$$\eta = -dt^2 + (dx^1)^2 + (dx^2)^2 + (dx^3)^2\ .$$

Recall that $dt^2 = dt \otimes dt$, $(dx^1)^2 = dx^1 \otimes dx^1$, etc. In the inertial coordinates $\{t, x^1, x^2.x^3\}$ the components of the Minkowski metric are constant.

Consider the metric

$$-2du\, dv + dy^2 + dz^2 \tag{9.38}$$

for $(u, v, y, z) \in \mathbb{R}^4$. The assertion is that this metric is again the Minkowski metric η, but in so-called 'double null' instead of inertial coordinates. The proof is straightforward. The components of the metric are

$$\begin{pmatrix} 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} ,$$

in particular the components are constant and do not depend on the position. Therefore we are able to go over to orthonormal coordinates; we refer to section A.4 for further details. We find that

$$u = \frac{1}{\sqrt{2}}\,(t+x)\,, \qquad v = \frac{1}{\sqrt{2}}\,(t-x) \tag{9.39}$$

yields the desired result. Indeed,

$$\begin{aligned} -2du\,dv + dy^2 + dz^2 &= -2\,\frac{1}{\sqrt{2}}\,(dt+dx)\frac{1}{\sqrt{2}}\,(dt-dx) + dy^2 + dz^2 \\ &= -dt^2 + dx^2 + dy^2 + dz^2\,. \end{aligned}$$

The coordinates (u, v, y, z) are called 'double null' for the reason that the coordinate lines defined by u and v are null lines. Clearly,

$$\eta(\partial_u, \partial_u) = 0\,, \qquad \eta(\partial_v, \partial_v) = 0\,,$$

which is obvious from (9.38). Hence, in double null coordinates, the light cone of a point (in 2-dimensional Minkowski space) is given by the coordinate lines through that point.

From (9.39) we have

$$\partial_t = \frac{\partial}{\partial t} = \frac{\partial u}{\partial t}\frac{\partial}{\partial u} + \frac{\partial v}{\partial t}\frac{\partial}{\partial v} = \frac{1}{\sqrt{2}}\left(\partial_u + \partial_v\right).$$

Using (9.38) we find that

$$\eta(\partial_t, \partial_t) = -2\,\frac{1}{\sqrt{2}}\frac{1}{\sqrt{2}} = -1\,,$$

hence ∂_t is a timelike unit vector, as expected.

Let us turn our attention to the metric

$$ds^2 = -x'^2\,dt'^2 + dx'^2 + dy'^2 + dz'^2\,, \tag{9.40}$$

where $(t', x', y', z') \in \mathbb{R}^4$ with $x' > 0$. The claim is that (9.40) is again the *Minkowski metric*, represented in an accelerated frame of reference.

To establish the claim consider the coordinate transformation

$$t' = \operatorname{artanh} \frac{t}{x}\,, \qquad x' = \sqrt{-t^2 + x^2}\,, \tag{9.41}$$

and $y' = y$, $z' = z$. The coordinates $\{t, x, y, z\}$ will then turn out to be standard inertial coordinates. We have

$$\left(\frac{\partial x'^\mu}{\partial x^\sigma}\right)_{\mu,\sigma} = \begin{pmatrix} \frac{\partial t'}{\partial t} & \frac{\partial t'}{\partial x} \\ \frac{\partial x'}{\partial t} & \frac{\partial x'}{\partial x} \end{pmatrix} = \begin{pmatrix} \frac{x}{-t^2+x^2} & -\frac{t}{-t^2+x^2} \\ \frac{-t}{\sqrt{-t^2+x^2}} & \frac{x}{\sqrt{-t^2+x^2}} \end{pmatrix} \tag{9.42}$$

and therefore

$$\begin{aligned} ds^2 &= -(-t^2 + x^2)\left(\frac{\partial t'}{\partial t}\, dt + \frac{\partial t'}{\partial x}\, dx\right)^2 + \left(\frac{\partial x'}{\partial t}\, dt + \frac{\partial x'}{\partial x}\, dx\right)^2 + dy^2 + dz^2 \\ &= \left[-(-t^2 + x^2)\left(\frac{\partial t'}{\partial t}\right)^2 + \left(\frac{\partial x'}{\partial t}\right)^2\right] dt^2 \\ &\qquad + \left[-2(-t^2 + x^2)\left(\frac{\partial t'}{\partial t}\frac{\partial t'}{\partial x}\right) + 2\left(\frac{\partial x'}{\partial t}\frac{\partial x'}{\partial x}\right)\right] dt\, dx \\ &\qquad\quad + \left[-(-t^2 + x^2)\left(\frac{\partial t'}{\partial x}\right)^2 + \left(\frac{\partial x'}{\partial x}\right)^2\right] dx^2 + dy^2 + dz^2 \\ &= -dt^2 + dx^2 + dy^2 + dz^2\,, \end{aligned}$$

hence (9.40) is indeed the Minkowski metric as claimed.

The coordinates (9.41) are called 'Rindler coordinates'. These coordinates do not cover the entire Minkowski space but merely a part of it, namely, the Cartesian product of the set of events that are spacelike separated from the origin in the $\langle t, x\rangle$ plane times the $\langle y, z\rangle$ plane. A 'Rindler observer' is an observer with world line $x' = \text{const}$, $y' = \text{const}$, $z' = \text{const}$; note that the tangent $\partial_{t'}$ is timelike as required. In inertial coordinates, a Rindler observer is represented by a hyperbolic world line, hence it is a uniformly accelerated observer.

The Minkowski metric in Rindler coordinates is useful in several contexts. Due to its association with uniformly accelerated observers there is a close connection with the equivalence principle. E.g., it is possible (and instructive) to reinvestigate the results of section 9.2 with the aid of (9.40).

An important lesson to learn from the considerations of this section is that it might to difficult to tell whether a metric is truly different from a metric that is already well-known or just a well-known metric 'in disguise', i.e., represented in unusual coordinates. At least in the case of the Minkowski metric it turns out that there is a good criterion, the vanishing or non-vanishing of the Riemann curvature tensor associated with the metric.[16] In anticipation of things to come, we conclude this section with the vague remark that, in general relativity, gravity is represented by metrics that are different from the Minkowski metric.

9.5 Geodesics

Consider two points p_1 and p_2 on a manifold with a Riemannian metric. What is the distance between p_1 and p_2?

Well, if c is a (differentiable) curve connecting the two points, then the length of c is

$$\int ds = \int_{\lambda_1}^{\lambda_2} g(w,w)^{1/2} d\lambda = \int_{\lambda_1}^{\lambda_2} \left(g_{ij}\left(x^k(\lambda)\right) \frac{dx^i}{d\lambda}\frac{dx^j}{d\lambda} \right)^{1/2} d\lambda \,, \tag{9.43}$$

where

$$w^i(\lambda) = \frac{dx^i(\lambda)}{d\lambda}$$

is the field of tangent vectors along the curve c, which is parametrized[17] as $\lambda \mapsto x^i(\lambda)$.[18]

Example. We use (9.37) in (9.43) to compute the circumference of a circle of latitude. This circle is represented by the curve $\vartheta = \text{const}$ parametrized by $\varphi \in [0, 2\pi)$. The tangent vector field is ∂_φ; the length of the tangent vectors is

$$g_{S^2}(\partial_\varphi, \partial_\varphi) = \sin^2\vartheta \,.$$

The length of the curve is computed through the path integral, i.e.,

$$s = \int ds = \int_0^{2\pi} \sqrt{g_{S^2}(\partial_\varphi, \partial_\varphi)}\, d\varphi = \int_0^{2\pi} \sqrt{\sin^2\vartheta}\, d\varphi = 2\pi \sin\vartheta \,.$$

[16] The metric is Minkowski in the former case; in the latter it is not.

[17] The choice of parametrization is irrelevant. Simple *exercise*: Show the invariance of (9.43) under reparametrizations.

[18] For simplicity we assume that the entire curve is contained in the domain of one chart with coordinates $(x^1, \dots, x^n)$. If two (or more) charts are necessary to cover the curve we divide the curve into two (or more) parts whose lengths we compute separately by (9.43); adding up we obtain the total length.

In connection with the remarks of section 9.3 we note that our computation is based on purely intrinsic quantities of the sphere (namely the metric and the tangent vector field of the curve); the ambient vector space did not enter our considerations.

The distance between two points p_1 and p_2 on a Riemannian manifold we take to be the length of the shortest curve connecting p_1 and p_2 (assuming that the lower bound in length is actually attained). Note that this curve need not be unique; e.g., there exists infinitely many curves of equal (shortest) distance between the north and the south pole of a sphere. In any case, the shortest curve is a (local) extremum of length; we call such a curve a *geodesic*. The notion of a geodesic is the generalization of a straight line in $\mathbb{R}^n$.

Let us turn to spacetimes, i.e., manifolds with Lorentzian metrics (where the signature is $(-+++)$ as usual). A curve is said to be *timelike* if the norm of its tangent is everywhere timelike, i.e., $g(v,v) < 0$; it is *null* if its tangent is everywhere null; it is *spacelike* if its tangent is everywhere spacelike; cf. definition 4.2.

Every spacelike curve has a length defined by its 'arc length' (9.43). The 'arc length' of timelike curves, on the other hand, represents *proper time*. Consider a timelike curve c,

$$\lambda \mapsto x^\mu(\lambda)\,,$$

which connects two events p_1 and p_2 that are represented by $x^\mu(\lambda_1)$ and $x^\mu(\lambda_2)$, respectively. Let

$$w^\mu(\lambda) = \frac{dx^\mu(\lambda)}{d\lambda}$$

denote the tangent vector field along c. The proper time that passes along this curve is

$$s = \int_{\lambda_1}^{\lambda_2} \big(-g(w,w)\big)^{1/2} d\lambda = \int_{\lambda_1}^{\lambda_2} \Big(-g_{\mu\nu}\big(x^\sigma(\lambda)\big)\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}\Big)^{1/2} d\lambda\,. \tag{9.44}$$

This is the straightforward generalization of equation (4.9) that gives proper time along a curve in Minkowski space.[19]

A curve that extremizes (9.44) we call a (timelike) *geodesic*. To derive the condition on a curve of being a geodesic, i.e., the *geodesic equation*, we regard (9.44)

[19] As in (9.43) we assume, for simplicity, that the entire curve can be represented by one set of local coordinates.

as a Lagrangian action and use variational analysis. Let

$$\mathcal{L}(x^\mu, \dot{x}^\mu) = \left(-g_{\mu\nu}(x^\sigma)\,\dot{x}^\mu \dot{x}^\nu\right)^{1/2}\,, \tag{9.45a}$$

where in this context (and in this context alone) we make use of the abbreviation $\dot{x}^\mu = dx^\mu/d\lambda$. (In general, we reserve the dot notation for differentiation w.r.t. proper time.[20]) Varying the action

$$s = \int \mathcal{L}\, d\lambda\,, \tag{9.45b}$$

where we keep the endpoints fixed, we obtain the Euler-Lagrange equations

$$\frac{\partial \mathcal{L}}{\partial x^\mu} - \frac{d}{d\lambda}\frac{\partial \mathcal{L}}{\partial \dot{x}^\mu} = 0\,. \tag{9.46}$$

For the first term we find

$$\frac{\partial \mathcal{L}}{\partial x^\mu} = -\frac{1}{2}\frac{1}{\mathcal{L}}\, g_{\sigma\lambda,\mu}\,\dot{x}^\sigma \dot{x}^\lambda\,,$$

the second term is

$$\frac{\partial \mathcal{L}}{\partial \dot{x}^\mu} = -\frac{1}{\mathcal{L}}\, g_{\mu\lambda}\,\dot{x}^\lambda\,,$$

hence the Euler-Lagrange equations read

$$-\frac{1}{2}\frac{1}{\mathcal{L}}\, g_{\sigma\lambda,\mu}\,\dot{x}^\sigma \dot{x}^\lambda + \frac{d}{d\lambda}\Big(\frac{1}{\mathcal{L}}\, g_{\mu\lambda}\,\dot{x}^\lambda\Big) = 0\,. \tag{9.47}$$

We choose to use a parametrization of the curve w.r.t. proper time (which is an 'affine parametrization', cf. the remark on page 177), whence

$$\frac{d}{ds} = \frac{1}{\mathcal{L}}\frac{d}{d\lambda}\,,$$

see (9.45). Accordingly, when we multiply (9.47) with $\mathcal{L}^{-1}$ we obtain

$$-\frac{1}{2}\, g_{\sigma\lambda,\mu}\,\frac{dx^\sigma}{ds}\frac{dx^\lambda}{ds} + \frac{d}{ds}\Big(g_{\mu\lambda}\,\frac{dx^\lambda}{ds}\Big) = 0\,.$$

Since $g_{\mu\lambda} = g_{\mu\lambda}(x^\kappa)$, this further results in

$$g_{\mu\lambda}\,\frac{d^2x^\lambda}{ds^2} + g_{\mu\lambda,\sigma}\,\frac{dx^\sigma}{ds}\frac{dx^\lambda}{ds} - \frac{1}{2}\, g_{\sigma\lambda,\mu}\,\frac{dx^\sigma}{ds}\frac{dx^\lambda}{ds} = 0\,.$$

[20] More generally, we use the dot in connection with affine parameters, see below.

Let us reintroduce the dot notation, but this time according to the standard convention that the overdot refers to differentiation w.r.t. proper time. The equation then becomes

$$g_{\mu\lambda}\,\ddot{x}^\lambda + g_{\mu\lambda,\sigma}\,\dot{x}^\sigma\dot{x}^\lambda - \frac{1}{2}\,g_{\sigma\lambda,\mu}\,\dot{x}^\sigma\dot{x}^\lambda = 0\,.$$

Using the fact that $\dot{x}^\lambda\dot{x}^\sigma$ is symmetric in λ and σ we write the second term as

$$g_{\mu(\lambda,\sigma)}\,\dot{x}^\sigma\dot{x}^\lambda\,.$$

Summarizing,

$$g_{\mu\lambda}\,\ddot{x}^\lambda + \frac{1}{2}\Big(g_{\mu\lambda,\sigma} + g_{\mu\sigma,\lambda} - g_{\lambda\sigma,\mu}\Big)\,\dot{x}^\sigma\dot{x}^\lambda = 0 \tag{9.48}$$

corresponds to the Euler-Lagrange equations. This equation is the condition that the curve extremizes 'arc length' (proper time), i.e., the geodesic equation.

Define

$$\Gamma_{\mu\nu\sigma} = \frac{1}{2}\Big(g_{\mu\nu,\sigma} + g_{\mu\sigma,\nu} - g_{\nu\sigma,\mu}\Big)\,.$$

As an aide memoir one can use the 'curly braces' notation

$$A_{\{ijk\}} := A_{ijk} - A_{jki} + A_{kij}\,;$$

the indices run over the cyclic permutations of (ijk), where positive and negative signs alternate.[21] We thereby obtain

$$\Gamma_{\mu\nu\sigma} = \frac{1}{2}\,g_{\{\mu\nu,\sigma\}} = \frac{1}{2}\,\Big(g_{\mu\nu,\sigma} - g_{\nu\sigma,\mu} + g_{\sigma\mu,\nu}\Big)\,.$$

Expressed in terms of $\Gamma_{\mu\nu\sigma}$, equation (9.48) becomes

$$g_{\mu\nu}\,\ddot{x}^\nu + \Gamma_{\mu\sigma\lambda}\,\dot{x}^\sigma\dot{x}^\lambda = 0 \tag{9.48$'$}$$

We define the **Christoffel symbols** as[22]

$$\Gamma^\mu{}_{\nu\sigma} = g^{\mu\lambda}\Gamma_{\lambda\nu\sigma} = \frac{1}{2}\,g^{\mu\lambda}\,\Big(g_{\lambda\nu,\sigma} + g_{\lambda\sigma,\nu} - g_{\nu\sigma,\lambda}\Big)\,. \tag{9.49}$$

[21] A note of caution: This notation is not particularly common in the literature.

[22] The Christoffel symbols (9.49) represent the so-called Levi-Cività connection associated with the metric g; see also footnote 26.

On multiplying (9.48′) with the inverse of the metric, it becomes the *geodesic equation*

$$\ddot{x}^\mu + \Gamma^\mu{}_{\sigma\lambda}\, \dot{x}^\sigma \dot{x}^\lambda = 0\,. \tag{9.50}$$

Let us summarize what we succeeded in showing: Let c be a timelike curve that is parametrized w.r.t. proper time (or, more generally, affinely parametrized[23]) Then c is a *geodesic* if and only if it satisfies the geodesic equation (9.50).

The case of spacelike geodesic is completely analogous. A spacelike curve that is affinely parametrized (which in the spacelike case means that it is parametrized w.r.t. arc length[24]) is a geodesic if and only if (9.50) holds.

Finally, consider null geodesics. Since the tangent field is a null vector field, the 'arc length' is zero. A straightforward analog of the extremization considerations is thus not available. However, we may simply resort to (9.50). A (parametrized) null curve is called an (affinely parametrized) null geodesic if (9.50) is satisfied.

Remark. Let us explain the terminology 'affine parametrization' for timelike and spacelike geodesics. Suppose that a timelike geodesic is parametrized w.r.t. proper time s. Let $t = \lambda_1 + \lambda_2 s$ with $\lambda_1, \lambda_2 \in \mathbb{R}$. It is common to refer to parameters t of this kind as *affine parameters*, since the transformation $s \mapsto t$ is obviously an affine transformation. The importance of affine parameters lies in the fact that the geodesic equation (9.50) is invariant under affine reparametrizations.

Example. Using an inertial frame of reference in Minkowski space, the Christoffel symbols vanish, i.e.,

$$\Gamma^\mu{}_{\nu\sigma} = 0\,.$$

Therefore the geodesic equation becomes $\ddot{x}^\mu = 0$, which yields the straight lines.

From (9.49) it is immediate that $\Gamma^\mu{}_{\nu\sigma}$ is *symmetric* in $(\nu\sigma)$.[25] It is important to note that *the Christoffel symbols do not form a tensor field.* We will elaborate on the transformation of the Christoffel symbols shortly; but first, another example.

[23] Proper time is a particular choice of 'affine parameter'. We refer to the remark on page 177.
[24] Or a multiple thereof; cf. footnote 23 and the remark on page 177.
[25] Note, however, that this symmetry requires the choice of a coordinate frame.

Example. Consider the (Riemannian) manifold S^2 with the standard metric (9.37). Set $x^1 = \vartheta$ and $x^2 = \varphi$; then

$$\Gamma^i{}_{jk} = \frac{1}{2}\, g^{il} \left(g_{lj,k} + g_{lk,j} - g_{jk,l} \right)$$

leads to

$$\Gamma^1{}_{jk} = \frac{1}{2}\, g^{1l} \left(g_{lj,k} + g_{lk,j} - g_{jk,l} \right) = \frac{1}{2}\, g^{11} \left(g_{1j,k} + g_{1k,j} - g_{jk,1} \right) ,$$
$$\Gamma^2{}_{jk} = \frac{1}{2}\, g^{2l} \left(g_{lj,k} + g_{lk,j} - g_{jk,l} \right) = \frac{1}{2}\, g^{22} \left(g_{2j,k} + g_{2k,j} \right) ,$$

since the metric is diagonal and independent of the second coordinate. We find

$$\begin{aligned}
\Gamma^1{}_{11} &= \frac{1}{2}\, g^{11} \left(g_{11,1} + g_{11,1} - g_{11,1} \right) = 0 \,, \\
\Gamma^1{}_{12} &= \frac{1}{2}\, g^{11} \left(g_{11,2} + g_{12,1} - g_{12,1} \right) = 0 \,, \\
\Gamma^1{}_{22} &= \frac{1}{2}\, g^{11} \left(g_{12,2} + g_{12,2} - g_{22,1} \right) = -\sin\vartheta\, \cos\vartheta \,, \\
\Gamma^2{}_{11} &= \frac{1}{2}\, g^{22} \left(g_{21,1} + g_{21,1} \right) = 0 \,, \\
\Gamma^2{}_{12} &= \frac{1}{2}\, g^{22} \left(g_{21,2} + g_{22,1} \right) = \frac{1}{\sin\vartheta}\, \cos\vartheta \,, \\
\Gamma^2{}_{22} &= \frac{1}{2}\, g^{22} \left(g_{22,2} + g_{22,2} \right) = 0 \,.
\end{aligned}$$

The geodesic equation (9.50) thus reads

$$\begin{aligned}
\ddot{\vartheta} - \sin\vartheta\, \cos\vartheta\, \dot{\varphi}\, \dot{\varphi} &= 0 \,, \\
\ddot{\varphi} + 2\, \frac{\cos\vartheta}{\sin\vartheta}\, \dot{\vartheta}\, \dot{\varphi} &= 0 \,.
\end{aligned} \tag{9.51}$$

Note the factor of 2 in the second equation which is due to the fact that $\Gamma^2{}_{21} = \Gamma^2{}_{12}$. Simple solutions of (9.51) are the circles $\varphi = \text{const}$ and $\vartheta = \pi/2$; for a simplification of the ODEs (9.51) we refer to the exercise course.

Remark. By equation (9.49), the Christoffel symbols can be computed directly from the metric. However, in many cases this is rather time-consuming. An alternative way to obtain the Christoffel symbols is to proceed in analogy with (9.45) et. seq., i.e., to derive the Euler-Lagrange equations of the Lagrangian

$$\mathcal{L}(x^\mu, \dot{x}^\mu) = \left(\pm g_{\mu\nu}(x^\sigma)\, \dot{x}^\mu \dot{x}^\nu \right)^{1/2} .$$

(Note that $\mathcal{L} \equiv \pm 1$ in the affine parametrization w.r.t. proper time or arc length.) As an example consider again the standard metric (9.37) on S^2 which yields

$$\mathcal{L} = \left(\dot{\vartheta}^2 + \sin^2\vartheta\ \dot{\varphi}^2\right)^{1/2} .$$

The Euler-Lagrange equations are

$$\frac{\partial \mathcal{L}}{\partial \vartheta} - \frac{d}{ds}\frac{\partial \mathcal{L}}{\partial \dot{\vartheta}} = 0\,, \qquad \frac{\partial \mathcal{L}}{\partial \varphi} - \frac{d}{ds}\frac{\partial \mathcal{L}}{\partial \dot{\varphi}} = 0\,.$$

Since $\mathcal{L} \equiv 1$ we may replace $\mathcal{L}$ by $\mathcal{L}^2$ in these equations to simplify matters, i.e.,

$$\frac{\partial \mathcal{L}^2}{\partial \vartheta} - \frac{d}{ds}\frac{\partial \mathcal{L}^2}{\partial \dot{\vartheta}} = 0\,, \qquad \frac{\partial \mathcal{L}^2}{\partial \varphi} - \frac{d}{ds}\frac{\partial \mathcal{L}^2}{\partial \dot{\varphi}} = 0\,.$$

We then obtain

$$2\sin\vartheta\ \cos\vartheta\ \dot{\varphi}^2 - 2\frac{d}{ds}\dot{\vartheta} = 0\,, \qquad -2\frac{d}{ds}\Big(\sin^2\vartheta\,\dot{\varphi}\Big) = 0\,,$$

which results in

$$\ddot{\vartheta} - \sin\vartheta\ \cos\vartheta\ \dot{\varphi}^2 = 0\,, \qquad \ddot{\varphi} + 2\,\frac{\cos\vartheta}{\sin\vartheta}\,\dot{\vartheta}\,\dot{\varphi} = 0\,.$$

From these geodesic equations, the Christoffel symbols can be read off.

Christoffel symbols and covariant derivative

By (9.49), the Christoffel symbols are a collection of numbers which we write as an array with three indices. However, *the Christoffel symbols do not form a tensor field*. This becomes explicit when we consider the transformation of the Christoffel symbols under a change of coordinates.

Consider the geodesic equation (9.50). Being the tangent vector field to a curve, $\dot{x}^\mu$ behaves like any other vector field under a change of coordinates, i.e.,

$$\dot{\bar{x}}^\mu = \frac{d\bar{x}^\mu}{ds} = \frac{\partial \bar{x}^\mu}{\partial x^\nu}\frac{dx^\nu}{ds} = \frac{\partial \bar{x}^\mu}{\partial x^\nu}\,\dot{x}^\nu\,, \tag{9.52a}$$

cf. (9.24). The expression

$$\ddot{x}^\mu = \frac{d^2 x^\mu}{ds^2}\,,$$

on the other hand, *is not a vector.* We simply note that

$$\ddot{\bar{x}}^\mu = \frac{d}{ds}\,\frac{d\bar{x}^\mu}{ds} = \frac{d}{ds}\Big(\frac{\partial \bar{x}^\mu}{\partial x^\lambda}\,\dot{x}^\lambda\Big) = \frac{\partial^2 \bar{x}^\mu}{\partial x^\lambda \partial x^\sigma}\,\dot{x}^\sigma \dot{x}^\lambda + \frac{\partial \bar{x}^\mu}{\partial x^\lambda}\,\ddot{x}^\lambda \tag{9.52b}$$

under the change of coordinates $x^\mu \mapsto \bar{x}^\mu = \bar{x}^\mu(x^\sigma)$.

Exercise. In chapter 7 we defined the acceleration as $a^\mu = \ddot{x}^\mu$ and treated it as a four-*vector*. Why is this valid in the context of special relativity? Argue with the aid of (9.52b).

Consider a geodesic c. Expressed w.r.t. the coordinate system $\{x^\mu\}$ this means that c is represented by the (affinely parametrized) curve $s \mapsto x^\mu(s)$ that satisfies the geodesic equation

$$\ddot{x}^\mu + \Gamma^\mu{}_{\sigma\lambda}(x^\kappa)\,\dot{x}^\sigma \dot{x}^\lambda = 0\,, \tag{9.53a}$$

cf. (9.50). Equivalently, we are able to use a different coordinate system $\{\bar{x}^\mu\}$ and represent c as $s \mapsto \bar{x}^\mu(s)$ with

$$\ddot{\bar{x}}^\mu + \bar{\Gamma}^\mu{}_{\sigma\lambda}(\bar{x}^\kappa)\,\dot{\bar{x}}^\sigma \dot{\bar{x}}^\lambda = 0\,. \tag{9.53b}$$

Insertion of (9.52) yields

$$\frac{\partial \bar{x}^\mu}{\partial x^\lambda}\,\ddot{x}^\lambda + \frac{\partial^2 \bar{x}^\mu}{\partial x^\lambda \partial x^\sigma}\,\dot{x}^\sigma \dot{x}^\lambda + \bar{\Gamma}^\mu{}_{\sigma\lambda}(\bar{x}^\kappa)\,\frac{\partial \bar{x}^\sigma}{\partial x^{\sigma'}}\,\frac{\partial \bar{x}^\lambda}{\partial x^{\lambda'}}\,\dot{x}^{\sigma'} \dot{x}^{\lambda'} = 0\,,$$

or, equivalently, on multiplication with $\partial x^{\mu'}/\partial \bar{x}^\mu$,

$$\ddot{x}^{\mu'} + \underbrace{\Big(\frac{\partial x^{\mu'}}{\partial \bar{x}^\mu}\,\frac{\partial^2 \bar{x}^\mu}{\partial x^{\sigma'} \partial x^{\lambda'}} + \bar{\Gamma}^\mu{}_{\sigma\lambda}(\bar{x}^\kappa)\,\frac{\partial x^{\mu'}}{\partial \bar{x}^\mu}\,\frac{\partial \bar{x}^\sigma}{\partial x^{\sigma'}}\,\frac{\partial \bar{x}^\lambda}{\partial x^{\lambda'}}\Big)}_{\Gamma^{\mu'}{}_{\sigma'\lambda'}}\,\dot{x}^{\sigma'} \dot{x}^{\lambda'} = 0\,. \tag{9.54}$$

Comparison with (9.53a) thus entails that

$$\Gamma^{\mu'}{}_{\sigma'\lambda'} = \frac{\partial x^{\mu'}}{\partial \bar{x}^\mu}\,\frac{\partial^2 \bar{x}^\mu}{\partial x^{\sigma'} \partial x^{\lambda'}} + \bar{\Gamma}^\mu{}_{\sigma\lambda}\,\frac{\partial x^{\mu'}}{\partial \bar{x}^\mu}\,\frac{\partial \bar{x}^\sigma}{\partial x^{\sigma'}}\,\frac{\partial \bar{x}^\lambda}{\partial x^{\lambda'}}\,, \tag{9.55a}$$

or, equivalently,

$$\bar{\Gamma}^\mu{}_{\sigma\lambda}(\bar{x}^\kappa) = \frac{\partial \bar{x}^\mu}{\partial x^\nu}\,\frac{\partial^2 x^\nu}{\partial \bar{x}^\sigma \partial \bar{x}^\lambda} + \Gamma^{\mu'}{}_{\sigma'\lambda'}(x^\kappa)\,\frac{\partial \bar{x}^\mu}{\partial x^{\mu'}}\,\frac{\partial x^{\sigma'}}{\partial \bar{x}^\sigma}\,\frac{\partial x^{\lambda'}}{\partial \bar{x}^\lambda}\,. \tag{9.55b}$$

From (9.55) we see that the Christoffel symbols are merely an array of real numbers and not a tensor; the first term in the transformation formula is ‘untensorial’, cf. (9.31).

Exercise. Show the equivalence of (9.55a) and (9.55b). This amounts to proving that

$$-\frac{\partial^2\bar{x}^\mu}{\partial x^\nu \partial x^\sigma}\frac{\partial x^\nu}{\partial \bar{x}^\lambda}\frac{\partial x^\sigma}{\partial \bar{x}^\rho} = \frac{\partial \bar{x}^\mu}{\partial x^\kappa}\frac{\partial^2 x^\kappa}{\partial \bar{x}^\lambda \partial \bar{x}^\rho} .$$

Use that

$$\frac{\partial \bar{x}^\mu}{\partial x^\kappa}\frac{\partial x^\kappa}{\partial \bar{x}^\nu} = \delta^\mu{}_\nu ,$$

and differentiate w.r.t., say, x^π.

Finally, let us briefly introduce the notion of **covariant derivative**.[26] Let $f = f(x^k)$ be a function; then the covariant derivative ∇_i of f is defined to coincide with the regular partial derivative, i.e.,[27]

$$\nabla_i f(x^k) = \partial_i f(x^k) . \tag{9.56a}$$

Let v be a vector field, i.e., $v = v^i(x^k)\partial_i$. Then the covariant derivative of v is a tensor field whose components we write as $\nabla_i v^j$; these are defined to be

$$\nabla_i v^j = \partial_i v^j + \Gamma^j{}_{ik} v^k . \tag{9.56b}$$

Let a be a covector field, i.e., $a = a_i(x^k)dx^i$. Then the covariant derivative is a tensor field whose components we write as $\nabla_i a_j$; these are defined to be

$$\nabla_i a_j = \partial_i a_j - \Gamma^k{}_{ij} a_k . \tag{9.56c}$$

The covariant derivative of tensor fields generalizes the covariant derivative of vector and covector fields in the obvious way: Each contravariant (i.e., upper) index behaves according to (9.56b), each covariant (i.e., lower) index behaves according to (9.56c); e.g.,

$$\nabla_i g_{jk} = \partial_i g_{jk} - \Gamma^l{}_{ij} g_{lk} - \Gamma^l{}_{ik} g_{jl} .$$

Let u be a vector field. The covariant derivative in the direction of u is

$$\nabla_u = u^i \nabla_i .$$

It is immediate from (9.56) that the covariant derivative ∇_u preserves the rank of the tensor: If, e.g., v is a vector, then so is $\nabla_u v$. (The components of $\nabla_u v$ are are $\nabla_u v^i$.) If, e.g., T is a tensor of rank (p, q), then so is $\nabla_u T$.

[26] For the abstract (and beautiful) definition of covariant derivatives in terms of connections we refer to the lecture course "Einführung in die Relativitätstheorie und Kosmologie II".

[27] We use Latin indices for a change to indicate the general nature of these considerations.

Several comments are in order. First, a comment on notation. It is customary to use the 'comma notation' for the partial derivative and the 'semicolon notation' for the covariant derivative, i.e., (9.56) is written as[28]

$$f_{;i} = f_{,i}\,, \qquad v^j{}_{;i} = v^j{}_{,i} + \Gamma^j{}_{ik} v^k\,, \qquad a_{j;i} = a_{j,i} - \Gamma^k{}_{ij} a_k\,. \tag{9.56$'$}$$

Second, a comment on consistency. In (9.27) we have seen that a covector field a acts on vector fields v to produce functions $a(v) = a_i v^i$. Let us compute $\nabla_i\big(a(v)\big)$. On the one hand, since this is a function, we obtain from (9.56a)

$$\nabla_i\big(a_j v^j\big) = \partial_i\big(a_j v^j\big) = a_{j,i} v^j + a_j v^j{}_{,i}\,.$$

On the other hand, we apply (9.56b) and (9.56c) and the Leibniz rule for derivative operators, i.e.,

$$\begin{aligned}\nabla_i\big(a_j v^j\big) &= \big(\nabla_i a_j\big) v^j + a_j\big(\nabla_i v^j\big) = \big(a_{j,i} - \Gamma^k{}_{ij} a_k\big) v^j + a_j\big(v^j{}_{,i} + \Gamma^j{}_{ik} v^k\big) \\ &= a_{j,i} v^j + a_j v^j{}_{,i}\,.\end{aligned}$$

Since this reproduces the original result, we find that the three definitions (9.56) are consistent (i.e., consistent with the Leibniz rule).

Third, the proof of the crucial claim: We show the tensorial character of the covariant derivative. Consider $\nabla_i v^j$; we prove that these are the components

[28] The last equation looks nicer when expressed as $a_{j;i} = a_{j,i} - \Gamma^k{}_{ji} a_k$; we simply use the symmetry of the Christoffel symbols in the two lower indices. Note, however, that this symmetry requires the choice of a coordinate frame.

of a tensor of rank $(1,1)$. We have

$$
\begin{aligned}
\bar{\nabla}_i \bar{v}^j &= \bar{\partial}_i \bar{v}^j + \bar{\Gamma}^j{}_{ik} \bar{v}^k = \frac{\partial}{\partial \bar{x}^i} \left(\frac{\partial \bar{x}^j}{\partial x^k}\, v^k \right) + \bar{\Gamma}^j{}_{ik} \bar{v}^k \\
&= \frac{\partial x^l}{\partial \bar{x}^i} \frac{\partial}{\partial x^l} \left(\frac{\partial \bar{x}^j}{\partial x^k}\, v^k \right) + \bar{\Gamma}^j{}_{ik} \bar{v}^k \\
&= \frac{\partial x^l}{\partial \bar{x}^i} \frac{\partial^2 \bar{x}^j}{\partial x^k \partial x^l}\, v^k + \frac{\partial x^l}{\partial \bar{x}^i} \frac{\partial \bar{x}^j}{\partial x^k} \frac{\partial v^k}{\partial x^l} + \bar{\Gamma}^j{}_{ik} \bar{v}^k \\
&= \frac{\partial x^l}{\partial \bar{x}^i} \frac{\partial^2 \bar{x}^j}{\partial x^k \partial x^l}\, v^k + \frac{\partial x^l}{\partial \bar{x}^i} \frac{\partial \bar{x}^j}{\partial x^k}\, \partial_l v^k + \bar{\Gamma}^j{}_{ik} \frac{\partial \bar{x}^k}{\partial x^l}\, v^l \\
&= \frac{\partial x^l}{\partial \bar{x}^i} \frac{\partial^2 \bar{x}^j}{\partial x^k \partial x^l}\, v^k + \frac{\partial x^l}{\partial \bar{x}^i} \frac{\partial \bar{x}^j}{\partial x^k}\, \partial_l v^k + \\
&\qquad + \left(\frac{\partial \bar{x}^j}{\partial x^{j'}} \frac{\partial x^{i'}}{\partial \bar{x}^i} \frac{\partial x^{k'}}{\partial \bar{x}^k}\, \Gamma^{j'}{}_{i'k'} - \frac{\partial \bar{x}^j}{\partial x^{j'}} \frac{\partial x^{i'}}{\partial \bar{x}^i} \frac{\partial x^{k'}}{\partial \bar{x}^k} \frac{\partial x^{j'}}{\partial \bar{x}^n} \frac{\partial^2 \bar{x}^n}{\partial x^{i'} \partial x^{k'}} \right) \frac{\partial \bar{x}^k}{\partial x^l}\, v^l \\
&= \frac{\partial x^l}{\partial \bar{x}^i} \frac{\partial^2 \bar{x}^j}{\partial x^k \partial x^l}\, v^k + \frac{\partial x^l}{\partial \bar{x}^i} \frac{\partial \bar{x}^j}{\partial x^k}\, \partial_l v^k + \\
&\qquad + \frac{\partial \bar{x}^j}{\partial x^{j'}} \frac{\partial x^{i'}}{\partial \bar{x}^i}\, \delta^{k'}{}_l\, \Gamma^{j'}{}_{i'k'}\, v^l - \frac{\partial x^{i'}}{\partial \bar{x}^i}\, \delta^j{}_n\, \delta^{k'}{}_l \frac{\partial^2 \bar{x}^n}{\partial x^{i'} \partial x^{k'}}\, v^l \\
&= \frac{\partial x^l}{\partial \bar{x}^i} \frac{\partial^2 \bar{x}^j}{\partial x^k \partial x^l}\, v^k + \frac{\partial \bar{x}^j}{\partial x^k} \frac{\partial x^l}{\partial \bar{x}^i}\, \partial_l v^k + \frac{\partial \bar{x}^j}{\partial x^{j'}} \frac{\partial x^{i'}}{\partial \bar{x}^i}\, \Gamma^{j'}{}_{i'l}\, v^l - \frac{\partial x^{i'}}{\partial \bar{x}^i} \frac{\partial^2 \bar{x}^j}{\partial x^{i'} \partial x^l}\, v^l \\
&= \frac{\partial \bar{x}^j}{\partial x^k} \frac{\partial x^l}{\partial \bar{x}^i} \left(\partial_l v^k + \Gamma^k{}_{ln}\, v^n \right) \\
&= \frac{\partial \bar{x}^j}{\partial x^k} \frac{\partial x^l}{\partial \bar{x}^i}\, \nabla_l v^k \,,
\end{aligned}
$$

which is what we intended to show. Here we have used the transformation of vectors, see (9.24), and the transformation rule for the Christoffel symbols in the form (9.55a).

The covariant derivative is intimately connected with the notion of *parallel transport*. Let c be a curve with tangent vector field u. A vector field v is called *parallelly transported* along c if

$$\nabla_u v = 0$$

along this curve. If the curve is represented by $\lambda \mapsto x^\mu(\lambda)$ in local coordinates, then $u^\mu = dx^\mu/d\lambda$ and $\nabla_u v$ reads

$$u^\mu\big(x^\kappa(\lambda)\big)\, \nabla_\mu v^\nu\big(x^\kappa(\lambda)\big)\,,$$

when written out explicitly.

Let c be a timelike curve and $s \mapsto x^\mu(s)$ an affine parametrization (i.e., s is proper time or a multiple thereof). The tangent vector field is

$$u^\mu(s) = \dot{x}^\mu(s) = \frac{dx^\mu(s)}{ds} \,.$$

The covariant derivative of this vector along the curve c itself is

$$\begin{aligned}\nabla_u u^\mu = u^\nu \nabla_\nu u^\mu = u^\nu \partial_\nu u^\mu + u^\nu {\Gamma^\mu}_{\nu\sigma} u^\sigma &= \frac{dx^\nu}{ds}\frac{\partial}{\partial x^\nu}\frac{dx^\mu}{ds} + {\Gamma^\mu}_{\nu\sigma}\dot{x}^\nu \dot{x}^\sigma \\ &= \frac{d^2x^\mu}{ds^2} + {\Gamma^\mu}_{\nu\sigma}\dot{x}^\nu\dot{x}^\sigma = \ddot{x}^\mu + {\Gamma^\mu}_{\nu\sigma}\dot{x}^\nu\dot{x}^\sigma \,.\end{aligned}$$

Comparing with (9.50) we conclude that a curve c is a *geodesic* if and only if its *tangent vector* is *parallelly propagated* along c, i.e., parallel to itself along the curve. The geodesic equation is simply $\nabla_u u = 0$ or

$$\nabla_{\dot{x}}\dot{x} = 0 \,. \tag{9.57}$$

We conclude this section with a property of parallel transport: Parallel transport respects lengths and angles. Let c be a curve and u its tangent field. Suppose that v and w are vector fields that are parallelly propagated along c. Then $g(v,v)$, $g(w,w)$, and $g(v,w)$ are constant along c. To see this we compute, e.g.,

$$\begin{aligned}\nabla_u g(v,w) &= \nabla_u\big(g_{\mu\nu}v^\mu w^\nu\big) = \big(\nabla_u g_{\mu\nu}\big)v^\mu w^\nu + g_{\mu\nu}\Big(\big(\nabla_u v^\mu\big)w^\nu + v^\mu \nabla_u w^\nu\Big) \\ &= \big(\nabla_u g_{\mu\nu}\big)v^\mu w^\nu \,.\end{aligned}$$

We then note that

$$\nabla_\sigma g_{\mu\nu} = 0 \,,$$

which is because

$$\begin{aligned}g_{\mu\nu;\sigma} &= g_{\mu\nu,\sigma} - {\Gamma^\kappa}_{\sigma\mu}g_{\kappa\nu} - {\Gamma^\kappa}_{\sigma\nu}g_{\mu\kappa} = g_{\mu\nu,\sigma} - \Gamma_{\nu\sigma\mu} - \Gamma_{\mu\sigma\nu} \\ &= g_{\mu\nu,\sigma} - \frac{1}{2}\Big(g_{\{\nu\sigma,\mu\}} + g_{\{\mu\sigma,\nu\}}\Big) \\ &= g_{\mu\nu,\sigma} - \frac{1}{2}\Big(g_{\nu\sigma,\mu} - g_{\sigma\mu,\nu} + g_{\mu\nu,\sigma} + g_{\mu\sigma,\nu} - g_{\sigma\nu,\mu} + g_{\nu\mu,\sigma}\Big) = 0 \,.\end{aligned}$$

In other words, the metric is a parallel tensor field, i.e., parallelly propagated along arbitrary curves. Continuing the above argument we find that

$$\nabla_u g(v,w) = \big(\nabla_u g_{\mu\nu}\big)v^\mu w^\nu = 0 \,,$$

from which we conclude that

$$g(v, w) = \text{const}$$

along c. The claim follows.

Remark. As a special case consider a geodesic c. Then the tangent u is by definition parallel along c. If v is another parallel vector field along c, then

$$g(u, v) = \text{const}$$

along c. In particular, $g(u, u) = \text{const}$, as expected, since c is affinely parametrized.

Exercise. If c satisfies the geodesic equation (9.57), then the causal character (timelike/null/spacelike) of c cannot change along the curve. Prove this.

9.6 A scalar relativistic theory of gravity

In section 7.5 we have made an attempt to formulate a relativistic theory of gravity. In this theory the gravitational field is basically represented by a scalar function ('gravitational potential') $V = V(x^\lambda)$ on Minkowski spacetime; we thus speak of a scalar theory of gravity. The potential $V(x^\lambda)$ gives rise to a gravitational field tensor

$$G_{\mu\nu\sigma} = G_{\mu\nu\sigma}(x^\lambda) = V_{,\mu}\,\eta_{\nu\sigma} - \eta_{\mu(\nu}V_{,\sigma)}\,, \tag{9.58}$$

see (7.80), and the motion of a test particle is described by Newton's second law, where the field tensor generates the force, i.e.,

$$m_\mathrm{i}\,\dot{u}^\mu = m_\mathrm{g}\,G^\mu{}_{\nu\sigma}\,u^\nu u^\sigma \tag{9.59}$$

or

$$m_\mathrm{i}\,\ddot{x}^\mu = m_\mathrm{g}\,G^\mu{}_{\nu\sigma}\,\dot{x}^\nu \dot{x}^\sigma\,. \tag{9.59$'$}$$

In complete analogy with the arguments of section 9.1 we find that imposing the equivalence principle leads to equality of inertial and gravitational mass, i.e., $m_\mathrm{i} = m_\mathrm{g}$. Therefore,

$$\ddot{x}^\mu = G^\mu{}_{\nu\sigma}\,\dot{x}^\nu \dot{x}^\sigma \tag{9.60}$$

holds irrespective of the particle's mass. Note that the dot stands for differentiation w.r.t. proper time s, i.e., a solution $s \mapsto x^\mu(s)$ of (9.60) is a world line parametrized by 'arc length'; this is equivalent to the condition $u_\mu u^\mu = -1$ (where $u^\mu = \dot{x}^\mu$).

Remark. To formulate a complete scalar theory of gravity we need a relativistic analog of the Poisson equation for the potential V. A reasonable equation (in the vacuum case) is $\square V = 0$. However, as mentioned in section (7.5), this scalar theory of gravity is falsified by experiments. (The 'correct' theory is general relativity.)

A quick glance at (9.60) reveals its similarity to the geodesic equation (9.50). Note that $G^{\mu}{}_{\nu\sigma}$ is symmetric in $(\nu\sigma)$ as are the Christoffel symbols of a Levi-Cività connection. The Christoffel symbols of which metric? Define the metric

$$g_{\mu\nu} = e^{2V}\,\eta_{\mu\nu}\,. \tag{9.61}$$

The associated Christoffel symbols are

$$\begin{aligned}\Gamma_{\mu\nu\sigma} &= \tfrac{1}{2}\,g_{\{\mu\nu,\sigma\}} = \tfrac{1}{2}\,\left(g_{\mu\nu,\sigma} - g_{\nu\sigma,\mu} + g_{\sigma\mu,\nu}\right) = \tfrac{1}{2}\,\big(2g_{\mu(\nu,\sigma)} - g_{\nu\sigma,\mu}\big)\\ &= -e^{2V}\,\big(V_{,\mu}\eta_{\nu\sigma} - 2\eta_{\mu(\nu}V_{,\sigma)}\big)\,,\end{aligned} \tag{9.62}$$

which closely resemble (9.58), Therefore we conjecture that (9.60) coincides with the geodesic equation in the spacetime with metric $g_{\mu\nu} = e^{2V}\,\eta_{\mu\nu}$.

Let us be precise. Consider a world line in Minkowski space that is parametrized w.r.t. proper time, i.e., $x^{\mu}(s)$. Note that proper time s is the proper time w.r.t. the Minkowski metric $\eta_{\mu\nu}$. Let $\bar{x}^{\mu}(\bar{s})$ denote the *same world line* in a different parametrization, i.e.,

$$x^{\mu}(s) = \bar{x}^{\mu}(\bar{s})\,,$$

where $\bar{s}$ denotes proper time w.r.t. the metric $g_{\mu\nu}$ given by (9.61).

And here's the **claim**: $x^{\mu}(s)$ satisfies the equation (9.60) in Minkowski space if and only if $\bar{x}^{\mu}(\bar{s})$ satisfies the geodesic equation in the spacetime with metric (9.61).

Let us *prove* this claim. Since we use two concepts of proper time in parallel, it is advisable to avoid the use of the overdot. Let s denote proper time w.r.t. $\eta_{\mu\nu}$ and $\bar{s}$ proper time w.r.t. $g_{\mu\nu}$; then

$$d\bar{s} = e^{V}\,ds\,. \tag{9.63}$$

To see this we simply note that $g(v,v) = e^{2V}\,\eta(v,v)$ for every vector v^{μ} and in particular for every timelike vector; if $\mathsf{x}(\lambda)$ denotes an arbitrary parametrization of a timelike world line, then, by (4.8) and (9.44),

$$d\bar{s} = \sqrt{-g\big(d\mathsf{x}/d\lambda, d\mathsf{x}/d\lambda\big)}\,d\lambda = e^{V}\,\sqrt{-\eta\big(d\mathsf{x}/d\lambda, d\mathsf{x}/d\lambda\big)}\,d\lambda = e^{V}\,ds\,.$$

A very intuitive way of deriving (9.63) is to use the line element notation, i.e.,

$$g_{\mu\nu} = e^{2V}\,\eta_{\mu\nu} \quad \Leftrightarrow \quad d\bar{s}^2 = e^{2V}\,ds^2 \,,$$

from which (9.63) can be read off immediately.

We set

$$u^\mu(s) = \frac{d}{ds}\,x^\mu(s) \quad \text{and} \quad \bar{u}^\mu(\bar{s}) = \frac{d}{d\bar{s}}\,\bar{x}^\mu(\bar{s}) \,.$$

Since

$$\frac{d}{d\bar{s}} = e^{-V}\,\frac{d}{ds} \,,$$

we find that

$$\bar{u}^\mu(\bar{s}) = e^{-V(x^\lambda(s))}\,u^\mu(s) \,.$$

Suppressing the arguments we simply write $\bar{u}^\mu = e^{-V}\,u^\mu$. Note that u^μ is normalized w.r.t. $\eta_{\mu\nu}$ while $\bar{u}^\mu$ is normalized w.r.t. $g_{\mu\nu}$, i.e.,

$$\eta_{\mu\nu}u^\mu u^\nu = -1 \,, \qquad g_{\mu\nu}\bar{u}^\mu\bar{u}^\nu = -1 \,.$$

This is consistent with the requirement that the world line be parametrized w.r.t. proper time s and $\bar{s}$, respectively.

Based on these preparations we obtain

$$\frac{d}{d\bar{s}}\,\bar{u}^\mu(\bar{s}) = e^{-V}\,\frac{d}{ds}\,\left[e^{-V}\,u^\mu(s)\right] = e^{-2V}\left[\frac{du^\mu(s)}{ds} - \frac{dV(x^\lambda(s))}{ds}\,u^\mu(s)\right] \,.$$

Furthermore,

$$\frac{dV(x^\lambda(s))}{ds} = \frac{\partial V}{\partial x^\mu}\,\frac{dx^\mu}{ds} = V_{,\mu}\,u^\mu(s) \,,$$

and hence

$$\frac{d}{d\bar{s}}\,\bar{u}^\mu(\bar{s}) = e^{-2V}\left[\frac{du^\mu(s)}{ds} - V_{,\tau}\,u^\tau(s)\,u^\mu(s)\right] \,. \tag{9.64}$$

The equation of motion (9.60) reads

$$\frac{du^\mu}{ds} = G^\mu{}_{\nu\sigma}\,u^\nu u^\sigma \,. \tag{9.65}$$

It is important to note that the indices are raised and lowered with $\eta_{\mu\nu}$ in this context; in particular,

$$G^\mu{}_{\nu\sigma} = \eta^{\mu\rho}G_{\rho\nu\sigma} \,.$$

Let us do the comparison with (9.62). Because

$$G^{\mu}{}_{\nu\sigma} = \eta^{\mu\rho} G_{\rho\nu\sigma} = \eta^{\mu\rho}\left(-e^{-2V}\Gamma_{\rho\nu\sigma} + \eta_{\rho(\nu}V_{,\sigma)}\right) = -e^{-2V}\eta^{\mu\rho}\Gamma_{\rho\nu\sigma} + \delta^{\mu}{}_{(\nu}V_{,\sigma)}\,,$$

and

$$g^{\mu\nu} = e^{-2V}\,\eta^{\mu\nu}\,,$$

where $g^{\mu\nu}$ is the inverse of $g_{\mu\nu}$, i.e., $g^{\mu\nu}g_{\nu\sigma} = \delta^{\mu}{}_{\sigma}$, we obtain

$$G^{\mu}{}_{\nu\sigma} = -g^{\mu\rho}\Gamma_{\rho\nu\sigma} + \delta^{\mu}{}_{(\nu}V_{,\sigma)} = -\Gamma^{\mu}{}_{\nu\sigma} + \delta^{\mu}{}_{(\nu}V_{,\sigma)}\,, \tag{9.66}$$

where the indices on the r.h.s. are raised and lowered with $g_{\mu\nu}$. An equation like (9.66) is extremely 'risky'; it is preferable to write

$$\eta^{\mu\rho}G_{\rho\nu\sigma} = -g^{\mu\rho}\Gamma_{\rho\nu\sigma} + \delta^{\mu}{}_{(\nu}V_{,\sigma)}$$

to avoid confusion about which indices are raised and lowered with which metric.[29]

Inserting (9.65) into (9.64) and using (9.66) we find

$$\begin{aligned}
\frac{d}{d\bar{s}}\,\bar{u}^{\mu}(\bar{s}) &= e^{-2V}\;\left[G^{\mu}{}_{\nu\sigma}\,u^{\nu}(s)u^{\sigma}(s) - V_{,\tau}\,u^{\tau}(s)\,u^{\mu}(s)\right] \\
&= e^{-2V}\;\left[-\Gamma^{\mu}{}_{\nu\sigma}\,u^{\nu}(s)\,u^{\sigma}(s) + \delta^{\mu}{}_{(\nu}V_{,\sigma)}\,u^{\nu}(s)u^{\sigma}(s) - V_{,\tau}\,u^{\tau}(s)\,u^{\mu}(s)\right] \\
&= -\Gamma^{\mu}{}_{\nu\sigma}\,\bar{u}^{\nu}(\bar{s})\,\bar{u}^{\sigma}(\bar{s}) + e^{-2V}\left[\delta^{\mu}{}_{\nu}V_{,\sigma}\,u^{\nu}(s)\,u^{\sigma}(s) - V_{,\tau}\,u^{\tau}(s)\,u^{\mu}(s)\right] \\
&= -\Gamma^{\mu}{}_{\nu\sigma}\,\bar{u}^{\nu}(\bar{s})\,\bar{u}^{\sigma}(\bar{s}) + e^{-2V}\left[u^{\sigma}(s)\,V_{,\sigma}\,u^{\mu}(s) - V_{,\tau}\,u^{\tau}(s)\,u^{\mu}(s)\right] \\
&= -\Gamma^{\mu}{}_{\nu\sigma}\,\bar{u}^{\nu}(\bar{s})\,\bar{u}^{\sigma}(\bar{s})\,.
\end{aligned}$$

This is the geodesic equation for $\bar{x}(\bar{s})$, i.e.,

$$\frac{d}{d\bar{s}}\,\bar{u}^{\mu}(\bar{s}) + \Gamma^{\mu}{}_{\nu\sigma}\,\bar{u}^{\nu}(\bar{s})\bar{u}^{\sigma}(\bar{s}) = 0 \tag{9.67}$$

or

$$\frac{d^2}{d\bar{s}^2}\,\bar{x}^{\mu}(\bar{s}) + \Gamma^{\mu}{}_{\nu\sigma}\,\frac{d\bar{x}^{\nu}(\bar{s})}{d\bar{s}}\frac{d\bar{x}^{\sigma}(\bar{s})}{d\bar{s}} = 0\,. \tag{9.67$'$}$$

This completes the proof of the claim. □

[29] A large number of mistakes in general relativity are due to this kind of confusion.

Let us recapitulate the essence of what we have shown. To formulate a scalar relativistic theory of gravity we consider a potential $V(x^\lambda)$ and a resulting tensor field (9.58) on Minkowski spacetime and use (9.60) as the equation motion of a particle in the gravitational field. However, there is an equivalent viewpoint. The gravitational field is not represented by a field on Minkowski spacetime but by a different spacetime, which is characterized by a metric that differs from the Minkowski metric; in our case, the metric (9.61). The motion of test particles is simply geodesic motion (free fall) w.r.t. this spacetime metric.

Unfortunately, the scalar theory of gravity we have discussed here does not represent our physical reality correctly. But we have come across a fundamental principle:

Gravitation is modeled by spacetime metrics. Minkowski spacetime represents a spacetime where gravitational interaction is absent—the Minkowski metric is flat. Gravity is present in curved spacetimes, which are spacetimes with metrics different from the Minkowski metric.

In these curved spacetimes, the motion of test particles (in the absence of other forces) is geodesic motion, i.e., free fall.

9.7 The equivalence principle

Suppose $(M, g_{\mu\nu})$ is a spacetime, i.e., M a four-dimensional manifold equipped with a Lorentzian metric. The motion of a freely falling test body is geodesic motion; this is irrespective of the composition and the mass of the test body.[30] The equivalence principle is thus automatically implemented. Let $\{z^\mu\}$ denote a local coordinate system[31] ; then the equation of motion is the geodesic equation, i.e.,

$$\nabla_{\dot{z}} \dot{z}^\mu = \ddot{z}^\mu + \Gamma^\mu{}_{\sigma\lambda}\, \dot{z}^\sigma \dot{z}^\lambda = 0\,. \tag{9.68}$$

In sections 9.1 and 9.2 we have taken the local equivalence between gravitation and acceleration to good account. Let us finally formalize this idea.

Take an arbitrary point $p \in M$. In a neighborhood of p we construct a system

[30] We avoid the subtleties connected with the concept of a 'test body'.

[31] We save the notation $\{x^\mu\}$ for the coordinate system of Riemann normal coordinates we construct below.

of coordinates that are called *Riemann normal coordinates*. Let us describe the procedure.[32] Choose an orthonormal basis of vectors at the point p, which we call $e_0|_p, \ldots, e_3|_p$. Choose a quadruple $(x^0, x^1, x^2, x^3) \in \mathbb{R}^4$ and define the vector $x|_p = x^\mu e_\mu|_p$. Solve the geodesic equation (9.68) with initial data $z(0) = p$ and $\dot{z}(0) = x|_p$ (i.e., $\dot{z}^\mu(0) = x^\mu$). The solution $t \mapsto z(t)$ is the unique geodesic passing through p whose tangent vector at p is $x|_p$. Flow along with this geodesic until the affine parameter has reached 1, i.e., follow the geodesic to the point $z(1)$. Finally, assign the point $z(1)$ on the spacetime the coordinates x^μ.

Remark. Suppose that we take $\lambda x|_p$ instead of $x|_p$ (with $\lambda \in \mathbb{R}$) as the initial vector. The geodesic $t \mapsto \bar{z}(t)$ with initial data $\bar{z}(0) = p$ and $\dot{\bar{z}}(0) = \lambda x|_p$ is merely an affine reparametrization of the original geodesic; regarded as geometric curves, the two coincide. Hence the point $\bar{z}(1)$ coincides with $z(\lambda)$.

Exercise. An alternative description of the construction of Riemann normal coordinates is the following: Take a point p and any point q of a neighborhood; let $t \mapsto z(t)$ be the unique geodesic joining p and q, where the parametrization is chosen to satisfy $z(0) = p$ and $z(1) = q$. The coordinates assigned to q are the components of $\dot{z}(0)$ w.r.t. the orthonormal basis $e_0|_p, \ldots, e_3|_p$ at p. Fill in the small gaps in the argumentation.

Using the procedure outlined above we are able to assign coordinates x^μ to each point in a neighborhood of p. The coordinates so constructed are the Riemann normal coordinates. These coordinates have convenient properties. By construction we have

$$\partial_\mu\big|_p = \frac{\partial}{\partial x^\mu}\Big|_p = e_\mu\big|_p \,.$$

Therefore, in these coordinates the metric is $ds^2 = g_{\mu\nu}\, dx^\mu dx^\nu$, where

$$g_{\mu\nu}\big|_p = \eta_{\mu\nu} \,, \tag{9.69}$$

because $g_{\mu\nu} = g(\partial_\mu, \partial_\nu)$. But Riemann normal coordinate achieve much more than (9.69). In these coordinates we have

$$g_{\mu\nu}(x^\sigma) = \eta_{\mu\nu} - \frac{1}{3}\, R_{\mu\sigma\nu\rho}\, x^\sigma x^\rho + \text{terms of third order} \,, \tag{9.70}$$

where $R_{\mu\sigma\nu\rho}$ is the *Riemann tensor*. (We refrain from giving a proof of (9.70).) From (9.70) we obtain (9.69) and

$$g_{\mu\nu,\sigma}\big|_p = 0 \,,$$

[32] To get a good picture of the construction it is useful to imagine a simple Riemannian manifold like the unit sphere instead of a Lorentzian manifold.

which implies that

$$\Gamma^{\mu}{}_{\nu\sigma}\big|_p = 0\,. \tag{9.71}$$

We conclude that, in Riemann normal coordinates in a neighborhood of a point p, the spacetime looks like Minkowski to zeroth and first order. In particular, the equation of motion of test particles (geodesic equation) is $\ddot{x}^{\mu}|_p = 0$ and $\ddot{x}^{\mu} \approx 0$ in a neighborhood of p because of (9.71).

There exists a slightly more difficult construction of coordinates in a neighborhood of a given timelike geodesic (instead of a point); these are the Fermi-Walker coordinates. The final statements are similar: In these coordinate, along the geodesic, we have

$$g_{\mu\nu} = \eta_{\mu\nu} \quad \text{and} \quad \Gamma^{\mu}{}_{\nu\sigma} = 0\,.$$

Therefore, in a sufficiently small neighborhood of the geodesic, the spacetime looks like Minkowski space. This is the mathematical formulation of the *equivalence principle*. We have used a freely falling frame of reference to establish local approximate equivalence of the curved spacetime and flat Minkowski space.

APPENDIX A

MATHEMATICAL BACKGROUND

A.1 Vector spaces and dual spaces

Conventions

Let us consider an abstract vector space V of dimension $\dim V = n$ over the field of real numbers $\mathbb{R}$. When we choose a basis

$$\{e_1, e_2, \ldots, e_n\}$$

of V, then each vector $v \in V$ admits a unique decomposition with respect to this basis, or, in other words, v can be written as a unique linear combination of the basis vectors. In this way, we assign to each vector v an n-tuple of real numbers which are called the components of the vector.[1]

To 'formalize' this we make use of *conventions* that are customary in several branches of mathematics and theoretical physics, and in particular in relativity:

- The components of a vector are denoted by *upper indices* (contravariant index notation).

[1] Thereby, the abstract vector space V becomes isomorphic to $\mathbb{R}^n$. Note, however, that this isomorphism is not canonical (which means 'not unique' in this context).

- In expressions that contain a particular index *twice*, once as a *subscript*, once as a *superscript*, it is implied that we sum over all possible values of that index (Einstein summation convention).

The decomposition of a vector v can thus be written as

$$v = v^i e_i \,. \tag{A.1}$$

The Einstein summation convention implies that summation over i is understood; equation (A.1) is merely a space saving way of writing $v = \sum_{i=1}^{n} v^i e_i$ or $v = v^1 e_1 + v^2 e_2 + \cdots + v^n e_n$. The real numbers $v^1, v^2, \ldots, v^n$ are the components of the vector v.

Remark. In several instances Greek indices are used instead of Latin indices. In relativity, the use of Greek indices is standard[2] in the case of *four dimensions*, i.e., $n = 4$. For example, we would write

$$v = v^\mu e_\mu \,. \tag{A.2}$$

In this context, Greek indices are assumed to run from *zero to three*, i.e., the basis is $\{e_0, e_1, e_2, e_3\}$, the components of the vector are v^0, v^1, v^2, v^3, and (A.2) means $v = v^0 e_0 + v^1 e_1 + v^2 e_2 + v^3 e_3$. (Note again that there's nothing deep in this—we are merely discussing conventions of writing up things.)

It is common not to distinguish between the abstract vector v and the collection of its components w.r.t. the chosen basis[3]; we therefore typically write

$$v = \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix} \,. \tag{A.3}$$

It is important to note that (A.3) tacitly assumes that a basis has been selected; *otherwise* (A.3) *does not make sense.*

Rather confusing, at first sight at least, is the *abstract index notation.* Adopting this notation (A.3) becomes

$$v^i = \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix} \,. \tag{A.4}$$

[2] Unfortunately not all relativists follows this convention—some stick to Latin indices.

[3] This is true when we consider only one (fixed) basis. However, when we deal with two different bases at the same time, a distinction is useful; see (A.31) below.

Here, v^i does not denote the i^{th} component of the vector, but the *collection of all components* $i = 1, 2, \ldots, n$, and hence the vector itself. In abstract index notation we would typically write "Consider a vector $v^i \in V$ and ...", where the superscript i is not an actual index but merely a dummy indicating the vector character of the object; this is particularly useful, when we deal simultaneously with vectors, covectors, and tensors, see below. In this script we use both the standard notation and the abstract index notation.

Remark. In relativity, in dimension four, we would write

$$v^\mu = \begin{pmatrix} v^0 \\ v^1 \\ v^2 \\ v^3 \end{pmatrix} \tag{A.5}$$

in abstract index notation. Again, μ does not denote any particular value, but the collection of all $\mu = 0, \ldots, 3$.

The dual space

Consider a vector space V of dimension n over the real numbers. We now define the *dual space* V^* associated with V.

Definition A.1. *The* **dual space** *V^* is the space of linear functionals on V. Elements of V^* are called* **covectors**.

Let $a \in V^*$; by definition, a is a map

$$a : V \to \mathbb{R} \tag{A.6a}$$

$$v \mapsto a(v) \in \mathbb{R}\,, \tag{A.6b}$$

that is linear, i.e.,

$$a(v + \lambda w) = a(v) + \lambda a(w) \quad \text{for all } v, w \in V,\ \lambda \in \mathbb{R}\,. \tag{A.6c}$$

As a matter of course, V^* is a vector space, since addition and scalar multiplication of elements of V^* are well-defined (and obey the vector space axioms). Namely, for $a_1 \in V^*$, $a_2 \in V^*$, and $\lambda \in \mathbb{R}$, $a_1 + \lambda a_2$ is defined as the map $v \mapsto a_1(v) + \lambda a_2(v)$, which is again in V^*.

In principle, a basis on V^* can be chosen independently of a basis on V; however, when the vector space V is equipped with a basis $\{e_1, e_2, \ldots, e_n\}$ it is convenient to use the so-called *dual basis* (or: *co-basis*)

$$\{e^1, e^2, \ldots, e^n\}\ ; \tag{A.7}$$

note the convention that co-basis vectors are indexed by superscripts. In order to define the co-basis vector e^i, since it is a map of the type (A.6), we must prescribe how it acts on vectors $v \in V$. We define

$$e^i(v) = v^i\ , \tag{A.8}$$

i.e., e^i applied to a vector yields the i^{th} component of the vector. Since $v = v^j e_j$, we have

$$\underbrace{e^i(v)}_{v^i} = e^i(v^j e_j) = v^j e^i(e_j)$$

for all v. We thus see that an equivalent definition of e^i is

$$e^i(e_j) = \delta^i{}_j\ . \tag{A.8$'$}$$

Note in passing that it is not difficult to prove that the dimension of V^* coincides with the dimension of V, as we have tacitly presupposed above.

For an arbitrary covector $a \in V^*$ we have the decomposition

$$a = a_i e^i\ , \tag{A.9}$$

where the Einstein summation convention is used. By convention, the covector components are denoted by *lower indices*: $a_1, a_2, \ldots, a_n$. These components are collected into a row vector, i.e.,

$$a = (a_1, a_2, \ldots, a_n)\ . \tag{A.10}$$

Remark. In abstract index notation we would write "Given a covector $a_i \in V^*$, then …", and

$$a_i = (a_1, a_2, \ldots, a_n)$$

instead of (A.10).

Let $v \in V$ and $a \in V^*$; by definition, $a(v)$ is a real number. Making use of the component decomposition it is simple to compute $a(v)$:

$$a(v) = a_i e^i(v) = a_i v^i\ , \tag{A.11}$$

where summation is always understood. As a special case, when a is applied to a basis vector $e_j \in V$, we obtain

$$a(e_j) = a_i e^i(e_j) = a_i \delta^i{}_j = a_j \,. \tag{A.12}$$

Hence when the covector a is applied to the basis vector e_j we obtain the j^{th} component of the covector, a_j. (This equation should be compared with equation (A.8).)

A.2 Transformation of (co)vectors

In this section we analyze how vectors and covectors transform[4] under a change of basis. To that end suppose that the vector space V is equipped with two different bases,

$$\{\hat{e}_1, \dots, \hat{e}_n\} \qquad \text{versus} \qquad \{\breve{e}_1, \dots, \breve{e}_n\} \,.$$

The decomposition of a vector v w.r.t. the first basis is

$$v = \hat{v}^i \hat{e}_i \,, \tag{A.13}$$

which we prefer to write as

$$v = \begin{pmatrix} \hat{v}^1 \\ \vdots \\ \hat{v}^n \end{pmatrix} \,. \tag{A.13$'$}$$

Analogously, the decomposition of v w.r.t. the second basis is

$$v = \breve{v}^i \breve{e}_i \,, \tag{A.14}$$

which we choose to write as

$$v = \begin{pmatrix} \breve{v}^1 \\ \vdots \\ \breve{v}^n \end{pmatrix} \,. \tag{A.14$'$}$$

It is important to keep in mind that (A.13′) and (A.14′) only make sense w.r.t. the bases we use. To avoid ambiguities it necessary to always specify the basis.

[4] Note that (co)vectors do *not* change under a change of basis. It is merely the components of the (co)vectors w.r.t. the different bases that differ (and can be transformed into each other).

For instance, we could write

$$v = \begin{pmatrix} \hat{v}^1 \\ \vdots \\ \hat{v}^n \end{pmatrix} \qquad \text{w.r.t.} \quad \{\hat{e}_1, \ldots, \hat{e}_n\} \tag{A.15}$$

and

$$v = \begin{pmatrix} \breve{v}^1 \\ \vdots \\ \breve{v}^n \end{pmatrix} \qquad \text{w.r.t.} \quad \{\breve{e}_1, \ldots, \breve{e}_n\} \,. \tag{A.16}$$

Remark. In relativity, bases correspond to observers. Hence, when we introduce a (four-)vector we must specify the observer, w.r.t. which we decompose the vector.

The decomposition of a covector $a \in V^*$ w.r.t. the co-bases $\{\hat{e}^1, \ldots, \hat{e}^n\}$ and $\{\breve{e}^1, \ldots, \breve{e}^n\}$ is analogous. We have

$$a = \hat{a}_i \hat{e}^i \qquad \text{and} \qquad a = \breve{a}_i \breve{e}^i \,, \tag{A.17}$$

or, equivalently,

$$a = (\hat{a}_1, \ldots, \hat{a}_n) \qquad \text{w.r.t.} \;\; \{\hat{e}^1, \ldots, \hat{e}^n\} \tag{A.17$'a$}$$

$$a = (\breve{a}_1, \ldots, \breve{a}_n) \qquad \text{w.r.t.} \;\; \{\breve{e}^1, \ldots, \breve{e}^n\} \,. \tag{A.17$'b$}$$

Now suppose that the two bases are related via a *basis transformation*, i.e.,

$$\breve{e}_i = A^j{}_i \hat{e}_j \,, \tag{A.18}$$

where the coefficients $A^j{}_i$ can be collected into a non-singular $(n \times n)$ matrix. We obtain

$$v = \hat{v}^i \hat{e}_i$$
$$v = \breve{v}^k \breve{e}_k = \breve{v}^k A^i{}_k \hat{e}_i = \left(A^i{}_k \breve{v}^k\right) \hat{e}_i \,,$$

therefore,

$$\hat{v}^i = A^i{}_k \breve{v}^k \,. \tag{A.19}$$

For the transformation of the dual basis we make the ansatz

$$\breve{e}^i = B^i{}_j \hat{e}^j \,. \tag{A.20}$$

We then obtain

$$\delta^j{}_i = \breve{e}^j(\breve{e}_i) = B^j{}_l \hat{e}^l(A^k{}_i \hat{e}_k) = B^j{}_l A^k{}_i \underbrace{\hat{e}^l(\hat{e}_k)}_{\delta^l{}_k} = B^j{}_k A^k{}_i \,,$$

which implies that B is the inverse of A,

$$B = A^{-1} \,, \qquad \text{i.e.,} \qquad B^i{}_k = (A^{-1})^i{}_k \,. \tag{A.21}$$

Finally, for covectors we get

$$\begin{aligned} a &= \hat{a}_i \hat{e}^i \\ a &= \breve{a}_k \breve{e}^k = \breve{a}_k B^k{}_i \hat{e}^i = \big(B^k{}_i \breve{a}_k\big)\hat{e}^i \,, \end{aligned}$$

so that

$$\hat{a}_i = B^k{}_i \breve{a}_k \,. \tag{A.22}$$

We conclude that

$$a(v) = \hat{a}_i \hat{v}^i = B^k{}_i \breve{a}_k A^i{}_l \breve{v}^l = \underbrace{B^k{}_i A^i{}_l}_{\delta^k{}_l} \breve{a}_k \breve{v}^l = \breve{a}_k \breve{v}^k \,, \tag{A.23}$$

i.e., we find consistency.

A.3 Bilinear forms

Non-degenerate symmetric bilinear forms

Let V be a finite-dimensional real vector space. In order to be able to define the concepts of orthogonality, the length of a vector, and related concepts, the vector space must be endowed with an additional geometric structure: a scalar product or a generalization thereof.

Definition A.2. *A non-degenerate symmetric bilinear form b on V is a map*

$$b : V \times V \to \mathbb{R} \tag{A.24}$$

that satisfies the conditions of

- **bilinearity**: *for all $u, v, w \in V$ and $\lambda \in \mathbb{R}$ we have*

$$\begin{aligned} b(u+\lambda v, w) &= b(u,w) + \lambda b(v,w) \\ b(u, v+\lambda w) &= b(u,v) + \lambda b(u,w)\ ; \end{aligned} \tag{A.25a}$$

- **symmetry**: *for all $v, w \in V$ there holds*

$$b(v,w) = b(w,v)\ ; \tag{A.25b}$$

- **non-degeneracy**: *the following implication for $v \in V$ holds:*

$$b(v,w) = 0 \quad \forall w \in V \quad \Rightarrow \quad v = 0\,. \tag{A.25c}$$

Since b is a map with two slots, we will occasionally write $b(\cdot,\cdot)$. This is reminiscent of the notation $\langle\cdot,\cdot\rangle$ that is employed for scalar products.

Example. A scalar product $\langle\cdot,\cdot\rangle$ is defined as a positive definite symmetric bilinear form. It is the prime example for a non-degenerate symmetric bilinear form. To show this, we must prove that positive definiteness implies non-degeneracy, cf. (A.25c): Thus, for $v \in V$ assume that $\langle v, w\rangle = 0\ \forall w \in V$. Since this holds for all w it holds necessarily also for $w = v$, i.e., $\langle v, v\rangle = 0$. However, the requirement of positive definiteness implies that $\langle v, v\rangle$ is always positive unless $v = 0$. Hence, from $\langle v, v\rangle = 0$ we conclude that $v = 0$. We have thus established (A.25c).

Definition A.3. *Let $\{e_1, e_2, \dots, e_n\}$ be a basis of V. The* **components** *of the bilinear form $b = b(\cdot,\cdot)$ are given by applying b to the basis vectors: we define b_{ij} as*

$$b_{ij} = b(e_i, e_j)\,. \tag{A.26}$$

Remark for experts. The resemblance between (A.12) and (A.26) is not a coincidence. In fact, a bilinear form is a co-tensor of rank 2, i.e., an element of $V^* \otimes V^*$, namely $b = b_{ij} e^i \otimes e^j$ (cf. $a = a_i e^i$). Its behavior thus naturally generalizes the behavior of covectors.

The components b_{ij} can be collected into an $(n \times n)$ matrix, which we call again b in slight abuse of notation,

$$b = \big(b_{ij}\big)_{i,j} = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{pmatrix}. \tag{A.27}$$

This matrix is obviously symmetric, because $b(\cdot,\cdot)$ is symmetric. Furthermore, it is not difficult to show that non-degeneracy is equivalent to the matrix b being non-singular.

Using the component representation of the bilinear form we obtain for $v, w \in V$:

$$b(v,w) = b(v^i e_i, w^j e_j) = v^i w^j b(e_i, e_j) = b_{ij} v^i w^j \, . \tag{A.28}$$

This is evidently the (tensor) analog of (A.11). In matrix notation, when we use (A.3) and (A.27) we can write (A.28) as

$$b(v,w) = v^{\mathrm{T}} b\, w = (v^1, \dots, v^n) \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{1n} & \cdots & b_{nn} \end{pmatrix} \begin{pmatrix} w^1 \\ \vdots \\ w^n \end{pmatrix} . \tag{A.28$'$}$$

Example. The standard scalar product on $\mathbb{R}^3$ can be written as

$$\langle v, w\rangle = \delta_{ij} v^i w^j = v^{\mathrm{T}}\, \mathbb{1} w = (v^1, v^2, v^3) \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix} \begin{pmatrix} w^1 \\ w^2 \\ w^3 \end{pmatrix} = v^{\mathrm{T}} w \, .$$

Example. Consider the two-dimensional plane with the bilinear form

$$b(v,w) = v^1 w^1 - 2v^1 w^2 - 2v^2 w^1 - 2v^2 w^2 = (v^1, v^2) \begin{pmatrix} 1 & -2 \\ -2 & -2 \end{pmatrix} \begin{pmatrix} w^1 \\ w^2 \end{pmatrix} , \tag{A.29}$$

the definition of which is given w.r.t. some chosen basis $\{e_1, e_2\}$. This bilinear form is clearly symmetric and non-degenerate, since the matrix is symmetric and non-singular.

Remark. Equation (A.28$'$) can be written in another form that is used frequently, namely as

$$b(v,w) = v^{\mathrm{T}} b\, w = \langle v, b\, w\rangle \, , \tag{A.28$''$}$$

where $\langle \cdot, \cdot \rangle$ denotes the standard scalar product w.r.t. the basis under consideration.[5] In quantum mechanics it is customary to write $\langle v|b|w\rangle$ instead of (A.28$''$).

[5]There is a slight subtlety involved here that we choose to suppress. In (A.28$''$), strictly speaking, b denotes the endomorphism on V whose matrix representation $\left(b^i{}_j\right)_{i,j}$ is given by (A.27).

Orthogonality

A non-degenerate symmetric bilinear form $b = b(\cdot,\cdot)$ can be regarded as a generalized scalar product, since it can be employed to define several geometric concepts that are usually associated with a scalar product in a similar way. The most important of these concepts is the concept of orthogonality.

Definition A.4. *Two vectors v and w in V are called (pseudo-)orthogonal, if $b(v, w) = 0$.*

Since $b(\cdot,\cdot)$ is in general not the standard scalar product, the concept of orthogonality differs from the one we are used to in Euclidean geometry. This is illustrated by the following example.

Example. Consider again the two-dimensional plane with the bilinear form given by (A.29). A simple calculation shows that the vectors

$$v = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \text{and} \qquad w = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

are orthogonal.

Several basic properties carry over from Euclidean geometry, an important one being the following:

Proposition A.5. *Let V be an n-dimensional vector space endowed with a non-degenerate symmetric bilinear form and let $v \in V$. Then the orthogonal complement of v is an $(n-1)$-dimensional subspace of V.*

Proof. The orthogonal complement of v is the set of all vectors w orthogonal to v, i.e., $\{w \,|\, b(v, w) = 0\}$. Define $\bar{v} = b\,v$, where b denotes the (symmetric) matrix (A.27). Since b is non-singular (so that $\ker b = 0$), it follows that $\bar{v} \neq 0$. According to (A.28″) we can write

$$b(v, w) = b(w, v) = \langle w, b\,v\rangle = \langle w, \bar{v}\rangle = \langle \bar{v}, w\rangle \,.$$

Therefore, $b(v, w) = 0$ if and only if $\langle \bar{v}, w\rangle = 0$, and the orthogonal complement of v coincides with the standard Euclidean orthogonal complement of $\bar{v}$. Since the latter is clearly a $(n-1)$-dimensional subspace of V, the proposition is established. $\square$

In Euclidean geometry, the length of a vector v is given as its norm $\|v\|$, where $\|v\| = \sqrt{\langle v, v\rangle}$. A non-degenerate symmetric bilinear form $b = b(\cdot,\cdot)$, however, does *not* define a norm in general. This is simply because there might exists vectors v such that $b(v, v) = 0$ or $b(v, v) < 0$.

Example. Consider again the two-dimensional plane with the bilinear form given by (A.29). There exist vectors v such that $b(v, v) < 0$; for example,

$$\text{for} \quad v = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \qquad \text{we have} \qquad b(v, v) = -2\,.$$

Although we are thus unable to define norms in the proper sense, we see that we can easily ascribe the "squared norm" $b(v, v)$ to each vector v. This concept is of central importance in non-Euclidean geometry and in applications.

Remark. Occasionally (and in particular in relativity) we speak of the *square* of a vector v. Despite being prone to confusion, it is also common to write v^2 for the expression $b(v, v)$. (This is reminiscent of the notation $\vec{v}^2 = \langle \vec{v}, \vec{v}\rangle$ in Euclidean space $\mathbb{R}^3$.)

Based on the above geometric implications of non-degenerate bilinear forms one often speaks of *pseudo-scalar products*. To emphasize this, one often goes so far as to use $\langle\cdot,\cdot\rangle$ to denote a bilinear form $b(\cdot,\cdot)$ even though it might not be positive definite. We conclude this section with a statement that completes Proposition A.5.

Proposition A.6. *Let V be an n-dimensional vector space endowed with a pseudo-scalar product $b(\cdot,\cdot)$. Let $v \in V$ and denote by $v^\perp$ the orthogonal complement of v. Then*

$$v \in v^\perp \quad \Leftrightarrow \quad b(v, v) = 0\,. \tag{A.30}$$

Accordingly, when $b(v, v) \neq 0$, then $v \notin v^\perp$ and $V = \langle v\rangle \oplus v^\perp$.

Proof. The proof is trivial when we recall that $v^\perp = \{w \mid b(v, w) = 0\}$. ☐

A.4 Transformation of bilinear forms

It is almost a tautological remark when we say that the component representation of a pseudo-scalar product $b(\cdot,\cdot)$, in either of its forms (A.28) or (A.28′), depends on the basis we choose. In the following we investigate how the component representation $(b_{ij})_{i,j}$ transforms under a change of basis.

Bilinear forms and basis transformations

Suppose we have two bases in V,

$$\{\hat{e}_1, \dots, \hat{e}_n\} \qquad \text{versus} \qquad \{\breve{e}_1, \dots, \breve{e}_n\} \,.$$

The decomposition of the vector v w.r.t. the first basis is[6]

$$v = \hat{v}^i \hat{e}_i \,, \quad \text{or equivalently} \qquad v \hookrightarrow \hat{v} = \begin{pmatrix} \hat{v}^1 \\ \vdots \\ \hat{v}^n \end{pmatrix} . \tag{A.31a}$$

Analogously, the decomposition of v w.r.t. the second basis is

$$v = \breve{v}^i \breve{e}_i \,, \quad \text{or equivalently} \qquad v \hookrightarrow \breve{v} = \begin{pmatrix} \breve{v}^1 \\ \vdots \\ \breve{v}^n \end{pmatrix} . \tag{A.31b}$$

Evidently, the analogous decompositions hold for the vector w. Also the bilinear form $b = b(\cdot, \cdot)$ possesses two component representations, which are given by

$$\hat{b}_{ij} = b(\hat{e}_i, \hat{e}_j) \qquad \text{versus} \qquad \breve{b}_{ij} = b(\breve{e}_i, \breve{e}_j) \,.$$

Accordingly, via (A.28) and (A.28′), $b(v, w)$ becomes

$$b(v, w) = \hat{b}_{ij} \hat{v}^i \hat{w}^j = \hat{v}^{\mathrm{T}} \hat{b}\, \hat{w} \,, \tag{A.32a}$$

$$b(v, w) = \breve{b}_{ij} \breve{v}^i \breve{w}^j = \breve{v}^{\mathrm{T}} \breve{b}\, \breve{w} \,. \tag{A.32b}$$

Now, the two bases are related via a *basis transformation*, i.e.,

$$\breve{e}_i = A^j{}_i \hat{e}_j \,,$$

where the coefficients $A^j{}_i$ can be collected into a non-singular $(n \times n)$ matrix. This implies that

$$\breve{b}_{ij} = b(\breve{e}_i, \breve{e}_j) = b(A^k{}_i \hat{e}_k, A^l{}_j \hat{e}_l) = A^k{}_i A^l{}_j b(\hat{e}_k, \hat{e}_l) = A^k{}_i A^l{}_j \hat{b}_{kl} \,, \tag{A.33}$$

or in matrix notation

$$\breve{b} = A^{\mathrm{T}} \hat{b}\, A \,. \tag{A.33′}$$

[6] Recall that the representation of a vector v as a column vector presupposes the choice of a basis, see (A.3). Although this is a trivial fact, it tends to be forgotten easily.

Based on (A.33) we can write (A.32) as

$$b(v,w) = \hat{b}_{ij}\hat{v}^i\hat{w}^j = A^k{}_i A^l{}_j \hat{b}_{kl}\,\breve{v}^i\breve{w}^j\ . \tag{A.34}$$

We see that when $\hat{b}_{ij}$ is used to compute $b(v,w)$ in hatted coordinates, then $A^k{}_i A^l{}_j \hat{b}_{kl}$ is used in breve coordinates. We have thus proved the following proposition:

Proposition A.7. *On an n-dimensional vector space consider a bilinear form whose component representation is $(b_{ij})_{i,j}$ w.r.t. some basis $\{e_1,\dots,e_n\}$. With respect to a different basis $\{\bar{e}_1,\dots,\bar{e}_n\}$, the component representation is given by*

$$\bar{b}_{ij} = A^k{}_i A^l{}_j b_{kl}\,, \tag{A.35}$$

which corresponds to $A^{\mathrm{T}} b\, A$ in matrix notation, where $A = (A^i{}_j)_{i,j}$ is the basis transformation matrix, $\bar{e}_i = A^k{}_i e_k$.

Remark. The transformation behavior (A.35) does not come as a surprise. In fact, it reflects the standard transformation of general co-tensors $T_{i_1\cdots i_n}$, since b_{ij} can be viewed as a co-tensor of rank 2.

Example. Consider again the two-dimensional plane that is endowed with the bilinear form $b = b(\cdot,\cdot)$, whose component representation w.r.t. some given basis $\{e_1, e_2\}$ is (A.29), i.e.,

$$\left(b_{ij}\right)_{i,j} = \begin{pmatrix} 1 & -2 \\ -2 & -2 \end{pmatrix}\ . \tag{A.36}$$

Now consider a second basis, $\{\bar{e}_1, \bar{e}_2\}$, which is related to the original basis via a basis transformation, i.e.,

$$\bar{e}_i = A^k{}_i e_k\ ,$$

where

$$\left(A^i{}_j\right)_{i,j} = \frac{1}{\sqrt{30}}\begin{pmatrix} \sqrt{2} & -2\sqrt{3} \\ 2\sqrt{2} & \sqrt{3} \end{pmatrix}\ .$$

W.r.t. the new basis $\{\bar{e}_1, \bar{e}_2\}$ the bilinear form exhibits the component representation

$$\left(\bar{b}_{ij}\right)_{i,j} = \left(A^k{}_i A^l{}_j b_{kl}\right)_{i,j} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\ . \tag{A.37}$$

The relevance of this particular example will become clear below.

Orthogonal transformations

Definition A.8. *Consider a vector space endowed with a bilinear form $b = b(\cdot,\cdot)$. A basis transformation $\left(A^i{}_j\right)_{i,j}$ that does not change the component representation of the bilinear form, i.e.,*

$$\left(A^k{}_i A^l{}_j b_{kl}\right)_{i,j} = b_{ij} \,, \qquad (\textit{or equivalently} \quad A^{\mathrm{T}} b\, A = b) \,, \tag{A.38}$$

*is called (pseudo-)*orthogonal *basis transformation w.r.t. $b(\cdot,\cdot)$.*

Example. Consider the two-dimensional plane with a bilinear form given by

$$\left(b_{ij}\right)_{i,j} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \tag{A.39}$$

w.r.t. some basis $\{e_0, e_1\}$ (where we follow the relativistic convention of indexing, for a change). A straightforward computation shows that the basis transformation determined by the matrix

$$\left(A^i{}_j\right)_{i,j} = \begin{pmatrix} \sqrt{2} & -1 \\ -1 & \sqrt{2} \end{pmatrix} .$$

is an example of an orthogonal basis transformation. This means that w.r.t. the basis $\{\bar{e}_0, \bar{e}_1\}$, given by $\bar{e}_i = A^j{}_i e_j$ $(i,j = 0,1)$, the bilinear form reads

$$\bar{b}_{ij} = \left(A^k{}_i A^l{}_j b_{kl}\right)_{i,j} = b_{ij} \,.$$

Remark. In our analysis we have focused on basis transformations on V and on the associated transformation of the component representation of a bilinear form $b(\cdot,\cdot)$. If we consider linear maps instead of basis transformations,[7] the discussion is similar: Consider a vector space V with bilinear form $b(\cdot,\cdot)$, a vector space $\overline{V}$ with bilinear form $\bar{b}(\cdot,\cdot)$, and a linear map $A : V \to \overline{V}$. Then A is called orthogonal, if $\bar{b}(Av, Aw) = b(v,w)$ for all $v, w \in V$. Furthermore, if $(A^i{}_p)_{i,p}$ is the matrix representing the linear map A (w.r.t. bases $\{e_1, \dots, e_n\}$ in V and $\{\bar{e}_1, \dots, \bar{e}_{\bar{n}}\}$ in $\overline{V}$), then $(A^i{}_p)_{i,p}$ satisfies $\left(A^i{}_p A^j{}_q \bar{b}_{ij}\right)_{i,j} = b_{pq}$. If $\left(\overline{V}, \bar{b}(\cdot,\cdot)\right) = \left(V, b(\cdot,\cdot)\right)$, then this reduces to (A.38).

[7] Linear maps are sometimes called active transformations, while basis transformations are called passive transformations. We have chosen to use the nomenclature that is more common in mathematics.

Orthonormal bases

In the context of scalar products, orthonormal bases play an important role. This is because the component representation of a scalar product $\langle\cdot,\cdot\rangle$ assumes its simplest form w.r.t. an orthonormal basis: if $\{e_1,\dots,e_n\}$ is an orthonormal basis w.r.t. $\langle\cdot,\cdot\rangle$, then the component representation of $\langle\cdot,\cdot\rangle$ is

$$\langle v,w\rangle = \delta_{ij} v^i w^j = v^{\mathrm{T}} \mathbb{1}\, w\,,$$

or equivalently $\langle v,w\rangle = v^{\mathrm{T}} w = \sum_{i=1}^n v^i w^i$. This is generalized to pseudo-scalar products by virtue of the following theorem.

Theorem A.9. *Let $b(\cdot,\cdot)$ be a non-degenerate symmetric bilinear form on V. There exist adapted bases $\{e_1,\dots,e_n\}$, which we call* orthonormal bases*, such that the component representation b_{ij} is diagonal and normalized, i.e.,*

$$b = \big(b_{ij}\big)_{i,j} = \operatorname{diag}\big(\underbrace{-1,-1,\dots,-1}_{n_-\ \text{times}},\underbrace{+1,+1,\dots,+1}_{n_+\ \text{times}}\big)\,, \tag{A.40}$$

where n_- and n_+ are characteristic of the bilinear form $b(\cdot,\cdot)$.

Proof. We perform a proof by induction. For a one-dimensional vector space, the statement of the theorem is trivial. Suppose that we have proved the theorem for all vector spaces of dimension less than n. To show that the theorem also holds for n-dimensional vector spaces, let $\big(V,b(\cdot,\cdot)\big)$ be n-dimensional. Take an arbitrary vector with non-zero square and call it e_1. (Such a vector exists, since $b(\cdot,\cdot)$ is non-degenerate.) Since $b(e_1,e_1)\neq 0$ we are able to rescale this vector to obtain either $b(e_1,e_1)=+1$ or $b(e_1,e_1)=-1$. By Proposition A.5 the orthogonal complement of e_1 is a $(n-1)$-dimensional vector space T (a subspace of V), which is naturally endowed with a symmetric bilinear form, namely the restriction $b(\cdot,\cdot)|_{T\times T}$ of the bilinear form $b(\cdot,\cdot)$. This bilinear form on T is non-degenerate as is shown by the following argument: Let $v\in T$. Assume that $b(v,w)=0$ for all $w\in T$; then $b(v,\lambda e_1+w)=0$ for all $\lambda\in\mathbb{R}$. Since $\langle e_1\rangle\oplus T=V$ by Proposition A.6, it follows that $b(v,w)=0$ for all $w\in V$. Hence $v=0$ by non-degeneracy of $b(\cdot,\cdot)$, which establishes non-degeneracy of $b(\cdot,\cdot)|_{T\times T}$.

Now, by the induction hypothesis, the orthogonal complement T possesses an orthonormal basis w.r.t. $b(\cdot,\cdot)|_{T\times T}$, which we denote by $\{e_2,\dots,e_n\}$. We have thus constructed an orthonormal basis $\{e_1,e_2,\dots,e_n\}$ of $\big(V,b(\cdot,\cdot)\big)$, such that, after a possible rearrangement of the basis vectors, we obtain (A.40). The fact that the number of minus-signs and plus-signs does not depend on the actual choice of basis follows from dimensional considerations. □

Definition A.10. *The* **signature** *of a non-degenerate symmetric bilinear form $b(\cdot,\cdot)$ is defined to be the distribution of minus-signs and plus-signs in the normal form* (A.40) *of b, i.e.,*

$$\operatorname{sign} b = \big(\underbrace{- - \ldots - -}_{n_- \text{ times}}\ \underbrace{+ + \ldots + +}_{n_+ \text{ times}}\big)\,. \tag{A.41}$$

Example. The signature of the bilinear form (A.36) is $(-+)$, since there exists an orthonormal basis, w.r.t. which it takes the form (A.37).

W.r.t. an orthonormal basis the pseudo-scalar product of two vectors v and w can be written as

$$b(v,w) = v^{\mathrm{T}} \begin{pmatrix} -1 & & & & & \\ & \ddots & & & & \\ & & -1 & & & \\ & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{pmatrix} w \;= -\sum_{i\le n_1} v^i w^i + \sum_{i>n_1} v^i w^i\,.$$

As with conventional scalar products, an orthonormal basis w.r.t. $b(\cdot,\cdot)$ is not unique. If we have one orthonormal basis $\{e_1,\ldots,e_n\}$, then other orthonormal bases can be obtained via orthogonal basis transformations, since these transformations do not change the form $b = \operatorname{diag}\big(-1,\ldots,-1,+1,\ldots,+1\big)$.

The signature of a non-degenerate symmetric bilinear form $b(\cdot,\cdot)$ can be obtained easily by computing the eigenvalues of its component representation b_{ij}. We formulate this as another theorem:

Theorem A.11. *The signature of $b(\cdot,\cdot)$ is given by the number of negative versus the number of positive eigenvalues of b, i.e.,*

$$\begin{aligned} \operatorname{sign} b =& \big(\underbrace{- - \ldots - -}_{n_- \text{ times}}\ \underbrace{+ + \ldots + +}_{n_+ \text{ times}}\big) \quad \Leftrightarrow \\ &\Leftrightarrow \quad b \textit{ possesses } n_- \textit{ negative and } n_+ \textit{ positive eigenvalues}\,. \end{aligned}$$

Proof. The proof of this theorem can also be regarded as an alternative proof of Theorem A.9. Starting from a given basis $\{e_1,\ldots,e_n\}$, where $b(\cdot,\cdot)$ is represented by b_{ij}, we explicitly construct an orthonormal basis $\{\bar e_1,\ldots,\bar e_n\}$.

To that end recall from (A.35) that with $\bar{e}_i = A^j{}_i e_j$ the components of b_{ij} transform according to $\bar{b}_{ij} = A^k{}_i A^l{}_j b_{kl}$, or, in matrix notation,

$$\bar{b} = A^{\mathrm{T}} b\, A\,. \tag{A.42}$$

Viewed as a matrix, b is symmetric and non-singular; therefore, b can be diagonalized by means of an orthogonal matrix O,

$$O^{-1} b\, O = O^{\mathrm{T}} b\, O = \operatorname{diag}\big(\lambda_1, \dots, \lambda_n\big)\,. \tag{A.43}$$

The eigenvalues λ_i are real (since b is symmetric) and non-zero (since b is not singular). Assume that there exist n_- negative eigenvalues and n_+ positive eigenvalues. Without loss of generality we may arrange the eigenvalues in such an order that $\lambda_1 < 0, \dots, \lambda_{n_-} < 0$ and $\lambda_{n_-+1} > 0, \dots, \lambda_n > 0$. Let us introduce the matrix

$$\Lambda = \begin{pmatrix} \sqrt{|\lambda_1|} & & \\ & \ddots & \\ & & \sqrt{|\lambda_n|} \end{pmatrix}\,;$$

the diagonal matrix $\operatorname{diag}(\lambda_1, \dots, \lambda_n)$ admits the following (unique) decomposition

$$\operatorname{diag}\big(\lambda_1, \dots, \lambda_n\big) = \Lambda\ \operatorname{diag}\big(\underbrace{-1, \dots, -1}_{n_-\text{ times}}, \underbrace{+1, \dots, +1}_{n_+\text{ times}}\big)\ \Lambda\,.$$

Combining this equation with (A.43) we find

$$\underbrace{\Lambda^{-1}\, O^{\mathrm{T}}}_{A^{\mathrm{T}}}\ \ b\ \ \underbrace{O\, \Lambda^{-1}}_{A} = \operatorname{diag}\big(\underbrace{-1, \dots, -1}_{n_-\text{ times}}, \underbrace{+1, \dots, +1}_{n_+\text{ times}}\big)\,.$$

Therefore, by defining $A = \big(A^i{}_j\big)_{i,j}$ to be $A = O\,\Lambda^{-1}$, we achieve

$$A^{\mathrm{T}} b\, A = \bar{b} = \operatorname{diag}\big(\underbrace{-1, \dots, -1}_{n_-\text{ times}}, \underbrace{+1, \dots, +1}_{n_+\text{ times}}\big)\,, \tag{A.44}$$

as intended. We have thus found a orthonormal basis $\{\bar{e}_1, \dots, \bar{e}_n\}$, in which the components of $b(\cdot,\cdot)$ are diagonal and normalized,

$$b(\bar{e}_i, \bar{e}_j) = \bar{b}_{ij} = \operatorname{diag}\big(\underbrace{-1, \dots, -1}_{n_-\text{ times}}, \underbrace{+1, \dots, +1}_{n_+\text{ times}}\big)\,.$$

This ends the proof of the theorem. □

A.5 Raising and lowering of indices

A non-degenerate symmetric bilinear form $b(\cdot,\cdot)$ generates a *canonical map* between V and its dual space V^*. Using the musical notation it is called $\flat$ ("flat") and its inverse is $\sharp = \flat^{-1}$ ("sharp"):

$$\flat : V \to V^* \tag{A.45a}$$
$$\sharp = \flat^{-1} : V \leftarrow V^* \tag{A.45b}$$

The map $\flat : V \to V^*$ is defined according to

$$\flat : V \to V^* \tag{A.46a}$$
$$v \mapsto \flat(v) := b(v,\cdot) \in V^* \; ; \tag{A.46b}$$

in other words, when v is a vector, $\flat(v)$ is a covector that takes arguments $w \in V$ according to

$$\flat(v)(w) = b(v,w) \; . \tag{A.47}$$

The map $\flat$ is a map of rank n, which relies on the non-degeneracy of $b(\cdot,\cdot)$. In other words, $\flat(v) \neq 0$ whenever $v \neq 0$.

Written in components, we see that this abstract definition lays the foundation for the procedure commonly known as "lowering of indices". Let us write

$$v = v^i e_i \qquad \text{and} \qquad \underbrace{\flat(v)}_{v^\flat} = v^\flat_i e^i \; . \tag{A.48}$$

To obtain the i^{th} component $v^\flat_i$ of the covector we use (A.12):

$$v^\flat_i = v^\flat(e_i) = \flat(v)(e_i) = b(v,e_i) = v^j b(e_j,e_i) = b_{ji}v^j = b_{ij}v^j \; .$$

It is customary to drop the $^\flat$ and simply write

$$v_i = b_{ij}v^j \; . \tag{A.49}$$

Although this notation is slightly misleading, it offers so many advantages that we will always use it. It is important, however, to always keep in mind that while $v^i \in V$, v_i is not a vector but a covector, $v_i \in V^*$. The abstract index notation is particularly useful in this context.

The counterpart of (A.49) is the procedure we call "raising of indices" that arises when we use the map $\sharp$ instead of $\flat$:

$$v^i = b^{ij}v_j \; . \tag{A.50}$$

Since $\sharp = \flat^{-1}$, $\left(b^{ij}\right)_{i,j}$ denotes the inverse matrix of b_{ij}, i.e.,

$$b^{ij} b_{jk} = \delta^i{}_k \ . \tag{A.51}$$

It suggests itself that in order to lower/raise indices of a general tensor we have to contract every index, i.e.,

$$T^{i_1 \cdots i_n}{}_{j_1 \cdots j_m} = b^{i_1 k_1} \cdots b^{i_n k_n} b_{j_1 l_1} \cdots b_{j_m l_m} \, T_{k_1 \cdots k_n}{}^{l_1 \cdots l_m} \ .$$

Remark. By definition, we obtain b^{ij} from b_{ij} when we take the inverse matrix. However, it is equally possible to compute b^{ij} from b_{ij} by raising both indices. To see that we compute

$$b^{ij} = \delta^i{}_l b^{jl} = \left(b^{ik} b_{kl}\right) b^{jl} = b^{ik} b^{jl} b_{kl} \ ,$$

which proves the claim.

Using raising and lowering of indices provides a means to write the pseudo-scalar product in a convenient form. Namely, instead of $b(v,w) = b_{ij} v^i w^j$ we are able to write

$$b(v,w) = v^i w_i = v_i w^i \ . \tag{A.52}$$

Accordingly, the square of a vector becomes

$$v^2 = b(v,v) = v^i v_i = v_i v^i \ . \tag{A.53}$$

In application in relativity, where we use Greek indices that run from zero to three, expressions of this kind are abundant:

$$v^\mu w_\mu \ , \quad v^2 = v^\mu v_\mu \ , \quad \text{etc.}$$

A.6 Affine geometry

Alternative A: Non-technical introduction

Consider an abstract vector space V. By virtue of the vector space axioms, elements of V (a.k.a. vectors) can be added and multiplied with scalars. In particular, there exists one element of V that is distinguished from all other elements: the *zero vector*. The world of vectors is thus not a world where equality reigns. Let us go over to affine spaces.

Definition A.12. *An* **affine space** *over a vector space V is a set A that can be identified with V provided that one (arbitrary) element is distinguished as the* **origin**. *The elements of an affine space are called* **points**

An affine space can thus be viewed as a "forgetful vector space"—a vector space that has lost its information on the zero vector. By actively choosing an origin, the affine space becomes a vector space again. Note, however, the immanent arbitrariness: *any* point can be chosen as the origin. As a consequence, *all points in an affine space are equal.* There is a price to pay, though:

Although, by definition, we can choose an origin o (which can be any point of A) and turn A into a vector space,

$$A \ni p \longleftrightarrow \overrightarrow{op} \in V ,$$

this does not provide a sensible way of defining addition and scalar multiplication. This is because the definition of these operations cannot be made independently of the concrete choice of origin. The complete randomness of the choice of the origin o thus prohibits that addition and scalar multiplication of points of A can be defined canonically.

Example. Consider the affine 2-plane $\mathbb{A}^2$ over $\mathbb{R}^2$. There is no canonical prescription defining addition of points $p, q \in \mathbb{A}^2$; neither does λp for $\lambda \in \mathbb{R}$ make sense. When we distinguish one point $o \in \mathbb{A}^2$, then $\mathbb{A}^2$ becomes a vector space with vectors $\overrightarrow{op}$, $\overrightarrow{oq}, \dots$. Although it is possible to add $\overrightarrow{op} + \overrightarrow{oq}$ the result is clearly not invariant under a change of origin (and thus pretty useless).

While points in A cannot be added nor multiplied with scalars, it is intuitively clear that we can "add" points of A and vectors of V, i.e., for any $p \in A$, $v \in V$, the "sum" $p + v$ is a well-defined element of A (and this procedure is independent of any choice of origin).

Appealing to our mathematical intuition once again, we see that for any pair of points $p, q \in A$ there exists a unique vector v such that $p + v = q$; one writes $v = q - p$ or $v = \overrightarrow{pq}$. Loosely speaking, we can say that (a copy of) the vector space V is attached to each point $p \in A$.

Alternative B: Technical introduction

We begin by introducing the concept of a *group action*, which is relevant for many applications in theoretical physics. Let X denote a set and G a group.

Definition A.13. *A group G is said to act on a set X (on the left) if every $g \in G$ induces a bijective map g_X on the set X, i.e.,*

$$g_X : X \to X \tag{A.54a}$$
$$x \mapsto g \cdot x \,, \tag{A.54b}$$

such that

$$g_1 \cdot (g_2 \cdot x) = (g_1 \cdot g_2) \cdot x \qquad \forall x \in X \,, \quad \forall g_1, g_2 \in G \,, \tag{A.55a}$$
$$e \cdot x = x \qquad \forall x \in X \,, \tag{A.55b}$$

where e denotes the identity element of the group G.

Remark. The conditions (A.55) ensure that the mapping $G \to \mathrm{Bij}(X)$, $g \mapsto g_X$ is a homomorphism of groups.

The notation
$$x \mapsto g \cdot x$$
is chosen to suggest that elements of X can be "multiplied" with elements of g (on the left). In some cases (when the group operation of G is denoted by a plus sign instead of by a dot) it is preferable to write

$$x \mapsto x + g$$

instead of $x \mapsto g \cdot x$, to suggest that one can "add" an element of G to an element of X. Accordingly, equation (A.55) then reads

$$(x + g_1) + g_2 = x + (g_1 + g_2) \,, \qquad x + 0 = x \,,$$

where now 0 denotes the identity element of G. (We have chosen an action on the right in this case; however, since $(G, +)$ is commonly us for abelian groups, this makes no actual difference.)

It is important to note that group action is neither multiplication nor addition in the standard sense (groups, rings, fields), since those standard algebraic operations are defined on one single set only, namely as maps $G \times G \to G$.

Example (Scalar multiplication in vector spaces)*.* Consider a vector space V over the field of the real numbers $\mathbb{R}$. The vector space axioms guarantee the existence of an action of (the multiplication group underlying) $\mathbb{R}$ on V. This is the scalar multiplication

$$v \mapsto \lambda \cdot v \qquad (\lambda \in \mathbb{R},\, v \in V) \,. \tag{A.56}$$

For all $\lambda, \mu \in \mathbb{R}$ we have

$$\lambda \cdot (\mu \cdot v) = (\lambda\mu) \cdot v \quad \text{and } 1 \cdot v = v\,.$$

It is standard to omit the dots and simply write $\lambda(\mu v) = (\lambda\mu)v$ and $1v = v$.

Example. The general linear group $\mathsf{GL}(n, \mathbb{R})$ (special linear group, orthogonal group,...) acts on the vector space $\mathbb{R}^n$ as a group of endomorphisms.

There exists several types of group actions. We merely discuss the type that is essential for our purposes: A group action is called *simply transitive*, if for any two elements $x_1, x_2 \in X$ there exists exactly one element $g \in G$ such that $g \cdot x_1 = x_2$ (or $x_1 + g = x_2$ in the '+' notation).

Definition A.14. *An* **affine space** *over a vector space V is a set A, on which the (abelian group underlying the) vector space V acts simply transitively. The elements of an affine space are called* points.

Hence, for all $p \in A$, for all $v \in V$, there exists the action

$$p \mapsto p + v\,. \tag{A.57}$$

Furthermore, since the action is assumed to be simply transitive, for each pair of points $p, q \in A$ there exists a unique vector v such that $p + v = q$. One typically writes $v = q - p$ or $v = \overrightarrow{pq}$.

It follows that, if we distinguish one point in A, which we call o (the *origin*), then each point $p \in A$ can be identified with $\overrightarrow{op} \in V$ and conversely,

$$A \ni p \;\longleftrightarrow\; \overrightarrow{op} \in V\,.$$

In this way, A can be identified with the underlying vector space V. However, this identification is not canonical, since the choice of origin is completely free.

Affine coordinates

A natural question to ask is how an affine space A can be endowed with a coordinate system. In this context a coordinate system—or chart— is simply a bijective map from $\mathbb{R}^n$ in A; it assigns each point p in A a real n-tuple ('its coordinates'). Since the affine space A is modeled over a vector space V, it suggests itself to use this structure in order to construct a coordinate system. Let us begin with the following definition.

Definition A.15. *Let A be an affine space over an n-dimensional vector space V. An* affine basis *is a tuple $\big(o, \{e_1, \dots, e_n\}\big)$, where o is a point in A ('the origin') and $\{e_1, \dots, e_n\}$ a basis of V.*

By using an affine basis we obtain coordinates on A in a simple way: Let $p \in A$. To assign coordinates to p, we first note that $p = o + x$, where $x = \overrightarrow{op} \in V$, i.e.,

$$A \ni p \;\longleftrightarrow\; x = \overrightarrow{op} \in V \longleftrightarrow\; x^i \in \mathbb{R}^n \,.$$

Obviously, the vector $x \in V$ is described by its components w.r.t. the basis, i.e., $x = x^i e_i$,

$$A \ni p \;\longleftrightarrow x = \overrightarrow{op} \in V \;\longleftrightarrow\; x^i \in \mathbb{R}^n \,.$$

Hence the coordinates of p are given by

$$A \ni p \;\longleftrightarrow\; x = \overrightarrow{op} \in V \;\longleftrightarrow\; x^i \in \mathbb{R}^n \,. \tag{A.58}$$

Definition A.16. *A coordinate map $\mathbb{R}^n \to A$ of the type* (A.58) *is called an affine chart or an* affine coordinate system.

Suppose we have two affine bases on an affine space A, i.e.,

$$\big(o, \{e_1, \dots, e_n\}\big) \qquad \text{versus} \qquad \big(\bar{o}, \{\bar{e}_1, \dots, \bar{e}_n\}\big) \,,$$

where the bases $\{e_1, \dots, e_n\}$ and $\{\bar{e}_1, \dots, \bar{e}_n\}$ on V are related by the basis transformation[8]

$$e_i = G^j{}_i \bar{e}_j \,, \tag{A.59a}$$

where $G^i{}_k$ is a non-singular matrix. Let v^i denote the components of some vector w.r.t. the basis $\{e_1, \dots, e_n\}$ and $\bar{v}^i$ its components w.r.t. $\{\bar{e}_1, \dots, \bar{e}_n\}$. Then (A.59a) induces the transformation

$$\bar{v}^i = G^i{}_k v^k \tag{A.59b}$$

on the vector components.

Now consider a point $p \in A$, whose affine coordinates w.r.t. the two affine bases are

$p \leftrightarrow x^i \in \mathbb{R}^n$	$p \leftrightarrow \bar{x}^i \in \mathbb{R}^n$
w.r.t. $\big(o, \{e_1, \dots, e_n\}\big)$	w.r.t. $\big(\bar{o}, \{\bar{e}_1, \dots, \bar{e}_n\}\big)$

[8]Previously, in section A.4, we typically used the form $\bar{e}_i = A^j{}_i e_j$ for a basis transformation and called $(A^k{}_l)_{k,l}$ the basis transformation matrix. The previous equation clearly corresponds to (A.59a) with $A = G^{-1}$.

respectively, where $x = x^i e_i = \overrightarrow{op}$ and $\bar{x} = \bar{x}^i \bar{e}_i = \overrightarrow{\bar{o}p}$. Hence the relation $\overrightarrow{\bar{o}p} = \overrightarrow{op} + \overrightarrow{\bar{o}o}$ can simply be rewritten as $\bar{x}^i \bar{e}_i = x^i e_i + \bar{a}^i \bar{e}_i$. The vector $\bar{a}^i \bar{e}_i$ represents the translation between o and $\bar{o}$, i.e., $\bar{a}^k \bar{e}_k = \overrightarrow{\bar{o}o}$. Using (A.59a) it is simple to convince oneself that the affine coordinates x^i and $\bar{x}^i$ are related by

$$\bar{x}^i = G^i{}_k x^k + \bar{a}^k \,. \tag{A.60}$$

Equation (A.60) represents a *change of affine coordinates*. It is illustrative to juxtapose (A.59b) and (A.60).

As a matter of course it is not necessary to use affine coordinates on A. Instead we can employ any set of *curvilinear coordinates* that come to mind: spherical coordinates, cylindrical coordinates, ... However, affine coordinates exhibit one characteristic property that makes them stand out from the crowd of all possible coordinate systems:

Proposition A.17. *Let A be an affine space over a real[9] vector space. A coordinate system $\varphi : \mathbb{R}^n \to A$ is an affine coordinate system if and only if all straight lines in A appear as straight lines in $\mathbb{R}^n$.*

Remark. We must append the following definitions: An m-dimensional *affine subspace* of A is a set $p + \langle v_1, v_2, \ldots, v_m \rangle$, where $\{v_1, \ldots, v_m\}$ is a set of m linearly independent vectors. One-dimensional affine subspaces are called (straight) lines.

For a proof of the proposition see any textbook on affine geometry.

Orthonormal affine coordinates

Remark. Suppose that the vector space V is endowed with a pseudo-scalar product $b(\cdot,\cdot)$. Under a change of basis (A.59a), the components transform according to

$$b_{ij} = G^k{}_i G^l{}_j \bar{b}_{kl} \,,$$

where b_{ij} are the components of the pseudo-scalar product w.r.t. the basis $\{e_1, \ldots, e_n\}$ and $\bar{b}_{ij}$ the components w.r.t. $\{\bar{e}_1, \ldots, \bar{e}_n\}$. This suggests to define a subclass of affine coordinate changes (A.60) that is of particular significance: (pseudo-)orthogonal coordinate changes. In this case, $G^i{}_j$ is a (pseudo-)orthogonal basis transformation so that the component representation b_{ij} remains unaffected by the coordinate change. The translation vector remains unspecified in this context.

[9] The assumption that the field be $\mathbb{R}$ is important in this context.

Definition A.18. *Let A be an affine space over an n-dimensional vector space V that is endowed with a pseudo-scalar product. An* orthonormal affine basis *is a tuple $\big(o, \{e_1, \dots, e_n\}\big)$, where o is a point in A and $\{e_1, \dots, e_n\}$ an orthonormal basis of V.*

The affine coordinate system of A that is obtained by using an orthonormal affine basis is called *orthonormal affine coordinate system.* W.r.t. these coordinates, since $\{e_1, \dots, e_n\}$ is an orthonormal basis of V, the pseudo-scalar product $b(\cdot, \cdot)$ on V assumes the normal form

$$\big(b_{ij}\big)_{i,j} = \operatorname{diag}\big(\underbrace{-1, -1, \dots, -1}_{n_- \text{ times}}, \underbrace{+1, +1, \dots, +1}_{n_+ \text{ times}}\big)\,,$$

see Theorem A.9.

A.7 Transformation of scalar fields

Let A be a abstract space (like an affine space or a manifold) and let

$$\begin{aligned} \psi : \mathbb{R}^n &\to A \\ (t,x) &\mapsto \psi(t,x) \end{aligned} \qquad\qquad \begin{aligned} \psi' : \mathbb{R}^n &\to A \\ (t',x') &\mapsto \psi'(t',x') \end{aligned}$$

be two different coordinate systems on A (where the fact that the coordinates are denoted by a tuple (t, x) instead of only x is of course completely irrelevant). A particular point $p \in A$ has thus two different coordinate representations

$$\begin{aligned} \mathbb{R}^n \xrightarrow{\;\psi\;} &A \xleftarrow{\;\psi'\;} \mathbb{R}^n \\ (t,x) \mapsto &\; p \;\mapsfrom (t',x')\,, \end{aligned}$$

and the coordinate transformation between the two charts is given by the map

$$\begin{aligned} \varphi = (\psi')^{-1} \circ \psi : \mathbb{R}^n &\to \mathbb{R}^n \\ (t,x) &\mapsto (t',x') \end{aligned}$$

that maps (t, x) into (t', x').

Consider a function

$$\begin{aligned} \Phi : A &\to \mathbb{R} \\ p &\mapsto \Phi(p) \end{aligned}$$

on the abstract space A. To obtain its coordinate representations we must compose Φ with the charts. W.r.t. to the coordinates given by the chart ψ, the function Φ looks like

$$\phi :(t,x) \xrightarrow{\psi} p = \psi(t,x) \xrightarrow{\Phi} \Phi(p) = \Phi(\psi(t,x)) = (\Phi \circ \psi)(t,x) \,.$$

W.r.t. to the coordinates given by the chart ψ' we obtain

$$\phi' :(t',x') \xrightarrow{\psi'} p = \psi'(t,x) \xrightarrow{\Phi} \Phi(p) = \Phi(\psi'(t',x')) = (\Phi \circ \psi')(t',x') \,.$$

Hence, in brief, the coordinate representations of the abstract function Φ are

$$\phi = \Phi \circ \psi \qquad \text{and} \qquad \phi' = \Phi \circ \psi' \,.$$

Therefore, we obtain

$$\phi = \Phi \circ \psi = \left(\phi' \circ (\psi')^{-1}\right) \circ \psi = \phi' \circ \left((\psi')^{-1} \circ \psi\right) = \phi' \circ \varphi \,,$$

i.e., the function ϕ and ϕ' behave under coordinate transformations φ according to

$$\phi = \phi' \circ \varphi \,,$$

which can also be written as $\phi(t,x) = \phi'(\varphi(t,x))$, or, since $(t',x') = \varphi(t,x)$, as

$$\phi(t,x) = \phi'(t',x') \,. \tag{A.61}$$

This is the standard transformation behavior of scalar fields.

APPENDIX B

(ANTI) SYMMETRIZATION

Consider an arbitrary tensor $T_{i_1\cdots i_k}$. Let $\mathcal{P}_{(1\cdots k)}$ denote the set of all $k!$ permutations of the k-tuple $(1,2,\ldots,k)$. Then the *symmetric part* of the tensor $T_{i_1\cdots i_k}$ is defined as

$$T_{(i_1\cdots i_k)} := \frac{1}{k!} \sum_{\sigma\in\mathcal{P}_{(1\cdots k)}} T_{i_{\sigma(1)}\cdots i_{\sigma(k)}} \,. \tag{B.1}$$

Likewise, the totally *antisymmetric part* is defined as

$$T_{[i_1\cdots i_k]} := \frac{1}{k!} \sum_{\sigma\in\mathcal{P}_{(1\cdots k)}} \operatorname{sgn}\sigma \; T_{i_{\sigma(1)}\cdots i_{\sigma(k)}} \,, \tag{B.2}$$

where $\operatorname{sgn}\sigma = \pm 1$ is the sign of the permutation.

Example. For a tensor A_{ij} we have

$$A_{(ij)} = \tfrac{1}{2}\Big(A_{ij} + A_{ji}\Big)\,, \qquad A_{[ij]} = \tfrac{1}{2}\Big(A_{ij} - A_{ji}\Big)\,. \tag{B.3}$$

Therefore,

$$A_{ij} = A_{(ij)} + A_{[ij]}\,, \tag{B.4}$$

which shows that every tensor of rank 2 (which corresponds to a matrix) can be uniquely decomposed into its symmetric plus its antisymmetric part.

Example. For a tensor $a_i b_j$ constructed from two covectors a_i and b_j we have

$$a_{(i}b_{j)} = \frac{1}{2}\left(a_i b_j + a_j b_i\right), \qquad a_{[i}b_{j]} = \frac{1}{2}\left(a_i b_j - a_j b_i\right). \tag{B.5}$$

Example. A tensor A_{ij} is symmetric if and only if $A_{ij} = A_{ji}$. Equivalently, we can write

$$A_{ij} = A_{ji} \quad \Leftrightarrow \quad A_{ij} = A_{(ij)} \quad \Leftrightarrow \quad A_{[ij]} = 0\,. \tag{B.6}$$

Analogously, a tensor A_{ij} is antisymmetric if and only if

$$A_{ij} = -A_{ji} \quad \Leftrightarrow \quad A_{ij} = A_{[ij]} \quad \Leftrightarrow \quad A_{(ij)} = 0\,. \tag{B.7}$$

Example. For a tensor A_{ijk} we have

$$A_{[ijk]} = \frac{1}{6}\left(A_{ijk} - A_{ikj} + A_{jki} - A_{jik} + A_{kij} - A_{kji}\right) \tag{B.8}$$

and the analogous result for $A_{(ijk)}$. Clearly, the analog of (B.4) does not hold.

Exercise. Let n be the dimension of the vector space and let $A_{i_1\cdots i_k}$ be a tensor of rank k (with $k \leq n$). Show that $A_{[i_1\cdots i_k]}$ has $\binom{n}{k}$ independent entries, while $A_{(i_1\cdots i_k)}$ has $\binom{n+k-1}{k}$.

Example. The tensor $A_{[ij]}$ has $\binom{n}{2}$ independent entries, while $A_{(ij)}$ has $\binom{n+1}{2}$. The sum of the two is n^2, which equals the number of coefficients of A_{ij}; compare this result with (B.4). In contrast, the tensor $A_{[ijk]}$ has $\binom{n}{3}$ independent entries, while $A_{(ijk)}$ has $\binom{n+2}{3}$. Show that the sum is $n(n^2+2)/3$, which is less than the n^3 coefficients of A_{ijk}.

When a symmetric object is contracted with an antisymmetric object, the result is zero. For instance, let A_{ijk} be totally antisymmetric and S^{ij} symmetric. Then

$$A_{ijk}S^{jk} = A_{i[jk]}S^{jk} = \frac{1}{2}A_{ijk}S^{jk} - \frac{1}{2}A_{ikj}\underbrace{S^{jk}}_{S^{kj}} = \frac{1}{2}A_{ijk}S^{jk} - \frac{1}{2}A_{ijk}S^{jk} = 0\,. \tag{B.9}$$

When an (anti)symmetrized object is contracted with an (anti)symmetric object, the (anti)symmetrization is superfluous. For instance, let A_{ijk} be totally antisymmetric and T^{ij} an arbitrary object. Then

$$A_{ijk}T^{[jk]} = \frac{1}{2}A_{ijk}T^{jk} - \frac{1}{2}\underbrace{A_{ijk}}_{-A_{ikj}}T^{kj} = \frac{1}{2}A_{ijk}T^{jk} + \frac{1}{2}A_{ijk}T^{jk} = A_{ijk}T^{jk}. \tag{B.10}$$

Example. A well-known totally antisymmetric tensor is the ϵ-tensor. In the case of a three-dimensional vector space we have

$$\epsilon_{ijk} = \begin{cases} +1 & \text{if} \quad \mathrm{sgn}(ijk) = +1 \\ -1 & \text{if} \quad \mathrm{sgn}(ijk) = -1\,, \end{cases} \tag{B.11a}$$

and $\epsilon_{ijk} = 0$ otherwise. In the case of an n-dimensional vector space, the definition is analogous, where ϵ now carries n indices. For example, in four dimensions,

$$\epsilon_{\alpha\beta\gamma\delta} = \begin{cases} +1 & \text{if} \quad \mathrm{sgn}(\alpha\beta\gamma\delta) = +1 \\ -1 & \text{if} \quad \mathrm{sgn}(\alpha\beta\gamma\delta) = -1\,, \end{cases} \tag{B.11b}$$

and $\epsilon_{\alpha\beta\gamma\delta} = 0$ otherwise. It is important to note that in an n-dimensional vector space, every totally antisymmetric tensor of rank n must be proportional to the ϵ-tensor, i.e., in n dimensions,

$$A_{[i_1\cdots i_n]} \propto \epsilon_{i_1\cdots i_n}\,. \tag{B.12}$$

Example. A generalization of the ϵ-tensor is the 'generalized Kronecker symbol':

$$\delta^{i_1\cdots i_k}_{j_1\cdots j_k} = \begin{cases} +1 & \text{if } (i_1\cdots i_k) \text{ is an } \textit{even permutation} \text{ of } (j_1\cdots j_k) \\ -1 & \text{if } (i_1\cdots i_k) \text{ is an } \textit{odd permutation} \text{ of } (j_1\cdots j_k) \\ 0 & \text{if } (i_1\cdots i_k) \text{ is } \textit{not a permutation} \text{ of } (j_1\cdots j_k) \end{cases} \tag{B.13}$$

Elementary considerations about permutations imply that

$$\delta^{i_1\cdots i_k}_{j_1\cdots j_k} = \det\left(\left(\delta^{i_p}_{j_q}\right)_{p,q}\right)\,, \tag{B.14}$$

hence

$$\delta^{i_1\cdots i_k}_{j_1\cdots j_k} = \sum_{\sigma\in\mathcal{P}_{(1\cdots k)}} \mathrm{sgn}\,\sigma\;\delta^{i_1}_{j_{\sigma(1)}}\cdots\delta^{i_k}_{j_{\sigma(k)}} \;=\; k!\,\delta^{i_1}_{[j_1}\cdots\delta^{i_k}_{j_k]}\,. \tag{B.15a}$$

Alternatively we obtain

$$\delta^{i_1\cdots i_k}_{j_1\cdots j_k} = k!\,\delta^{[i_1}_{j_1}\cdots\delta^{i_k]}_{j_k} \;=\; k!\,\delta^{[i_1}_{[j_1}\cdots\delta^{i_k]}_{j_k]}\,. \tag{B.15b}$$

In some contexts the nested variants of (B.15) are useful. We merely consider a particular example,

$$\delta^{\alpha\mu\nu}_{\kappa\sigma\tau} = 2\delta^{\alpha}_{\kappa}\delta^{\mu}_{[\sigma}\delta^{\nu}_{\tau]} + 4\delta^{\alpha}_{[\sigma}\delta^{[\mu}_{\tau]}\delta^{\nu]}_{\kappa}\,. \tag{B.16}$$

The proof of this relation is not difficult:

$$\delta^{\alpha\mu\nu}_{\kappa\sigma\tau} = 6\delta^{\alpha}_{[\kappa}\delta^{\mu}_{\sigma}\delta^{\nu}_{\tau]} \stackrel{(B.8)}{=} 2\delta^{\alpha}_{\kappa}\delta^{\mu}_{[\sigma}\delta^{\nu}_{\tau]} + 2\delta^{\alpha}_{[\sigma}\delta^{\mu}_{\tau]}\delta^{\nu}_{\kappa} + 2\delta^{\alpha}_{[\tau}\delta^{\mu}_{\hat{\kappa}}\delta^{\nu}_{\sigma]} \ ,$$

where by convention, hatted indices are excluded from the antisymmetrization; for instance, $A_{[\tau\hat{\kappa}\sigma]} = \frac{1}{2}\big(A_{\tau\kappa\sigma} - A_{\sigma\kappa\tau}\big)$; continuing the calculation we get

$$\delta^{\alpha\mu\nu}_{\kappa\sigma\tau} = 2\delta^{\alpha}_{\kappa}\delta^{\mu}_{[\sigma}\delta^{\nu}_{\tau]} + 2\delta^{\alpha}_{[\sigma}\delta^{\mu}_{\tau]}\delta^{\nu}_{\kappa} + 2\,\underbrace{\delta^{\alpha}_{[\tau}\delta^{\nu}_{\sigma]}}_{-\delta^{\alpha}_{[\sigma}\delta^{\nu}_{\tau]}}\,\delta^{\mu}_{\kappa} = 2\delta^{\alpha}_{\kappa}\delta^{\mu}_{[\sigma}\delta^{\nu}_{\tau]} + 4\delta^{\alpha}_{[\sigma}\delta^{[\mu}_{\tau]}\delta^{\nu]}_{\kappa} \ ,$$

which proves the claim. Finally we note that $\epsilon_{i_1\cdots i_k}$ is a special case of the generalized Kronecker symbol, because

$$\epsilon_{i_1\cdots i_k} = \delta^{1\cdots k}_{i_1\cdots i_k} \ . \tag{B.17}$$

Example. A common source of confusion is the ϵ-tensor with upstairs indices. On the one hand, there exists the tensor $\epsilon^{i_1\cdots i_k}$ defined in analogy to (B.11), i.e.,

$$\epsilon^{i_1\cdots i_k} = \begin{cases} +1 & \text{if} \quad \mathrm{sgn}(i_1\cdots i_k) = +1 \\ -1 & \text{if} \quad \mathrm{sgn}(i_1\cdots i_k) = -1 \ , \end{cases} \tag{B.18}$$

and $\epsilon^{i_1\cdots i_k} = 0$ otherwise. However, if the vector space is endowed with a metric (scalar product), then the symbol $\epsilon^{i_1\cdots i_k}$ is often used to denote the tensor that is generated by raising the indices of $\epsilon_{i_1\cdots i_k}$. For instance, in Minkowski space, in contrast to (B.18), the symbol $\epsilon^{\alpha\beta\gamma\delta}$ might be used to denote

$$\epsilon^{\alpha\beta\gamma\delta} = \eta^{\alpha\alpha'}\eta^{\beta\beta'}\eta^{\gamma\gamma'}\eta^{\delta\delta'}\epsilon_{\alpha'\beta'\gamma'\delta'} \ . \tag{B.19}$$

In particular

$$\epsilon^{0123} = \eta^{0\alpha'}\eta^{1\beta'}\eta^{2\gamma'}\eta^{3\delta'}\epsilon_{\alpha'\beta'\gamma'\delta'} = \det\eta = -1 \ . \tag{B.20}$$

We conclude that

$$\epsilon^{\alpha\beta\gamma\delta} = \begin{cases} -1 & \text{if} \quad \mathrm{sgn}(\alpha\beta\gamma\delta) = +1 \\ +1 & \text{if} \quad \mathrm{sgn}(\alpha\beta\gamma\delta) = -1 \ , \end{cases} \tag{B.21}$$

and $\epsilon^{\alpha\beta\gamma\delta} = 0$ otherwise. Let us summarize this notational confusion with the 'equation'

$$\epsilon^{\alpha\beta\gamma\delta} = -\epsilon^{\alpha\beta\gamma\delta} \ , \tag{B.22}$$

where the l.h.s. is the tensor generated by raising the indices of $\epsilon_{\alpha\beta\gamma\delta}$ and the r.h.s. is the 'normal' ϵ-tensor with upstairs indices. Which of two conventions is followed is (hopefully) clear from the context.

Example. Often one is confronted with contractions of ϵ-tensors. Consider $\epsilon_{i_1\cdots i_k}$ and let $\epsilon^{i_1\cdots i_k}$ be the normal epsilon tensor with upstairs indices defined by (B.18). Let $k = p + q$. The fundamental relation is

$$\epsilon_{i_1\cdots i_p j_1\cdots j_q}\epsilon^{i_1\cdots i_p l_1\cdots l_q} = p!\,\delta^{l_1\ldots l_q}_{j_1\cdots j_q}\,, \tag{B.23}$$

where the factor $p!$ enters, because there are $p!$ permutations of the p-tuple $(i_1\cdots i_p)$ that are summed over. In the special case of four dimensions we thus obtain the following relations:

$$\begin{aligned}
\epsilon_{\nu_1\nu_2\nu_3\nu_4}\ \epsilon^{\nu_1\nu_2\nu_3\nu_4} &= 4! && \text{(B.24a)}\\
\epsilon_{\nu_1\nu_2\nu_3\alpha}\ \epsilon^{\nu_1\nu_2\nu_3\beta} &= 3!\,\delta^{\beta}_{\alpha} && \text{(B.24b)}\\
\epsilon_{\nu_1\nu_2\alpha_1\alpha_2}\ \epsilon^{\nu_1\nu_2\beta_1\beta_2} &= 2!\,\delta^{\beta_1\beta_2}_{\alpha_1\alpha_2} && \text{(B.24c)}\\
\epsilon_{\nu\alpha_1\alpha_2\alpha_3}\ \epsilon^{\nu\beta_1\beta_2\beta_3} &= \quad\delta^{\beta_1\beta_2\beta_3}_{\alpha_1\alpha_2\alpha_3} && \text{(B.24d)}\\
\epsilon_{\alpha_1\alpha_2\alpha_3\alpha_4}\epsilon^{\beta_1\beta_2\beta_3\beta_4} &= \quad\delta^{\beta_1\beta_2\beta_3\beta_4}_{\alpha_1\alpha_2\alpha_3\alpha_4} && \text{(B.24e)}
\end{aligned}$$

When we use the ϵ-tensor (B.21), which is the negative of the tensor used above, we obtain the following relations:

$$\begin{aligned}
\epsilon_{\nu_1\nu_2\nu_3\nu_4}\ \epsilon^{\nu_1\nu_2\nu_3\nu_4} &= -4! &&= -24 && \text{(B.25a)}\\
\epsilon_{\nu_1\nu_2\nu_3\alpha}\ \epsilon^{\nu_1\nu_2\nu_3\beta} &= -3!\,\delta^{\beta}_{\alpha} &&= -6\,\delta^{\beta}_{\alpha} && \text{(B.25b)}\\
\epsilon_{\nu_1\nu_2\alpha_1\alpha_2}\ \epsilon^{\nu_1\nu_2\beta_1\beta_2} &= -2!\,\delta^{\beta_1\beta_2}_{\alpha_1\alpha_2} &&= -4\,\delta^{\beta_1}_{[\alpha_1}\delta^{\beta_2}_{\alpha_2]} && \text{(B.25c)}\\
\epsilon_{\nu\alpha_1\alpha_2\alpha_3}\ \epsilon^{\nu\beta_1\beta_2\beta_3} &= -\,\delta^{\beta_1\beta_2\beta_3}_{\alpha_1\alpha_2\alpha_3} &&= -6\,\delta^{\beta_1}_{[\alpha_1}\delta^{\beta_2}_{\alpha_2}\delta^{\beta_3}_{\alpha_3]} && \text{(B.25d)}\\
\epsilon_{\alpha_1\alpha_2\alpha_3\alpha_4}\epsilon^{\beta_1\beta_2\beta_3\beta_4} &= -\,\delta^{\beta_1\beta_2\beta_3\beta_4}_{\alpha_1\alpha_2\alpha_3\alpha_4} &&= -24\,\delta^{\beta_1}_{[\alpha_1}\delta^{\beta_2}_{\alpha_2}\delta^{\beta_3}_{\alpha_3}\delta^{\beta_4}_{\alpha_4]} && \text{(B.25e)}
\end{aligned}$$

Here we have used (B.15). The relations (B.25) are useful in many contexts.

Example. If (anti)symmetrization involves two identical objects, the additional symmetry results in simplifications. The simplest example is

$$a_{[i}a_{j]} = 0\,, \tag{B.26}$$

which is proved straightforwardly: $a_{[i}a_{j]} = \frac{1}{2}\big(a_i a_j - a_j a_i\big) = 0$. Analogously,

$$a_{(i}a_{j)} = a_i a_j\,, \tag{B.27}$$

because $a_i a_j$ is already symmetric. For tensors of higher order things get more complicated; we merely consider one example. Let F_{ab} be antisymmetric, i.e., $F_{[ab]} = F_{ab}$. Then

$$F_{[ab}F_{cd]} \;=\; F_{a[b}F_{cd]} \;. \tag{B.28a}$$

The latter can be written as

$$F_{a[b}F_{cd]} = \frac{1}{3}\left(F_{ab}F_{cd} + 2F_{a[c}F_{d]b}\right) \;. \tag{B.28b}$$

Let us prove (B.28a). (Recall that hatted indices are excluded from the antisymmetrization; for example, in $[ab\hat{c}d]$ the antisymmetrization involves only the indices $[abd]$, hence $A_{[ab\hat{c}d]} = \frac{1}{6}\big(A_{ab\hat{c}d} - A_{ba\hat{c}d} + A_{bd\hat{c}a} + \dots\big)$.)

$$\begin{aligned}
F_{[ab}F_{cd]} &= \frac{1}{4}\left(F_{a[b}F_{cd]} - F_{[b\hat{a}}F_{cd]} + F_{[bc}F_{\hat{a}d]} - F_{[bc}F_{d]a}\right)\\
&= \frac{1}{4}\Big(F_{a[b}F_{cd]} - F_{[cd}\underbrace{F_{b]a}}_{-F_{\hat{a}b]}} + F_{a[d}F_{bc]} - F_{[bc}\underbrace{F_{d]a}}_{-F_{\hat{a}d]}}\Big)\\
&= \frac{1}{4}\Big(F_{a[b}F_{cd]} + F_{a[b}F_{cd]} + F_{a[d}F_{bc]} + F_{a[d}F_{bc]}\Big) = F_{a[b}F_{cd]}
\end{aligned}$$

We have only used elementary considerations about permutations; furthermore, we have employed the antisymmetry of F_{ab} and the fact that cyclic permutations of a triple do not change signs. The proof of (B.28b) is simpler, since

$$F_{a[b}F_{cd]} = \frac{1}{3}\Big(F_{ab}F_{[cd]} + F_{a[c}F_{d]b} + F_{a[d}\underbrace{F_{\hat{b}c]}}_{-F_{c]b}}\Big)$$

and the result follows.