28.3.2022

Special relativity

(notes for "Relativity" a.a. 2021/22)

Fiorenzo Bastianelli

1 Introduction

By the beginning of 1900, the study of Nature had produced two great and sufficiently self-consistent theories:

1) Newtonian mechanics,

2) classical electrodynamics.

However, these two theories did not seem to be compatible with each other.

Newtonian mechanics was known to be consistent for changes of the inertial reference frames as given by *Galilean transformations*. An inertial frame K is used to measure the time t and the position $\vec{x}$ in Cartesian coordinates. We indicate it schematically by the four coordinates, $K = (t, \vec{x}) = (t, x, y, z) = (t, x^1, x^2, x^3)$. Similarly, one can consider a second inertial frame $K' = (t', \vec{x}\,')$. If the system K' moves with velocity v along the positive x direction of K, see fig. 1, the relation that connects the coordinates of these two inertial frames is given by the following Galilean transformation

$$\begin{cases} t' = t \\ x' = x - vt \\ y' = y \\ z' = z \,. \end{cases} \tag{1}$$

The equation $t' = t$ tells that time is absolute, in the sense that it flows independently of the state of motion of the observer.

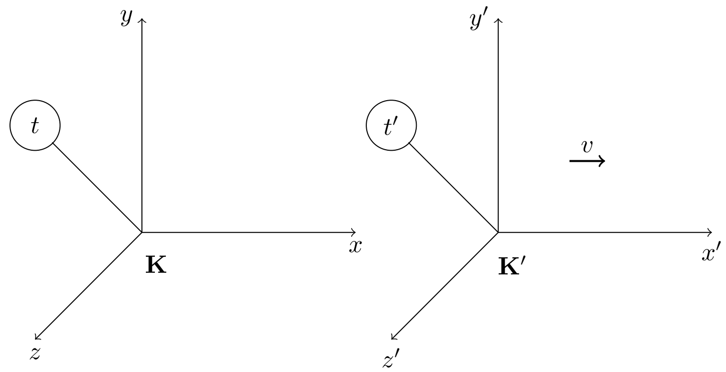


Figure 1: *The inertial frames K and K' with their own clock for measuring time.*

The above transformation leaves Newton's equations *invariant in form*. Let us verify this crucial statement on the equations of motion of a free particle of mass m

$$m\frac{d^2\vec{x}(t)}{dt^2} = 0 \qquad \longleftrightarrow \qquad m\frac{d^2\vec{x}\,'(t')}{dt'^2} = 0 \,. \tag{2}$$

This situation was described nicely by Galileo (and perhaps even by Giordano Bruno) with the image of a ship in constant motion with respect to mainland (see Galileo in "Dialogo sopra i due massimi sistemi del mondo" (1632)).

On the other hand, Newton's equations are not form invariant under a transformation with $x' = x + \frac{1}{2}gt^2$, that gives

$$m\frac{d^2\vec{x}(t)}{dt^2} = 0 \qquad \longrightarrow \qquad m\frac{d^2\vec{x}'(t')}{dt'^2} = m\vec{g} \tag{3}$$

with $\vec{g} = (g, 0, 0)$. In the new frame, there appears a *fictitious force*, also known as *inertial force*, i.e. a force that is not related to interactions with other particles. In this case there is no symmetry between the two frames, therefore they are not equivalent. The first frame is inertial, while the second one is not, because of the presence of fictitious forces.

In general, one refers to a *symmetry* in the case where the equations of motion are unchanged in form under suitable transformations of the dynamical variables. The appropriate mathematical language to describe symmetries is *group theory*, which has become quite important in modern studies of physical theories (see appendix A for the mathematical definition of a group).

Classical Newtonian mechanics is invariant under the *Galilean group*, which includes the symmetry described above for the equations of motion of free particles. More generally, Galilean invariance is maintained if the forces are consistent with this symmetry principle. An example is the gravitational force due to Newton's law of universal gravitation, which describes the gravitational forces produced by massive bodies: N massive particles of masses m_k and positions $\vec{x}_k$, with $k = 1, .., N$, are subject to the gravitational forces

$$m_k\frac{d^2\vec{x}_k}{dt^2} = G\sum_{l\neq k} m_k m_l \frac{\vec{x}_l - \vec{x}_k}{|\vec{x}_l - \vec{x}_k|^3} \ . \tag{4}$$

These equations are invariant under the transformations of the Galileo group, which take the general form

$$\begin{cases} t' = t + \tau \\ \vec{x}' = R\vec{x} - \vec{v}t + \vec{a} \end{cases} \tag{5}$$

where τ, $\vec{v}$ and $\vec{a}$ are constant, and R is a constant orthogonal matrix that describes rotations of the coordinate axes of three-dimensional space. The 10 free parameters describing this general transformation (1 parameter the time translation τ, 3 parameters for the space translations $\vec{a}$, 3 angles specifying an arbitrary rotation given by the orthogonal matrix R, and the 3 components of the velocity $\vec{v}$ specifying the proper Galilean transformation) are the 10 Lie parameters of the Galilean group.

On the other hand, electrodynamics is synthesized by Maxwell's equations, which in suitable units[1] are written as

$$\begin{aligned} \vec{\nabla}\cdot\vec{E} &= \rho \ , \qquad & \vec{\nabla}\times\vec{B} - \frac{1}{c}\frac{\partial\vec{E}}{\partial t} &= \frac{1}{c}\vec{J} \\ \vec{\nabla}\cdot\vec{B} &= 0 \ , \qquad & \vec{\nabla}\times\vec{E} + \frac{1}{c}\frac{\partial\vec{B}}{\partial t} &= 0 \ . \end{aligned} \tag{6}$$

These equations are not invariant under the Galilean transformation (1). Rather, these equa-

[1] Lorentz-Heaviside units.

tions were found to be invariant for another type of transformation, the Lorentz transformation

$$\begin{cases} t' = \dfrac{t - \frac{v}{c^2}x}{\sqrt{1 - \frac{v^2}{c^2}}} \\ x' = \dfrac{x - vt}{\sqrt{1 - \frac{v^2}{c^2}}} \\ y' = y \\ z' = z \end{cases} \tag{7}$$

where c is the velocity of light in vacuum, which enters the Lorentz transformation as a universal constant of nature. Interpreting this new transformation as the correct one linking the two inertial frames, one now recognizes that time is relative to the frame of reference. In particular, the concept of *simultaneity* looses its absolute character and it has a precise meaning only within a specific reference frame.

It was Einstein who made it clear that the symmetry properties of electrodynamics were the correct ones. The Lorentz transformation is recognized as the correct one that relates the two reference frames of fig. 1. The Galilean transformation emerges only as an approximation when $v \ll c$. Then, Einstein concluded that it was necessary to change the newtonian mechanics to a new *relativistic mechanics* to make it compatible with electrodynamics and with the correct transformation laws linking the various inertial frames.

The key points on which relativistic mechanics stands are:

- the physical laws are identical in all inertial frames,
- the "speed of light" is the same in all inertial frames.

These two fundamental properties of special relativity have been and continue to be verified experimentally with great precision. They can be used to rederive axiomatically the Lorentz transformations and their properties. As we shall see, in relativity by "speed of light" one should more generally understand it as the maximal speed of propagation possible, which numerically equals to $c \sim 300\,000$ Km/s.

2 Consequences of Lorentz transformations

As anticipated, the Lorentz transformation in eq. (7) shows that time t has no longer an absolute meaning, but must be considered as a parameter relative to the chosen frame of reference, just like the spatial coordinates $\vec{x}$. The concept of simultaneity loses its absolute character, and has a precise meaning only within a specific reference frame: forgetting this fact often leads to paradoxes. As we shall see, a consequence of the Lorentz transformation is that the speed of light is the same in the two inertial frames K and K', and thus must be considered as a fundamental constant of nature.

To study better the consequences of the Lorentz transformation, let us introduce the definitions of β (velocity measured in units of the speed of light) and γ (the relativistic Lorentz factor)

$$\beta \equiv \frac{v}{c}, \qquad \gamma \equiv \frac{1}{\sqrt{1-\beta^2}} = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}$$

that assume the values $0 \leq \beta < 1$ and $1 \leq \gamma < \infty$, since c is an upper limit for velocities.

With these notations, the Lorentz transformation and its inverse read

$$\begin{cases} t' = \gamma(t - \frac{\beta}{c}x) \\ x' = \gamma(x - \beta ct) \\ y' = y \\ z' = z \end{cases} , \qquad \begin{cases} t = \gamma(t' + \frac{\beta}{c}x') \\ x = \gamma(x' + \beta ct') \\ y = y' \\ z = z' \end{cases} \tag{8}$$

with the second one derivable form the first one by simple algebra, though symmetry considerations indicates that it is obtained by swapping primed and unprimed coordinates and reversing the velocity $v \to -v$ ($\beta \to -\beta$).

Immediate consequences of the Lorentz transformation are:
i) sum of velocities,
ii) length contraction,
iii) time dilation.

i) Sum of velocities

Let us consider a particle with velocity V_x along the x direction in frame K, see fig. 2.

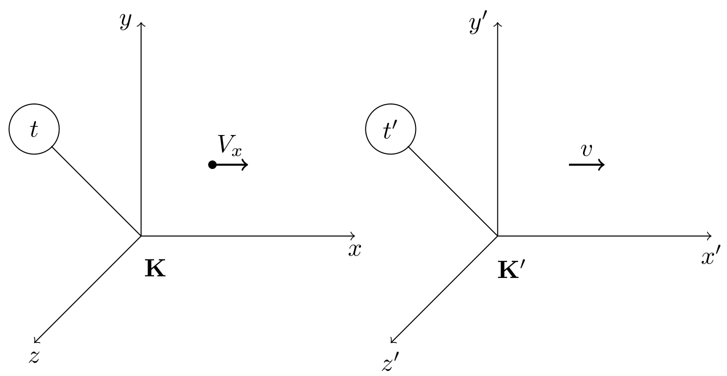


Figure 2: *Particle with velocity V_x in frame K.*

From the definition of velocity we find in systems K and K'

$$V_x = \frac{dx}{dt} , \qquad V'_x = \frac{dx'}{dt'}$$

and from the Lorentz transformation in (8) it follows that

$$dx' = \gamma(dx - vdt) , \qquad dt' = \gamma(dt - \frac{v}{c^2}dx)$$

so that

$$V'_x = \frac{dx'}{dt'} = \frac{\gamma(dx - vdt)}{\gamma(dt - \frac{v}{c^2}dx)} = \frac{V_x - v}{1 - \frac{vV_x}{c^2}} .$$

Thus,

$$\boxed{V'_x = \frac{V_x - v}{1 - \frac{vV_x}{c^2}}} . \tag{9}$$

Note that if one set $V_x = c$ then $V'_x = c$, the velocity of light is invariant under a Lorentz transformation.

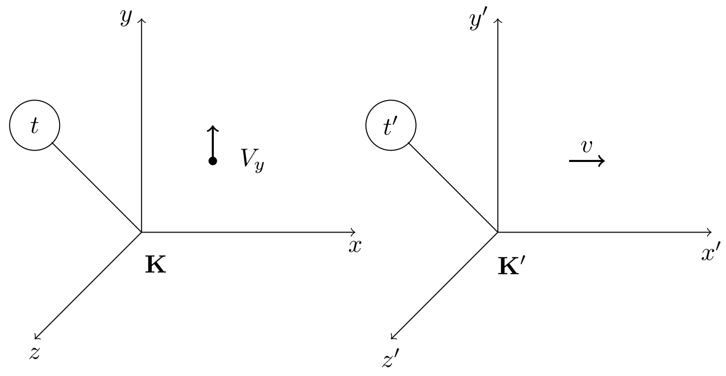


Figure 3: *Particle with velocity V_y in frame K.*

One can proceed in a similar way if the particle has a component of the velocity along the y direction, as in fig. 3. By definition $V_y = \frac{dy}{dt}$ and $V'_y = \frac{dy'}{dt'}$, and from the Lorentz transformation (8) it follows that

$$dy' = dy \ , \qquad dt' = \gamma(dt - \frac{v}{c^2}dx)$$

so that

$$V'_y = \frac{dy'}{dt'} = \frac{dy}{\gamma(dt - \frac{v}{c^2}dx)} = \frac{V_y}{\gamma(1 - \frac{vV_x}{c^2})} = \frac{V_y}{1 - \frac{vV_x}{c^2}}\sqrt{1 - \frac{v^2}{c^2}}$$

and thus

$$\boxed{V'_y = \frac{V_y}{1 - \frac{vV_x}{c^2}}\sqrt{1 - \frac{v^2}{c^2}}} \ . \tag{10}$$

ii) Length contraction

Let us consider in the K' system an object at rest of length $L_0 = x'_2 - x'_1$ placed along the x axis, as in figure 4. This length is called *proper length*, as it is the length measured in the frame

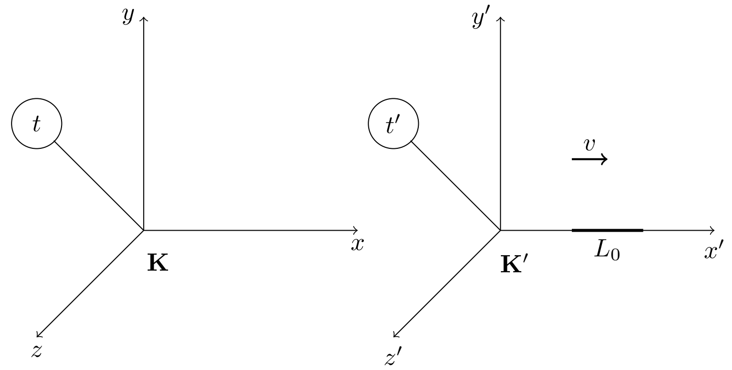


Figure 4: *An object of length L_0 along the x' axis of the K' system.*

where the object is at rest. In the system K the object is seen to move with velocity v along the positive x axis, and to measure its length one must measure simultaneously the position of its extremes, say at a fixed time t, and calculate the length L by

$$L = x_2(t) - x_1(t)$$

(remember that simultaneity is a concept defined within a single frame of reference). Now one can ask: how are L and L_0 related? From the Lorentz transformation we obtain

$$L_0 = x'_2 - x'_1 = \frac{x_2(t) - vt}{\sqrt{1 - \frac{v^2}{c^2}}} - \frac{x_1(t) - vt}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{x_2(t) - x_1(t)}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{L}{\sqrt{1 - \frac{v^2}{c^2}}} = \gamma L \ .$$

We have used the direct Lorentz transformation, the first one in (8), which makes it easy to impose the requirement that the extremes of the object are measured simultaneously at time t in frame K where the object is seen in motion. Therefore, the length seen in the frame where the object is moving with velocity v results contracted

$$\boxed{L = \gamma^{-1} L_0} \ . \tag{11}$$

Indeed, since $v \leq c$, one has $\gamma \geq 1$ and thus $L \leq L_0$.

iii) Time dilation

Let us now consider two events E'_1 and E'_2 in the K' system that occur at the same spatial point, for example $E'_1 = (t'_1, 0, 0, 0)$ and $E'_2 = (t'_2, 0, 0, 0)$, as depicted in fig. 5

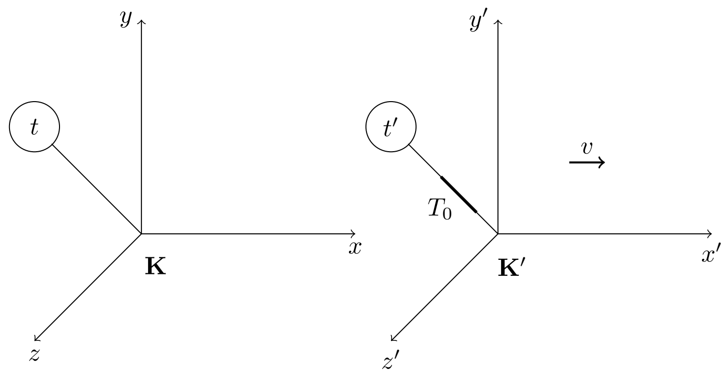


Figure 5: *Time dilation.*

These two events are separated by a time interval $T_0 = t'_2 - t'_1$. Now this lapse of time measured in frame K is given by

$$T = t_2 - t_1 = \gamma t'_2 - \gamma t'_1 = \gamma (t'_2 - t'_1) = \gamma T_0$$

where we have used the inverse Lorentz transformations, the second one in (8), where it is easy to impose the condition that the time interval is associated to events that have the same spatial coordinates in the K' system, which we took as the origin (for example one may imagine a particle at rest in the spatial origin of K' that decays in a certain interval of time). Therefore

$$\boxed{T = \gamma T_0} \ . \tag{12}$$

Since $T \geq T_0$, one calls this phenomenon "time dilation". The time that flows in the frame at rest with the object is said *proper time*. It is the shortest possible time measured for the observed phenomenon, since in all other frames this time is necessarily dilated.

To exemplify these phenomena, let us consider the decay of a relativistic particle produced by the cosmic rays in the upper atmosphere. In particular, consider a muon, a particle whose decay time is of the order of $\tau \sim 10^{-6}$ s (the time of decay is defined as the average time taken by the particle to decay in its rest frame). Suppose the muon was produced in the upper atmosphere (say at about 20 Km from the soil) with a very high speed so that $\gamma = 10^3$. One could ask whether this muon has the possibility of reaching the surface of the earth, if moving downward. First of all, one can immediately realize that this muon travels essentially at the speed of light. This is seen calculating

$$\beta = \sqrt{1 - \frac{1}{\gamma^2}} = \sqrt{1 - 10^{-6}} \sim 1 \ .$$

An observer at rest on the earth must take into account the time dilation, hence the decay time of the muon is seen dilated. Thus, before decaying the particle is seen to travel for a distance

$$L = v\tau\gamma \sim c\tau\gamma \sim 300\,\text{Km} \tag{13}$$

sufficient to reach the earth (one can estimate the thickness of the atmosphere to about 10-20 Km, where the troposphere ends). Forgetting the factor for time dilation would give a distance of about 300 m, insufficient to reach the earth.

On the other hand, an observer at rest with the muon does not experience the effect of time dilation, but sees the earth approaching with a speed close to that of light, and observes the thickness of the atmosphere contract by a factor of γ^{-1}. Thus, this second observer shares the same conclusion of the first one, the muon will reach the surface of the earth.

Exercizes

It is always good to try to solve some exercises. A few ones are found here
`https://phas.ubc.ca/~mcmillan/rqpdfs/1_relativity.pdf` any other source is fine as well.
This link with animations may also be useful
`https://physics.nyu.edu/~ts2/Animation/special_relativity.html`
An example of a paradox that arises by abusing the concepy of simultaneity of events is the ladder paradox, described here `https://en.wikipedia.org/wiki/Ladder_paradox`
See also a nice animation here about the relativity of simultaneity `https://en.wikipedia.org/wiki/Relativity_of_simultaneity`

3 Minkowski spacetime

Let us verify again that the Lorentz transformation guarantees that the speed of light is identical in inertial frames.

Exercise: *A flash of light emitted at time $t = 0$ at point $x = y = z = 0$ of system K describes a spherical wave front with coordinates (t, x, y, z) related by*

$$-c^2t^2 + x^2 + y^2 + z^2 = 0 \ .$$

Using the Lorentz transformation, verify that in the system K' the wave front remains spherical and described by the equation

$$-c^2t'^2 + x'^2 + y'^2 + z'^2 = 0$$

with the same constant c. Thus, light travels at the same speed c in both reference frames.

The exercise suggets that "*ct*" may be used as a coordinate with the dimensions of length to represent time

$$(ct, x, y, z,) = (x^0, x^1, x^2, x^3) = (x^0, \vec{x}) = x^\mu$$

where $\mu = (0, 1, 2, 3)$. Note that the upper index is not a power, but indicates which coordinates one considers. The four coordinates x^μ are the coordinates of the relativistic spacetime, called *Minkowski space*, and x^μ is called *position four-vector* as it indicates the position of a point in spacetime. Points in spacetime are also called "events": an event is described by the space-time coordinates of something that happens at time $t \equiv x^0/c$ and at the spatial point $\vec{x}$.

The previous exercise also shows that the quantity

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

is invariant under the Lorentz transformation. The exercise asked to prove it for $s^2 = 0$, but the proof is identical also for $s^2 \neq 0$.

The quantity s^2 is a *scalar*: it means that it is invariant under a Lorentz transformation and therefore it can be calculated at will using the coordinates x^μ or the coordinates x'^μ (just like the modulus square of a usual vector, which can be calculated with the pythagorean theorem using the components of the vector along a set of cartesian axes or with the components in rotated cartesian axes). The quantity s^2 is interpretable as the squared invariant distance of the event with coordinates x^μ from the origin of the reference system of spacetime (with coordinates $x_0^\mu = (0, 0, 0, 0)$). In general, given two events with coordinates x^μ and y^μ, the square of the invariant distance is given by

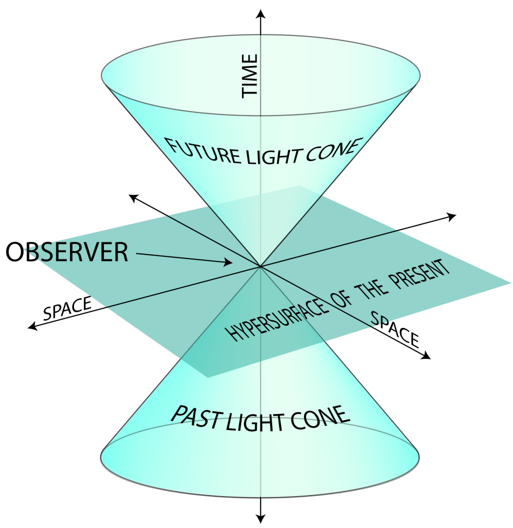
$$s^2 = -(x^0 - y^0)^2 + (x^1 - y^1)^2 + (x^2 - y^2)^2 + (x^3 - y^3)^2 \; .$$

The choice of the origin of the reference system is in general arbitrary, and has no particular significance (this is formalized by saying that spacetime is an affine space rather than a vector space).

For events connected by the propagation of light one finds that $s^2 = 0$, as seen in the previous exercise. More generally, one has the following classification

$$\begin{array}{ll} s^2 < 0 & \text{(timelike distances)} \\ s^2 = 0 & \text{(lightlike distances)} \\ s^2 > 0 & \text{(spacelike distances).} \end{array}$$

This classification is useful because it does not depend on the choice of the inertial frame: it is an invariant classification (see fig. 6). Here is another graphic representation of space time



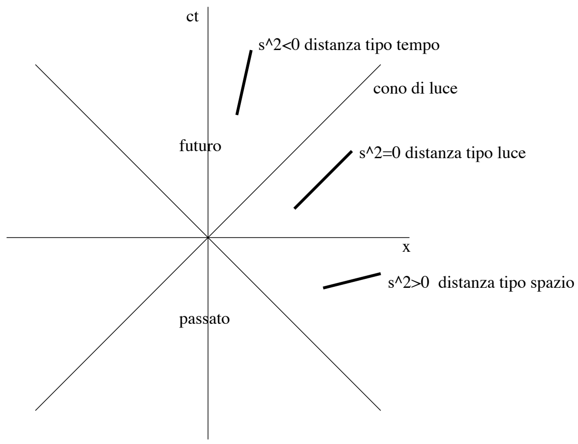
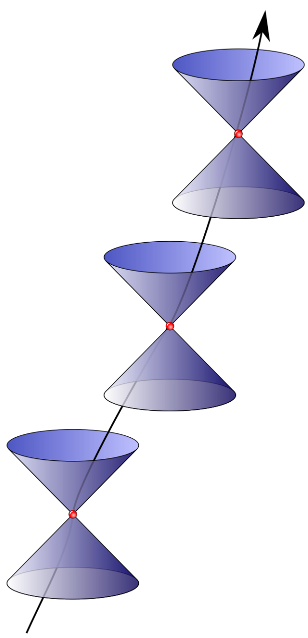


Figure 6: *Minkowski space: the coordinates* (ct, x) *are shown, while the coordinates* y *and* z *are neglected. The lightcone with respect to the origin* $(0, 0)$ *is indicated: the upper part inside the lightcone describes the absolute future of the origin, the lower part describes its absolute past. Timelike, lightlike, and spacelike distances are also shown.*

Let us analyze again the concept of proper time, and recognized it as a relativistic invariant. For any object, its proper time τ is the time that flows in its rest frame. An infinitesimal lapse of proper time $d\tau$ for an object at rest in the K' reference system is therefore given by

$$d\tau = dt' = \gamma^{-1} dt = dt\,\sqrt{1 - \frac{v^2}{c^2}} = \sqrt{dt^2 - \frac{dx^2 + dy^2 + dz^2}{c^2}} = \sqrt{-\frac{ds^2}{c^2}} \tag{14}$$

where in the last step we used the definition of the squared minkowskian length, which is negative for timelike distances. This length is Lorentz invariant, and therefore its value is calculable in any reference system. From (14) it follows that proper time is a relativistic invariant.
A particle in motion describes a line in spacetime, called *worldline*. At any point the worldline must be contained inside the lightcone of that point, since the speed of light can never be exceeded. The proper time measures (invariantly) the length of the worldline of the particle, as $d\tau = \frac{1}{c}\sqrt{-ds^2}$. Here is a picture



4 Lorentz transformations and tensor formalism

We rewrite the Lorentz transformation (7) as

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \tag{15}$$

where

$$\beta \equiv \frac{v}{c}, \qquad \gamma \equiv \frac{1}{\sqrt{1-\beta^2}} = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}$$

with $0 \leq \beta < 1$ and $1 \leq \gamma < \infty$ (see a plot of γ in fig. 7).

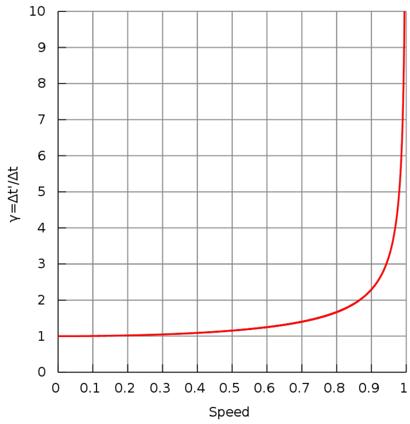


Figure 7: *Lorentz factor γ as a function of speed β. Notice that for small speeds (say $\beta < 0.1$) γ is approximately 1.*

It can be written more compactly using various notations

$$x' = \Lambda x \tag{16}$$

$$x'^{\mu} = \sum_{\nu=0}^{3} \Lambda^{\mu}{}_{\nu} x^{\nu} \tag{17}$$

$$x'^{\mu} = \Lambda^{\mu}{}_{\nu} x^{\nu} \tag{18}$$

where Λ is the matrix of the Lorentz transformation in (15) and $\Lambda^{\mu}{}_{\nu}$ denotes its matrix elements

$$\Lambda = \begin{pmatrix} \Lambda^0{}_0 & \Lambda^0{}_1 & \Lambda^0{}_2 & \Lambda^0{}_3 \\ \Lambda^1{}_0 & \Lambda^1{}_1 & \Lambda^1{}_2 & \Lambda^1{}_3 \\ \Lambda^2{}_0 & \Lambda^2{}_1 & \Lambda^2{}_2 & \Lambda^2{}_3 \\ \Lambda^3{}_0 & \Lambda^3{}_1 & \Lambda^3{}_2 & \Lambda^3{}_3 \end{pmatrix} = \Lambda^{\mu}{}_{\nu} \; .$$

In the matrix elements $\Lambda^{\mu}{}_{\nu}$, the *first index* is the *row index* and is conventionally placed at the top, the *second index* is the *column index* and is conventionally placed at the bottom. Specifying the precise position of an index is important, as it carries information which otherwise is lost and formulas may become ambiguous. The matrix notation (16) is handy if one deals with vectors and matrices only, in which case the usual algebraic product between matrices and vectors is used (the row-by-column multiplication). In (18) the Einstein's summation convention is used: indices repeated twice are automatically summed over all their possible values. Explicitly

$$\Lambda^\mu{}_\nu x^\nu = \Lambda^\mu{}_0 x^0 + \Lambda^\mu{}_1 x^1 + \Lambda^\mu{}_2 x^2 + \Lambda^\mu{}_3 x^3$$

which gives four different equations (obtained by setting $\mu = 0, 1, 2, 3$). Here the index μ is the free index, while the two indices ν are summed over. In Einstein's convention one index must be up and the other one down (this is going to be useful since it guarantees a manifest invariance from a group theory point of view, as we shall see). Since the index is summed over, one may rename it to a different letter, so that $\Lambda^\mu{}_\nu x^\nu = \Lambda^\mu{}_\rho x^\rho = x^\sigma \Lambda^\mu{}_\sigma$ are all equalities (numbers commute). The renaming is useful, as in a single expression one does not want to see more then two identical indices ($\Lambda^\mu{}_\nu x^\nu$ is fine, while $\Lambda^\mu{}_\mu x^\mu$ is ambiguous and should not be used). Again a word of caution: one should not confuse upper indices with a power: bearing this in mind will help resolving notational ambiguities.

The square of the minkowskian distance s^2, which we have recognized to be a Lorentz invariant, can be written also as follows

$$\begin{aligned} s^2 = -c^2t^2 + x^2 + y^2 + z^2 &= \begin{pmatrix} ct, & x, & y, & z \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \\ &= x^T \eta\, x = \sum_{\mu=0}^{3} \sum_{\nu=0}^{3} x^\mu \eta_{\mu\nu} x^\nu = x^\mu \eta_{\mu\nu} x^\nu = \eta_{\mu\nu} x^\mu x^\nu \end{aligned} \tag{19}$$

where $\eta_{\mu\nu}$ is the *Minkowski metric*, i.e. the matrix η with components $\eta_{\mu\nu}$

$$\eta = \begin{pmatrix} \eta_{00} & \eta_{01} & \eta_{02} & \eta_{03} \\ \eta_{10} & \eta_{11} & \eta_{12} & \eta_{13} \\ \eta_{20} & \eta_{21} & \eta_{22} & \eta_{23} \\ \eta_{30} & \eta_{31} & \eta_{32} & \eta_{33} \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} .$$

The metric allows us to evaluate the squared modulus of a four-vector. Thus, $\eta_{00} = -1$, $\eta_{11} = \eta_{22} = \eta_{33} = 1$, while $\eta_{\mu\nu} = 0$ if $\mu \neq \nu$. The squared modulus of a four-vector in Minkowski space generalizes that of a vector in euclidean space. The row and column indices of the metric are defined to be lower indices, by convention. Note that the matrix η is symmetric, so one can swap rows and columns leaving the matrix unchanged ($\eta^T = \eta$), i.e. $\eta_{\mu\nu} = \eta_{\nu\mu}$.

Now we are in a position to define the set of all Lorentz transformations. By definition, they are given by those transformations that leave the minkowskian distance invariant. In matrix notation

$$\begin{aligned} s^2 &= x^T \eta\, x \qquad &\text{(squared modulus of a four-vector)} \\ x' &= \Lambda\, x \qquad &\text{(arbitrary Lorentz transformation)} \end{aligned}$$

so that

$$s^2 = s'^2 \qquad \Rightarrow \qquad x^T \eta\, x = x'^T \eta\, x' = x^T \Lambda^T \eta\, \Lambda\, x \qquad \Rightarrow \qquad \Lambda^T \eta\, \Lambda = \eta$$

where the last equation follows since x is an arbitrary four-vector. Therefore all possible Lorentz transformations are those defined by matrices Λ satisfying

$$\Lambda^T \eta\, \Lambda = \eta \tag{20}$$

with η the Minkowski metric. In tensor notation this relation is written as

$$\eta_{\mu\nu} \Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta = \eta_{\alpha\beta} \ . \tag{21}$$

It can be rederived by computing in this notation

$$s^2 = \eta_{\mu\nu} x'^\mu x'^\nu = \eta_{\mu\nu} (\Lambda^\mu{}_\alpha x^\alpha)(\Lambda^\nu{}_\beta x^\beta) = \eta_{\mu\nu} \Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta x^\alpha x^\beta = \eta_{\alpha\beta} x^\alpha x^\beta$$

and the last step implies (21).

Exercise: *Check that the matrix in* (15) *satisfies the relation* (20)*.*

The set of Lorentz transformations form a group, *the Lorentz group*, often denoted by

$$O(3,1) = \{\text{real } 4 \times 4 \text{ matrices}\, \Lambda \,|\, \Lambda^T \eta \Lambda = \eta\} \ . \tag{22}$$

The fundamental property (20) that characterizes the Lorentz group can be used to deduce that the Lorentz group (prove that the above set of matrices satisfies the axioms of a group) is a Lie group parameterized by 6 variables: 3 angles defining the rotation of the spatial cartesian axes plus the 3 components of the relative velocity $\vec{v}$ between the two inertial frames. There are also discrete transformations, spatial inversion (parity) and time inversion, which will be discussed later.

Technical note and comparison with the rotation group

We are going to use matrices, so let us review some of their properties. For square matrices one can define a product and several other operations. Let us review this using 2×2 matrices, the extension to higher dimensions being obvious. At first we do not make a distinction between upper and lower indices, and just use lower indices

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \ . \tag{23}$$

The product of two such matrices

$$C = AB$$

is defined as usual by the row-by-column multiplication rule

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \begin{pmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{pmatrix}$$

most simply written as

$$C_{ij} = \sum_{k=1}^{2} A_{ik} B_{kj}$$

or in Einstein convention as $C_{ij} = A_{ik} B_{kj}$. For such matrices one can define the transposed matrix by exchanging rows and columns. The transpose of the matrix (23) is then

$$A^T = \begin{pmatrix} A^T_{11} & A^T_{12} \\ A^T_{21} & A^T_{22} \end{pmatrix} = \begin{pmatrix} A_{11} & A_{21} \\ A_{12} & A_{22} \end{pmatrix}$$

i.e.

$$A^T_{ij} = A_{ji} \ .$$

For products of matrices one finds that

$$\begin{aligned} (AB)^T &= B^T A^T \\ (AB)^{-1} &= B^{-1} A^{-1} \\ \det AB &= \det A \det B \end{aligned} \tag{24}$$

where we recall that the inverse of a matrix exists only if its determinant in nonvanishing. To familiarize with these notations let us observe the way one writes the following product

$$A = B^T C \qquad \longleftrightarrow \qquad A_{ij} = B^T_{ik} C_{kj} = B_{ki} C_{kj} \ .$$

Square matrices define linear operators on vector spaces. Denote a vector $\vec{x} \in \mathbb{R}^2$ by

$$\vec{x} = x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \ .$$

It is transformed to a new vector $\vec{x}\,' = x'$ by the matrix A

$$x' = Ax \qquad \longleftrightarrow \qquad \begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \qquad \longleftrightarrow \qquad x'_i = A_{ij} x_j \tag{25}$$

where in the last expression we have used again the Einstein convention.

Rotations in euclidean space

A rotation is performed by an orthogonal matrix $R = \left(\begin{smallmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{smallmatrix}\right)$

$$\begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

see fig. 8. The orthogonality condition is $R^T R = \mathbb{1}$ or equivalently $R^T = R^{-1}$.

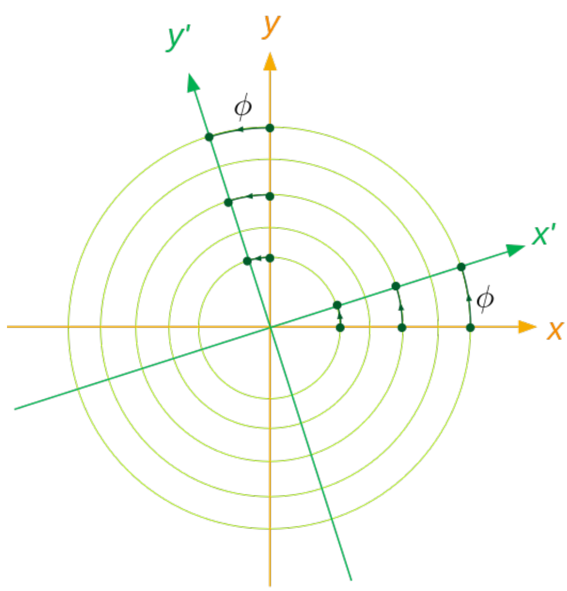


Figure 8: *Rotation in the euclidean plane.*

Orthogonal 2×2 matrices R define the orthogonal group

$$O(2) = \{\text{real } 2 \times 2 \text{ matrices } R \,|\, R^T R = \mathbb{1}\}$$

which is the group of symmetries that leaves invariant the squared modulus of a vector

$$s^2 = \vec{x} \cdot \vec{x} = x^T x = (x_1)^2 + (x_2)^2 \ . \tag{26}$$

To verify it, let us impose invariance under an arbitrary transformation $x' = Rx$

$$s'^2 = x'^T x' = (Rx)^T Rx = x^T R^T R\, x = x^T x = s^2$$

which is true for any x if and only if R is orthogonal

$$R^T R = \mathbb{1} \ . \tag{27}$$

The $O(2)$ group contains also discrete transformations that reverse the direction of one axis (in two dimensions reversing both axes is equivalent to a rotation). The basic parity transformation that reverses the first axis is generated by the orthogonal matrix

$$P = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

which has $\det P = -1$. Parity transformations can be eliminated by requiring that the determinant of R be one. From the defining property of orthogonality, one computes

$$\det R^T R = \det \mathbb{1} = 1 \quad \to \quad (\det R)^2 = 1 \quad \to \quad \det R = \pm 1$$

so that one may define the true rotation group of the euclidean plane by restricting the determinant of the orthogonal matrices to be 1

$$SO(2) = \{\text{real } 2 \times 2 \text{ matrices}\, R \,|\, R^T R = \mathbb{1} \,, \det R = 1\} \ .$$

This group is called the *special orthogonal group* in 2 dimensions. One may verify indeed that it is a group. It is the subgroup of $O(2)$ that excludes the parity transformation.

This discussion extends easily to arbitrary dimensions. The rotation group in three euclidean dimensions in $SO(3)$, a non-abelian Lie group depending continuously on three angles. More generally the rotation group in N euclidean dimensions is $SO(N)$, a Lie group depending on $\frac{1}{2}N(N-1)$ angles, which is non-abelian for $N \geq 3$ (this is seen noticing that SO(3) is non-abelian and is a subgroup of $SO(N)$ for $N > 3$).

On top of vectors, simple geometrical objects are the so-called *tensors*, defined by their simple transformation rules under rotations. A vector of $\mathbb{R}^N$ is described by its components x_i that for any matrix $R \in SO(N)$ transforms as

$$x_i \quad \xrightarrow{R} \quad x'_i = R_{ij} x_j \ .$$

A tensor of rank 2 is an object with components T_{ij} (there are N^2 independent components) that under the rotation transforms as

$$T_{ij} \quad \xrightarrow{R} \quad T'_{ij} = R_{im} R_{jn} T_{mn}$$

where each index is rotated by the matrix R. Similarly, a tensor of rank p is described by components $T_{i_1 \ldots i_p}$ transforming as

$$T_{i_1 \ldots i_p} \quad \xrightarrow{R} \quad T'_{i_1 \ldots i_p} = R_{i_1 j_1} \ldots R_{i_p j_p} T_{j_1 \ldots j_p} \ .$$

Before closing the discussion about rotations in euclidean space, let us rewrite the definition of the squared modulus of a vector (26) by pointing out the presence of the euclidean metric $\mathbb{1}$ with components the Kronecker delta δ_{ij}

$$s^2 = x^T x = x^T \mathbb{1}\, x = x_i \delta_{ij} x_j = \delta_{ij} x_i x_j \ . \tag{28}$$

Similarly, the orthogonality condition (27) can be written equivalently as

$$R^T \mathbb{1} R = \mathbb{1} \tag{29}$$

This rewriting shows the similarity between the rotation group of euclidean space and the Lorentz group of Minkowski space.

Lorentz transformations of Minkowski space

The Lorentz transformations are by definition those that leave invariant the modulus squared of a four-vector x^μ, defined with the Minkowski metric $\eta_{\mu\nu}$

$$s^2 = \eta_{\mu\nu} x^\mu x^\nu \ .$$

The matrices Λ of Lorentz transformations must obey

$$\Lambda^T \eta \Lambda = \eta \qquad \text{i.e.} \qquad \eta_{\mu\nu} \Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta = \eta_{\alpha\beta} \ .$$

The similarity with the rotation group of euclidean space is evident: the metric $\mathbb{1}$ is substituted by the Minkowski metric η. Lorentz transformations generalize to Minkowski space the rotations in euclidean space, and can be viewed as a rotation by an imaginary angle $i\phi$, as in fig. 9.

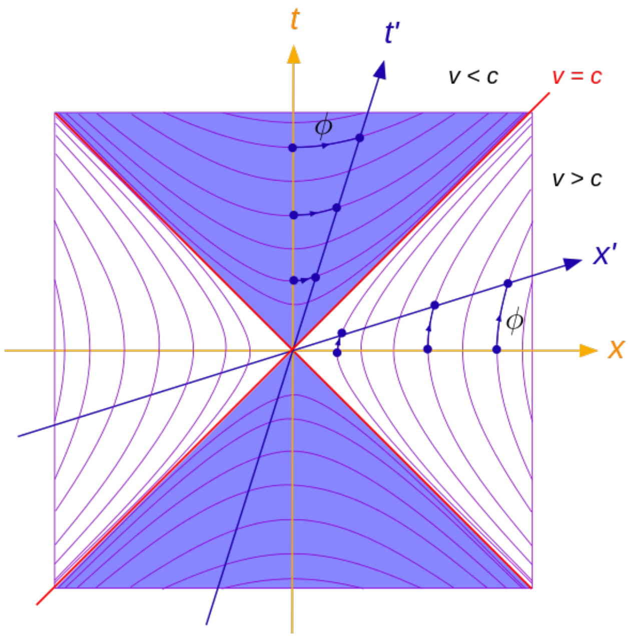


Figure 9: *Lorentz transformation as a rotation by an imaginary angle $i\phi$.*

To see that, let us rewrite the nontrivial part of the standard Lorentz transformation as

$$\begin{pmatrix} ct' \\ x' \end{pmatrix} = \begin{pmatrix} \gamma & -\beta\gamma \\ -\beta\gamma & \gamma \end{pmatrix} \begin{pmatrix} ct \\ x \end{pmatrix}$$

$$\begin{pmatrix} \gamma & -\beta\gamma \\ -\beta\gamma & \gamma \end{pmatrix} = \begin{pmatrix} \cosh\phi & -\sinh\phi \\ -\sinh\phi & \cosh\phi \end{pmatrix} = \begin{pmatrix} \cos i\phi & i\sin i\phi \\ i\sin i\phi & \cos i\phi \end{pmatrix}$$

where

$$\gamma = \cosh\phi = \frac{e^\phi + e^{-\phi}}{2} \ , \qquad \beta\gamma = \sinh\phi = \frac{e^\phi - e^{-\phi}}{2} \ ,$$

$$\cosh^2\phi - \sinh^2\phi = 1 \qquad \longleftrightarrow \qquad \gamma^2 - \beta^2\gamma^2 = 1 \ .$$

Notice that ϕ is an additive variable for the compositions of boosts in the same direction, $\phi_{12} = \phi_1 + \phi_2$, while the velocity v is not. The variable ϕ is often called *rapidity*.

Returning to the tensor notation, notice that the vertical position of indices is now important: indices are defined to be up on four-vectors like x^μ (*contravariant vectors*). The Einstein convention then fixes their position on the Lorentz matrices $\Lambda^\mu{}_\nu$ and on the Minkowski metric $\eta_{\mu\nu}$. Notice also that the product $\Lambda_1\Lambda_2$ of two Lorentz transformations Λ_1 and Λ_2 is written in components as

$$(\Lambda_1\Lambda_2)^\mu{}_\nu = (\Lambda_1)^\mu{}_\rho(\Lambda_2)^\rho{}_\nu \ .$$

Then, it is useful to define a four-vector with a lower index (*covariant vector*) using the Minkowski metric as

$$x_\mu \equiv \eta_{\mu\nu}x^\nu$$

so that the squared modulus of a four-vector x^μ is more simply written as

$$s^2 = x_\mu x^\mu \tag{30}$$

with the Minkowski metric absorbed into the definition of the covariant vector x_μ. One way of interpreting the invariance of (30) is to notice that the transformation property of a covariant vector is compensated by the transformation property of a covariant vector, so that they form a scalar product.

Denoting the components of the inverse Minkowski metric by

$$\eta^{\mu\nu} \equiv (\eta^{-1})^{\mu\nu} \qquad \longrightarrow \qquad \eta\eta^{-1} = \mathbb{1} \qquad \longleftrightarrow \qquad \eta_{\mu\nu}\eta^{\nu\rho} = \delta^\rho_\mu$$

one reobtains the contravariant vector x^μ from the covariant one by

$$x^\mu = \eta^{\mu\nu}x_\nu \ .$$

Simply said, the metric allows to lower and raise indices.

These definitions are self-consistent: as an example one may verify that

$$x^\mu = \eta^{\mu\nu}x_\nu = \eta^{\mu\nu}(\eta_{\nu\lambda}x^\lambda) = (\eta^{\mu\nu}\eta_{\nu\lambda})x^\lambda = \delta^\mu_\lambda x^\lambda = x^\mu \ .$$

In general, *scalars*, *four-vectors* and *four-tensors* (or simply *vectors* and *tensors*) are defined as quantities that transform in a very precise way under Lorentz transformations

$$\begin{aligned}
s' &= s && \text{(scalar)} \\
x'^\mu &= \Lambda^\mu{}_\nu\, x^\nu && \text{vector} \\
F'^{\mu\nu} &= \Lambda^\mu{}_\alpha\Lambda^\nu{}_\beta\, F^{\alpha\beta} && \text{tensor of rank 2} \\
T'^{\mu\nu\rho} &= \Lambda^\mu{}_\alpha\Lambda^\nu{}_\beta\Lambda^\rho{}_\gamma\, T^{\alpha\beta\gamma} && \text{tensor of rank 3} \\
&\dots
\end{aligned} \tag{31}$$

As a last observation, we point out that with our conventions on raising and lowering indices with the metric we may write the Lorentz transformation of a covariant 4-vector p_μ as

$$p'_\mu = \Lambda_\mu{}^\nu p_\nu \qquad \text{with} \qquad \Lambda_\mu{}^\nu \equiv \eta_{\mu\alpha}\Lambda^\alpha{}_\beta\eta^{\beta\nu} = (\Lambda^{T,-1})_\mu{}^\nu \tag{32}$$

(indeed from $\Lambda^T\eta\Lambda = \eta$ one finds $\Lambda^T\eta\Lambda\eta^{-1} = \mathbb{1}$), so that $p_\mu x^\nu$ is a scalar.

Examples:
i) a scalar is the Minkowskian distance s^2
ii) a vector is the position four-vector x^μ or, as we shall see, the four-momentum p^μ

iii) a two-index tensor is the electromagnetic field tensor $F^{\mu\nu}$, an antisymmetric tensor with six independent components ($F^{\mu\nu} = -F^{\nu\mu}$), written (in Gaussian or Heaviside-Lorentz units) as

$$F^{\mu\nu} = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & B_3 & -B_2 \\ -E_2 & -B_3 & 0 & B_1 \\ -E_3 & B_2 & -B_1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & B_z & -B_y \\ -E_y & -B_z & 0 & B_x \\ -E_z & B_y & -B_x & 0 \end{pmatrix} . \qquad (33)$$

Let us work out in more detail the transformation properties of the electromagnetic field. Assuming $F^{\mu\nu}$ to be a tensor under the Lorentz group, one knows that it transforms as

$$F'^{\mu\nu} = \Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta \, F^{\alpha\beta} \, . \qquad (34)$$

Let us first prove that if $F^{\mu\nu}$ is antisymmetric, also $F'^{\mu\nu}$ remains antisymmetric

$$\begin{aligned} F'^{\mu\nu} &= \Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta F^{\alpha\beta} \\ &= \Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta (-F^{\beta\alpha}) \\ &= -\Lambda^\nu{}_\beta \Lambda^\mu{}_\alpha F^{\beta\alpha} = -F'^{\,\nu\mu} \, . \end{aligned}$$

The two index tensor $F^{\mu\nu}$ can be viewed also as a matrix, and using matrices this property reads

$$F' = \Lambda F \Lambda^T$$

and if $F^T = -F$ then also

$$F'^T = (\Lambda F \Lambda^T)^T = \Lambda F^T \Lambda^T = \Lambda(-F)\Lambda^T = -\Lambda F \Lambda^T = -F' \, .$$

Let us now examine the transformation rules of the electromagnetic field under the usual Lorentz transformation determined by

$$\Lambda^\mu{}_\nu = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad i.e. \qquad \begin{array}{l} \Lambda^0{}_0 = \Lambda^1{}_1 = \gamma \\ \Lambda^1{}_0 = \Lambda^0{}_1 = -\beta\gamma \\ \Lambda^2{}_2 = \Lambda^3{}_3 = 1 \\ \text{other components vanish.} \end{array}$$

Evaluating the various components in (34) one finds

$$\begin{aligned} E'_x &= F'^{01} = \Lambda^0{}_\alpha \Lambda^1{}_\beta F^{\alpha\beta} = \Lambda^0{}_\alpha (\Lambda^1{}_0 F^{\alpha 0} + \Lambda^1{}_1 F^{\alpha 1}) \\ &= \Lambda^0{}_1 \Lambda^1{}_0 F^{10} + \Lambda^0{}_0 \Lambda^1{}_1 F^{01} = (\Lambda^0{}_0 \Lambda^1{}_1 - \Lambda^0{}_1 \Lambda^1{}_0) F^{01} \\ &= (\gamma^2 - \beta^2\gamma^2) E_x \\ &= E_x \\ E'_y &= F'^{02} = \Lambda^0{}_\alpha \Lambda^2{}_\beta F^{\alpha\beta} = \Lambda^0{}_\alpha \Lambda^2{}_2 F^{\alpha 2} = \Lambda^0{}_\alpha F^{\alpha 2} \\ &= \Lambda^0{}_0 F^{02} + \Lambda^0{}_1 F^{12} = \gamma E_y - \beta\gamma B_z \\ &= \gamma(E_y - \beta B_z) \\ E'_z &= F'^{03} = \Lambda^0{}_\alpha \Lambda^3{}_\beta F^{\alpha\beta} = \Lambda^0{}_\alpha \Lambda^3{}_3 F^{\alpha 3} = \Lambda^0{}_\alpha F^{\alpha 3} \\ &= \Lambda^0{}_0 F^{03} + \Lambda^0{}_1 F^{13} = \gamma E_z - \beta\gamma(-B_y) \\ &= \gamma(E_z + \beta B_y) \end{aligned}$$

and similarly

$$
\begin{aligned}
B'_x &= F'^{23} = \Lambda^2{}_\alpha \Lambda^3{}_\beta F^{\alpha\beta} = \Lambda^2{}_2 \Lambda^3{}_3 F^{23} = F^{23} = B_x \\
B'_y &= F'^{31} = \Lambda^3{}_\alpha \Lambda^1{}_\beta F^{\alpha\beta} = \Lambda^3{}_3 \Lambda^1{}_\beta F^{3\beta} = \Lambda^1{}_\beta F^{3\beta} \\
&= \Lambda^1{}_0 F^{30} + \Lambda^1{}_1 F^{31} = -\beta\gamma(-E_z) + \gamma B_y \\
&= \gamma(B_y + \beta E_z) \\
B'_z &= F'^{12} = \Lambda^1{}_\alpha \Lambda^2{}_\beta F^{\alpha\beta} = \Lambda^1{}_\alpha \Lambda^2{}_2 F^{\alpha 2} = \Lambda^1{}_\alpha F^{\alpha 2} = \\
&= \Lambda^1{}_0 F^{02} + \Lambda^1{}_1 F^{12} = -\beta\gamma E_y + \gamma B_z \\
&= \gamma(B_z - \beta E_y)
\end{aligned}
$$

which are summarized as follows

$$
\begin{array}{ll}
E'_x = E_x & B'_x = B_x \\
E'_y = \gamma(E_y - \beta B_z) & B'_y = \gamma(B_y + \beta E_z) \\
E'_z = \gamma(E_z + \beta B_y) & B'_z = \gamma(B_z - \beta E_y) \ .
\end{array}
$$

Now, let us apply the concept of lowering indices on tensors and calculate

$$F_{\mu\nu} = \eta_{\mu\alpha}\eta_{\nu\beta}F^{\alpha\beta} \ .$$

Since $\eta_{00} = -1$ and $\eta_{11} = \eta_{22} = \eta_{33} = 1$, with other components vanishing, we find

$$F_{0i} = -F^{0i} \ , \qquad F_{ij} = F^{ij}$$

i.e. in $F_{\mu\nu}$ the electric field has changed sign with respect to $F^{\mu\nu}$ while the magnetic $\vec{B}$ did not

$$
F_{\mu\nu} = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & B_z & -B_y \\ E_y & -B_z & 0 & B_x \\ E_z & B_y & -B_x & 0 \end{pmatrix} \ . \tag{35}
$$

It is now easy to evaluate the scalar

$$F^{\mu\nu}F_{\mu\nu} = 2(\vec{B}^2 - \vec{E}^2)$$

The correct contraction of indices insures that its value is invariant under Lorentz transformations. This scalar is proportional to the lagrangian density of the electromagnetic field.

4.1 Summary

Let us recapitulate some properties of the tensor formalism we have been developing so far.

A way of interpreting the invariance generated by the "contraction" of indices (the sum of two indices in Einstein convention), as in $s^2 = x^\mu x_\mu$, is to say that the transformation of the contravariant four-vector x^μ is compensated by the transformation of the covariant four-vector x_μ, hence the "contraction" of the indices in $x^\mu x_\mu$ produces a scalar. Similarly, $F^{\mu\nu}A_\mu B_\nu$ is a scalar, $F^{\mu\nu}B_\nu$ is a contravariant four-vector, etc., if the original quantities transform tensorially as indicated by their indices.

In general, the positions of indices on tensors indicate the properties of transformation under Lorentz transformations, and contracted indices can be ignored as they behave like a scalar.

Taking into account the metric that lowers and raises indices, and applying this also to the matrices of the Lorentz transformations, we can write the Lorentz transformation of a covariant vector in the form

$$x'_\mu = \Lambda_\mu{}^\nu x_\nu \qquad \text{with} \qquad \Lambda_\mu{}^\nu = \eta_{\mu\alpha}\Lambda^\alpha{}_\beta \eta^{\beta\nu} = (\Lambda^{T,-1})_\mu{}^\nu \tag{36}$$

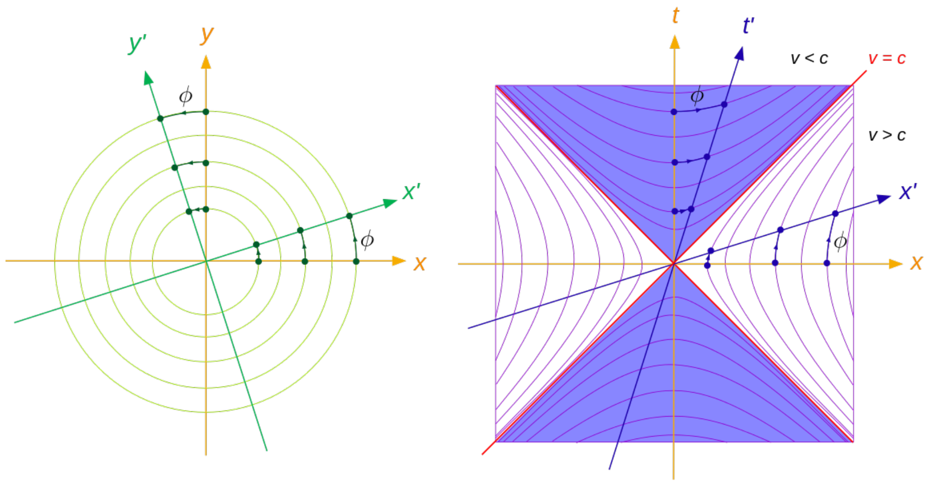
where in the first equality on the right the first index is lowered and the second one is raised, while the second equality follows from the defining property $\Lambda^T\eta\Lambda = \eta$ (which one can rewrite as $\Lambda^T\eta\Lambda\eta^{-1} = \mathbb{1}$, so that $\eta\Lambda\eta^{-1} = \Lambda^{T,-1}$). We can check again the invariance of $x_\mu x^\mu$

$$\begin{aligned} x_\mu x^\mu \;\to\; x'_\mu x'^\mu &= \Lambda_\mu{}^\alpha x_\alpha \Lambda^\mu{}_\beta x^\beta = \Lambda_\mu{}^\alpha \Lambda^\mu{}_\beta x_\alpha x^\beta \\ &= (\eta\Lambda\eta^{-1})_\mu{}^\alpha \Lambda^\mu{}_\beta x_\alpha x^\beta = (\Lambda^{T,-1})_\mu{}^\alpha \Lambda^\mu{}_\beta x_\alpha x^\beta \\ &= (\Lambda^{-1})^\alpha{}_\mu \Lambda^\mu{}_\beta x_\alpha x^\beta = (\Lambda^{-1}\Lambda)^\alpha{}_\beta x_\alpha x^\beta \\ &= \delta^\alpha{}_\beta x_\alpha x^\beta = x_\alpha x^\alpha = x_\mu x^\mu \;. \end{aligned}$$

Tensorial equations are those which equate tensors of the same rank, and therefore objects with identical transformation properties. Often, a tensor is equated to the zero tensor, the tensor that vanishes in all reference frames.

5 Proper orthochronous Lorentz group, Poincaré group

So far, we have considered the Lorentz group $O(3,1)$, defined in eq. (22), which generalizes the orthogonal group $O(3)$ to a spacetime with three spacelike directions and a timelike one, see the following picture for a comparison.



These are Lie groups, i.e. groups that depend continuously on some parameters. For the Lorentz group these parameters are the three components of the relative velocity $\vec{v}$ between the two inertial frames, and three angles $\vec{\theta}$ which describe a possible rotation of the spatial axes (e.g. the Euler angles). It is therefore a six-parameter Lie group. When these parameters vanish we have the identity transformation. Then, by continuously varying these parameters one can reach all the Lorentz matrices continuously connected to identity. We can indicate these matrices with $\Lambda(\vec{v},\vec{\theta})$. They have determinant $\det\Lambda(\vec{v},\vec{\theta}) = 1$ and component $\Lambda^0{}_0(\vec{v},\vec{\theta}) \geq 1$. In general, one may deduce that

$$\det(\Lambda^T\eta\,\Lambda) = \det(\eta) \qquad \to \qquad \det(\Lambda) = \pm 1 \;.$$

Furthermore, evaluating the 00 component of (21) one finds

$$\eta_{\mu\nu}\Lambda^\mu{}_0\Lambda^\nu{}_0 = \eta_{00} \qquad \to \qquad -(\Lambda^0{}_0)^2 + \sum_{i=1}^{3}(\Lambda^i{}_0)^2 = -1$$

i.e.

$$(\Lambda^0{}_0)^2 = 1 + \sum_{i=1}^{3} (\Lambda^i{}_0)^2 \geq 1 \qquad \rightarrow \qquad \Lambda^0{}_0 \geq 1 \;\text{ or }\; \Lambda^0{}_0 \leq -1 \;.$$

Since the identity transformation has unit determinant and $\Lambda^0{}_0 = 1$, by continuity the transformations of the part continuosly connected to the identity must have $\det \Lambda(\vec{v}, \vec{\theta}) = 1$ and $\Lambda^0{}_0 \geq 1$.

These matrices form a subgroup, called the *proper orthochronous Lorentz group*, and denoted by $SO^+(3,1)$. Relativistic invariance usually means only invariance under this subgroup.

There are discrete transformations, spatial inversion (or parity) and temporal inversion, which are not connected to identity, but belong to the full Lorentz group $O(3,1)$.

The spatial inversion (denoted by P) is defined by

$$x^{\mu\prime} = P^\mu{}_\nu x^\nu \;, \qquad P^\mu{}_\nu = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

and changes the orientation of the spatial axes. It belongs to the Lorentz group ($P^T \eta P = \eta$) and has $\det P = -1$.

Similarly, the time inversion (denoted by T) is defined by

$$x^{\mu\prime} = T^\mu{}_\nu x^\nu \;, \qquad T^\mu{}_\nu = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

changes the direction of the timelike axis, belongs to the Lorentz group ($T^T \eta T = \eta$) and has $\det T = -1$.

By composing the matrices of the proper and orthochronous Lorentz group $SO^+(3,1)$ with the discrete transformations P and T one obtain the elements of the disconnected parts of the entire Lorentz group $O(3,1)$, which in total has four disconnected components.

In addition to the possibility of changing the reference system with Lorentz transformations it is possible to make a different choice of the origin. This corresponds to the possibility of operating 3 translations in space and 1 translation in time. Adding these transformations to the Lorentz group one obtains the *Poincaré group*, a Lie group with ten parameters, that transform the coordinates of spacetime the following way

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

The ten parameters correspond to the 4 parameters a^μ which define a space-time translation plus the 6 Lorentz group parameters contained in a generic $\Lambda^\mu{}_\nu$. It is sometimes called the inhomogeneous Lorentz group and indicated by $ISO^+(3,1)$. It can be proved (Noether's theorem) that the invariance under spacetime translations is related to the conservation of energy and momentum. Similarly, the invariance under Lorentz transformations implies the conservation of other 6 quantities (which include the 3 components of angular momentum).

6 Relativistic mechanics

Let us now introduce notions of relativistic mechanics.

6.1 Relativistic definition of energy and momentum: the four-momentum

We have already introduced the concept of proper time. The infinitesimal proper time of an object in motion with velocity $\vec{v}$ in an arbitrary inertial frame can be written as

$$d\tau = dt\,\sqrt{1-\frac{v^2}{c^2}} = \frac{1}{c}\sqrt{-dx_\mu dx^\mu}\ . \tag{37}$$

It is a relativistic invariant (a scalar), as evident once written in the last form. We use it to introduce the concept of four-momentum of massive particles.

6.1.1 Massive particles

A particle in motion describes a trajectory in spacetime, the worldline, see fig. 10. The worldline

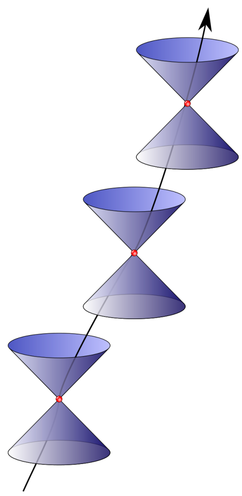


Figure 10: *Worldline of a massive particle, contained inside the lightcone of any of its points.*

can be parameterized in various ways, for example using the time t which labels the position of the particle at that time $\vec{x}(t)$. Another way is to use the proper time τ, which has the advantage of being a scalar, so to indicate by $x^\mu(\tau)$ the worldline of the particle: for each value of the proper time τ of the particle, the functions $x^\mu(\tau)$ tells us the position of the particle in spacetime by giving the position $\vec{x}(\tau)$ at time $x^0(\tau)$.

So let us consider a massive particle with worldline $x^\mu(\tau)$ parameterized by the proper time. The four-velocity is by definition the following four-vector

$$u^\mu(\tau) \equiv \frac{dx^\mu(\tau)}{d\tau} = \left(\frac{cdt(\tau)}{d\tau}, \frac{d\vec{x}(\tau)}{d\tau}\right) = \left(\frac{c}{\sqrt{1-\frac{v^2}{c^2}}}, \frac{\vec{v}}{\sqrt{1-\frac{v^2}{c^2}}}\right) = (\gamma c, \gamma\vec{v}) \tag{38}$$

where we used equation (37). This is a four-vector since $d\tau$ is a scalar and dx^μ a four-vector, so that transforms as $u'^\mu = \Lambda^\mu{}_\nu u^\nu$. Its square is given by

$$u^\mu u_\mu = -(u^0)^2 + \vec{u}\cdot\vec{u} = -c^2 \tag{39}$$

which shows that it is a timelike 4-vector. Since we know that $u^\mu u_\mu$ is a Lorentz invariant, we could have calculated it in the particle rest frame, where $u^\mu = (c, 0)$, see (38), and it follows immediately that $u^\mu u_\mu = -c^2$.

The *four-momentum* of the particle is defined by

$$p^\mu = m u^\mu \tag{40}$$

where m is the mass of the particle. This mass is by definition a scalar quantity assigned to the particle. Sometimes it is called invariant mass, rest mass, or proper mass to differentiate it from other (often useless) definitions. Therefore, p^μ is also a four-vector, and under a Lorentz transformation it transforms as

$$p'^\mu = \Lambda^\mu{}_\nu p^\nu \ .$$

The square of p^μ is easily computable

$$p^\mu p_\mu = m^2 u^\mu u_\mu = -m^2 c^2 \tag{41}$$

If no force acts on the particle, the four-velocity and the four-momentum are constant four-vectors. To get a bit more familiar with these relativistic definitions, let us show how the usual non-relativistic definitions of energy and momentum of a free particle get generalized in relativistic mechanics

$$p^\mu = m\frac{dx^\mu(\tau)}{d\tau} = \left(mc\frac{dt(\tau)}{d\tau}, m\frac{d\vec{x}(\tau)}{d\tau}\right) = \left(\frac{mc}{\sqrt{1-\frac{v^2}{c^2}}}, \frac{m\vec{v}}{\sqrt{1-\frac{v^2}{c^2}}}\right) = \left(\frac{E}{c}, \vec{p}\right) = (p^0, \vec{p}) \ . \tag{42}$$

We have identified p^0 with the energy E (divided by c for dimensional reasons), while the spatial components $\vec{p}$ of the four-vector are identified with the relativistic definition of linear momentum. We justify these identifications by looking at the non-relativistic limit $v \ll c$.

Energy E: from (42) one finds

$$E = \frac{mc^2}{\sqrt{1-\frac{v^2}{c^2}}} \ . \tag{43}$$

For $v = 0$ we see that special relativity assigns in a natural way a rest energy proportional to the mass $E = mc^2$. For $v \ll c$ we can expand[2] in $\frac{v}{c}$, which is a small number compared to 1,

$$E = mc^2\left(1-\frac{v^2}{c^2}\right)^{-\frac{1}{2}} = mc^2\left(1+\frac{1}{2}\frac{v^2}{c^2}+...\right) = mc^2 + \frac{1}{2}mv^2 + ... \tag{44}$$

This shows that the non-relativistic definition of kinetic energy is reproduced at velocities which are much smaller compared to that of light. An appropriate definition of *kinetic energy* T in the relativistic case is

$$T = E - mc^2 \ . \tag{45}$$

Linear momentum $\vec{p}$: from (42) we see that the linear momentum for $v \ll c$ becomes

$$\vec{p} = \frac{m\vec{v}}{\sqrt{1-\frac{v^2}{c^2}}} \qquad \rightarrow \qquad \vec{p} = m\vec{v} \ . \tag{46}$$

Note that massive particles cannot reach the speed of light, otherwise they would acquire infinite energy and momentum. The velocity $v = c$ is a limit velocity, theoretically unattainable for massive particles, since no physical phenomenon can deliver infinite energy to a particle.

---

[2] Recall the following Taylor expansion valid for $x \ll 1$

$$(1+x)^\alpha = 1 + \alpha x + \frac{1}{2}\alpha(\alpha-1)x^2 + \cdots \ .$$

These relativistic definitions can be justified more rigorously starting from an action principle that takes care of special relativity. The correct action for a free particle is proportional to the integral of the proper time of the particle, so that relativistic invariance is guaranteed

$$S[\vec{x}(t)] = -mc^2 \int dt \sqrt{1 - \frac{\dot{\vec{x}} \cdot \dot{\vec{x}}}{c^2}} \tag{47}$$

where of course $\dot{\vec{x}} = \frac{d\vec{x}}{dt} = \vec{v}$ is the speed of the particle. The proportionality constant $(-mc^2)$ is chosen to obtain the correct non-relativistic limit of the lagrangian $L = -mc^2\sqrt{1 - \frac{v^2}{c^2}}$, which for small velocities becomes $L_{NR} = \frac{mv^2}{2} - mc^2 + \cdots$. Since the lagrangian L does not depend on the position $\vec{x}$, but only on the velocity $\vec{v} = \dot{\vec{x}}$, the conjugate momentum

$$\vec{p} \equiv \frac{\partial L}{\partial \dot{\vec{x}}} = \frac{m\dot{\vec{x}}}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{48}$$

is conserved, as a consequence of the Euler-Lagrange equations $\frac{d\vec{p}}{dt} = 0$. Also the Hamiltonian

$$H = \vec{p} \cdot \dot{\vec{x}} - L = \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{49}$$

which coincides with the energy of the particle is conserved. This derivation does not highlight the fact that energy and momentum are the components of a four-vector, but we have already verified that to be the case. An action principle that will keep this property manifest will be discussed later on.

If external forces are present, in a relativistic theory the equations of motion can be written in the form

$$\frac{dp^\mu}{d\tau} = f^\mu \tag{50}$$

where f^μ is called *four-force*, the relativistic generalization of the concept of force, that transforms as a four-vector. To interpret its components, let us evaluate

$$f^\mu = \frac{dp^\mu}{d\tau} = \frac{dt}{d\tau}\frac{dp^\mu}{dt} = \gamma\frac{dp^\mu}{dt}$$

and realize that

$$f^0 = \frac{dp^0}{d\tau} = \frac{1}{c}\frac{dE}{d\tau} = \frac{\gamma}{c}\frac{dE}{dt}$$

is proportional to the power $W = \frac{dE}{dt}$ (energy variation in unit of time) so that

$$\boxed{f^\mu = (f^0, \vec{f}) = \left(\gamma\frac{W}{c}, \gamma\vec{F}\right)} \tag{51}$$

where $\vec{F} = \frac{d\vec{p}}{dt}$ is the usual definition of force. An example of a relativistic equation is that of a charged particle of mass m and charge e subjected to the Lorentz force due to an electromagnetic field $F^{\mu\nu}(x)$, which takes the form

$$\frac{dp^\mu}{d\tau} = \frac{e}{c}F^{\mu\nu}(x)\frac{dx_\nu}{d\tau} \tag{52}$$

which indeed equates tensors of the same type. The right hand side is the expression of the Lorentz four-force.

Thus, free massive particles of mass m have a four-momentum p^μ which satisfies

$$p^\mu p_\mu = -m^2c^2 \qquad \longleftrightarrow \qquad E^2 = \vec{p}^{\,2}c^2 + m^2c^4$$

from which we can find the energy as a function of momentum

$$E = \sqrt{\vec{p}^{\,2}c^2 + m^2c^4}\,. \tag{53}$$

6.2 Massless particles

We have seen that massive particles cannot reach the speed of light exactly. However, particles may travel at the speed of light provided they have zero mass. This interpretation is consistent with special relativity. Special relativity predicts that particles that travel at the speed of light must always travel at that speed, which is necessarily the same in all inertial frames. This property also tells that there is no frame at rest with a massless particle, and thus there is no concept of proper time for them. In fact from (37) we see that the presumed proper time would vanish, and thus cannot flow and used to parameterize the worldline of massless particles. The worldline line of massless particles is lightlike and must necessarily lie on a lightcone.

Nevertheless, one can assign a four-momentum to massless particles. In such a case the relativistic invariant $p^\mu p_\mu$ vanishes and can be used to derive a relationship between energy and momentum

$$p^\mu p_\mu = 0 \qquad \Rightarrow \qquad -\frac{E^2}{c^2} + \vec{p}\cdot\vec{p} = 0 \qquad \Rightarrow \qquad E = |\vec{p}|c\,. \tag{54}$$

This formula is approximately valid also for particles traveling at speeds which are very close to that of light, and which consequently have energies much greater than their mass, $E \gg mc^2$ (ultra-relativistic particles). Note that the concept of ultra-relativistic particle depends on the chosen reference system. Let us try to verify this statement. We have seen that for a particle of mass m

$$p^\mu = \left(\frac{E}{c}, \vec{p}\right) = \left(\gamma mc, \gamma m\vec{v}\right)$$

so that

$$\vec{p} = \gamma m\vec{v} = \frac{E\vec{v}}{c^2}\,.$$

In the limit $v \to c$

$$|\vec{p}| = \frac{E}{c}$$

which reproduces eq. (54) of massless particles.

We conclude this section by recalling that it has never been possible to interpret consistently "tachyons" (hypothetical particles that travel at speeds greater than c).

6.3 The conservation of four-momentum

The previous definition of four-momentum is appropriate as it is this quantity that satisfies precise conservation laws. In fact, one can prove that:

The total four-momentum is conserved in relativistic interactions which are invariant under translations in space and time (all fundamental interactions have this property)

$$P^\mu_{total,\,initial} = P^\mu_{total,\,final}\,. \tag{55}$$

This corresponds to four conservation laws (the four possible values of the index μ in (55): the conservation of energy and the conservation of the three components of the linear momentum. This is an application of the famous Noether's theorem, that relates conserved charges to the Lie symmetries of a dynamical system, but we will not present here the proof.

Let us consider an application: the process of a particle of mass M at rest, which decays into two particles of mass m_1 and m_2. We use the conservation of the total four-momentum to calculate the energies of the final particles. The four-momentum of the initial particle is given by

$$p^\mu = (Mc, 0)$$

while we indicate with

$$p_1^\mu = (E_1/c, \vec{p}_1)\ , \qquad p_2^\mu = (E_2/c, \vec{p}_2)$$

the four-momentum of the final particles. The conservation law reads

$$p^\mu = p_1^\mu + p_2^\mu$$

which reduces to

$$\begin{aligned} &Mc^2 = E_1 + E_2 \\ &\vec{p}_1 + \vec{p}_2 = 0\ . \end{aligned}$$

First notice that since $E_1 \geq m_1 c^2$ and $E_2 \geq m_2 c^2$, necessarily $M \geq m_1 + m_2$ for the process to take place. To continue, consider the second equation which implies

$$\vec{p}_1^{\,2} = \vec{p}_2^{\,2} \qquad \to \qquad E_1^2 - m_1^2 c^4 = E_2^2 - m_2^2 c^4$$

and combined with the first equation (the conservation of energy) gives

$$E_1 = \frac{M^2 + m_1^2 - m_2^2}{2M} c^2\ , \qquad E_2 = \frac{M^2 + m_2^2 - m_1^2}{2M} c^2\ .$$

Note that the final particles are monoenergetic.

The β decay of a neutron is instead a three-body decay: $n \to p + e^- + \bar{\nu}_e$. Originally, the neutrino was not known to exist, but since the final electron did not show a monoenergetic spectrum, Pauli in 1930 assumed conservation of energy and hypothesized the existence of a third particle produced in the decay, the neutrino, discovered experimentally in 1956.

To investigate decay processes, it is convenient to define the concept of *invariant mass*: in processes with many particles emerging in the final state, the invariant mass of N such particles is defined by

$$M = \frac{1}{c}\sqrt{-p^\mu p_\mu}\ , \qquad p^\mu = \sum_{i=1}^{N} p_i^\mu \tag{56}$$

where p_i^μ with $i = 1, ..., N$ is the four-momentum of each of the N particles. It corresponds to the mass of a hypotetical particle produced in the collision, that later decays into the N particles. The case with $N = 2$ is the one used the most.

To describe scattering of two into two particles, as in figure 11, the use of the Mandelstam variables is useful

$$\begin{aligned} s &= -(p_1 + p_2)^2 \\ t &= -(p_2 - p_3)^2 \\ u &= -(p_1 - p_3)^2 \end{aligned} \tag{57}$$

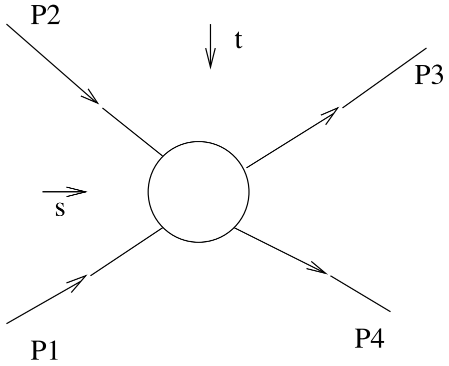


Figure 11: *Scattering process. The Mandelstam variables s and t are indicated.*

where on the right side we used the minkowskian square, and p_n is the four-momentum of the n-th particle with mass m_n, oriented as in figure 11

These Mandelstam variables are not independent, but satisfy (setting $c = 1$ for simplicity)

$$s + t + u = m_1^2 + m_2^2 + m_3^2 + m_4^2 \tag{58}$$

where $(p_1)^2 = -m_1^2$, etc. To prove this relationship it is necessary to use the conservation of the total four-momentum, which in the present case takes the form

$$p_1^\mu + p_2^\mu = p_3^\mu + p_4^\mu \ . \tag{59}$$

If further particles are produced in the collision, the variables t and u lose their immediate meaning, while the variable s continues to be useful: $\sqrt{s}$ corresponds to the total energy present in the "center of mass" frame, which is available for the creation of new particles.

By definition the "center of mass" (CM) frame is the one in which the total spatial momentum is zero. Another useful reference frame is the "laboratory frame" (LAB), in which one particle is at rest (target) while the other one has a non-zero momentum (projectile). The use of relativistic invariants is often convenient for studying scattering processes. This is exemplified by the following exercise

Exercise: *Calculate the minimum kinetic energy (threshold energy) that a proton must have to interact with another proton at rest and generate in addition a proton-antiproton pair*

$$p + p \quad \rightarrow \quad p + p + p + \bar{p} \ .$$

To solve the exercise, consider first the CM frame. The minimum energy that can trigger the creation of the proton-antiproton pair is the one for which all final particles are at rest in the CM frame, so that the final total four-momentum is $P_f'^{\mu} = (4m_p, \vec{0})$ with modulus square $P_f'^2 = -16m_p^2$. In the LAB frame the initial total four-momentum is $P_i^\mu = (m_p, \vec{0}) + (E, \vec{p}) = (m_p + E, \vec{p})$, and $P_i^2 = -2m_p^2 - 2m_p E$. Using the conservation of the total four-momentum, and the relativistic invariance of its square, one can equate $P_f'^2 = P_i^2$, so that $E = 7m_p$. The corresponding kinetic energy of the moving proton (projectile) is therefore $T = E - m_p = 6m_p \sim 5.6\,\text{GeV}$.

7 Maxwell's equations

Let us now use the tensor formalism to write the Maxwell equations in tensorial form and show their Lorentz invariance (or more properly Lorentz covariance).

As a preparation, we first comment on the derivative operator

$$\partial_\mu \equiv \frac{\partial}{\partial x^\mu} \,. \tag{60}$$

Since the spacetime coordinates x^μ behave as a contravariant four-vector ($x'^\mu = \Lambda^\mu{}_\nu x^\nu$), the derivative operator behaves like a covariant 4-vector that transforms as

$$\frac{\partial}{\partial x'^\mu} = \Lambda_\mu{}^\nu \frac{\partial}{\partial x^\nu} \qquad i.e. \qquad \partial'_\mu = \Lambda_\mu{}^\nu \partial_\nu \tag{61}$$

remembering that $\Lambda_\mu{}^\nu = (\Lambda^{T,-1})_\mu{}^\nu$, as in eq. (36), so that

$$\partial_\mu x^\mu = \partial'_\mu x'^\mu = 4$$

is indeed a scalar. More generally,

$$\partial_\mu x^\nu = \delta^\nu_\mu$$

the Kronecker delta, which is recognized here to be an invariant tensor (a tensor that under the transformation laws, as dictated by its index structure, remains the same). To practice, let us calculate the derivative of the invariant s^2

$$\partial_\mu s^2 \equiv \frac{\partial s^2}{\partial x^\mu} = \frac{\partial}{\partial x^\mu}(\eta_{\nu\lambda} x^\nu x^\lambda) = \eta_{\nu\lambda}\delta^\nu_\mu x^\lambda + \eta_{\nu\lambda} x^\nu \delta^\lambda_\mu = 2x_\mu \tag{62}$$

which equates tensors of the same rank (i.e. with the same index structure). Tensorial equations are those which equate tensors of the same rank, and therefore objects with identical transformation properties. They take the same form in all inertial frames, and very often contain the zero tensor on the right side (as in (64)). Finally, notice that from $x'^\mu = \Lambda^\mu{}_\nu x^\nu$ we find

$$\frac{\partial x'^\mu}{\partial x^\nu} = \Lambda^\mu{}_\nu \,.$$

We are now ready to verify that Maxwell equations can be written in a tensorial form. The first line of Maxwell's equations in (6), the equations with sources, can be written as

$$\boxed{\partial_\mu F^{\mu\nu} = -\frac{1}{c} J^\nu} \tag{63}$$

where $F^{\mu\nu}$ is the electromagnetic field strength in (33) and $J^\mu = (J^0, \vec{J}) = (c\rho, \vec{J})$ is the 4-vector charge-current density that is interpreted as the source of the electromagnetic field. To verify it, let us evaluate (63) for different values of the free index $\nu = (0, i)$ with $i = (1, 2, 3)$. For $\nu = 0$ we find

$$\begin{aligned}\partial_\mu F^{\mu 0} &= \partial_0 F^{00} + \partial_i F^{i0} = \partial_1 F^{10} + \partial_2 F^{20} + \partial_3 F^{30} = -(\partial_1 E_1 + \partial_2 E_2 + \partial_3 E_3) \\ &= -\vec{\nabla} \cdot \vec{E} = -\frac{1}{c} J^0 = -\rho\end{aligned}$$

which we recognize as the Gauss law $\vec{\nabla} \cdot \vec{E} = \rho$. For $\nu = 1$ we find

$$\begin{aligned}\partial_\mu F^{\mu 1} &= \partial_0 F^{01} + \partial_i F^{i1} = \partial_0 E_1 + \partial_2 F^{21} + \partial_3 F^{31} = \frac{1}{c}\partial_t E_1 - \partial_2 B_3 + \partial_3 B_2 \\ &= \Big(\frac{1}{c}\frac{\partial \vec{E}}{\partial t} - \vec{\nabla} \times \vec{B}\Big)_1 = -\frac{1}{c} J_1\end{aligned}$$

which corresponds to the first component of the vectorial equation $\vec{\nabla}\times\vec{B}-\frac{1}{c}\frac{\partial\vec{E}}{\partial t}=\frac{1}{c}\vec{J}$. Similarly for $\nu = 2, 3$. Analogously, one may check that the other Maxwell's equations in (6), the homogeneous ones, can be written as

$$\boxed{\partial_\mu F_{\nu\lambda} + \partial_\nu F_{\lambda\mu} + \partial_\lambda F_{\mu\nu} = 0}\ . \tag{64}$$

The sum with the cyclic permutation of indices makes the left hand side completely antisymmetric, and thus with only 4 independent components.[3] Evaluating (64) with $(\mu, \nu, \lambda) = (1, 2, 3)$, and recalling (35), one finds

$$\partial_1 F_{23} + \partial_2 F_{31} + \partial_3 F_{12} = \partial_1 B_1 + \partial_2 B_2 + \partial_3 B_3 = 0$$

and thus $\vec{\nabla}\cdot\vec{B} = 0$. Similarly, setting $(\mu, \nu, \lambda) = (0, 1, 2)$ the equation becomes

$$\partial_0 F_{12} + \partial_1 F_{20} + \partial_2 F_{01} = \frac{1}{c}\partial_t B_3 + \partial_1 E_2 - \partial_2 E_1 = 0$$

that corresponds to the third component of the vector equation $\vec{\nabla}\times\vec{E}+\frac{1}{c}\frac{\partial\vec{B}}{\partial t} = 0$, etc..

Now we can study the Lorentz transformation properties of Maxwell's equations

$$\begin{aligned} \partial_\mu F^{\mu\nu} &= -\frac{1}{c} J^\nu \\ \partial_\mu F_{\nu\lambda} + \partial_\nu F_{\lambda\mu} + \partial_\lambda F_{\mu\nu} &= 0\ . \end{aligned} \tag{65}$$

If the density current is a 4-vector, and declaring $F_{\mu\nu}$ to be a tensor, we see immediately that these equations involve only tensors, and their covariance under Lorentz transformations is manifest: they must have the same form in all inertial frames.

Let us also notice that

$$\partial_\mu \partial_\nu F^{\mu\nu} = 0$$

since derivatives commute while $F^{\mu\nu}$ is antisymmetric, so that in the sums all terms cancel pairwise. Then, applying the derivative ∂_ν to both members of the first equation in (65) one finds

$$\boxed{\partial_\nu J^\nu = 0}\ . \tag{66}$$

This is a "*continuity equation*" that describes the local conservation of the electric charge

$$\partial_\mu J^\mu = \partial_0 J^0 + \vec{\nabla}\cdot\vec{J} = \frac{\partial \rho}{\partial t} + \vec{\nabla}\cdot\vec{J} = 0\ . \tag{67}$$

Defining the total electric charge Q at time t by

$$Q(t) = \int d^3x\, \rho(t, \vec{x}) \tag{68}$$

one finds that it is conserved as a consequence of the continuity equation

$$\frac{dQ}{dt} = \int d^3x\, \frac{\partial \rho}{\partial t} = -\int d^3x\, \vec{\nabla}\cdot\vec{J} = 0 \tag{69}$$

where we have assumed that the density current $\vec{J}$ vanishes sufficiently fast at spatial infinity, and used Gauss theorem.

[3] It is easy to count that the number of independent components of a totally antisymmetric tensor $A_{\mu\nu\lambda}$: they are $\frac{4\times3\times2}{3!} = 4$, and the tensorial equation $A_{\mu\nu\lambda} = 0$ contains only 4 independent equations.

The four-potential A_μ

The second equation in (65) indicates that magnetic monopoles do not exists (this is seen more easily making explicit the role of the magnetic field, as in the second line of (6)). It can be interpreted as an integrability condition, that has a solution of the form

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{70}$$

where A_μ is called the four-potential. To verify this statement one may substitute it back into the equation and check that

$$\partial_\mu(\partial_\nu A_\lambda - \partial_\lambda A_\nu) + \partial_\nu(\partial_\lambda A_\mu - \partial_\mu A_\lambda) + \partial_\lambda(\partial_\mu A_\nu - \partial_\nu A_\mu) = 0$$

where terms cancel two by two as derivatives commute. Identifying the components of the four-potential in terms of the electric potential ϕ and vector potential $\vec{A}$ by $A^\mu = (A^0, \vec{A}) = (\phi, \vec{A})$, so that $A_\mu = (A_0, \vec{A}) = (-\phi, \vec{A})$, one obtains the usual formulas for the electric and magnetic fields in terms fo the potentials

$$\begin{aligned} E_i &= F_{i0} = \partial_i A_0 - \partial_0 A_i = -\partial_i \phi - \frac{1}{c}\partial_t A_i = \Big(-\vec{\nabla}\phi - \frac{1}{c}\frac{\partial \vec{A}}{\partial t}\Big)_i \\ B_i &= \frac{1}{2}\epsilon_{ijk}F_{jk} = \frac{1}{2}\epsilon_{ijk}(\partial_j A_k - \partial_k A_j) = \epsilon_{ijk}\partial_j A_k = (\vec{\nabla}\times\vec{A})_i \end{aligned} \tag{71}$$

i.e.

$$\vec{E} = -\vec{\nabla}\phi - \frac{1}{c}\frac{\partial \vec{A}}{\partial t}\,, \qquad \vec{B} = \vec{\nabla}\times\vec{A}\,. \tag{72}$$

Gauge symmetry

The four-potential $A_\mu(x)$ is not unique. One could also use a different potential $A'_\mu(x)$ defined by

$$A'_\mu(x) = A_\mu(x) + \partial_\mu\theta(x) \tag{73}$$

where $\theta(x)$ is an arbitrary function of spacetime, as the electromagnetic fields remains invariant

$$F'_{\mu\nu} = \partial_\mu A'_\nu - \partial_\nu A'_\mu = \partial_\mu A_\nu + \partial_\mu\partial_\nu\theta - \partial_\nu A_\mu - \partial_\nu\partial_\mu\theta = \partial_\mu A_\nu - \partial_\nu A_\mu = F_{\mu\nu}\,.$$

The transformation (73) is a *local symmetry*, a Lie symmetry with parameter $\theta(x)$ that depends in an arbitrary way on the spacetime point. It is also called *gauge symmetry*. It is related to the group $U(1)$, the group of phases, as one can write the transformation (73) in an equivalent way as

$$A'_\mu(x) = A_\mu(x) - ie^{-i\theta(x)}\partial_\mu e^{i\theta(x)} \tag{74}$$

where $e^{i\theta(x)} \in U(1)$ for each x. Electromagnetism is said to be a gauge theory based on the gauge group $U(1)$, and $A_\mu(x)$ is also called the *gauge potential*.

The generalization of this theory to more general groups (in particular to non-abelian groups) is the basis of the Standard Model of elementary particles, where the gauge group is given by the product of three independent groups, $SU(3)\times SU(2)\times U(1)$.

Action

We have noticed that the homogeneous equations (64) are solved by using the potential A_μ. With it, one is able to introduce an action functional[4]

$$S[A_\mu] = \frac{1}{c}\int d^4x\Big(-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{c}A_\mu J^\mu\Big) \ . \tag{75}$$

Imposing the least action principle, one finds the remaining Maxwell equations

$$\partial^\mu F_{\mu\nu} = -\frac{1}{c}J_\nu \ . \tag{76}$$

As we already noticed, the conservation of the current ($\partial_\mu J^\mu = 0$) is necessary for the consistency of Maxwell equations. It is also needed to preserve the gauge invariance of the action. The action is useful to study the quantization of the electromagnetic field.

8 Action for a relativistic particle and the Lorentz force

We have described the action principle for a relativistic particle in (47). We have noticed that the action must be proportional to the proper time to ensure relativistic invariance. Now, we proceed in a covariant manner using tensor calculus (to keep Lorentz symmetry manifest) at the price of introducing a local symmetry, which is associated to a change of parameterization of the worldline, as we shall see. This price is worth paying, as the manifest Lorentz invariance can be used to introduce in a simple way interactions that are guaranteed to be consistent with special relativity. In particular, we describe the coupling of a charged particle to electromagnetism deriving the Lorentz force.

Let us denote the infinitesimal proper time of a particle that travels for an infinitesimal distance dx^μ in spacetime with

$$d\tau = \frac{1}{c}\sqrt{-dx^\mu dx_\mu} \ . \tag{77}$$

Now, we rewrite the action (47) in the following form

$$S = -mc^2\int d\tau = -mc\int \sqrt{-dx^\mu dx_\mu} = -mc\int d\lambda\sqrt{-\frac{dx^\mu}{d\lambda}\frac{dx_\mu}{d\lambda}} \tag{78}$$

by using an arbitrary parameter λ for identifying the worldline $x^\mu(\lambda)$ of the particle. This parameter can be chosen at will, and this arbitrariness constitutes a local symmetry of the action (gauge symmetry)[5]. It allows to present the action in a manifestly Lorentz invariant form.

The choice of the parameter we made earlier in (47) was to use $\lambda = t$. We now choose the proper time τ to parametrize the worldline $x^\mu(\tau)$. Indicating with $\dot{x}^\mu = \frac{dx^\mu}{d\tau}$ we write the action as

$$S[x^\mu] = -mc\int d\tau\sqrt{-\dot{x}^\mu\dot{x}_\mu} \tag{79}$$

to be interpreted as a functional of the four functions $x^\mu(\tau)$. The coupling to the electromagnetic field is obtained using the 4-potential $A^\mu = (A^0, \vec{A}) = (\phi, \vec{A})$ and introducing the interaction term in the Lagrangian

$$L_{int} = \frac{e}{c}\, A_\mu(x(\tau))\, \dot{x}^\mu(\tau) \tag{80}$$

[4]The integration in spacetime can be written as $\int dt d^3x = \frac{1}{c}\int dx^0 d^3x = \frac{1}{c}\int d^4x$.

[5]Its detailed study is pedagogically useful also for learning the basics of relativistic strings, nowadays used as a model for a quantum theory of gravity.

which evidently is a scalar under Lorentz transformations, and therefore consistent with Lorentz invariance. The constant e represents the electric charge of the particle. The Lorentz invariance, manifest in the tensor formalism, has lead us to the correct interaction term with minor efforts. Other simple options, such as $A_\mu(x(\tau))\, x^\mu(\tau)$ are easily ruled out, for example by gauge invariance.

The total action

$$S[x^\mu] = \int d\tau \left[-mc\sqrt{-\dot{x}^\mu \dot{x}_\mu} + \frac{e}{c} A_\mu(x) \dot{x}^\mu \right] \tag{81}$$

can be used to find the equations of motion (using the least action principle $\delta S = 0$). One finds

$$\boxed{\frac{dp^\mu}{d\tau} = \frac{e}{c} F^{\mu\nu} \dot{x}_\nu} \tag{82}$$

where p^μ is the momentum conjugate to x^μ

$$p^\mu = \frac{mc\dot{x}^\mu}{\sqrt{-\dot{x}^2}} = m\dot{x}^\mu$$

(the last equality follows from the definition of the proper time). The spatial part of the (82) contains the usual Lorentz force. It is given here in terms of a four-force f^μ, recall eqs. (50) and (51). If we go back to using the time t rather than τ, we find the standard expression of the Lorentz force

$$\begin{aligned}
\frac{dp^i}{dt} &= \frac{e}{c} F^{i\nu} \frac{dx_\nu}{dt} = \frac{e}{c} F^{i0} \frac{dx_0}{dt} + \frac{e}{c} F^{ij} \frac{dx_j}{dt} \\
&= eE^i + \frac{e}{c} \epsilon^{ijk} B^k \frac{dx_j}{dt} = eE^i + \frac{e}{c} \epsilon^{ijk} \frac{dx_j}{dt} B^k \\
&= eE^i + \frac{e}{c} (\vec{v} \times \vec{B})^i
\end{aligned}$$

that is

$$\boxed{\frac{d\vec{p}}{dt} = e\vec{E} + \frac{e}{c} \vec{v} \times \vec{B}}\ . \tag{83}$$

Exercise: *Derive the Lorentz equations* (82) *from the least action principle* ($\delta S = 0$)*.*

We vary the coordinate functions $x^\mu(\tau) \to x^\mu(\tau) + \delta x^\mu(\tau)$ and study the variation of the action (81)

$$\delta S[x^\mu] \equiv S[x^\mu + \delta x^\mu] - S[x^\mu] = \int d\tau \left[-mc \frac{1}{2} \frac{1}{\sqrt{-\dot{x}^2}} (-2\dot{x}_\mu \delta \dot{x}^\mu) + \frac{e}{c} A_\mu(x) \delta \dot{x}^\mu + \frac{e}{c} \delta A_\mu(x) \dot{x}^\mu \right] .$$

In the first term we integrate a $\frac{d}{d\tau}$ by parts to free δx^μ from it, neglecting the total derivative (the boundary term) which vanishes by the boundary conditions, to find

$$\int d\tau \left[mc \frac{1}{\sqrt{-\dot{x}^2}} \dot{x}_\mu \delta \dot{x}^\mu \right] = -\int d\tau \frac{d}{d\tau} \left(\frac{mc\dot{x}_\mu}{\sqrt{-\dot{x}^2}} \right) \delta x^\mu .$$

The second term is treated similarly, using the chain rule

$$\int d\tau \left[\frac{e}{c} A_\mu(x) \delta \dot{x}^\mu \right] = -\int d\tau \left[\frac{e}{c} \frac{d}{d\tau} A_\mu(x) \right] \delta x^\mu = -\int d\tau \left[\frac{e}{c} \dot{x}^\nu \partial_\nu A_\mu(x) \right] \delta x^\mu .$$

The third term produces

$$\int d\tau \left[\frac{e}{c}\delta A_\mu(x)\dot{x}^\mu\right] = \int d\tau \left[\frac{e}{c}\delta x^\nu \partial_\nu A_\mu(x)\dot{x}^\mu\right] = \int d\tau \left[\frac{e}{c}\partial_\mu A_\nu(x)\dot{x}^\nu\right]\delta x^\mu$$

where we have renamed indices and factored δx^μ. Adding the three pieces we get

$$\delta S[x^\mu] = \int d\tau \left[-\frac{d}{d\tau}\left(\frac{mc\dot{x}_\mu}{\sqrt{-\dot{x}^2}}\right) + \frac{e}{c}(\partial_\mu A_\nu - \partial_\nu A_\mu)\dot{x}^\nu\right]\delta x^\mu \ .$$

Setting $\delta S[x^\mu] = 0$, and for arbitrary variations $\delta x^\mu(\tau)$, it produces the equation of motion of the particle with the Lorentz four-force

$$-\frac{d}{d\tau}\left(\frac{mc\dot{x}_\mu}{\sqrt{-\dot{x}^2}}\right) + \frac{e}{c}(\partial_\mu A_\nu - \partial_\nu A_\mu)\dot{x}^\nu = 0 \ .$$

Denoting $p^\mu = \frac{mc\dot{x}_\mu}{\sqrt{-\dot{x}^2}}$ (i.e. $p^\mu = m\dot{x}_\mu$ if τ is taken as the proper time) and $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, it becomes

$$\boxed{\frac{dp^\mu}{d\tau} = \frac{e}{c}F^{\mu\nu}\dot{x}_\nu} \ .$$

Electromagnetic current of a pointlike charge

We can write the interaction term in (81)

$$S_{int} = \int d\tau \, \frac{e}{c}A_\mu(x(\tau))\dot{x}^\mu(\tau)$$

in the form

$$S_{int} = \int d^4x \, \frac{1}{c^2}A_\mu(x)J^\mu(x)$$

which enters (75) to recognize the current density $J^\mu(x)$ due to a pointlike charge. We can use the four-dimensional Dirac delta function

$$\delta^4(x) = \delta(x^0)\delta^3(\vec{x}) = \delta(x^0)\delta(x^1)\delta(x^2)\delta(x^3)$$

to write $A_\mu(x(\tau)) = \int d^4x \, A_\mu(x)\delta^4(x - x(\tau))$ so that

$$\begin{aligned} S_{int} &= \int d\tau \left[\frac{e}{c}A_\mu(x(\tau))\,\dot{x}^\mu(\tau)\right] = \int d^4x \int d\tau \left[\frac{e}{c}A_\mu(x)\delta^4(x - x(\tau))\,\dot{x}^\mu(\tau)\right] \\ &= \int d^4x \, A_\mu(x) \, \frac{e}{c} \int d\tau \; \delta^4(x - x(\tau))\,\dot{x}^\mu(\tau) \end{aligned}$$

from which we recognize the 4-current of the point charge

$$J^\mu(x) = ec \int d\tau \; \delta^4(x - x(\tau))\,\dot{x}^\mu(\tau) \tag{84}$$

with $\dot{x}^\mu(\tau)$ its 4-velocity. To clarify it, it is useful to exploit the arbitrariness of the reparameterization of the worldline using t instead of τ to find

$$\begin{aligned} J^\mu(x) &= ec \int d\tau \; \delta^4(x - x(\tau))\frac{dx^\mu}{d\tau} = ec \int dt \; \delta^4(x - x(t))\frac{dx^\mu}{dt} \\ &= ec \int dt \; \delta(ct - x^0(t))\delta^3(\vec{x} - \vec{x}(t))\frac{dx^\mu}{dt} = e\,\delta^3(\vec{x} - \vec{x}(t))\frac{dx^\mu}{dt} \\ &= \Big(ec\,\delta^3(\vec{x} - \vec{x}(t)), \; e\,\frac{d\vec{x}}{dt}\delta^3(\vec{x} - \vec{x}(t))\Big) = (ec\,\delta^3(\vec{x} - \vec{x}(t)), \; e\vec{v}\,\delta^3(\vec{x} - \vec{x}(t))) \\ &= (c\rho, \vec{J}) \end{aligned}$$

and recognize the charge density $\rho(t,\vec{x}) = e\delta^3(\vec{x}-\vec{x}(t))$ and the current density $\vec{J}(t,\vec{x}) = e\vec{v}(t)\delta^3(\vec{x}-\vec{x}(t))$ of the pointlike particle of charge e.

Summary

The system of a pointlike charge with coordinates $x^\mu(\tau)$ interacting with the electromagnetic field $A_\mu(x)$ is described by the manifestly Lorentz invariant action

$$S[x^\mu(\tau), A_\mu(x)] = \int d\tau \left[-mc\sqrt{-\dot{x}^\mu \dot{x}_\mu} + \frac{e}{c} A_\mu(x(\tau))\dot{x}^\mu\right] - \frac{1}{4c}\int d^4x\ F_{\mu\nu}F^{\mu\nu} \tag{85}$$

which leads to the coupled equations

$$\begin{aligned} \partial_\mu F^{\mu\nu}(x) &= -e\int d\tau\ \delta^4(x - x(\tau))\,\dot{x}^\nu(\tau) \\ \frac{d}{d\tau}\left(\frac{mc\dot{x}^\mu}{\sqrt{-\dot{x}^2}}\right) &= \frac{e}{c}F^{\mu\nu}(x(\tau))\dot{x}_\nu \end{aligned} \tag{86}$$

These equations are quite complex. Most of the times one assumes the motion of the charge as given and solves for the electromagnetic field, or alternatively one assumes the electromagnetic field as given and solves for the motion of the particle. In general, the coupled problem, which includes also phenomena of radiation reaction, is much more complex.

9 More on Maxwell's equations

The free Maxwell equations (i.e. the Maxwell equations in vacuum) are given by

$$\partial_\mu F^{\mu\nu} = 0 \tag{87}$$

that in terms of A_μ become

$$\partial_\mu \partial^\mu A^\nu - \partial^\nu \partial_\mu A^\mu = 0\ . \tag{88}$$

The solutions for A_μ are not unique, even fixing suitable initial data, because of the arbitrariness related to the gauge symmetry. This arbitrariness allows one to fix suitable auxiliary conditions (gauge-fixing conditions). Given a gauge potential $A_\mu(x)$, one can find an equivalent gauge potential $A'_\mu(x) = A_\mu(x) + \partial_\mu\theta(x)$, describing the same electric and magnetic fields, with the function $\theta(x)$ chosen in such a way so that $A'_\mu(x)$ now satisfies certain properties i.e. certain gauge-fixing conditions. Dropping the prime on A_μ, let us present the most commonly used gauge choices:

• Lorenz gauge: $\partial_\mu A^\mu = 0$

which reduces the equations of motion to

$$\partial_\mu \partial^\mu A_\nu = 0 \tag{89}$$

recognized as the D'Alembert wave equation with c as the speed of the waves.

• Coulomb gauge: $A_0 = 0\ ,\ \vec{\nabla}\cdot\vec{A} = 0$

so that the equations of motion reduce again to a wave equation

$$\partial_\mu \partial^\mu A_i = 0 \tag{90}$$

with A_i constrained by the Coulomb gauge, so that only two independent solutions exist (they can be taken as the two circular polarizations of light, also known as *helicities*).

Let us introduce a useful notation for the wave operator of D'Alembert

$$\Box = \partial^\mu \partial_\mu = \eta^{\mu\nu}\partial_\mu \partial_\nu = -\partial_0\partial_0 + \nabla^2 = -\frac{1}{c^2}\frac{\partial^2}{\partial t^2} + \nabla^2 \tag{91}$$

so that the equation in the Lorenz gauge is often written as

$$\Box A_\mu = 0 \ . \tag{92}$$

Let us postpone the analysis of its plane wave solution, as it is useful to discuss first a similar equation for the case of the Klein Gordon field.

9.1 Energy-momentum tensor of the electromagnetic field

The energy-momentum tensor of the electromagnetic field in vacuum is defined by

$$\boxed{T^{\mu\nu} = F^{\mu\alpha}F^{\nu}{}_{\alpha} - \frac{1}{4}\eta^{\mu\nu}F^{\alpha\beta}F_{\alpha\beta}} \ . \tag{93}$$

It is symmetric ($T^{\mu\nu} = T^{\nu\mu}$), conserved ($\partial_\mu T^{\mu\nu} = 0$) and with vanishing trace ($T^\mu{}_\mu = 0$). It describes the density of energy and momentum. The total energy and momentum of an electromagnetic field configuration is given by

$$P^\mu = \int d^3x \ T^{0\mu} \ . \tag{94}$$

In particular, the energy density, which we will calculate shortly, is given by

$$T^{00} = \frac{1}{2}(\vec{E}^2 + \vec{B}^2) \ . \tag{95}$$

We recognize that this expression is not Lorentz invariant (while we know that $\vec{B}^2 - \vec{E}^2$ is).

Exercise: *Verify the conservation of the stress tensor*

We compute

$$\partial_\mu T^{\mu\nu} = (\partial_\mu F^{\mu\alpha})F^{\nu}{}_{\alpha} + F^{\mu\alpha}(\partial_\mu F^{\nu}{}_{\alpha}) - \frac{1}{4}\partial^\nu(F^{\alpha\beta}F_{\alpha\beta})$$

the first term vanishes because of Maxwell's equations, the remainder (lowering the index ν for simplicity and renaming indices) becomes

$$\begin{aligned}
\partial_\mu T^\mu{}_\nu &= F^{\mu\alpha}(\partial_\mu F_{\nu\alpha}) - \frac{1}{2}F^{\alpha\beta}(\partial_\nu F_{\alpha\beta}) \\
&= F^{\alpha\beta}(\partial_\alpha F_{\nu\beta}) - \frac{1}{2}F^{\alpha\beta}(\partial_\nu F_{\alpha\beta}) \\
&= \frac{1}{2}F^{\alpha\beta}\Big(2\partial_\alpha F_{\nu\beta} - \partial_\nu F_{\alpha\beta}\Big) \\
&= \frac{1}{2}F^{\alpha\beta}\Big(\partial_\alpha F_{\nu\beta} - \partial_\beta F_{\nu\alpha} - \partial_\nu F_{\alpha\beta}\Big) \\
&= -\frac{1}{2}F^{\alpha\beta}\Big(\partial_\alpha F_{\beta\nu} + \partial_\beta F_{\nu\alpha} + \partial_\nu F_{\alpha\beta}\Big) = 0
\end{aligned}$$

which vanishes because of the Bianchi identities (the homogeneous Maxwell equations).

Let's now calculate some components of the energy-momentum tensor. Remembering

$$F^{\mu\nu} = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & B_3 & -B_2 \\ -E_2 & -B_3 & 0 & B_1 \\ -E_3 & B_2 & -B_1 & 0 \end{pmatrix}$$

we write

$$T^{00} = F^{0\alpha}F^{0}{}_{\alpha} - \frac{1}{4}\eta^{00}\, 2(\vec{B}^2 - \vec{E}^2) = \vec{E}^2 + \frac{1}{2}(\vec{B}^2 - \vec{E}^2) = \frac{1}{2}(\vec{E}^2 + \vec{B}^2)$$

that corresponds to the energy density. Similarly the density of momentum is given by

$$T^{0i} = F^{0\alpha}F^{i}{}_{\alpha} = F^{0j}F^{i}{}_{j} = F^{0j}F^{ij} = E_j\epsilon_{ijk}B_k = \epsilon_{ijk}E_jB_k = (\vec{E}\times\vec{B})_i$$

known as the Poynting vector, defined by $\vec{S} = \vec{E}\times\vec{B}$.

The energy-momentum tensor $T^{\mu\nu}$ appears as a source for the gravitational equations of general relativity.

10 Klein-Gordon equation

A very useful relativistic wave equation is known as the Klein-Gordon equation. It describes the dynamics of a scalar field $\phi(x)$ and it has the form

$$(\Box - \mu^2)\phi(x) = 0 \tag{96}$$

where $\Box \equiv \eta^{\mu\nu}\partial_\mu\partial_\nu = \partial^\mu\partial_\mu = -\frac{1}{c^2}\frac{\partial^2}{\partial t^2} + \nabla^2$ is the d'Alembertian. It is manifestly a *relativistic equation*, and is used to describe possible scalar forces associated with a potential ϕ. In that respect, we will soon derive the so-called Yukawa potential, originally introduced to describe the nuclear forces that keep protons and neutrons bound inside the atomic nucleus. The parameter μ is the inverse of a length and in a quantum interpretation is related to the mass m of a scalar particle (the quanta of the field)

$$\mu = \frac{mc}{\hbar}\,, \qquad\qquad \frac{1}{\mu} = \frac{\hbar}{mc} = [length] \tag{97}$$

with the latter recognized as the Compton wavelength of a particle of mass m.

The Klein-Gordon equation can be obtained from the quantization of a relativistic particle. We have seen that for a relativistic particle with 4-momentum p^μ

$$p^\mu p_\mu = -m^2c^2 \quad\Longrightarrow\quad -\frac{E^2}{c^2} + \vec{p}^{\,2} = -m^2c^2 \quad\Longrightarrow\quad E^2 = \vec{p}^{\,2}c^2 + m^2c^4\,. \tag{98}$$

Since the energy must be positive (or, more generally, bounded from below) one finds

$$E = \sqrt{\vec{p}^{\,2}c^2 + m^2c^4}\,. \tag{99}$$

This relation could be used to find a relativistic extension of the Schroedinger equation of quantum mechanics. Applying the substitution

$$E \to i\hbar\frac{\partial}{\partial t}\,, \quad \vec{p} \to -i\hbar\vec{\nabla} \tag{100}$$

as used in the derivation of the Schroedinger equation, one gets a very complicated equation with

$$i\hbar\frac{\partial}{\partial t}\phi(\vec{x},t) = \sqrt{-\hbar^2c^2\nabla^2 + m^2c^4}\;\phi(\vec{x},t) \tag{101}$$

which is quite difficult to interpret. This equation was soon abandoned. Klein and Gordon proposed a simpler equation for the quantum mechanics of relativistic particles: using directly the quadratic relation

$$E^2 = \vec{p}^{\,2}c^2 + m^2c^4 \tag{102}$$

they obtained

$$\left(-\frac{1}{c^2}\frac{\partial^2}{\partial t^2} + \nabla^2 - \frac{m^2c^2}{\hbar^2}\right)\phi(\vec{x},t) = 0 \qquad \text{i.e.} \qquad (\Box - \mu^2)\phi(x) = 0 \tag{103}$$

known as the Klein-Gordon equation.

Solutions of the Klein-Gordon equation

Simple solutions of the Klein-Gordon equation are the plane waves

$$\phi_{\vec{p}}(x) = e^{i(\vec{k}\cdot\vec{x}-\omega t)} = e^{\frac{i}{\hbar}(\vec{p}\cdot\vec{x}-Et)} = e^{\frac{i}{\hbar}p_\mu x^\mu} \qquad \text{with} \qquad E = cp^0 = \sqrt{\vec{p}^{\,2}c^2 + m^2c^4}\;, \quad \vec{p}\in\mathbb{R}^3 \tag{104}$$

which can be related to quanta consisting of scalar particles of mass m. Here $\vec{k} = \frac{\vec{p}}{\hbar}$ is the (angular) wave-number vector, and $\omega = \frac{E}{\hbar}$ the angular frequency of the plane wave. They are related to the energy and momentum through the Einstein and de Broglie relations, namely $E = h\nu = \hbar\omega$ where the period $T = \frac{1}{\nu}$ is inverse of the frequency, and $|\vec{p}| = \frac{h}{\lambda}$ with the wavelength $\lambda = \frac{2\pi}{k}$ given by the inverse of the wave-number $\frac{k}{2\pi}$. These relations are expressed covariantly by considering the wave 4-vector $k^\mu = (k^0, \vec{k}) = (\frac{\omega}{c}, \vec{k})$ which is related to the 4-momentum by $p^\mu = \hbar k^\mu$.

The verification that the ansatz (104) solves the wave equation is immediate.

However, there is another class of solutions, those with negative energies $E = -\sqrt{\vec{p}^{\,2}c^2 + m^2c^4}$. How to interpret them? They signal the existence of *antiparticles*, particles with the same mass (and spin), but opposite charge. The existence of antiparticles is understood in general as a consequence of special relativity and quantum mechanics. The proper mathematical framework that allows for a consistent interpretation of the negative energy solutions in terms of antiparticle takes the name of "quantum field theory" (QFT).

A general solution of the free Klein-Gordon equation can be written as a linear combination (in the sense of a Fourier integral) of plane waves with both positive and negative energies

$$\phi(x) = \int \frac{d^3p}{(2\pi)^3}\frac{1}{2E_p}\Big(a(\vec{p})\,e^{-iE_pt+i\vec{p}\cdot\vec{x}} + b^*(\vec{p})\,e^{iE_pt-i\vec{p}\cdot\vec{x}}\Big) \tag{105}$$

and

$$\phi^*(x) = \int \frac{d^3p}{(2\pi)^3}\frac{1}{2E_p}\Big(b(\vec{p})\,e^{-iE_pt+i\vec{p}\cdot\vec{x}} + a^*(\vec{p})\,e^{iE_pt-i\vec{p}\cdot\vec{x}}\Big) \tag{106}$$

where we have set $\hbar = 1$, $E_p = \sqrt{\vec{p}^{\,2}c^2 + m^2c^4}$, while the factor $\frac{1}{2E_p}$ is conventional. For real fields ($\phi^* = \phi$) the different Fourier coefficients coincide, $a(\vec{p}) = b(\vec{p})$.

Yukawa potential

Let us consider the equation with an external source

$$(\Box - \mu^2)\phi(x) = J(x) \tag{107}$$

where $J(x) = g\delta^3(\vec{x})$ is the source of a pointlike charge localized at the origin of the spatial coordinates ($\delta^3(\vec{x})$ is the Dirac's delta). If we consider a time-independent solution, the equation simplifies and becomes

$$(\nabla^2 - \mu^2)\phi(\vec{x}) = g\delta^3(\vec{x}) \ . \tag{108}$$

For $\mu = 0$ one recognizes the Poisson equation of electrostatics, with solution the Coulomb potential of a pointlike charge g

$$\phi(\vec{x}) = -\frac{g}{4\pi}\frac{1}{r} \ . \tag{109}$$

For $\mu \neq 0$ the solution can be computed using the Fourier transform and is known as the Yukawa potential

$$\phi(\vec{x}) = \int \frac{d^3p}{(2\pi)^3}\,\frac{(-g)}{\vec{p}^2 + \mu^2}\, e^{i\vec{p}\cdot\vec{x}} = -\frac{g}{4\pi}\frac{e^{-\mu r}}{r} \ . \tag{110}$$

Even if one is not familiar with the techniques of Fourier transforms, one may verify that it satisfies the equation for $r \neq 0$ (it is enough to use the laplacian written in spherical coordinates, $\nabla^2 = \frac{1}{r^2}\partial_r r^2 \partial_r$ + derivatives on the angles). Furthermore, the singular behavior at $r = 0$ is related to the intensity of the pointlike charge, as dictated by the Coulomb case that emerges in the limit $\mu \to 0$. See fig. 12 for a graphical sketch of the Yukawa and Coulomb potential.

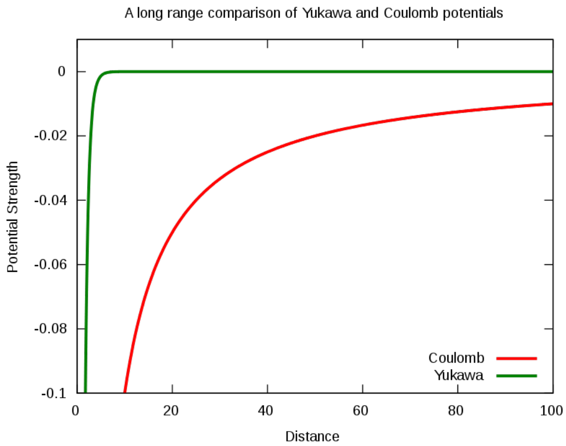


Figure 12: *Comparison between the Yukawa potential and the Coulomb potential.*

11 Electromagnetic waves

After having described the Klein-Gordon equation, let us return to the study of the electromagnetic waves. The introduction of the four-potential A_μ solves half of the Maxwell equations. The remaining ones in vacuum take the form

$$\partial^\mu F_{\mu\nu} = \partial^\mu(\partial_\mu A_\nu - \partial_\nu A_\mu) = 0 \tag{111}$$

and are gauge invariant under the gauge transformation

$$\delta A_\mu = \partial_\mu \theta \tag{112}$$

with θ an arbitrary function of spacetime. The gauge freedom allows to require the Lorenz gauge for the potential, $\partial^\mu A_\mu = 0$, and in this gauge the equations simplify to

$$\begin{aligned}\Box A_\mu &= 0 \\ \partial^\mu A_\mu &= 0 \ .\end{aligned} \tag{113}$$

Plane wave solution are found using the ansatz (up to an overall normalization) by setting

$$A_\mu(x) = \epsilon_\mu(k)\, e^{ik\cdot x} + c.c. \tag{114}$$

where $\epsilon_\mu(k)$ is an arbitrary polarization depending on the wave vector k^μ, and the exponent contains the Lorentz invariant phase $k \cdot x = k_\mu x^\mu = \eta_{\mu\nu} k^\mu x^\nu = -k^0 x^0 + \vec{k} \cdot \vec{x}$. The notation *c.c.* stands for complex conjugation, and makes the solution real. Plugging this ansatz into the equations (113) allows one to find a solution if

$$k^\mu k_\mu = 0 \ , \qquad k^\mu \epsilon_\mu(k) = 0 \ . \tag{115}$$

The second equation says that a linear combination of the possible four independent polarization vectors vanishes. Thus, only three polarizations $\epsilon_\mu(k)$ are possible. However, one of these polarizations is not physical, the one with $\epsilon_\mu(k) \sim k_\mu$. It does not carry any electric and magnetic fields, and thus no energy and momentum. It can be removed by a gauge transformation. The gauge transformations that removes it has the form (112) with

$$\theta(x) \sim e^{ik\cdot x} \tag{116}$$

that satisfies $\Box\theta(x) = 0$ and thus does not ruin the Lorenz gauge condition. The gauge transformation becomes

$$\delta A_\mu = \partial_\mu \theta \sim i k_\mu\, e^{ik\cdot x} \tag{117}$$

and shows that the polarization $\epsilon_\mu(k) \sim k_\mu$ is not physical, as can be removed by an appropriate gauge transformation. Thus, only two physical polarizations remain.

Let us exemplify this considering the motion along the z axis. We can take (setting $c = 1$)

$$k^\mu = (k^0, \vec{k}) = (\omega, 0, 0, \omega) \tag{118}$$

which solves $k^\mu k_\mu = 0$ and produces the phase $e^{ik\cdot x} = e^{i\omega(z-t)}$. It describes a wave moving along the z axis. The two expected polarizations can then be taken as given by

$$\begin{aligned}\epsilon^1_\mu &= (0, 1, 0, 0) \\ \epsilon^2_\mu &= (0, 0, 1, 0)\end{aligned} \tag{119}$$

which indeed satisfy

$$k^\mu \epsilon^i_\mu = 0 \ , \qquad \epsilon^i_\mu \neq \alpha k_\mu \ . \tag{120}$$

Considering for example the solution with ϵ^1_μ, plugging it into (114), and multiplying with an arbitrary amplitude A_0 one finds

$$\begin{aligned}\vec{A} &= A_0 \cos(\omega z - \omega t)\, \hat{x} \\ \vec{E} &= -\frac{\partial \vec{A}}{\partial t} = E_0 \sin(\omega z - \omega t)\, \hat{x} \\ \vec{B} &= \vec{\nabla} \times \vec{A} = B_0 \sin(\omega z - \omega t)\, \hat{y}\end{aligned} \tag{121}$$

where $E_0 = B_0 = \omega A_0$, and $\hat{x}, \hat{y}, \hat{z}$ are the usual unit vectors.

The above plane waves do not carry angular momentum. Plane waves carrying angular momentum are obtained using the circular polarization defined by

$$\epsilon^{\pm}_{\mu} = \epsilon^{1}_{\mu} \pm i\epsilon^{2}_{\mu} \, . \tag{122}$$

They are also said to correspond to the helicity $h = \pm 1$. In a quantum interpretation they are related to photons carrying angular momentum $\pm\hbar$ along the direction of motion (helicity), with a wavefunction of the form

$$A_{\mu}(x) = \epsilon^{\pm}_{\mu}(k) e^{ik\cdot x} = \epsilon^{\pm}_{\mu}(k) e^{\frac{i}{\hbar} p\cdot x} \tag{123}$$

where $p^{\mu} = \hbar k^{\mu}$ is the 4-momentum of the photon.

A Appendix: Definition of a group

The concept of *group* gives the appropriate mathematical language for describing the symmetries of a physical system.

A group G is defined as a set of elements, $G = \{g\}$, that satisfies the following properties:

1) *composition* law: for any two elements $g_1, g_2 \in G$ then $g_1 \cdot g_2 = g_3 \in G$,

2) *identity* element: $\exists e \in G$ such that $g \cdot e = e \cdot g = g$ for any $g \in G$

3) *inverse* element: if $g \in G$ then $\exists g^{-1} \in G$, such $g \cdot g^{-1} = g^{-1} \cdot g = e$,

4) *associativity*: for any three elements $g_1, g_2, g_3 \in G$ then $(g_1 \cdot g_2) \cdot g_3 = g_1 \cdot (g_2 \cdot g_3)$.

Discrete groups are those groups that contain a finite number of elements, for example $Z_2 \equiv \{1, -1\}$ with the usual law of multiplication defines a group with two elements. *Lie groups* are those groups whose elements depend continuously on some parameters, for example the rotations around the z axis form a Lie group whose elements are parametrized by an angle $\phi \in [0, 2\pi]$. The *abelian groups* are those groups whose elements commute under the law of composition: $g_1 \cdot g_2 = g_2 \cdot g_2$ for any $g_1, g_2 \in G$. If that does not happen, the group is said to be *non-abelian*.

Examples of discrete groups are:

• the cyclic group Z_n, generated by the powers of an element a of the group, $Z_n = \{e, a, a^2, ..., a^{n-1}\}$, with the relation $a^n = a^0 = e$ (the group is isomorphic to the n-th root of unity, $e^{\frac{2\pi i}{n}k}$ with $k = 0, 1, ..., n-1$);

• the group of permutations of n elements, denoted by S_n, which contains $n!$ elements.

Examples of Lie groups are:

• $O(N)$, the orthogonal group, i.e. the group of real, orthogonal, $N \times N$ matrices, describing the invariances of the scalar product $x^T x$ with $x \in \mathbb{R}^N$;

• $SO(N)$, the special orthogonal group, i.e. the group of real, orthogonal, $N \times N$ matrices with unit determinant.

• $U(1) = \{z \in \mathbb{C} \mid |z| = 1\} = \{e^{i\theta} \mid \theta \in [0, 2\pi]\}$, describing the invariances of $w^* w$ with $w \in \mathbb{C}$). It is isomorphic to $SO(2)$.