PREPARED WITH THE JHEP LaTeX STYLE Hilary Term 2025/2026

Special Relativity

University of Oxford, Mathematics, Part B course

Luis Fernando Alday

Mathematical Institute
Radcliffe Observatory Quarter
Woodstock Road
Oxford, OX2 6GG, United Kingdom

alday@maths.ox.ac.uk

Course Information

These are lecture notes for the part B Special Relativity course given at the Mathematical Institute, University of Oxford.

The official course overview reads as follows:

> The first part of course sets the basic principles of special relativity. Constancy of the speed of light. Lorentz transformations; time dilation, length contraction, the relativistic Doppler effect. Index notation, four-vectors, four-velocity and four-momentum; equivalence of mass and energy; particle collisions and four-momentum conservation; equivalence of mass and energy; four-acceleration and four-force. The second part of the course is devoted to the consequences of special relativity: electromagnetism and relativity; and relativistic observations in astronomy and cosmology.

And the "learning outcomes" are describe as:

> Students will be able to describe the fundamental phenomena of relativistic physics within the algebraic formalism of four-vectors. They will be able to solve simple problems involving Lorentz transformations. They will acquire a basic understanding of how the four-dimensional picture completes and supersedes the physical theories studied in first and second-year work. They will also understand how this new picture explains basic observations in astronomy and cosmology.

The material shall be presented in a slightly different order than the one in the course overview. In particular, relativistic observations in astronomy and cosmology will be presented before going into the full detail of the structure of space-time in four dimensions. This will hopefully make the basic concepts more clear through specific applications. Students in the course are not assumed to be well versed in Maxwell's equations. These will be presented in a self-contained manner and discussed in relation to Special relativity. However, students are strongly encouraged to take this course together with the part B Electromagnetism course.

Resources

All the presented material has been borrowed from other sources. In particular, I have heavily borrowed from the lecture notes by Chris Beem, for a part A course, of which the present notes are an expanded version, together with the following two books.

- N. M. J. Woodhouse, *Special Relativity*, Springer (2002).

The primary textbook reference for this course. I strongly advise students to read this book. If you are unfamiliar with electromagnetism, you can mostly skip the corresponding chapters without missing too much in what follows.

- A. P. French, *Special Relativity*, (MIT Introductory Physics Series).

Classic book of Special Relativity used in various Physics courses. It contains plenty of experimental examples leading to/confirming special relativity.

In addition, anyone learning special relativity should at least *look* at Einstein's original papers (or their translations into English),

- A. Einstein, *On the electrodynamics of moving bodies,* Annalen Phys. 17 (1905) 891–921.
- A. Einstein, *Does the inertia of a body depend on its energy content?,* Annalen Phys. 18 (1905) 639-–641.

You can also read a synthesized treatment of the subject (along with the general theory of relativity) in the 1916 book by the same author,

- A. Einstein, *Relativity: The Special and General Theory*, Henry Holt and Company (1920).

Another excellent set of lectures notes are those by David Tong at the University of Cambridge. Special relativity is more conceptually challenging than it is technically difficult. For this reason it may be useful to look at a variety of different presentations of the material.

Contents

Introduction: Relativity as Reality[1]

The theory of relativity, as formulated by Albert Einstein (1879-1955), has had a dramatic impact since its inception in 1905. It has fundamentally transformed our understanding of space and time. Perhaps only Darwin's theory of evolution by natural selection has made such an impression on the public as an example of how scientific analysis can explode accepted assumptions. This revolution in thought came to its climax against the background of the catastrophe of the First World War. Between the wars, relativity was denounced by the dictatorships. But it has outlasted them.

Crucially, relativity describes the real world. This may not be obvious from the words which form the title of this course. Both the words 'special' and 'relativity' are opaque, even misleading. 'Special' means in practice that it doesn't include gravity, which needs 'General' relativity. (Later, we shall see a more precise definition of this distinction). But even the word 'Relativity' has lent itself to the notion that it claims that 'everything is relative' and has something to do with a twentieth century loss of certainty. Actually, relativity is based solidly on real measurements, more solid and consistent than anything before. In several ways it gives a more *absolute* account of physical quantities than pre-1905 physics could.

To get started on this, we will go back to the basis of Prelims mechanics: Newton's laws of motion. It is essential to note that these laws only holds in an *inertial coordinate system*, or ICS, which can be characterised as the coordinates of a non-accelerating observer. All the Prelims work rested on this idea, and we are not going to change it! However, we are going to put a new emphasis on the idea that *all such ICS's are equally valid*, and that none of them is preferred.

Galileo and Newton were familiar with this idea. Apart from anything else, Galileo had to explain how it could be that the earth is in rapid motion around the sun, and yet we feel unaware of this and live our lives as if it were fixed in space. Galileo gave a vivid illustration

[1]Introduction taken almost wholesale from Andrew Hodges' 2014-2015 lecture notes.

of the principle by describing how fish, swimming in a bowl on board a ship, seem unaffected by the motion of the ship.

This is the Galilean principle of relativity, embodied in Prelims mechanics. But you may already have been struck by two things which contradict it:

- The Lorentz force law: A particle of charge q moving with velocity $\mathbf{v}$ through an electric field of strength $\mathbf{E}$ and a magnetic field of strength $\mathbf{B}$ is subject to a force

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \wedge \mathbf{B}) \ . \tag{0.1}$$

  In which ICS should the velocity be measured in?

- You have probably encountered the statement that the speed of light is c, a fundamental constant. Speed compared with what? Measured in which ICS?

1 Newtonian Mechanics

1.1 Relativity in Newtonian mechanics

In order to describe the motion of a particle in Newtonian mechanics it is necessary to first choose a reference frame.

Definition 1. In Newtonian mechanics, a *reference frame* is a choice of spatial origin, together with a set of perpendicular (right handed) Cartesian coordinate axes.

We will usually denote a reference frame by O, from observer. We imagine the observer O as defining the spatial origin of the frame, so that locations are measured with respect to such an observer. Given such a reference frame, the position, velocity and acceleration of a particle at time t, relative to that reference frame, will then be given by vectors $\mathbf{r}, \mathbf{v}, \mathbf{a}$ whose components are given by

$$\begin{aligned}\mathbf{r} &= (x(t), y(t), z(t)), \\ \mathbf{v} &= (\dot{x}(t), \dot{y}(t), \dot{z}(t)), \\ \mathbf{a} &= (\ddot{x}(t), \ddot{y}(t), \ddot{z}(t)),\end{aligned}$$

where a dot denotes derivative with respect to time. The law of inertia, or Newton's first law, says that there exists a class of frames of reference relative to which the motion of a particle not subject to any force is in a straight line at constant speed.

Definition 2. A frame in which the law of inertia holds is called an *inertial frame*. An inertial frame, together with a choice of zero for time, determines an inertial coordinate system (ICS).

Furthermore, in an inertial frame we also have Newton's second law: the acceleration of a particle of inertial mass m is related to the force acting on it by

$$\mathbf{F} = m\mathbf{a}. \tag{1.1}$$

This equation is non-trivial if the force can be independently determined by other means. Note that in Newtonian mechanics the inertial mass is a property of the particle, independent of motion and the same in any inertial frame.

Consider now two inertial frames, O and O', which agree in their choice of time for simplicity. The location of a particle in their respective coordinate systems is given by $(x(t), y(t), z(t))$ and $(x'(t), y'(t), z'(t))$, related in general by

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = H \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} + T, \tag{1.2}$$

where H is a 3×3 proper orthogonal matrix

$$H^T H = H H^T = 1, \quad \det H = 1,$$

representing a rotation of the axes, while T is a column vector, representing a translation. In principle H and T can be arbitrary (smooth) functions of time. Suppose now that the motion of the particle is such that $\ddot{x} = \ddot{y} = \ddot{z} = 0$, and hence no force is acting on it. Since O' is also an inertial frame, we also require $\ddot{x}' = \ddot{y}' = \ddot{z}' = 0$. This statement is true for a general (rectilinear) trajectory provided

$$\dot{H} = 0, \quad \ddot{T} = 0,$$

so that the two ICSs O, O' are related by (1.2) with constant $H, \dot{T}$. We can combine this transformation with a shift in the origin of the time coordinate so that

$$t = t' + \text{constant}.$$

The final result is a *Galilean transformation*, the general transformation between two inertial frames in Newtonian mechanics. Ensembling time and space coordinates into four-component vectors we have the following definition

Definition 3. A Galilean transformation is a coordinate transformation of the form

$$\begin{pmatrix} t \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ v_1 & H_{11} & H_{12} & H_{13} \\ v_2 & H_{21} & H_{22} & H_{23} \\ v_3 & H_{31} & H_{32} & H_{33} \end{pmatrix} \begin{pmatrix} t' \\ x' \\ y' \\ z' \end{pmatrix} + T, \tag{1.3}$$

where $v = (v_i)$ is a constant column vector of length 3, $H = (H_{ij})$ a constant 3×3 proper orthogonal matrix, and T a constant column vector of length 4.

Galilean transformations preserve Newton's laws, in the sense that if two coordinate systems are related by a Galilean transformation, then the laws hold in one system if and only if they hold in the other.[2] The Galilean principle of relativity is that, as far the laws of mechanics are concerned, all inertial frames are on an equal footing. In particular, Newton's laws hold in all inertial frames and we cannot distinguish any frame as 'at rest'.

Every Galilean transformation is a combination of a rotation, a boost and a translation, which correspond to the following special cases:

- Rotations. They correspond to $T = 0$ and $v = 0$ and are parametrised by H. The frames are at rest relative to each other, while the axes are rotated.
- Boosts. They correspond to $H = 1$ and $T = 0$ and are parametrised by v. The axes are parallel to each other and the origins coincide at $t = t' = 0$. The frame R' moves with constant velocity (v_1, v_2, v_3) relative to R.
- Translations. They correspond to $H = 1$ and $v = 0$ and are parametrised by T. The coordinate systems are related by a translation of the origin and a shift of the zero of t.

1.2 Galilean space-time

The first notion we introduce in order to discuss the structure of Galilean space-time is the notion of an event.

Definition 4. An *event* is a specific point in space at a specific moment in time.

In a given coordinate system, an event is labelled by a time t and spatial coordinates (x, y, z). In another coordinate system the same event will be labelled by a time t' and spatial coordinates (x', y', z'). Under changes of coordinates, the representation of an event may change - and for inertial frames they will be related by a Galilean transformation - but the event itself is an invariant notion, independent of the given choice of coordinate system.

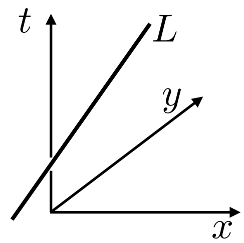
Given two events A, B, the following statements are invariant, in the sense that if they are true in an inertial frame, then they are also true in all other inertial frames.

- A and B are simultaneous.
- B happens time t after A.
- A and B are simultaneous and separated by a distance D.

On the other hand, statements like "A and B happen in the same place (at different times)" or "A and B, happening at different times, are separated by a distance D" are not invariant. They can be true in one frame, but false in others (convince yourself this is the case).

Next, we introduce the notion of space-time. Taken together, the set of all events constitutes the entirety of space and time

[2] We are assuming that the force between particles is a function of the relative position among the particles.



Definition 5. *Space-time* is the set of all possible events.

As a topological space we can think of Galilean space-time as the direct product of three-dimensional space with an additional time dimension, $\mathbb{R}_t \times \mathbb{R}^3$. This gives us, for example, a notion of continuous paths in space-time.

Definition 6. The continuous path through space-time traced out by a particle is its *world-line*.

As for the case of events, we think of space-time and world-lines, as something independent of a given choice of coordinates. Once we choose a given (inertial) frame, we can then draw events and world-lines by using a set of axis, with time commonly pointing upwards, in what we call a space-time diagram. For instance, the world-line of a free particle is then simply a straight-line. See figure.

1.3 Conservations laws in Newtonian mechanics

A fundamental result in Newtonian dynamics is the existence of conserved quantities. We will consider 'scattering processes'. In a scattering process we imagine a set of incoming particles, well separated from each other, with masses m_i and velocities $\mathbf{v}_i$, with $i = 1, \cdots, k$, with respect to an inertial frame. These particles then interact in a bounded region of space and result in a set of outgoing particles, also well separated from each other, with masses m_i and velocities $\mathbf{v}_i$, with $i = k+1, \cdots, n$. The number of incoming and outgoing particles is not necessarily the same, as one may consider processes where two particles join into one, or one fragments into two. In a scattering process we have the following three conservation laws

- Conservation of mass.
$$\sum_{i=1}^{k} m_i = \sum_{j=k+1}^{n} m_j.$$

- Conservation of linear momentum.
$$\sum_{i=1}^{k} m_i \mathbf{v}_i = \sum_{j=k+1}^{n} m_j \mathbf{v}_j.$$

- Conservation of energy. If we consider elastic scattering, then the kinetic energy is conserved, so that
$$\frac{1}{2}\sum_{i=1}^{k} m_i v_i^2 = \frac{1}{2}\sum_{j=k+1}^{n} m_j v_j^2.$$

Except the mass, note that the total linear momentum and kinetic energy are not the same for every inertial frame, but once we have chosen an inertial frame, the claim is they are the same before and after the collision.

2 From Galileo and Newton to Einstein

To recap, in Newtonian mechanics we have the following principles.

- There is a preferred class of frames of reference - the inertial frames.
- A frame is inertial if and only if Newton's first law holds for every particle not influenced by any forces.
- An inertial frame, together with a choice of zero for t, determines an inertial coordinate system t, x, y, z.
- Newton's laws hold for any mechanical system in an inertial frame. All inertial frames are on equal footing.
- The inertial coordinate systems of two inertial frames are related by a Galilean transformation.
- Mass is independent of motion, and is conserved.

We will now discuss strong evidence that suggests that, while Newtonian mechanics gives an excellent description of Nature in most 'reasonable' circumstances, it is not universally valid. In particular, it needs improving when discussing very fast particles. Fast compared to what? let's see.

2.1 Maxwell's equations

In the following we present the basic elements of Maxwell's theory of electromagnetism. The discussion will be self-contained but brief. For a (much more) detailed account please see the part B course on electromagnetism, which students are very encouraged to take. The theory is built on two basic objects:

- Charged particles.
- Electric and magnetic fields. Vector quantities $\mathbf{E}$ and $\mathbf{B}$ that depend on position and time.

The charge e of a particle, which can be positive or negative, is an intrinsic property of the particle. It determines the strength of the particle interaction with the electric and magnetic fields. In particular, a particle of charge e and velocity $\mathbf{v}$ in the presence of electric and magnetic fields $\mathbf{E}$ and $\mathbf{B}$ is exerted a force given by the Lorentz law

$$\mathbf{F} = e(\mathbf{E} + \mathbf{v} \wedge \mathbf{B}),$$

where $\mathbf{E}, \mathbf{B}$ are measured at the position of the particle. Compare this to the force $\mathbf{F} = m\mathbf{g}$ felt by a particle of mass m by a gravitational field $\mathbf{g}$. As mentioned in the introduction, the Lorentz force law already raises some questions: with respect to which inertial frame should we measure the particle's velocity? In the classical formulation of electromagnetism one assumes that there is a preferred frame 'at rest' called the luminiferous aether, and we are measuring things respect to this frame. We shall return to this point.

The Lorentz force law describes how charged particles react to fields. In particular, we can use a moving charge to 'measure' the electric and magnetic fields at different points in space. On the other hand, charges also produce electric and magnetic fields. A stationary point charge e generates an electric field, but no magnetic field, given by

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0}\frac{e\mathbf{r}}{r^3} \tag{2.1}$$

where $\mathbf{r}$ is the position vector from the charge and $r = \mathbf{r}$ is the distance to the charge. ϵ_0 is a positive constant. A point charge moving with velocity $\mathbf{v}$ will in addition generate a magnetic field, given by

$$\mathbf{B} = \frac{\mu_0}{4\pi}\frac{e\mathbf{v} \wedge \mathbf{r}}{r^3}, \tag{2.2}$$

where μ_0 is a second positive constant. In SI units the charge is measured in Coulombs, the magnetic field in Teslas and the electric field in volts per metre. In these units

$$\epsilon_0 = 8.9 \times 10^{-12}, \quad \mu_0 = 1.3 \times 10^{-6}. \tag{2.3}$$

In addition we have the principle of superposition: the magnetic and electric fields generated by a collection of charges is the sum of the electric and magnetic fields generated by the individual charges. In order to write down the Maxwell's equation it will be important to pass to a continuum limit, where we define charge and current density, for moving charges inside a small volume V by

$$\rho = \lim_{V\to 0}\frac{\sum_i e_i}{V}, \quad \mathbf{J} = \lim_{V\to 0}\frac{\sum_i e_i\mathbf{v}_i}{V}$$

The electric and magnetic fields generated by a continuous distribution of charge ρ and current density $\mathbf{J}$, both generally a function of position and time, satisfy the Maxwell's equations

Maxwell's equations

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \tag{2.4}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{2.5}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{2.6}$$

$$\nabla \times \mathbf{B} = \mu_0 \left(\mathbf{J} + \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \right) \tag{2.7}$$

The derivation of the Maxwell's equations from (2.1) and (2.2) plus the superposition principle is non-trivial and one needs to make some extra assumptions. On the other hand, these equations work extremely well in nature, even in situations where the speed of the particles involved is very high, and as we will see, are consistent in a very non-trivial way.

2.2 Consistency of the Maxwell's equations

For a given charge and current distributions $\rho, \mathbf{J}$, Maxwell's equations give eight equations (two vector equations plus two scalar equations) for six unknowns. This is an over-constrained system, and usually over-constrained systems do not have a solution, unless something special happens. Taking the divergence of the third equation we obtain

$$\frac{\partial}{\partial t} (\nabla \cdot \mathbf{B}) = 0, \tag{2.8}$$

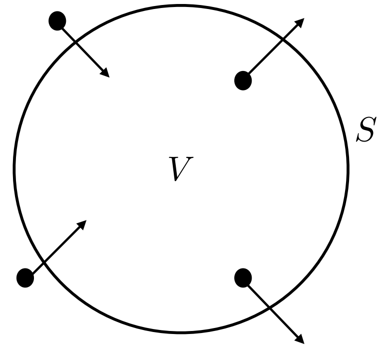
which is consistent with the second equation. However, by taking the divergence of the fourth equation and substituting in the first equation we obtain

$$\begin{aligned} 0 &= \nabla \cdot \nabla \times \mathbf{B} \\ &= \mu_0 \epsilon_0 \frac{\partial}{\partial t} \nabla \cdot \mathbf{E} + \mu_0 \nabla \cdot \mathbf{J} \\ &= \mu_0 \left(\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} \right). \end{aligned}$$

Hence the Maxwell's equations only have a solution provided

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0 \tag{2.9}$$

This is actually true! and it is equivalent to the statement that charge is neither created nor destroyed. Indeed, consider a region V with surface S with charges moving through.



The total charge inside V is $Q = \int_V \rho dV$, so that

$$\begin{aligned}
\frac{dQ}{dt} &= \int_V \frac{\partial \rho}{\partial t} dV \qquad \text{allowing } \tfrac{\partial \rho}{\partial t} \neq 0 \text{ for the moment} \\
&= \text{rate of increase of } Q \\
&= \text{rate charge goes in} - \text{rate charge goes out} \\
&= -\int_S \mathbf{J} \cdot d\mathbf{S} \\
&= -\int_V \nabla \cdot \mathbf{J} dV \qquad \text{by divergence theorem}
\end{aligned}$$

and so

$$\int_V \left(\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} \right) dV = 0 \tag{2.10}$$

this is to be true for all regions V, then

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0 \tag{2.11}$$

which is the *charge conservation equation* or continuity equation.

2.3 Electromagnetic potentials

The Maxwell's equations can be written in terms of potentials. Assume we are interested in a suitable (simply-connected, etc) region of space, then the Maxwell equations imply, first:

$$\nabla \cdot \mathbf{B} = 0 \Rightarrow \exists \mathbf{A} \text{ such that } \mathbf{B} = \nabla \times \mathbf{A};$$

next $\frac{\partial \mathbf{B}}{\partial t} = \nabla \times \frac{\partial \mathbf{A}}{\partial t}$, so that

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = \nabla \times \left(\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} \right) = 0,$$

whence

$$\exists \phi \text{ such that } \mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} = -\nabla \phi$$

i.e.

$$\mathbf{B} = \nabla \times \mathbf{A} \tag{2.12}$$

$$\mathbf{E} = -\nabla \phi - \frac{\partial \mathbf{A}}{\partial t} \tag{2.13}$$

$\mathbf{A}$ and ϕ are denoted the electromagnetic potentials. Note that we have the freedom to modify them slightly without changing the electric and magnetic fields. This freedom has the name of *gauge transformation*:

$$\mathbf{A} \to \mathbf{A} + \nabla \zeta \tag{2.14}$$

$$\phi \to \phi - \frac{\partial \zeta}{\partial t} \tag{2.15}$$

This *gauge transformation* can be exploited to simplify the potentials. Plugging the expression for the fields in terms of their potentials into the Maxwell equations we obtain

$$\nabla^2 \phi + \frac{\partial}{\partial t} \nabla \cdot \mathbf{A} = -\frac{1}{\epsilon_0} \rho \tag{2.16}$$

$$\nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} - \nabla \left(\nabla \cdot \mathbf{A} + \frac{1}{c^2} \frac{\partial \phi}{\partial t} \right) = -\mu_0 \mathbf{J} \tag{2.17}$$

where we have introduced $\frac{1}{c^2} = \epsilon_0 \mu_0$. This is a nice set of equations for the potentials, but they are coupled. You can show that the gauge transformations can be used in order to choose potentials such that

$$\nabla \cdot \mathbf{A} + \frac{1}{c^2} \frac{\partial \phi}{\partial t} = 0 \tag{2.18}$$

For this particular choice we obtain

$$\nabla^2 \phi - \frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2} = -\frac{1}{\epsilon_0} \rho \tag{2.19}$$

$$\nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mu_0 \mathbf{J} \tag{2.20}$$

These equations, supplemented with our gauge choice (2.18) are completely equivalent to Maxwell's equations.

2.4 Light

We will look for solutions to the source-free Maxwell's equations:

$$\nabla \cdot \mathbf{E} = 0, \qquad \nabla \cdot \mathbf{B} = 0$$
$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \qquad \nabla \times \mathbf{B} = \frac{1}{c^2}\frac{\partial \mathbf{E}}{\partial t}$$

We would like to decouple the equations for $\mathbf{E}$ and $\mathbf{B}$. Calculate

$$\begin{aligned} \nabla \times (\nabla \times \mathbf{E}) &= \nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\nabla^2 \mathbf{E} && (2.21) \\ &= -\nabla \times \frac{\partial \mathbf{B}}{\partial t} = -\frac{\partial}{\partial t}\nabla \times \mathbf{B} = -\frac{1}{c^2}\frac{\partial^2 \mathbf{E}}{\partial t^2} && (2.22) \end{aligned}$$

so that

$$\nabla^2 \mathbf{E} - \frac{1}{c^2}\frac{\partial^2 \mathbf{E}}{\partial t^2} = 0, \tag{2.23}$$

which is the *wave-equation* with speed $c = (\epsilon_0 \mu_0)^{-1/2}$. Similarly (exercise)

$$\nabla^2 \mathbf{B} - \frac{1}{c^2}\frac{\partial^2 \mathbf{B}}{\partial t^2} = 0. \tag{2.24}$$

We now look for solutions to those equations. Consider first the scalar version

$$\nabla^2 F - \frac{1}{c^2}\frac{\partial^2 F}{\partial t^2} = \frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial y^2} + \frac{\partial^2 F}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 F}{\partial t^2} = 0 \tag{2.25}$$

and try $F = f(\mathbf{k}\cdot\mathbf{r} - wt)$, with $\mathbf{r} = (x, y, z)$ and constant $w, \mathbf{k}$. Plugging this into the equation we obtain a solution provided $w^2 = c^2 k^2$. Note that F is constant on surfaces

$$\mathbf{k} \cdot \mathbf{r} - wt = \text{const}$$

at a fixed time t this is the equation of a plane, so these solutions are called plane waves. As t increases the plane wave fronts move in the direction of $\mathbf{k}$, with speed c. If f is of the form

$$f(\mathbf{k} \cdot \mathbf{r} - wt) \sim e^{i(\mathbf{k}\cdot\mathbf{r} - wt)}$$

then the waves are called *harmonic waves* with a single frequency, or monochromatic plane waves. Both the real and imaginary parts of the exponential above are solutions of the wave equation. Inspired by this discussion we now try

$$\begin{aligned} \mathbf{E} &= \mathbf{e} E_0 e^{i(\mathbf{k}\cdot\mathbf{r} - wt)} \\ \mathbf{B} &= \mathbf{b} B_0 e^{i(\mathbf{k}\cdot\mathbf{r} - wt)}, \end{aligned}$$

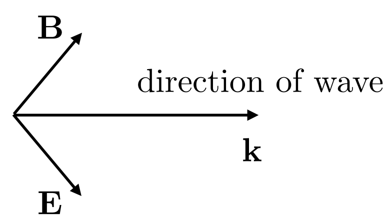
with the understanding that the electric and magnetic fields are the real part of the solutions. Here $\mathbf{e}$ and $\mathbf{b}$ are unit vectors characterising the directions of the fields and E_0, B_0 are constants. Plugging these expressions into $\nabla \cdot \mathbf{E} = 0$ and $\nabla \cdot \mathbf{B} = 0$ we obtain

$$\mathbf{e} \cdot \mathbf{k} = \mathbf{b} \cdot \mathbf{k} = 0.$$

This means both $\mathbf{E}$ and $\mathbf{B}$ are perpendicular to the direction of propagation of the wave. Such a wave is called a transverse wave. The other two Maxwell's equations imply

$$(\mathbf{k} \times \mathbf{e})\, E_0 - w\mathbf{b}B_0 = 0,$$
$$(\mathbf{k} \times \mathbf{b})\, B_0 + \frac{1}{c^2}\mathbf{e}E_0 = 0.$$

These imply the expected relation $c^2k^2 = w^2$, show that $\mathbf{e}$ and $\mathbf{b}$ are perpendicular to each other and fix the relative magnitudes $B_0 = \frac{1}{c}E_0$. See figure.



These plane waves are nothing but light. Plugging the values for ϵ_0 and μ_0 we see that Maxwell's theory of electromagnetism predicts $c \sim 3 \times 10^8 m/s$, which exactly agrees with the observed value for the velocity of light!

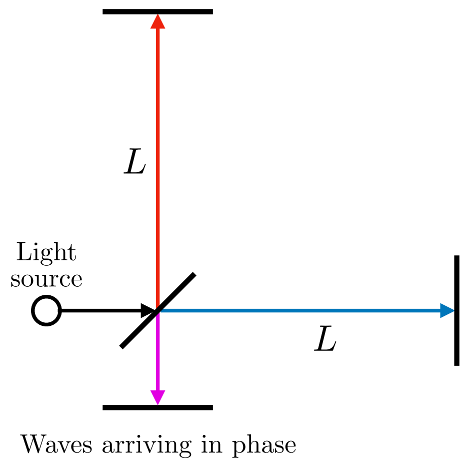
The solution we have found describes a wave with a single, precise, frequency, and hence delocalised in space, so that it evenly spreads over all space. It is sometimes convenient to see light as made out of *photons*: lumps of lights, localised in space, obtained by super-imposing plane waves around a certain frequency w, which we associate to the frequency of the photon. Photons are then (massless) particles that move to the speed of light. In quantum mechanics a photon of frequency w carries an energy $E = \hbar w$, with $\hbar = 1.05 \times 10^{-34} m^2 kg/s$.

It follows from Maxwell's equations that photons, the particles of light, move with velocity c, whatever direction they travel. This is very puzzling. In Newtonian/Galilean mechanics velocity is additive. If we have two frames of reference in relative motion, then the velocity of a particle relative to the first is the vector sum of its velocity relative to the second and the velocity of the second frame relative to the first frame. This should apply also to photons, but this means that Maxwell's equations could not hold in all frames. If they did, then the speed of photons would be the same in all frames.

As a way out of this puzzle, people at that time postulated the existence of a luminiferous aether: a medium that defines an absolute rest, filling the whole universe, and in which light propagates at speed c. The earth moves in this aether (as it goes around the sun) but its velocity is very small compared to that of light, for this effect to be measurable. This changed with the Michelson-Morley experiment.

2.5 The Michelson-Morley Experiment

The idea of the Michelson-Morley experiment is the following. Imagine first we are in a system which is at rest respect to the aether, see figure.

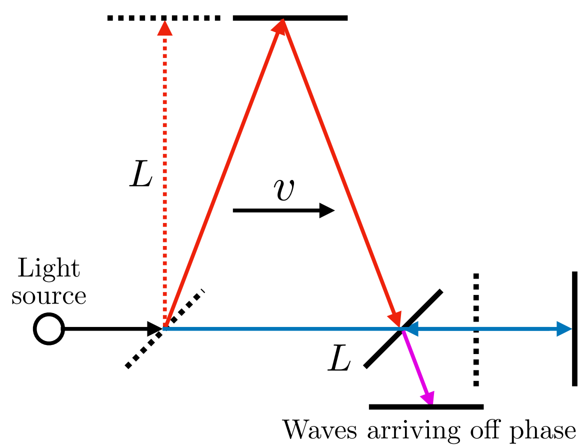


Light is produced by a light source, and then split into two rays, red and blue in the figure, perpendicular to each other. Each ray travels a distance L and then is reflected back by a reflecting mirror, traveling a total distance of $2L$. On their arrival the interference pattern they produce is measured by an interferometer. Because we are assuming the system is at rest respect to the aether, both rays travel at the speed of light, and take exactly the time

$$T_{red} = T_{blue} = \frac{2L}{c}$$

to travel to the reflecting mirrors and back. When they meet they are perfectly in phase.

Let us now suppose that we are in a system moving at speed v, in the 'blue' direction, with respect to the aether, see figure.



Let us first consider the blue ray. On the way to the reflecting mirror, the reflecting mirror is moving away at speed v, so that the speed respect to the mirror is $c - v$. On the way back, the interferometer is moving towards the ray, so that the speed respect to the interferometer is $c + v$. The time that is takes the blue ray to travel the total path is then

$$T_{blue} = \frac{L}{c+v} + \frac{L}{c-v}.$$

Let us now consider the red ray. Because now the reflecting mirror is moving sideways, the ray has to travel a bigger distance (see figure). Let us denote $\frac{1}{2}T_{red}$ the time that it takes to the ray to hit the mirror. In that time the mirror has moved $\frac{v}{2}T_{red}$ to the right (let's say in the x direction). Hence, by Pythagoras theorem the distance traveled by the light is

$$d = \sqrt{L^2 + \frac{v^2}{4} T_{red}^2}$$

But by definition, T_{red} is the time that takes light to cover twice that distance, so that we get the following equation

$$T_{red} = \frac{2\sqrt{L^2 + \frac{v^2}{4} T_{red}^2}}{c},$$

from which we obtain

$$T_{red} = \frac{2L}{\sqrt{c^2 - v^2}}.$$

So that the blue and red rays take different times to travel to the reflecting mirrors and back. The travel time difference for small velocities compared to the speed of light, namely $v \ll c$ is given by

$$T_{blue} - T_{red} = \frac{Lv^2}{c^3} + \cdots$$

To find the 'distance' difference between the two paths, we simply multiply by the speed of light

$$\Delta\lambda = L\frac{v^2}{c^2}$$

What happens if we put the expected numbers? the speed of light is $c \sim 3 \times 10^8 m/s$. The speed of earth as it orbits around the sun is $v \sim 3 \times 10^4 m/s$, so that

$$\frac{v^2}{c^2} \sim 10^{-8}$$

In their experiment Michelson and Morley used $L = 10m$, so that $\Delta\lambda$ is still incredibly small (how do we know, for instance, that both arms of the experiment are exactly the same length L?). To overcome these obstacles Michelson and Morley were not only very careful, but they crucially rotated their apparatus, looking for changes in the interference pattern. When the apparatus is rotated 90 degrees, the roles of the blue and red rays are interchanged, so the distance difference between the two configurations should be

$$2\Delta\lambda = 2L\frac{v^2}{c^2} \sim 2 \times 10^{-7} m$$

This is a small distance, but it should be compared to the wavelength of the light used in the experiment, which is $\lambda = 5 \times 10^{-7} m$, so that the expected 'fringe shift' between before and after rotating the apparatus is

$$n = \frac{2\Delta\lambda}{\lambda} \sim 0.4$$

Which is easily observable. To the great surprise and disappointment of Michelson and Morley they did not observe any change at all! Further refinements of the experiment gave the same negative results. All experimental observations pointed to the same:

Observation (experiment) 1 (Speed of light)**.** In all inertial frames, the measured speed of light (propagating in vacuum) is a constant independent of the direction of propagation and of the speed of the light source.

This remarkable experimental fact sets the stage for the downfall of Galilean relativity and its subsequent replacement due to Einstein.

2.6 Postulates of special relativity

The universality of the speed of light in all inertial frames is obviously incompatible with the behavior of inertial frames in Newtonian physics. In his famous 1905 paper, *On the electrodynamics of moving bodies*, Albert Einstein explained how this incompatibility should be resolved. He proposed to modify the postulates of Newtonian physics as follows,

- There is a preferred class of frames of reference, called *inertial frames*, in which the law of inertia holds. These frames execute uniform relative motion.
- An inertial frame, together with a choice of zero for time, determines an *inertial coordinate system (ICS)*.
- The laws of physics hold in any inertial frame (and in particular take the same form in any ICS).
- In all inertial frames, the speed of light (propagating in vacuum) is a constant independent of the direction of propagation and the speed of the light source.

So we have promoted the universality of the speed of light to the level of a postulate, at the expense of removing the rule that ICSs be related by Galilean transformations. We have also (with foresight) removed the postulate that inertial mass be independent of the state of motion of an object.

3 Special theory of relativity

In this section we will show how to derive Einstein's special theory of relativity from the four postulates above. For simplicity we will restrict our attention to motion in one spatial dimension (along with time), which we call $(1 + 1)$ dimensions. We will return to the full glory of $(3 + 1)$-dimensional space-time at a later point.

3.1 Lorentz transformations

Having revoked the rule that ICSs be related by Galilean transformations, we need to determine how ICSs *should* be related. It turns out that the requirement that the the speed of light be agreed upon in all such frames will be sufficient for us to determine the correct form of the coordinate transformations, which are called *Lorentz transformations.*

The algebraic manipulations and arguments required to derive the form of Lorentz transformations are not technically very difficult. However, the result implies the need for a deeper re-examination of our assumptions about space and time, distance and duration. In this section we will derive the formula for Lorentz transformations in two different ways. Our first derivation will be pragmatic and utilitarian.

Let us consider two inertial observers[3] O and O' travelling in one spatial dimension such that O' is moving at velocity v according to O. They pass each other at an event E and then move directly away from each other. They both set their clocks to zero at E. These two observers will coordinatize space-time using ICSs (x, t) and (x', t'), respectively, and our immediate task is to determine how these coordinate systems are related. In particular, we want functions f and g such that

$$x' = f(x, t) \ , \qquad t' = g(x, t) \ . \tag{3.1}$$

Because the observers synchronized their clocks when they coincided at E, they agree on the origin of their respective ICSs,

$$f(0, 0) = g(0, 0) = 0 \ . \tag{3.2}$$

From our first postulate, we deduce that straight lines in the ICS of observer O will have to transform to straight lines in the ICS of O' and vice versa. What this means is that the relevant changes of coordinates will have to be *linear* transformations,

$$x' = \alpha_1 x + \alpha_2 t \ , \qquad t' = \beta_1 x + \beta_2 t \ , \tag{3.3}$$

for some real constants $\alpha_{1,2}$ and $\beta_{1,2}$, which will depend on the relative velocity v. Given that we have stated that O' moves with velocity v according to O, the line $x = vt$ must map to the line $x' = 0$. Imposing this requirement gives us

$$x' = \gamma(x - vt) \ , \qquad t' = \beta_1 x + \beta_2 t \ . \tag{3.4}$$

For some coefficient γ that can in principle depend on the relative velocity v — let us write $\gamma = \gamma_v$. The fourth postulate implies that a light beam will have to be observed by both O and O' as moving with speed c. Let us consider two such beams, starting at the origin at $t = t' = 0$ and moving, with speed c, towards the positive/negative real axis. The fourth postulate then implies the following two conditions

- The line $x = c\,t$ must map to the line $x' = c\,t'$.
- The line $x = -c\,t$ must map to the line $x' = -c\,t$.

This gives the following equations

$$\begin{aligned} \gamma_v(c\,t - v\,t) &= c(\beta_1 c\,t + \beta_2 t) \\ \gamma_v(-c\,t - v\,t) &= -c(-\beta_1 c\,t + \beta_2 t) \end{aligned}$$

[3] You can think of observers as attached to/defining the origin of the reference system they move with, so that distances are measured with respect to them.

From which we can solve for β_1 and β_2 in terms of γ_v and c. We obtain

$$x' = \gamma_v(x - vt) \ , \qquad t' = \gamma_v(-\frac{v}{c^2}x + t). \tag{3.5}$$

We only need to determine the factor γ_v. This can be done as follows. We could have considered instead the transformation $x = f'(x', t'), t = g'(x', t')$. This should be given by the same expressions, but replacing $v \to -v$, since O is moving at velocity $-v$ according to O'. This leads to

$$x = \gamma_{-v}(x' + vt') \ , \qquad t = \gamma_{-v}(\frac{v}{c^2}x' + t'). \tag{3.6}$$

Plugging these two expressions into (3.5), or viceversa, we see these four equations are consistent provided

$$\gamma_{-v}\gamma_v\left(1 - \frac{v^2}{c^2}\right) = 1. \tag{3.7}$$

We can further argue that γ_v must be an *even* function of v, *i.e.*, $\gamma_v = \gamma_{-v}$. This follows from performing the same change of frame from O to O', but where we choose ICSs $(\tilde{x}, t)$ and $(\tilde{x}', t)$ for O and O' with $\tilde{x} = -x$, $\tilde{x}' = -x'$, in which case the relative velocity is $-v$ instead of v. Following the same manipulations as above, we get

$$\tilde{x}' = \gamma_{-v}(\tilde{x} + vt) \quad \Longrightarrow \quad x' = \gamma_{-v}(x - vt) \ . \tag{3.8}$$

Combined with (3.4), this tells us that $\gamma_v = \gamma_{-v}$. Combining then this with (3.7) we obtain

$$\gamma = \sqrt{\frac{1}{1 - \frac{v^2}{c^2}}} \ . \tag{3.9}$$

From now on we will suppress the subscript v on γ as we have done here whenever the relative velocity between the inertial frames in question is clear from the context. The quantity γ is called the *Lorentz factor* for the change of frame. This factor is ubiquitous in the equations of SR, and was in fact introduced by the Dutch physicist Hendrik Lorentz before Einstein's 1905 paper. We can then write our final result

$$\begin{aligned} x' &= \gamma(x - vt) \ , & x &= \gamma(x' + vt') \ , \\ ct' &= \gamma\left(ct - \frac{v}{c}x\right) \ , & ct &= \gamma\left(ct' + \frac{v}{c}x'\right) \ . \end{aligned} \tag{3.10}$$

It is often useful to write this Lorentz transformation in matrix form

$$\begin{pmatrix} ct' \\ x' \end{pmatrix} = \gamma \begin{pmatrix} 1 & -v/c \\ -v/c & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \end{pmatrix} \ . \tag{3.11}$$

Similarly, for the transformation from O' to O we have

$$\begin{pmatrix} ct \\ x \end{pmatrix} = \gamma \begin{pmatrix} 1 & v/c \\ v/c & 1 \end{pmatrix} \begin{pmatrix} ct' \\ x' \end{pmatrix} \ . \tag{3.12}$$

Note that for $v/c \to 0$, the Lorentz factor $\gamma \to 1$, and we recover the coordinate transformation for a Galilean boost.

Simultaneity in special relativity

In Newtonian mechanics there is an absolute notion of time, time passes equally for all inertial observers. In particular, this implies that there is an absolute notion of simultaneity: all inertial observers will agree whether events are simultaneous or not. In special relativity, things are not so simple and we see that the Lorentz transformations intertwine time and space together. Let us look at the following example.

Example 1. In a given ICS, consider two simultaneous events separated by $4 \times 10^7 m$. What is the time difference between them in a second ICS, moving with respect to the first at a speed $v = 3 \times 10^4 m/s$?

Let's denote the coordinates of the first ICS (t', x') and choose the time of the events as $t' = 0$. The coordinates of the events in O' are then $(0, x'_0)$ and $(0, x'_0 + d)$ with $d = 4 \times 10^7 m$. The Lorentz transformations then give, for the times of the two events in O:

$$ct_1 = \gamma \frac{v}{c} x'_0, \quad ct_2 = \gamma \frac{v}{c}(x'_0 + d)$$

So that the time difference is

$$\Delta t = \gamma \frac{v}{c^2} d = \frac{v}{c\sqrt{c^2 - v^2}} d \sim \frac{v}{c^2} d \sim 0.00001 s$$

A really small time difference, but nonetheless non-zero! the speed of the example represents the speed of earth in its orbit around the sun, while the distance is the distance between the poles.

Space-time diagrams

In order to really understand what is going on with these Lorentz transformations, it will be useful to make use of *space-time diagrams* as a visual aid. We already introduced them in Newtonian mechanics, but in dealing with Special Relativity they will be much more useful. For now we are working in (1+1) dimensions, so we will be drawing (x, t) space-time diagrams. These are simply pictures of space and time laid out on the plane where we are careful to make good, consistent choices. We can then view our Lorentz transformations as defining various lines of constant position/time coordinates on a fixed, invariant space-time.

The rules we will adopt for drawing space-time diagrams will be as follows,

- We choose our axes so that the horizontal coordinate is *space*, or x in a reference ICS, and the vertical coordinate is *time multiplied by the speed of light*, or ct in the same ICS.
- We choose units for t and x such that ct and x have the same scale. That is, the units are years and light-years, or seconds and light-seconds.
- Then all light-rays travel at a slope of angle $\pi/4$, and importantly *all observers agree that these lines are light-paths.*

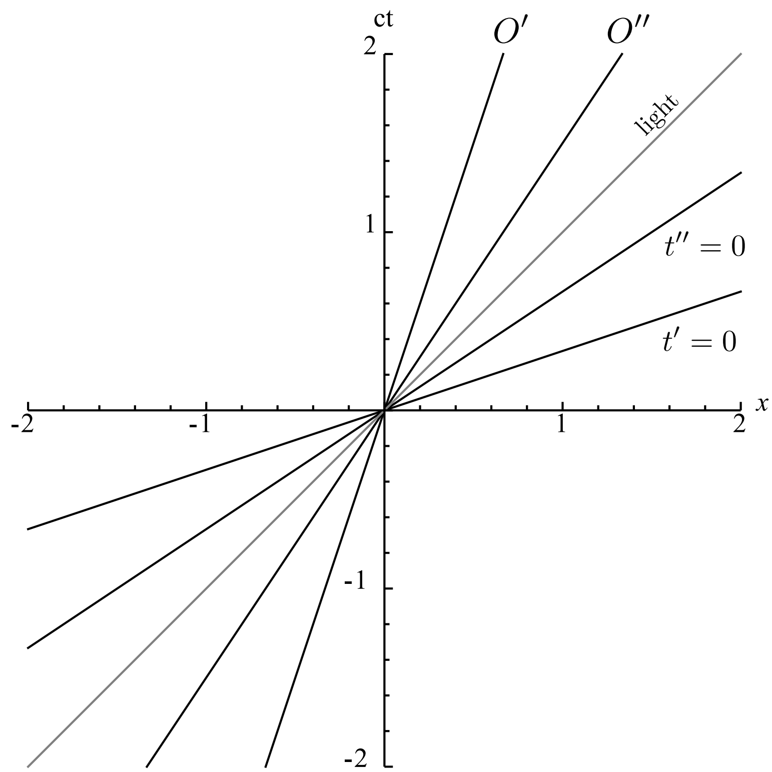


Figure 1. A space-time diagram showing lines of constant position and time coordinates for frames O and O' that are boosted relative to the the stationary frame of the diagram.

- We will sometimes draw lines indicating simultaneity, *i.e.*, $t =$ constant or $t' =$ constant, but must always remember that these are only the lines of simultaneity for a particular ICS.

In Figure 1 we show a space-time diagram on which we include a number of inertial observers, along with the lines of constant time in their respective ICSs. We immediately observe a certain pattern, which will be useful to remember in the future to help draw space-time diagrams more quickly and accurately:

- For any observer O', the lines of constant x' and constant t' (drawn in any ICS) are *pseudo-orthogonal*, which means that the angles they make to the horizontal, in the diagram, are α and $\pi/2 - \alpha$. (Contrast ordinary orthogonality, where the angles are α and $\pi/2+\alpha$.) The lines of slope $\pi/4$, corresponding to light rays, are pseudo-orthogonal to themselves, which does not have any Euclidean analogue.

There is a symmetry to this picture which is absent in the Galileo-Newton picture. For Galileo-Newton it is obvious that your notion of HERE (staying in the same place for different times) depends upon your velocity. But in relativistic geometry, so does the notion of NOW (being at different places for the same instant.)

3.2 Operational definition of inertial coordinates

In his 1905 paper, Einstein did not simply try to find a formula that would fit the constraints as we have done above. Instead, he performed a careful – almost philosophical – consideration of what it is we mean by *space* and *time*, *length* and *duration*. Indeed, the incompatibility of Galilean relativity with the constancy of the speed of light gives us a good indication that our intuitive understanding of these concepts requires refinement, and we can see that the Lorentz transformations that we have derived have the surprising property that they intermix space and time more thoroughly than we are accustomed to.

In Newtonian mechanics it is taken for granted that an inertial frame comes equipped with an *a priori* intertial coordinate system that agrees with what would be "measured" by an inertial observer in such a frame. If we want, we can think of an inertial observer as carrying with them a set of rigid measuring rods and a clock, with which they coordinatize space-time. An ICS obeying the rules of Galilean relativity then follows immediately if we assume a universal notion of time (*i.e.*, all clocks tick at the same speed) and an absolute notion of distance (*i.e.*, rigid rods have the same length regardless of their state of motion).

In special relativity, we are not willing to make these assumptions. Instead, to define the ICS associated to an inertial observer, we will proceed *operationally*, and explain how an inertial observer could (in principle) set up a coordinate system on space-time using only a clock (which they keep on their person) and a beam of light. From this construction and the universality of the speed of light, we will be able to re-derive the Lorentz transformation formula (3.10).

The radar method

Let us abandon our preconceived notions about coordinates in space and time and see what we can construct using the axiom that the speed of light is the same for all inertial observers. In a given inertial frame, how can we assign space and time coordinates to the various space-time events? It is easy enough to assign coordinates to the points on the worldline of the observer themself. The observer will keep a clock (we are assuming that reliable clocks exist!), and take readings from the clock as they progress through time, thus assigning coordinates $(t, 0, 0, 0)$ for $t \in \mathbb{R}$ (remember that the observer is, by definition, at the spatial origin).

To assign coordinates to a distant event A, we can imagine performing the following operation. Suppose that there is a reflecting mirror that we know will be present at the event in question. We send a light beam that arrives at A, reflects off the mirror and returns to us where we detect it (we think of the event as the light hitting the reflecting mirror). We keep careful track of the amount of time Δt that passes between when we emit the light and when it returns. We will then assign to the event A a position coordinate given by $c\Delta t/2$ in the direction we emitted the light, and a time coordinate equal to $\Delta t/2$.

Example 2. Consider one spatial dimension. At time t_0 the observer sends a light beam towards the positive direction $x > 0$. The beam is reflected back at the event A and the observer receives it at time $t_1 > t_0$. The coordinates we assign to the event A are then

$$(t_A, x_A) = \left(\frac{t_0 + t_1}{2}, c\frac{t_1 - t_0}{2}\right)$$

This (hypothetical) operation allows any inertial observer to, in principle, assign time and space coordinates to any and all events in space time (assuming they have lived forever and will live forever). We can see, though, that the procedure for comparing coordinate systems, including notions of time, between inertial observers will be substantially more complicated in this relativistic world than it was in the old days of Newton.

Lorentz transforms from the radar method

We again consider two inertial observers O and O' travelling in one spatial dimension with constant relative velocity. They pass each other at event E and then move directly away from each other. They both set their clocks to zero at E.

By using the radar method they both set up ICSs on space-time (we'll again call these coordinate systems (x, t) and (x', t')). Now we want to see that these coordinate systems are related by the Lorentz transformations we derived before.

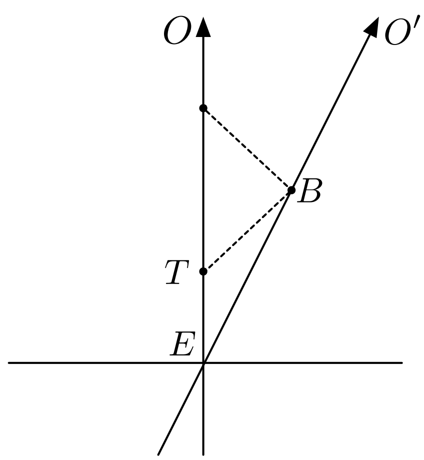


Figure 2. Comparing clocks with the radar method. The dotted lines are light rays.

Consider, as in Figure 2, a beam of light emitted by O towards O' at time T (as measured by the clock carried by O). Suppose the light arrives at O' at time $T' = kT$ (as measured by the clock carried by O'). The quantity k is called Bondi's k-factor. Since neither observer is accelerating, k is constant. On the assumption that only their relative velocity is observable, k must be a function of the relative velocity of O and O'.

Now consider the beam of light returning from O' to O, which allows O to make the "radar" observation. At what time does it hit O? By the assumption regarding the k-factor, it must be at time $kT' = k^2T$. Hence O assigns a spatial coordinate to B equal to $\frac{1}{2}c(k^2-1)T$ (from half the there-and-back time multiplied by the speed c) and a time coordinate to B equal to $\frac{1}{2}(k^2+1)\,T$ (the time half way between sending and receiving). So that the event B has coordinates in the O system given by

$$(t_B, x_B) = \left(\frac{1}{2}(k^2+1)\,T, \frac{1}{2}c(k^2-1)T\right)$$

But B is in the world-line of O', thus O reckons the (relative) speed of O' (as distance/time) to be

$$v = \frac{c(k^2-1)}{k^2+1} \ . \tag{3.13}$$

We can solve for k in terms of the relative velocity v:

$$k = \sqrt{\frac{c+v}{c-v}} \ . \tag{3.14}$$

Hence we also know the time T' in terms of the relative velocity:

$$T' = kT = \sqrt{\frac{c+v}{c-v}}\,T \ . \tag{3.15}$$

Observer O reckons that the amount of time elapsed from E to B is equal to $\Delta t = \frac{1}{2}(k^2+1)T = \frac{c}{c-v}T$. On the other hand, observer O' observes that the amount of time that has passed from E to B is $T' = kT = \sqrt{\frac{c+v}{c-v}}\,T$. We recognize the ratio of these two times, which is independent of T, as the Lorentz factor we found before,

$$\frac{\text{Time measured by } O}{\text{Time measured by } O'} = \frac{1}{\sqrt{1-v^2/c^2}} \equiv \gamma \ . \tag{3.16}$$

Again, we can see directly from this formula that as $c \to \infty$ (or $v/c \to 0$), the times become equal and we recover the Galileo-Newton assumption that time measurement will be the same for all observers. Thus the correction effect embodied in this formula is a direct consequence of our dumping the Newtonian assumption of absolute simultaneity, and replacing it by the assumption that the speed of light is the same for all observers.

To determine the form of a more general Lorentz transformation, we could consider the setup in Figure 3, with observers O and O' both assigning coordinates to a given event B in their respective ICSs. We can immediately see that the coordinates assigned to B by observer O will be

$$t = \frac{1}{2}(T_1 + kT_2) \ , \qquad x = \frac{1}{2}c(kT_2 - T_1) \ . \tag{3.17}$$

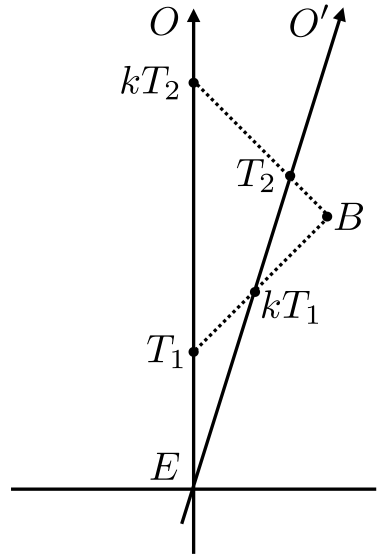


Figure 3. Two inertial observers coordinatize the same space-time event B using the radar method. Times displayed on the worldlines of the observers correspond to those observers' local measurements.

On the other hand, the coordinates assigned by observer O' will be

$$t' = \frac{1}{2}(kT_1 + T_2) \ , \qquad x' = \frac{1}{2}c(T_2 - kT_1) \ . \tag{3.18}$$

Here, again, $k = \sqrt{(c+v)/(c-v)}$. It is now a matter of algebra to solve for the relationship between (x,t) and (x',t') and recover the expression for the Lorentz transform given in (3.10).

3.3 Some relativistic effects

We now know how to relate the inertial coordinates of observers in relative motion. This is already enough for us to derive a number of surprising and famous conclusions about how distance, duration, and speed are distorted for objects in motion. However, before we dive in, a few words of warning.

Observing versus 'reckoning'

There is potential for confusion at this point as we emphasize the strange behaviors of spacetime in a relativistic setting. Since the finite speed of light plays an important role in uncovering the structure of special relativity, one might think that the "weirdness" that we will observe (length contraction, time dilation, etc.) is a consequence of what we *see* being delayed due to the time light takes to get from the event of interst to our eyes. This is not the case! Indeed, in the book of Woodhouse (and in our discussion of the radar method above), the word *reckon* is used rather than *observe* to make it clear that the assignment of relativistic inertial coordinates to spacetime events is the result of a *computation*, and is not the same

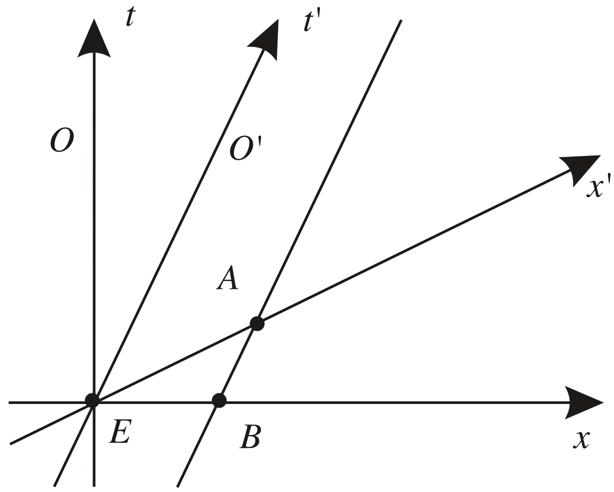


Figure 4. Space-time diagram demonstrating the phenomenon of length contraction.

as what a person would actually see (with the exception of events that take place along the worldline of the observer).

In fact, the question of what an observer actually does see adds an *additional* layer of complication. To understand how a relativistic world should really look, we need to use the framework we have developed above, and then further account for the time-delay in the propagation of light signals from a distant event to our eyes. Let us simply mention here that there is in fact a simple result due to Roger Penrose that was not discovered until 1959 that a spherical object always presents a circular outline in the sky to any inertial observer, whatever the speed. This is in contrast to the distorted shape we would 'reckon' that a circular object has when it is in motion.

3.3.1 Length contraction

A famous consequence of special relativity is the phenomenon of 'length contraction'. We consider two inertial observers O and O' whose inertial coordinate systems are related by (3.10). Now further supposed that observer O' carries with them a metal rod that occupies the x'-axis between $x' = 0$ and $x' = L$. So according to O', this rod has length equal to L. We should now ask what length the rod will have according to observer O.

In the ICS of observer O', the worldlines of the ends of the rod are given by $x'(t') = 0$ and by $x'(t') = L$. Consequently, in the ICS of observer O, they are therefore given parametrically by the two lines

$$(1) \qquad \begin{pmatrix} ct \\ x \end{pmatrix} = \gamma \begin{pmatrix} 1 & v/c \\ v/c & 1 \end{pmatrix} \begin{pmatrix} ct' \\ 0 \end{pmatrix} = \gamma \begin{pmatrix} ct' \\ vt' \end{pmatrix} , \tag{3.19}$$

$$(2) \qquad \begin{pmatrix} ct \\ x \end{pmatrix} = \gamma \begin{pmatrix} 1 & v/c \\ v/c & 1 \end{pmatrix} \begin{pmatrix} ct' \\ L \end{pmatrix} = \gamma \begin{pmatrix} ct' + Lv/c \\ vt' + L \end{pmatrix} . \tag{3.20}$$

Now it comes the important point. For O', L is defined as the distance between simultaneous events $(0, t')$ and (L, t'), but the same two events, would not be simultaneous in the ICS of O. For O, the length of the rod is defined as the distance of simultaneous events with respect to O. Let us choose $t = 0$ for simplicity. At time $t = 0$, the first end of the rod is at $x = 0$. But the other end of the rod only arrives at $t = 0$ when $t' = -Lv/c^2$, at which time the x-coordinate of that end of the rod is $\gamma(-Lv^2/c^2 + L) = L\sqrt{1 - v^2/c^2}$.

Thus O reckons that the rod, at $t = 0$, has length $L\sqrt{1 - v^2/c^2} = L/\gamma$. This is also known as the Lorentz or Fitzgerald-Lorentz contraction, because Lorentz (and slightly earlier, the Irish mathematician Fitzgerald) had put forward this formula in pre-relativity days, in the course of attempts to reconcile Maxwell's equations with the unobservability of the aether. They thought of it as a physical effect which shortens moving objects, and this idea has lingered on.

From the relativistic standpoint (*i.e.*, that of real physics), it should not be considered that anything actually contracts or shortens at all. The two observers are simply measuring different things. See figure 4 for the relevant space-time diagram. What O' measures as the length is the separation between events E and A, which are simultaneous in their ICS. What O measures instead is the separation between E and B, simultaneous in their ICS.

3.3.2 Time dilation

A second famous consequence of relativistic geometry has to do with the way time is measured by observers moving relative to one another. Again, let us take inertial observers O and O' who have set up their respective ICSs and synchronized their clocks at E where they coincide. Let's say that O' is carrying a clock that ticks with period δ. Thus, the ticks occur at the events whose coordinates in the O' ICS are $(ct', x') = (nc\delta, 0)$ for $n = 0, 1, 2, \ldots$. On the other hand, these events will have coordinates in the O ICS given by $(ct, x) = \gamma(nc\delta,\ n\delta v)$. Thus, the period of the ticks of the clock, according to O, is given by $\gamma\delta > \delta$, so it seems that the procession of time is *dilated* by a factor of γ.

Example 3. An astronaut is travelling directly away from earth with speed $v = \sqrt{3}/2c$, so that

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

So if the astronaut reckons that one hour passes between two evens in her spaceship, then an observer on earth reckons that two hours pass.

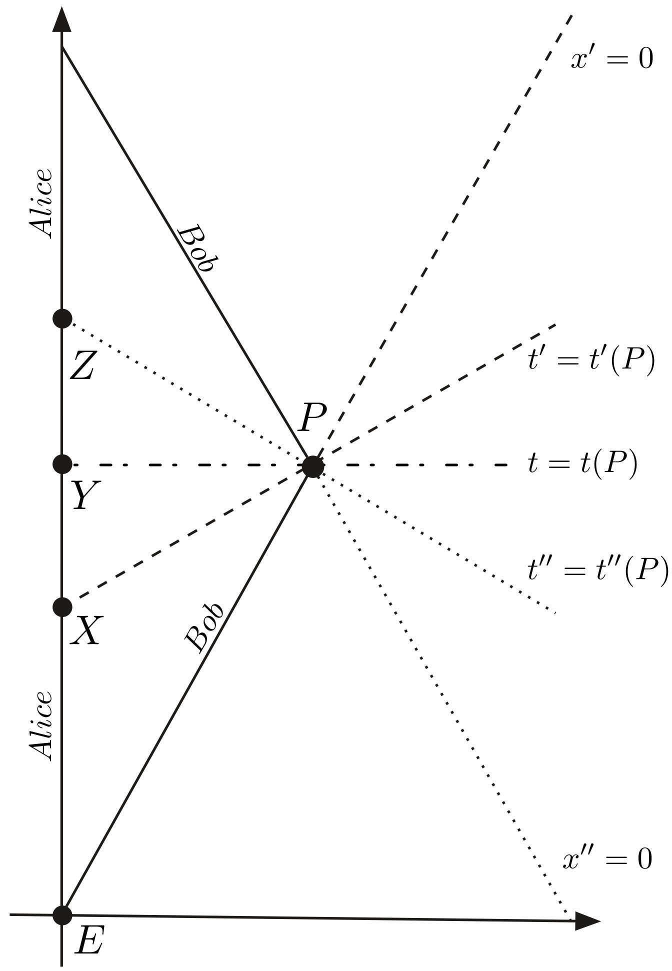


Figure 5. Space-time diagrams showing the twin paradox phenomenon. Dotted lines are lines of constant O' coordinates, dashed lines are lines of constant O'' coordinates. The dot-dashed line is a line of constant time in the original O frame of Alice.

This observation leads to the famous *twin paradox* (which, surely, is not a paradox). Suppose we have a pair of twins, Alice and Bob. Alice stays put on her home planet while Bob takes a trip on a spaceship, flying away at fixed speed v to a nearby planet. Alice watches Bob leave and after a time T in her reference frame he has reached the planet. At this point, Bob turns around and heads back to Earth, again at speed v. When he returns, he finds that Alice has aged by $T_A = 2T$. Bob, on the other hand, has only aged by $T_B = 2T/\gamma$. Bob is now younger than Alice, and in fact for large γ he could be quite a bit younger.

So far, we have just applied our time dilation formula, so where is the paradox? This comes about when we consider things from the point of view of Bob. He sits in his spaceship and watches Earth (and Alice) recede behind him at speed, v. From his perspective, it should be Alice who is younger. Surely things should be symmetric between the two.

The resolution is that there is, of course, no true symmetry between Bob's and Alice's experiences. Alice stayed in a fixed inertial frame for the whole time, while Bob did not. He had to turn around, meaning he had to accelerate and break the symmetry of the situation.

To understand this in detail, we should draw some space-time diagrams. We do this in Alice's frame in Figure 5. In the diagram, Bob starts out sitting at $x = vt$, or $x' = 0$ (we will call this the O' frame). Alice sits at $x = 0$ (the O frame). Bob's turning point is the event P, after which his new inertial motion defines the O'' frame. We've included lines of simultaneity for the event P in all three frames. The event Y corresponds to when Alice thinks that Bob is at P. The event X is where Bob thinks Alice was when he arrived at P. However, the event Z is where Bob thinks Alice was when he left P in his new inertial frame.

Here we immediately see the resolution of the paradox. In changing from the frame O' to the frame O'', Bob's conception of *when* Alice is has to jump from X to Z. In reality, Bob cannot instantaneously change from O' to O'', he needs to accelerate at some finite rate. While performing this acceleration, Alice will seem to age rapidly in correspondence with her transit from X to Z.

3.3.3 Velocity addition

Finally, we consider the problem of *velocity addition*. That is, suppose that we have three inertial observers, O, O', and O''. Let's say that O' appears to be traveling with velocity u in the frame of O, while O'' appears to be traveling with velocity v in the frame of O'. The problem of velocity addition is to determine the velocity w at which O'' travels in the frame of O. This is illustrated in Figure 6.

The trick, as in the previous examples, is to identify events whose coordinates in the various ICSs will encode the answer to our question. In particular, take two events on the worldline of O'', call them E and A. In the ICS of O', we have

$$t'(E) = x'(E) = 0 \ , \qquad t'(A) = T' \ , \ x'(A) = X' = vT' \ . \tag{3.21}$$

Now we want the coordinates of these same events in the ICS of observer O. Applying the Lorentz formula, we have

$$\begin{aligned} t(E) &= 0 \ , & x(E) &= 0 \ , \\ ct(A) &= \gamma_u \left(cT' + \frac{u}{c} X' \right) \ , & x(A) &= \gamma_u \left(X' + uT' \right) \ . \end{aligned}$$

Substituting the expression $X' = vT'$ and doing a bit of algebra, we recover the velocity addition formula,

$$w = \frac{x(A)}{t(A)} = \frac{u + v}{1 + \frac{uv}{c^2}} \ . \tag{3.22}$$

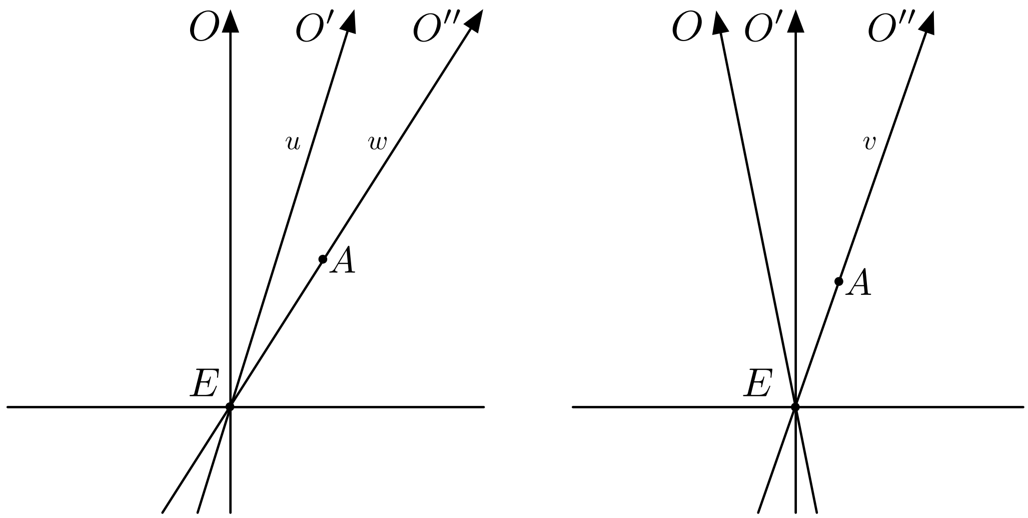


Figure 6. Space-time diagrams showing the configuration of observers relevant for the velocity addition problem. The velocity addition formula determines the relative velocity w in terms of u and v.

As usual, we see that in the limit of small velocities the Galilean rule for addition of velocity is recovered. Additionally, we happily see that for any choices of $u, v < c$, then we will also have $w < c$, so a change of frame of reference will never result in an object traveling faster than the speed of light. Furthermore, if we take $u \to c$ (or $v \to c$) we see that $w \to c$. This is simply a manifestation of the fact that the speed of light is the same in all inertial frames, and independent on the velocity of the source. We cannot make photons go faster with respect to us by simply moving in the opposite direction. We will see a simpler way to derive this formula below, when we discuss the group structure of Lorentz transformations.

3.4 The Lorentz group in 1+1 dimensions

The (1+1)-dimensional Lorentz transformations whose equation we derived above are actually *hyperbolic rotations* of the two-dimensional (t, x)-plane. It turns out that, like the ordinary Euclidean rotations in two dimensions, these hyperbolic rotations form a group that can be simply characterized in terms of its action on a certain preferred matrix. In this section we develop the formal characterization of the Lorentz group in these terms.

3.4.1 Rapidity and group structure

We start with a definition,

Definition 7. The *rapidity* associated with a relative speed u with $|u| < c$ is given by

$$\phi(u) = \mathrm{Tanh}^{-1}(u/c) \ . \tag{3.23}$$

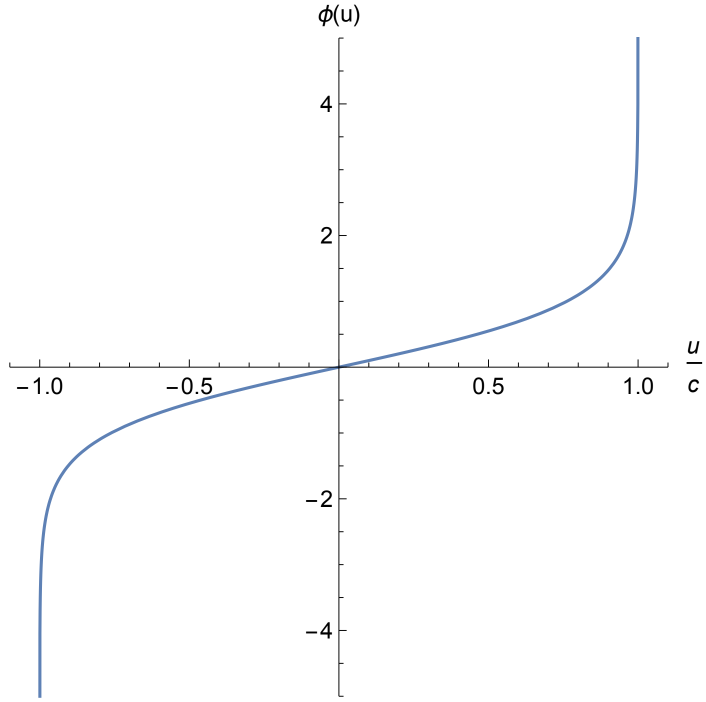


Figure 7. Rapidity as a function of velocity. The map to rapidity maps the set of physical velocities (with magnitude less than the speed of light) to the entire real line.

The rapidity as a function of the velocity (over c) is displayed in Fig. 7. For physical values of the velocity, the map to rapidity is one-to-one, so with no ambiguity we can work in terms of rapidity if we wish. In terms of the rapidity, we have the following nice expressions for various quantities that appear in the expressions for Lorentz transformations,

$$\begin{aligned} \gamma(u) &= \cosh\phi \ , \\ u\gamma(u)/c &= \sinh\phi \ , \\ k(u) &= \exp\phi \ , \end{aligned} \tag{3.24}$$

and in particular, the matrix that implements the Lorentz transformation on space-time coordinates now has the following simple form,

$$\gamma(u)\begin{pmatrix} 1 & u/c \\ u/c & 1 \end{pmatrix} = \begin{pmatrix} \cosh\phi & \sinh\phi \\ \sinh\phi & \cosh\phi \end{pmatrix} \ . \tag{3.25}$$

The 2×2 matrices of this form constitute a group, since

$$\begin{pmatrix} \cosh\phi & \sinh\phi \\ \sinh\phi & \cosh\phi \end{pmatrix}\begin{pmatrix} \cosh\psi & \sinh\psi \\ \sinh\psi & \cosh\psi \end{pmatrix} = \begin{pmatrix} \cosh(\phi+\psi) & \sinh(\phi+\psi) \\ \sinh(\phi+\psi) & \cosh(\phi+\psi) \end{pmatrix} \ . \tag{3.26}$$

The inverse of a transformation with rapidity ϕ is just the transformation with rapidity $-\phi$.

We see that under the group multiplication, rapidities simply add. This provides us with a more elegant way of solving the velocity addition problem from 3.3.3. If observers O, O', and O'' are arranged such that O' moves with velocity u relative to O, and O'' move with velocity v relative to O', then the velocity w of O'' relative to O obeys

$$\phi(w) = \phi(u) + \phi(v) \ , \tag{3.27}$$

and after some mostly painless algebra we recover the velocity addition formula,

$$w = \frac{u+v}{1+uv/c^2} \ . \tag{3.28}$$

3.4.2 Orthogonal transformations in two dimensions

The two-dimensional Lorentz transformations are clearly in some sense analogous to Euclidean rotations in two space dimensions, SO(2), which has elements given by 2×2 matrices of the form

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \ , \qquad \theta \in [0, 2\pi) \ . \tag{3.29}$$

An obvious difference is that SO(2) is compact (a circle) while the Lorentz transformations form a non-compact set (isomorphic to the real line under addition). We will take advantage of this analogy to find a simple recharacterization of the two-dimensional Lorentz group.

Recall that the *orthogonal group*, O(2), can be characterized as the linear transformations of the plane that preserve the Euclidean norm,

$$(H\mathbf{x})^{\mathrm{T}} \cdot (H\mathbf{x}) = \mathbf{x}^{\mathrm{T}} \cdot \mathbf{x} \ , \qquad \forall \mathbf{x} \in \mathbb{R}^2 \ . \tag{3.30}$$

Equivalently, these are the matrices that obey

$$H^{\mathrm{T}} H = I \ , \tag{3.31}$$

where I is the 2×2 identity matrix. For reasons that will become clear in a moment, we can insert an identity matrix in between the matrix H and its transpose and get the expression

$$H^{\mathrm{T}} I H = I \ . \tag{3.32}$$

Thus, we can equivalently describe the orthogonal group as the set of matrices that send the identity matrix I to itself under the action by conjugation given in (3.32).

We can further specialize to the *special orthogonal group*, $SO(2)$, which is the subgroup of the orthogonal group consisting of orthogonal matrices that preserve the orientation of the plane. These are precisely the orthogonal matrices with positive determinant, so we have the simple characterization of SO(2) as the 2×2 matrices H such that

- $H^{\mathrm{T}} I H = I$,

- $\det H > 0$.

This formulation of the rotation group will admit a straightforward generalization that defines the two-dimensional Lorentz group.

3.4.3 Pseudo-orthogonal transformations in two dimensions

Let us introduce a 2×2 matrix of indefinite signature as follows,

$$g = \begin{pmatrix} +1 & 0 \\ 0 & -1 \end{pmatrix} . \tag{3.33}$$

Then it is easy to see that conjugating g by a Lorentz transformation $L(\phi)$ of rapidity ϕ leaves it invariant,

$$\begin{aligned} L(\phi)^{\mathrm{T}} g L(\phi) &= \begin{pmatrix} \cosh\phi & \sinh\phi \\ \sinh\phi & \cosh\phi \end{pmatrix} \begin{pmatrix} \cosh\phi & \sinh\phi \\ -\sinh\phi & -\cosh\phi \end{pmatrix} \\ &= \begin{pmatrix} \cosh^2\phi - \sinh^2\phi & 0 \\ 0 & \sinh^2\phi - \cosh^2\phi \end{pmatrix} \\ &= g \ . \end{aligned} \tag{3.34}$$

We then introduce the following definition

Definition 8. The *Lorentz group* in $(1+1)$ dimensions, $\mathrm{O}(1,1)$, is the group of 2×2 matrices L that satisfy

$$L^{\mathrm{T}} g L = g \ .$$

It is a straightforward exercise to verify that the matrices satisfying the condition in the definition do indeed form a group.

It turns out (as you will explore in an exercise below) that not all elements of the $(1+1)$-dimensional Lorentz group are of the form given in (3.25). We introduce two further specializations of the Lorentz group. The first is analogous to the special orthogonal group in the context of Euclidean transformations,

Definition 9. The *proper Lorentz group* in $(1+1)$ dimensions, $\mathrm{SO}(1,1)$, is the subgroup of the $(1+1)$-dimensional Lorentz group whose elements also obey $\det L = 1$.

The Lorentz transformations of the form (3.25) are clearly proper.

Definition 10. The *orthochronous Lorentz group* in $(1+1)$ dimensions, $\mathrm{O}^+(1,1)$, is the subgroup of the $(1+1)$-dimensional Lorentz group whose top left entry is greater than zero.

Again, the Lorentz transformations (3.25) are clearly orthochronous. Thus the transformations we have been considering all belong to the intersection of the previous two groups, which we define as follows.

Definition 11. The *proper, orthochronous Lorentz group* in $(1+1)$ dimensions, $\mathrm{SO}^+(1,1)$, is the group of 2×2 matrices L with entries $L^a{}_b$, where a and b run over $(0,1)$, that satisfy the following three conditions:

- $L^{\mathrm{T}} g L = g$,
- $L^0{}_0 > 0$,
- $\det L = 1$.

The Lorentz transformations we have been considering so far are indeed proper and orthochronous.

Exercise 1. Show that the most general element of the proper, orthochronous Lorentz group is of the form

$$L(\phi) = \begin{pmatrix} \cosh\phi & \sinh\phi \\ \sinh\phi & \cosh\phi \end{pmatrix} , \tag{3.35}$$

for some rapidity ϕ. Thus the abstract characterization of proper, orthochronous Lorentz transformations agrees with the notion of Lorentz transformations that we had previously derived.

Exercise 2. Show that there are exactly four disjoint families of 2×2 matrices that define elements of the full Lorentz group $\mathrm{O}(1,1)$, one of which is the proper orthochronous Lorentz group. Characterize the action of the other three families on two-dimensional spacetime.

Finally, recall that the orthogonal group in two dimensions has the good property that when acting on vectors it leaves invariant the Euclidean norm. The Lorentz group has an analogous property. In particular, for any Lorentz transformation L of rapidity ϕ, we have

$$\begin{aligned}(ct)^2 - x^2 &= (ct, x)\ g \begin{pmatrix} ct \\ x \end{pmatrix} \\ &= \left(ct', x'\right) L^{\mathrm{T}} g L \begin{pmatrix} ct' \\ x' \end{pmatrix} \\ &= \left(ct', x'\right)\ g \begin{pmatrix} ct' \\ x' \end{pmatrix} \\ &= (ct')^2 - (x')^2 \ .\end{aligned} \tag{3.36}$$

Thus we recover a notion that generalizes the notion of distance in Euclidean space to the case of "Lorentzian" space, where we act with Lorentz transformations instead of rotations.

What we have written above is the *invariant interval* between the origin and an event with coordinates (ct, x).

More generally, we can look at the invariant interval between two events with coordinates (ct_1, x_1) and (ct_2, x_2),

$$c^2(t_2 - t_1)^2 - (x_2 - x_1)^2 = c^2(t_2' - t_1')^2 - (x_2' - x_1')^2 \ . \tag{3.37}$$

Notice that this invariant interval between two events is zero precisely when they are connected by the worldline of a light ray, *i.e.*, to travel between them you would need to move at the speed of light. In particular, invariance of the above interval implies that the speed of light will be c in all inertial frames.

4 Special relativity is astrophysics, astronomy and cosmology

4.1 Observation of time dilation with cosmic-ray muons

In the following we will describe an experiment, performed by Rossi and Hall in 1941, that shows that indeed 'moving clocks run slow', according to our discussion of time dilation. The experiment uses muons generated by cosmic rays entering the earth's atmosphere. Muons are unstable particles, similar to the electron, but with about 200 times its mass. Muons decay into electrons and neutrinos, through the process

$$\mu \to e + \nu_\mu + \bar{\nu}_e.$$

Unstable particles, and muons in particular, provide very precise clocks. Given a sample of muons (with a large number of them), half of them will decay within $1.5\mu s$ (1.5×10^{-6} seconds) - pretty long by subatomic standards. Given a sample of muons then, we can measure how much time has passed by measuring how many muons are left. For instance, if 1/4 of the initial muons are left, then $3\mu s$ have passed. The essential stages of the Rossi-Hall experiment are the following

1. Muons are produced by cosmic rays, then they enter the atmosphere and travel predominantly downward, with speeds very close to c.
2. The number of muons is counted (with a detector) at the top of a mountain. The counting apparatus recorded the arrival of 536 muons/hour, on average, at an altitude of 2000m.
3. The number of muons is counted again at sea level.

What do we expect the counting of step 3 be? Since the speed of the muons is very close to c, it takes them about $6.5\mu s$ to cover the distance from the top of the mountain to sea level. In that time we would expect only about 5% of the original muons to have survived, namely about 25. The measured result, however, is very different!

Observation (experiment) 2 (Rossi-Hall experiment)**.** The number of muons detected at sea level exceeds 400 muons/hour.

According to the clock represented by the moving muons themselves, the journey lasted only about $0.7\mu s$! the explanation for this time dilation factor of about 9 is of course, that the muons are moving very fast. Solving for $\gamma_v = 9$ we obtain $v \approx 0.994c$, extremely close to the speed of light.

Another interpretation of the time dilation experiment

In discussing the Rossi-Hall experiment we have considered everything from the point of view of the frame of reference attached to the laboratory, where the mountain is fixed. In this frame the clock represented by a group of fast moving muons runs slow. It is illuminating to consider things from the point of view of the muons themselves. In their reference frame (also inertial) first the top of the mountain flashes past, and then, a little later, the ground arrives.

Suppose that the vertical distance traveled by the muons is H measured in the frame O attached to earth ($H = 2000m$). The muons have speed v down through the atmosphere. The duration of the journey as measured in O is then $\Delta t = H/v \approx H/c$. According to the time dilation equation the journey takes time $\Delta t' = \Delta t(1 - v^2/c^2)^{1/2}$ as measured in the muons rest frame O'. From the point of view of the muons this time must also be measurable as a certain vertical distance H' (measured in O') divided by the speed at which the mountain moves upward past the muons. This speed is of course v, because it is an essential feature of relativity that there should be agreement about the magnitude of the relative velocity of the two frames: only the sign of the velocity changes, depending on the point of view. Thus we have

$$\Delta t = \frac{H}{v}, \quad \Delta t' = \frac{H'}{v}$$

hence

$$\frac{H'}{H} = \frac{\Delta t'}{\Delta t} = (1 - v^2/c^2)^{1/2} \to H' = H(1 - v^2/c^2)^{1/2}$$

Which is nothing but the Lorentz contraction formula! so, according to the muons the mountain moves at the same speed v, much is much shorter than $2000m$, so that they complete the journey in $0.7\mu s$.

We see then that a result upon which observers in all frames agree - what fraction of a group of unstable particles survives between one event and another - may be attributed to time dilation or to Lorentz contraction, according to one's point of view.

4.2 The relativistic Doppler effect

Let us start by reviewing the familiar, acoustical, Doppler effect. In this case in order to propagate sound waves need a medium, which we take to be, for instance, air. Consider a one-dimensional problem in which a source S and a receiver R are moving along the same line. Relative to the air, let the speeds of S and R be u_s and u_r respectively, in the positive direction. See figure 8. Let the source be emitting a signal of frequency ν and period $\tau = 1/\nu$.

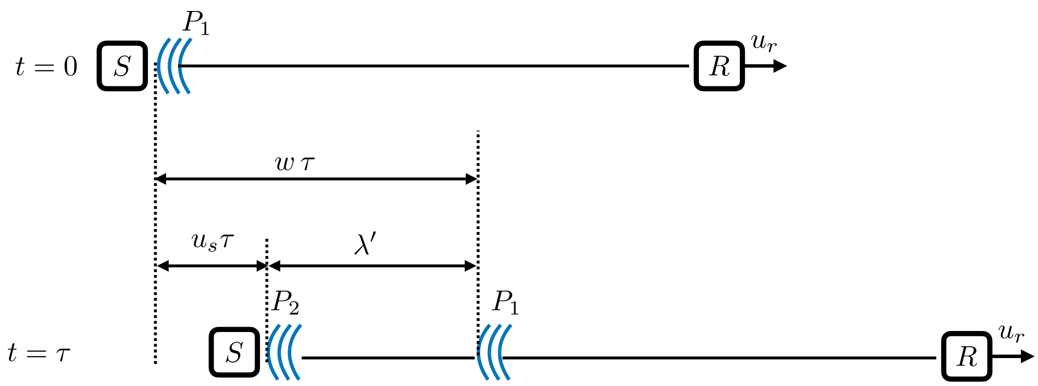


Figure 8. Doppler effect for radiation/sound from a moving source

We assume that the signal is in the form of very brief pulses separated by τ. Each pulse travels through the air at the speed of sound w. A pulse P_1 is emitted at time $t = 0$ and a second pulse P_2 is emitted at $t = \tau$, as shown. During the time τ the first pulse travels a distance $w\tau$ and the source moves a distance $u_s\tau$. Thus the distance between P_1 and P_2, which we can call the effective wavelength λ' is given by

$$\lambda' = (w - u_s)\tau = \frac{w - u_s}{\nu}$$

Now, the speed of the pulses relative to R is $w - u_r$, so that the time interval between the arrival of P_1 and P_2 at R is given by τ', where

$$\tau' = \frac{\lambda'}{w - u_r} = \frac{w - u_s}{\nu(w - u_r)}$$

The reciprocal of τ' defines an effective frequency ν', so that

$$\nu' = \frac{1}{\tau'} = \nu\frac{1 - u_r/w}{1 - u_s/w}$$

Note that for a given relative velocity $u_r - u_s$ between the source and the receiver, the value of ν' still depends on the velocities. This is due to the fact that for sound the medium defines an absolute rest. Since for light there is no medium, and no absolute rest, the result should only depend on the relative velocity, as we will see.

Let us now turn to the corresponding problem in special relativity. Suppose the source is located at the origin of a reference frame S, and that an observer moves relative to S at velocity v, so that is at rest in the system S', see figure 4.3 The source emits pulses of light, each separated by time intervals τ and which travel to the speed of light c. Suppose the first pulse is sent out at $t = 0$, when the observer is at the position $x = x_0$ (in the reference system

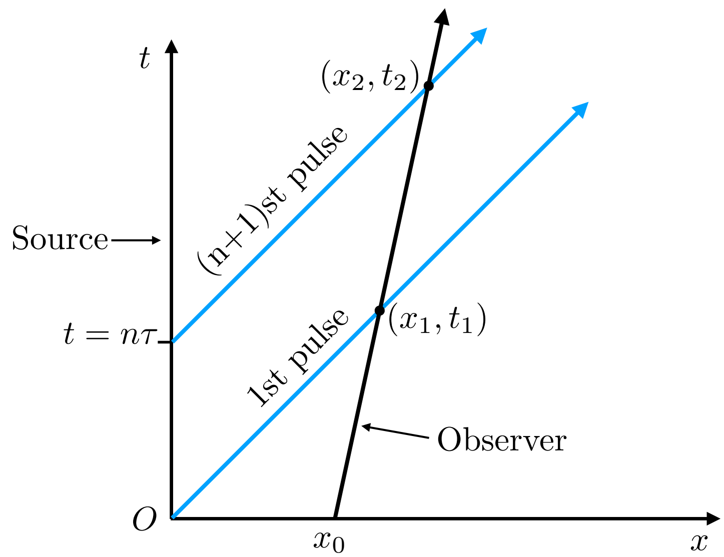


Figure 9. Relativistic Doppler effect for light

of the source S). The $(n+1)th$ pulse is sent out at $t = n\tau$. This will have covered n periods of vibration, so that the measured frequency of the source in S is $\nu = 1/\tau$.

What does the observer record? In the system S, the intersections of the worldline of the observer with the worldlines of the pulses (which correspond to the events "the observer has received the pulses") have coordinates (x_1, t_1) and (x_2, t_2), see figure, with

$$x_1 = ct_1 = x_0 + vt_1 \tag{4.1}$$

$$x_2 = c(t_2 - n\tau) = x_0 + vt_2. \tag{4.2}$$

Therefore

$$t_2 - t_1 = \frac{cn\tau}{c-v}, \qquad x_2 - x_2 = \frac{vcn\tau}{c-v}. \tag{4.3}$$

Now we can use the Lorentz transformations to compute the corresponding time interval measured in S'

$$\begin{aligned} t_2' - t_1' &= \gamma\left((t_2 - t_1) - v(x_2 - x_1)/c^2\right) \\ &= \gamma\left(\frac{cn\tau}{c-v} - \frac{v}{c^2}\frac{vcn\tau}{c-v}\right) = \frac{\gamma cn\tau}{c-v}\left(1 - \frac{v^2}{c^2}\right). \end{aligned}$$

Since this time interval covers n periods of the signal as received by the observer, the apparent period τ' is given by

$$\tau' = \frac{t_2' - t_1'}{n} = \frac{\gamma c\tau}{c-v}\left(1 - \frac{v^2}{c^2}\right) \tag{4.4}$$

Using $\gamma = (1 - v^2/c^2)^{-1/2}$ we obtain

$$\tau' = \left(\frac{1+\beta}{1-\beta}\right)^{1/2} \tau \tag{4.5}$$

where we have introduced $\beta = v/c$ as the measure of 'relativistic effects'. In terms of frequencies

$$\nu' = \left(\frac{1-\beta}{1+\beta}\right)^{1/2} \nu. \tag{4.6}$$

The most dramatic manifestation of the relativistic Doppler effect is the red shift of distant galaxies. The spectrum of a galaxy, being a synthesis from all the different radiating objects in it, is closed to being a continuous smear of frequencies (like white light). However, astrophysicists are able to distinguish a few prominent dark lines, *i.e.* narrow gaps in the frequency of light we receive from the galaxies. Two lines in particular are called the H and K absorption lines. These are dark lines in the ultraviolet spectrum of stars, which are caused by the absorption of photons by singly ionized calcium atoms present in galaxies. The H and K lines correspond to light of wavelength 3968.5 Å (H) and 3933.7 Å (K). An angstrom Å is a unit of length equal to 1×10^{-10} meters. The wavelength λ is related to the frequency ν by $\nu = \frac{c}{\lambda}$.

Measuring the location of the absorption lines for different galaxies astrophysicists saw that those values were shifted towards the red (lower frequencies, or bigger wavelength), from which it was deduced that the galaxies are moving away from us! Using (4.6), astrophysicists are able to measure the precise velocities at which they are moving away, the following table shows a few examples.

Galaxy	Distance (light years)	Velocity ($\times 10^7 m/s$)
Virgo	0.4×10^8	0.12
Ursa Major	5.0×10^8	1.4
Corona Borealis	7.0×10^8	2.14
Bootes	1.3×10^9	3.90
Hydra	2.0×10^9	6.10

The most remarkable feature of these results, known as *Hubble's law*, is the linear relation between the velocity of recession and the distance for remote galaxies. The *Hubble's law* is one of the pillars of modern cosmology.

4.3 Stellar aberration

In the next section we will present a systematic treatment of the Lorentz transformations in $3+1$ dimensions. In order to discuss stellar aberration, however, we need a very particular case, where two inertial frames are moving with respect to each other with velocity v in the

$x-$direction (exactly as described in the previous section), but in addition we will add a coincident coordinate y. The Lorentz transformations from S' to S are then given by

$$\begin{aligned} t &= \gamma(t' + vx'/c^2) \\ x &= \gamma(x' + vt') \\ y &= y' \end{aligned} \tag{4.7}$$

Our first task is to find the corresponding transformations for velocities. Suppose that an object has velocity components u'_x, u'_y as measured in S'. By definition of velocity we have

$$u'_x = \frac{dx'}{dt'}, \qquad u'_y = \frac{dy'}{dt'}$$

To work out the velocities in the system S take (4.7) and differentiate. We obtain

$$\begin{aligned} dt &= \gamma(1 + vu'_x/c^2)dt' \\ dx &= \gamma(u'_x + v)dt' \\ dy &= u'_y dt' \end{aligned} \tag{4.8}$$

Hence we have the following transformation rules

$$\begin{aligned} u_x &= \frac{dx}{dt} = \frac{u'_x + v}{1 + vu'_x/c^2}, \qquad u'_x = \frac{u_x - v}{1 - vu_x/c^2} \\ u_y &= \frac{dy}{dt} = \frac{u'_y/\gamma}{1 + vu'_x/c^2}, \qquad u'_y = \frac{u_y/\gamma}{1 - vu_x/c^2} \end{aligned} \tag{4.9}$$

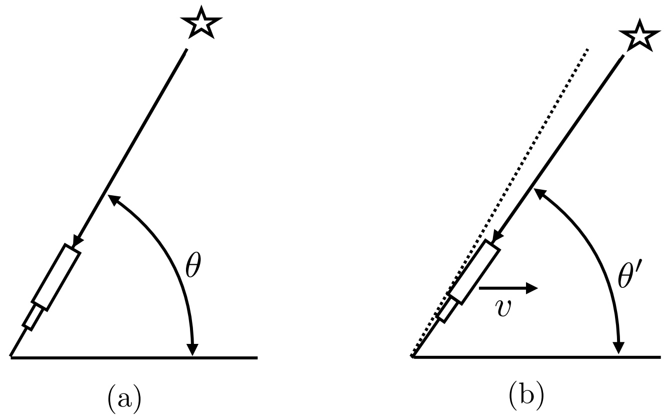
The equations in the first line represent the relativistic laws for the addition of two velocities, as derived above.

Stellar aberration

With (4.9) at hand we have a means of relating the *directions* of a given rectilinear motion, as described in two different frames. One interesting application of this result is to the problem of stellar aberration. This is the change in apparent direction of a distant object, such as a star, due to the motion of earth as it orbits around the sun, see figure. Let frame S be the rest frame of the sun, and the distant start, this is (a) in the figure. Let S' be the frame of the earth, traveling with the orbital velocity v relative to S, this is (b) in the figure. Suppose the direction of a star, measured from the plane of the earth's orbit, is θ in frame S and θ' in frame S'. Then we have $u_x = -c\cos\theta$ and $u_y = -c\sin\theta$ for the velocity components of the incoming photons, as measured in S. Using equations (4.9) we can then find

$$u'_x = -\frac{c\cos\theta + v}{1 + \frac{v}{c}\cos\theta} \tag{4.10}$$

$$u'_y = -\frac{c\sin\theta}{\gamma(1 + \frac{v}{c}\cos\theta)} \tag{4.11}$$



You can explicitly check that $(u'_x)^2 + (u'_y)^2 = c^2$ so that the direction θ' as observed by S' can be obtained as

$$\cos\theta' = -\frac{u'_x}{c} = \frac{\cos\theta + v/c}{1 + v/c\cos\theta} = \frac{\cos\theta + \beta}{1 + \beta\cos\theta}, \tag{4.12}$$

where recall $\beta = v/c$. As earth orbits around the sun β changes with the seasons, in the range $\beta \in [-10^{-4}, 10^{-4}]$. Such aberrations effect were actually observed, already in the late 1600s! way before special relativity.

5 Four dimensional space-time

We have so far restricted our attention to motion in one spatial dimension plus time, or (1+1) dimensions. We will now extend what we've found to $1 + 3$ dimensions, with one temporal and three spatial dimensions.

5.1 The Lorentz group in 1+3 dimensions

Pseudo-orthogonality again

We proceed by analogy with the relationship between the orthogonal and Lorentz groups in $(1 + 1)$ dimensions. We define the matrix $g = \mathrm{diag}(1, -1, -1, -1)$, which now generalizes the identity matrix in four-dimensions to the case of indefinite signature, with the number of minus signs corresponding to the number of spatial dimensions (as we did in $(1 + 1)$ dimensions). Our definition of Lorentz transformations will be as linear transformations on (ct, x, y, z) that preserve the matrix g, or equivalently, that preserve the four-dimensional invariant interval,

$$c^2t^2 - x^2 - y^2 - z^2 = (c\,t, x, y, z)\;\; g \begin{pmatrix} c\,t \\ x \\ y \\ z \end{pmatrix} \;. \tag{5.1}$$

As we have seen in $1+1$ dimensions, this invariance will ensure that the speed of light is c in all inertial frames. Suppose we have a matrix L defining the transformation

$$\begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} = L \begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} . \tag{5.2}$$

We then require the pseudo-orthogonality relation

$$L^{\mathrm{T}} g L = g . \tag{5.3}$$

Note that we have not yet introduced an analogue of the three-vector notation $\mathbf{x}$, and are obliged to write out (ct', x', y', z'). There is a reason for this: the proper description of four-vectors needs some careful notation for the components, which we shall introduce later on.

Proposition 1. The matrices L satisfying this condition form a group. This is the $(1+3)$-dimensional Lorentz groups $\mathrm{O}(1,3)$.

Proof. The proof of this statement is identical to the proof of the analogous statement in $(1+1)$ dimensions. □

As in $(1+1)$ dimensions, the set of matrices satisfying (5.3) has multiple components. There are matrices such as $\mathrm{diag}(-1,1,1,1)$, which reverse the direction of time, as well as matrices such as $\mathrm{diag}(1,-1,1,1)$, which act as an orientation-changing reflection in space.

Definition 12. The *proper, orthochronous Lorentz group* in $(1+3)$ dimensions, $\mathrm{SO}^{+}(1,3)$, is the group of 4×4 matrices L with entries $L^a{}_b$, where a and b run over $(0,1,2,3)$, that satisfy the following three conditions:

- $L^{\mathrm{T}} g L = g$,
- $L^0{}_0 > 0$,
- $\det L = 1$.

As you saw in $(1+1)$ dimensions (if you did the exercises), the second condition says that $\frac{\partial t}{\partial t'}$ is positive, and serves to rule out time-reversing transformations. Then, given that this $L^0{}_0$ entry is positive, the last condition ensures that the change in spatial coordinates preserves orientation rather than giving rise to a reflection.

Note that $L^{\mathrm{T}} g L = g$, written out in components rather than expressed as matrix multiplication, means

$$\sum_{ab} L^a{}_c g_{ab} L^b{}_d = g_{cd} , \tag{5.4}$$

where $g_{00} = +1$ and $g_{11} = g_{22} = g_{33} = -1$. The proper, orthochronous Lorentz transformations form a subgroup, $\mathrm{SO}^+(1,3) \subset \mathrm{O}(1,3)$. This subgroup can be characterized as the component of the full Lorentz group which is continuously connected with the identity. In this course we shall only be concerned with this subgroup.

Other approaches to the definition of Lorentz transformations

You might be suspicious that we have been too quick in generalizing the criterion of preserving the invariant interval from $(1+1)$ to $(1+3)$ space-time dimensions. Indeed, in the former case we saw that the Lorentz transformation rules follow from a very concrete analysis of the coordinates constructed by inertial observers using the radar method. While an analogous radar-method-based computation in more dimensions quickly gets out of hand, a more direct argument is possible, and it is presented in Woodhouse (pp. 79-80), where the argument is divided into two parts. First we use the fact that all observers will agree on the speed of light, which means that we must have

$$c^2t'^2 - x'^2 - y'^2 - z'^2 = 0 \iff c^2t^2 - x^2 - y^2 - z^2 = 0\ . \tag{5.5}$$

Woodhouse shows that this weaker condition is satisfied only by matrices which are of the form αL, where $\alpha > 0$ and L is a Lorentz transformation matrix obeying the rules we have outlined above. The further restriction to $\alpha = 1$ comes from the requirement that the time dilation factor depends only on the magnitude of relative velocity and so is the same between observers O and O' as between O' and O.

Let us define the *standard Lorentz transformation* in 1+3 dimensions as the result of performing a boost in the x-direction while keeping y and z unchanged. Under such a transformation, the x and t coordinates must transform as they did in $(1+1)$ dimensions, while the y and z transform trivially

$$y' = y\ , \qquad z' = z\ . \tag{5.6}$$

Thus, the standard Lorentz transformation is implemented by a 4×4 matrix of form

$$\begin{pmatrix} \gamma & \gamma v/c & 0 & 0 \\ \gamma v/c & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{5.7}$$

It is easy to check that this matrix satisfies the above conditions for a proper orthochronous Lorentz transformation. It is similarly easy to check that a pure rotation not involving the time coordinate, so a matrix of the form

$$\begin{pmatrix} 1 & 0 \\ 0 & H \end{pmatrix}\ , \tag{5.8}$$

where $H \in \mathrm{SO}(3)$, is also a proper orthochronous Lorentz transformation. Since the proper orthochronous Lorentz transformations form a group, everything obtained by the composition of the standard Lorentz transformations and pure rotations is inside the group of proper orthochronous Lorentz transformations.

It is only slightly more difficult to check the converse: that every L satisfying the conditions for the proper orthochronous Lorentz group can be written as a composition of pure rotations and a special Lorentz transformation. Specifically, we claim that we can always write

$$L = \begin{pmatrix} 1 & 0 \\ 0 & H \end{pmatrix} \begin{pmatrix} \gamma & \gamma v/c & 0 & 0 \\ \gamma v/c & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & H' \end{pmatrix} , \tag{5.9}$$

for some v and for some $H, H' \in \mathrm{SO}(3)$. The proof is left as an exercise.[4]

The group of rotations in three-space is a three-dimensional group. (There are nine entries H_{ij} for an orthogonal matrix, constrained by six conditions.) The space of Lorentz transformations is six-dimensional (ten conditions on sixteen matrix entries). You can think of these six dimensions as a choice of boost vector (three-dimensional), along with an orthogonal Cartesian basis for the spatial directions in the boosted frame (three-dimensional).

Exercise 3. The parameterization of a general Lorentz transformation given in (5.9)is naively seven-dimensional (two SO(3) matrices plus a choice of rapidity for the standard Lorentz matrix). How do you reconcile this with the Lorentz group being six-dimensional?

5.2 Poincaré transformations

So far we have restricted our attention to transformations of coordinates where both frames share the same origin. This is an unnecessarily restrictive assumption. The generalization to the situation where the origin is changed takes a simple form,

$$\begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} = L \begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} + \begin{pmatrix} T^0 \\ T^1 \\ T^2 \\ T^3 \end{pmatrix} , \tag{5.10}$$

where the T^a are constants. This transformation relates the origin in frame O' to the point T in frame O. Similarly, the transformation that relates the point T' in frame O' to the origin

[4] Hint: given the entries of L, you can find a pure rotation which, when left-multiplying L, has the effect of sending the last two entries in the 0-column of L to zero, while leaving the 0-row of L unchanged. Similarly there is another rotation which when right-multiplying L sends the last two entries in the 0-row of L to zero, while leaving the first column unaffected. The result of applying both of these can, with some work, be shown to be the product of the special Lorentz transformation with a rotation just in the yz-plane. This fact is tantamount to what is required.

in frame O is given by

$$\begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} = L \begin{pmatrix} ct' - T'^0 \\ x' - T'^1 \\ y' - T^2 \\ z' - T'^3 \end{pmatrix} , \tag{5.11}$$

These mappings are called *Poincaré transformations,* or *inhomogeneous Lorentz transformations.* Our original Lorentz transformations, with $T^a = 0$, are also called *homogeneous Lorentz transformations* when we are being very careful about terminology.

Poincaré group. The Poincaré transformations also form a group, known (unsurprisingly) as the *Poincaré group.* This group turns out to be the *semi-direct product* of the Lorentz group $\mathrm{SO}^+(1,3)$ extended by the translation group of four-dimensional spacetime $\mathbb{R}^{1,3}$,

$$\mathrm{ISO}^+(1,3) \equiv \mathrm{SO}^+(1,3) \ltimes \mathbb{R}^{1,3} . \tag{5.12}$$

This is the (identity component of) the *isometry group* of the four-dimensional Minkowski spacetime that we will meet below. The Poincaré group is ten dimensional. Six dimensions corresponding to homogeneous Lorentz transformations, plus four dimensions corresponding to translations in space-time.

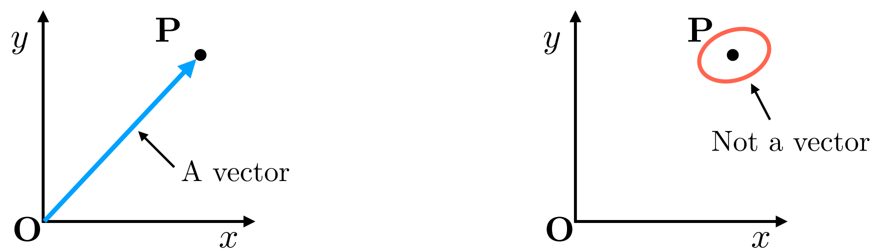
5.3 Four-vectors and Minkowski space

It would obviously be nice to introduce some streamlined vector notation such as X instead of writing out the four coordinates (ct, x, y, z). This turns out to be more than just a question of notation. It opens the way to the enormously important insight of the mathematician Hermann Minkowski (1864-1909), who in 1907-8 reformulated what Einstein had achieved in discarding the concept of absolute simultaneity. Minkowski saw that instead of thinking of physical reality as a direct product of one-dimensional time with a three-dimensional space, we should think in terms of a *four-dimensional space-time.* The methods and formulas developed by Einstein in 1905 could be seen as aspects of a new four-dimensional geometry, which Minkowski defined. Every development in physics since then has built on this idea. Einstein himself embraced this concept and took it much further with the theory of general relativity, so as to include gravity. Unfortunately Minkowski did not live to see this flowering of his geometric ideas. But Minkowski's name is remembered vividly, and the four-dimensional space-time of special relativity is called Minkowski space, or M. The second half of this course is largely devoted to the development of Minkowski's geometric viewpoint and its application to problems in special relativity.

5.3.1 Preliminaries about four-vectors

The general idea of a four-dimensional vector space is nothing new. The Prelims definition of a vector space, with linear transformations, bases, dual spaces, dual bases, matrix representations, and so on, applies here and will be assumed. But we have to be rather careful about what a four-vector is and we cannot just write down $X = (ct, x, y, z)$.

In Newtonian dynamics in three dimensions we may typically have written down $\mathbf{r} = (x, y, z)$, and thought of the vector $\mathbf{r}$ as giving the coordinates of a point. Actually, it should be thought of as the *displacement* of the point from the origin O, or its coordinates *relative to* O. In fact, when you learned about vectors at school, you may have used the notation $\overrightarrow{OP}$ for the position of a point P relative to an origin O. See figure.



We need to be careful with this distinction, so we shall not call the coordinate set (ct, x, y, z) a vector. The reason is simple: we need to allow for a change of coordinates in which the origin changes, and under such transformations (the Poincaré transformations, or inhomogeneous Lorentz transformations), the coordinate set changes according to (5.10). We shall reserve the term *four-vector* for entities which transform according to

$$X = LX' \; , \tag{5.13}$$

even when the origin is changed. This means that *displacements* or the *relative* coordinates of space-time events will be four-vectors. Formally, if P is at (ct_1, x_1, y_1, z_1), and Q is at (ct_2, x_2, y_2, z_2), then the displacement

$$\overrightarrow{PQ} = (\, ct_2 - ct_1 \; , \; x_2 - x_1 \; , \; y_2 - y_1 \; , \; z_2 - z_1 \,) \tag{5.14}$$

transforms correctly and so is an example of a four-vector.

In what follows in this section, you can visualize four-vectors as being such displacement vectors — just like the vectors $\overrightarrow{AB}$ you may first have used at school. However, we shall be moving on later to consider four-vectors for velocity, momentum, acceleration, force, and other things, in just the same way as we do with three-vectors.

Next, we need to pay attention to a vital difference between the three-vectors of Euclidean space, and the new four-vectors. The distinction is very simple: Euclidean space has a norm defined by length, inducing an inner product, and so the existence of *orthonormal bases*. In Euclidean geometry we always use such orthonormal bases, typically written as $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$, or more generally as $\{\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3}\}$. In Minkowski geometry we have no such orthonormal bases.

When we have an orthonormal basis we can, rather lazily, forget the distinction between a vector space and its dual vector space. The existence of the inner product gives a natural identification between the two spaces (while for a general vector space there is no such identification). This is made very explicit by the definition of the dot product in three-dimensional Euclidean geometry.

Consider the expression $\mathbf{u} \cdot \mathbf{v}$. This can actually be thought of in three different ways.

- The dot defines an inner product structure mapping a pair of vectors to a real number. The formula for it is $\sum_i u_i v_i$, where u_i and v_i are the components of $\mathbf{u}$ and $\mathbf{v}$ in some orthonormal basis.
- $\mathbf{v}$ is a vector, while $\mathbf{u}\cdot$ is an element of the dual vector space. The dot therefore defines a mapping from the vector space to the dual vector space.

  More precisely, given the vector space V of three-dimensional vectors, in linear algebra we would define the *dual vector space* V^* as the space of linear maps $\varphi : V \to \mathbb{R}$ from V to the real numbers. In the present case the set of linear maps is labelled by three-dimensional vectors, such that $\varphi_{\mathbf{u}}(\mathbf{v}) = \mathbf{u} \cdot \mathbf{v}$.
- The same as above, but with $\mathbf{u}$ and $\mathbf{v}$ reversed.

In practice we don't need to distinguish these interpretations, as the formula $\sum_i u_i v_i$ is the same however we think of it. The underlying reason for this is that if $\mathbf{u}$ has components (u_1, u_2, u_3) in some orthonormal basis, then the operation of dotting with $\mathbf{u}$, which is an element of the dual space, has the *same components* (u_1, u_2, u_3) in the dual basis. So the numbers (u_1, u_2, u_3) may refer to either the vector space or its dual, and we don't need to draw the distinction. In other words, if we use an orthonormal basis for the vector space, and its dual basis for the dual space, the dot is a map between the two spaces which is represented by the *identity matrix*.

The reason why this is rather lazy is that there is a real difference between a vector space and its dual, often with an important meaning in physical application. For instance, take the formula for a directional derivative, $\mathbf{p} \cdot \nabla\phi$. The $\mathbf{p}$ is a displacement vector, but the $\nabla\phi$ is a gradient, which is essentially different. Physically, suppose $\phi(x, y, z)$ is a function giving the temperature in a region of space. Then $\mathbf{p}$ has the dimensions of a length, but $\nabla\phi$ is measured in degrees *per* length, *i.e.*, with the dimensions of inverse length. If we change units from centimetres to metres, $\mathbf{p}$ decreases by a factor of 100, but $\nabla\phi$ increases by 100, while $\mathbf{p} \cdot \nabla\phi$ is invariant.

It would actually be more correct to use a different notation for elements of the vector space and elements of its dual. This can be done by using *upper and lower indices*. We use upper indices for vectors, for which displacement vectors are the model. So, in this notation we would for instance write $\mathbf{r} = (x, y, z) \to (x^1, x^2, x^3)$, for the displacement between the origin and the point with coordinates (x, y, z). We use lower indices for dual vectors. An example would be the gradient $\nabla\phi = (\frac{\partial\phi}{\partial x^1}, \frac{\partial\phi}{\partial x^2}, \frac{\partial\phi}{\partial x^3}) = (\partial_1\phi, \partial_2\phi, \partial_3\phi)$. Summation of indices must always involve one upper index and one lower index, reflecting the definition of the dual vector space as the space of linear maps acting on the vector space. So for instance, for the directional derivative example we would write $\mathbf{p} \cdot \nabla\phi = \sum_i p^i \partial_i \phi$.

If we adopt this more careful notation, we can rewrite the formula for $\mathbf{u} \cdot \mathbf{v}$ in one of a number of possible ways:

- $\sum_{ij} u^i \delta_{ij} v^j$, which expresses the idea of the dot as an inner product mapping a pair of vectors to a real number. The dot has matrix representation δ_{ij}, *i.e.*, the identity matrix
- $\sum_i u_i v^i$, which expresses the idea of the vector $\mathbf{v}$ being acted on by the operation 'dotting with $\mathbf{u}$', which is an element of the dual vector space and so has a lower index
- $\sum_i u^i v_i$, vice versa
- $\sum_i (\sum_j u^j \delta_{ij}) v^i$, which expresses the idea of mapping the vector $\mathbf{u}$ into the corresponding dual vector, and then having it act on $\mathbf{v}$.
- $\sum_i u^i (\sum_j v^j \delta_{ij})$, vice versa.

This may seem ridiculously fussy, which is why we don't bother with the distinction in Euclidean 3d-geometry, but in Minkowski space we shall find that we have to take it seriously. Fortunately, a streamlined notation enables us to deal with all the issues without too much difficulty.

5.3.2 Four-vector algebra

We are now ready to define the algebra and geometry of four-vectors. The most fundamental thing to remember is that a four-vector is not just a list of numbers, its components. It is a specification of what those components will be in every admissible coordinate system. In Euclidean space, the admissible coordinates are those defined by choices of mutually orthogonal x, y, and z directions, with different choices being related by rotations. In Minkowski space, the admissible coordinates are those defined by choices of mutually pseudo-orthogonal t, x, y, and z directions, with different choices being related by proper orthochronous inhomogeneous Lorentz transformations.

Suppose O has coordinates t, x, y, z and O' has coordinates t', x', y', z'. Defining $x^0 = ct$, $x^1 = x$, $x^2 = y$, and $x^3 = z$, we have the following transformation law for the inertial coordinates,

$$x^a = \sum_{b=0}^{3} L^a{}_b x'^b + T^a \ , \qquad a = 0, 1, 2, 3 \ . \tag{5.15}$$

We will call a quantity X a *four-vector* if the components of X in the two inertial coordinate systems are related by

$$X^a = \sum_{b=0}^{3} L^a{}_b X'^b \ , \qquad a = 0, 1, 2, 3 \ . \tag{5.16}$$

We immediately see that the coordinates x^a themselves are not four-vectors due to the inhomogeneous term in (5.15), but displacement vectors of the form $X^a = x^a - y^a$, will satisfy the criterion.

We can define a bilinear map g from pairs of four-vectors to the real numbers, an analogue of the dot product in Euclidean space:

$$g(X,Y) = \sum_{a,b=0}^{3} g_{ab} X^a Y^b \ , \tag{5.17}$$

where $g_{ab} = \mathrm{diag}(1,-1,-1,-1)$ is the invariant matrix from before.

Proposition 2 (Lorentz invariance of bilinear map)**.** The bilinear map $g(X,Y)$ is independent of the coordinate system, *i.e.*, it is Lorentz invariant.

Proof. This follows from the definition of the Lorentz transformations and the definition of four-vectors. First recall that the condition $L^{\mathrm{T}} g L = g$ means, in terms of components, that $\sum_{ab} L^a{}_c g_{ab} L^b{}_d = g_{cd}$.

Then we have

$$g(X,Y) = \sum_{a,b} g_{ab} X^a Y^b = \sum_{a,b} g_{ab} \sum_c L^a{}_c X'^c \sum_d L^b{}_d Y'^d = \sum_{c,d} g_{cd} X'^c Y'^d = g(X',Y') \ ,$$

as required. □

Proposition 3. The bilinear map is symmetric, $g(X,Y) = g(Y,X)$, but does not define an inner product.

Proof. The symmetry is obvious. The crucial fact is that $g(X,X)$ does not satisfy the criterion for an inner product, that it is non-negative and zero only if X is zero. □

Example 4. Compute $g(X,X)$ for: i.- $X = (1,0,0,0)$, ii.- $X = (0,1,0,0)$ and iii.- $X = (1,1,0,0)$.

In the first case we get $g(X,X) = 1$. In the second case $g(X,X) = -1$. In the third case $g(X,X) = 0$.

Definition 13. $g(X,Y)$ is said to define a *pseudo-inner product* on the vector space of four-vectors. It also defines a notion of *pseudo-orthonormal basis*. These are a set of four-vectors $\{e_0, e_1, e_2, e_3\}$ with the property $g(e_0,e_0) = 1, g(e_1,e_1) = g(e_2,e_2) = g(e_3,e_3) = -1, g(e_0,e_1) = g(e_0,e_2) = g(e_0,e_3) = g(e_1,e_2) = g(e_1,e_3) = g(e_2,e_3) = 0$.

In any admissible coordinate system, the vectors $(1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0), (0, 0, 0, 1)$ have this property.

Such a basis defines a dual basis for the dual vector space. We shall use lower indices for expressing the components of dual vectors with respect to that dual basis. If X is a vector and T a dual vector, then the composition $T(X)$ is written in components as $\sum_a T_a X^a$. This definition follows simply from the concept of vector space duality, and has got nothing to do with the existence of g. However, g induces a mapping from vectors to dual vectors, which is called 'lowering the index'. If X^a is a four-vector, then we can define a corresponding dual vector X_a by

$$X_a = \sum_b g_{ab} X^b \tag{5.18}$$

Now $g(X, Y)$ can be written as $\sum_a X_a Y^a$, in which X_a is a dual vector acting on the vector Y^a. Of course we could equally well write it as $\sum_a X^a Y_a$.

It is very useful to streamline the notation by dropping the summation symbol $\sum_{b=0}^{3}$. This is called the *Einstein summation convention*. We can also drop the conditions such as that in (5.15) which tells us that the statement applies for $a = (0, 1, 2, 3)$.

Summation convention: When an index in a term is repeated, once as an upper index and once as a lower index, a sum over 0, 1, 2, 3 is implied.

Range convention: An index which is not repeated is a free index. Any equation is understood to hold for all values of the free indices over the range $0, 1, 2, 3$.

With these conventions, the relation between co-ordinates is expressed simply by

$$x^a = L^a{}_b x'^b + T^a\,, \tag{5.19}$$

and the relation between components of vectors by

$$X^a = L^a{}_b X'^b\,. \tag{5.20}$$

The condition for L to be Lorentz is

$$L^a{}_c g_{ab} L^b{}_d = g_{cd}\,. \tag{5.21}$$

The pseudo-inner-product is defined as

$$g(X, Y) = g_{ab} X^a Y^b = X^0 Y^0 - X^1 Y^1 - X^2 Y^2 - X^3 Y^3\,, \tag{5.22}$$

in which two summations (one over a, one over b) are left implicit by the summation convention.

Lowering the index is written simply as

$$X_a = g_{ab} X^b\,, \tag{5.23}$$

and now $g(X,Y) = X_a Y^a = X^a Y_a$.

Remark: It should be clear by now why we write the L matrix entries as $L^a{}_b$. This means we adhere to the convention about summing over indices only when one is upper and the other is lower.

The value of all this machinery lies in the following fact. If the components of X^a are (X^0, X^1, X^2, X^3), then the components of X_a are $(X^0, -X^1, -X^2, -X^3)$. We no longer have the comfy Euclidean situation where the vectors and dual vectors can be identified with each other! These pesky minus signs come into every line of algebra we do, and have to be tracked very carefully. Fortunately, the machinery of the indices and the rules for combining four-vectors and the g matrix, if obeyed carefully, takes care of everything.

We can also define the pseudo-inner-product on dual vectors, and for this we need

$$g^{ab} = \mathrm{diag}(1,-1,-1,-1)\,. \tag{5.24}$$

Written out as a matrix this looks the same as g_{ab}, but it is in a different space as it acts on dual vectors. It should be thought of as the inverse matrix of g_{ab}, defined by[5]

$$g^{ab} g_{bc} = \delta^a{}_c.$$

With this definition,

$$g^{ab} X_b = X^a\ , \tag{5.25}$$

and now we can freely raise and lower indices.

Exercise 4. Check that this definition of lowering and raising indices is consistent, i.e that $g^{ab} g_{bc} X^c = X^a$.

Exercise 5. Show that dual vectors transform as $X_a = X'_b (L^{-1})^b{}_a = g_{ac} L^c{}_d g^{db} X'_b$.

Proposition 4. A Lorentz transformation can be characterized by specifying a pseudo-orthonormal basis. The rows of the L matrix can be read as sequence of four vectors, giving such a basis. So can the four columns.

Proof. The argument is just the same as for the rotation matrices in Euclidean space. This property is just another way of reading the defining property of the Lorentz transformation matrix. $\square$

[5]Note that in $3+1$ dimensions the Kronecker delta has an index up and an index down, and it's defined as $+1$ if the indices agree, and zero otherwise. In Euclidean space the distinction between up and down is not important, as discussed.

5.4 Classification of four-vectors

Using the pseudo-norm on four-vectors, we can classify four-vectors in a coordinate invariant way.

Definition 14. A four-vector X is said to be *timelike, spacelike, or null* according as $g(X,X) > 0, g(X,X) < 0$, or $g(X,X) = 0$.

Examples: The four-vectors with components $(1,0,0,0), (0,1,0,0)$, and $(1,1,0,0)$ (in some ICS) are respectively timelike, spacelike, and null. Note that a null four-vector need not be the zero four-vector.

A (non-zero) four-vector whose spatial part vanishes in some ICS must be timelike; and a (non-zero) four-vector whose temporal part vanishes in some ICS must be spacelike. The converses of these statements are the following two propositions.

Proposition 5. If X is timelike, there exists an ICS in which $X^1 = X^2 = X^3 = 0$.

Proof. As X is timelike, it has components of the form $(P, p\,\mathbf{e})$, where $\mathbf{e}$ is a unit spatial vector and $|P| > |p|$. Now consider the four four-vectors:

$$\frac{1}{\sqrt{P^2 - p^2}}(P, p\,\mathbf{e}), \quad \frac{1}{\sqrt{P^2 - p^2}}(p, P\,\mathbf{e}), \quad (0, \mathbf{q}), \quad (0, \mathbf{r}) \ , \tag{5.26}$$

where $\mathbf{q}, \mathbf{r}$ are chosen so that $(\mathbf{e}, \mathbf{q}, \mathbf{r})$ form an orthonormal triad in Euclidean space.

These four-vectors form a pseudo-orthonormal basis and so define an explicit Lorentz transformation to new coordinates. In these new coordinates the original four-vector has components $(1, 0, 0, 0)$. □

An alternative argument, in which we first do a rotation and then a standard velocity transformation, is given in Woodhouse (pp. 90-91). By a similar argument, we also have:

Proposition 6. If X is spacelike, then there exists an ICS in which $X^0 = 0$.

In the case of timelike and null vectors (but not spacelike vectors), the sign of the time component X^0 is invariant.

Proposition 7. Suppose that X is timelike or null. If $X^0 > 0$ in some ICS, then $X^0 > 0$ in every ICS.

Proof. Since rotations do not alter X^0, it is sufficient to consider what happens to X^0 under a standard Lorentz transformation. One need only note that if X^0 is positive, $|X^1| < |X^0|$, and $|u| < c$, the expression $\gamma(u)(X^0 + uX^1/c)$ is necessarily positive. □

Definition 15. A timelike or null vector X is said to be *future-pointing* if $X^0 > 0$ in some (and hence every) ICS, and *past-pointing*, if $X^0 < 0$.

Some analogous statements for null vectors are left to the worksheet.

The space of four-vectors is illustrated in Figure 10, where the X^0 axis is vertical and one spatial dimension is suppressed. The null vectors lie on the cone

$$(X^0)^2 - (X^1)^2 - (X^2)^2 - (X^3)^2 = 0 \ , \tag{5.27}$$

which has its vertex at the origin.

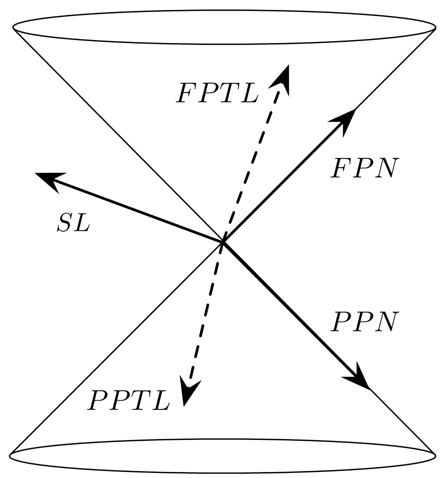


Figure 10. *The space of four-vectors, illustrating future-pointing timelike, future-pointing null, space-like, past-pointing null, past-pointing timelike.*

This is called the *light cone* and is of fundamental importance in space-time geometry.

Consider a free particle moving in a straight line between two events A and B. If the displacement vector $B - A$ is timelike, it means the particle is moving at a speed v, smaller than the speed of light. We then say that B is inside the light-cone of A. If the displacement vector is null, it means the particle is moving at the speed of light. We then say that B is on the light-cone of A. In these two cases we can say whether A is in the future of B or vice versa.

If the displacement vector is spacelike, then the particle is moving at a speed faster than the speed of light. We then say that B is outside the light-cone of A. It's important to note that we cannot make a definition of 'future-pointing' and 'past-pointing' for spacelike vectors, because the sign of the time component of a spacelike vector is not Lorentz invariant; it will be different in different ICSs.

This means that a motion which is going into the future faster than light in one frame, will appear as infinitely fast in another, and as going faster than light into the past in yet another. The ability to travel faster than light would imply the ability to travel into the past (and lead to the famous *Back to the Future* paradoxes of causality).

This does not mean that spacelike vectors are unrelated to physical effects. For instance, suppose you sweep the face of the Moon with a laser spotlight. The point where the laser hits the Moon's surface may well move faster than light across it. This means that there is an ICS in which this spot is moving backwards in the time coordinate. This does not lead to any paradox as the motion of the spot of light is not conveying any information faster than light.

5.5 Triangle inequality

A fundamental feature of Euclidean geometry is the triangle inequality, which states that for any vectors $\mathbf{x}$ and $\mathbf{y}$ we will have

$$|\mathbf{x}+\mathbf{y}| \leqslant |\mathbf{x}| + |\mathbf{y}| \ . \tag{5.28}$$

This inequality encapsulates the fact that a straight line is the shortest distance between any two points in Euclidean space. We can find an analogous statement in Minkowski geometry.

Proposition 8 (Minkowski triangle inequality)**.** If U and V are future-pointing, timelike four-vectors, then $U+V$ is also future-pointing timelike and satisfies

$$\sqrt{g(U+V,U+V)} \geqslant \sqrt{g(U,U)} + \sqrt{g(V,V)} \ . \tag{5.29}$$

Proof. If U is future-pointing timelike, then there is an ICS in which it has components $(U^0,0,0,0)$, with $U^0>0$. By a further rotation of the (x,y,z) coordinates, there is an ICS in which U takes this form and V has components $(V^0,V^1,0,0)$, with $V^0>0$ and $|V^0|>|V^1|$. $(U+V)$ is then timelike since $g(U+V,U+V)=((U^0+V^0)^2-(V^1)^2)>(V^0)^2-(V^1)^2>0$. Also $U^0+V^0>0$ so it is future-pointing timelike.

The inequality to be shown then becomes

$$\sqrt{(U^0+V^0)^2-(V^1)^2} \geqslant U^0 + \sqrt{(V^0)^2-(V^1)^2} \tag{5.30}$$

Squaring both sides, this is equivalent to

$$(U^0+V^0)^2-(V^1)^2 \geqslant (U^0)^2 + 2U^0\sqrt{(V^0)^2-(V^1)^2} + (V^0)^2-(V^1)^2 \ , \tag{5.31}$$

which, after cancelling terms from both sides, is equivalent to

$$2U^0V^0 \geqslant 2U^0\sqrt{(V^0)^2-(V^1)^2} \ , \tag{5.32}$$

which is obviously true. Equality occurs if and only if $V^1=0$, *i.e.*, U and V are proportional. □

The pseudo-norm of a future-pointing timelike vector is the amount of time that will be measured on the clock of an inertial observer traveling along that same vector. Thus, the inequality tells us that the straight-line path between time-like separated events A and B, *i.e.*, the path maintaining constant velocity, is the one that takes the *longest* according to an observer following that path. What is the shortest path? There is no lower limit: by traveling at speeds nearer and nearer to c the proper time can be arbitrarily small.

5.6 Invariant operators in $3+1$ dimensions

Let us start with a brief discussion in three spatial dimensions. Consider the three partial derivatives

$$\partial_x = \frac{\partial}{\partial x}, \quad \partial_y = \frac{\partial}{\partial y}, \quad \partial_z = \frac{\partial}{\partial z}$$

It turns out they transform as the components of a vector operator ∇. In other words, consider the coordinate transformation

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = H \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} + T \tag{5.33}$$

with H a proper orthogonal matrix and T a constant column vector, it is possible to show

$$\begin{pmatrix} \partial_x \\ \partial_y \\ \partial_z \end{pmatrix} = H \begin{pmatrix} \partial_{x'} \\ \partial_{y'} \\ \partial_{z'} \end{pmatrix} \tag{5.34}$$

By acting with ∇ on a scalar field f or a vector field $\mathbf{X}$ we can form the familiar differential operators from vector calculus $\mathrm{grad} f = \nabla f$, $\mathrm{div}\,\mathbf{X} = \nabla \cdot \mathbf{X}$, $\mathrm{curl}\,\mathbf{X} = \nabla \times \mathbf{X}$. These operators are invariant, in the sense that they are the same regardless of the (right-handed) Cartesian coordinate system in which they are applied.

We will now present the corresponding operators in $3+1$ dimensions. Let (t, x, y, z) be the coordinates of an ICS, we then define

$$\partial_a = \frac{\partial}{\partial x^a}, \quad a = 0, 1, 2, 3.$$

where $x^0 = ct$ and $x^1 = x, x^2 = y, x^3 = z$. For a function f of the coordinates introduce the following notation

$$\partial f = (\partial_0 f, \partial_1 f, \partial_2 f, \partial_3 f)$$

Now consider an inhomogeneous Lorentz transformation (5.10) and define $\partial' f$ in the same way, where $\partial'_0 = \frac{\partial f}{\partial x'^0}$ and so on. Then we have the following

Lemma 1. For any function of spacetime $\partial' f = \partial f\, L$.

Which can be easily proved. From this we will construct a four-vector. By using the pseudo-orthogonality relation $L^{-1} = gL^T g$ we then have

$$\partial f g = \partial' f g L^T \to (\partial f g)^T = L\left(\partial' f g\right)^T$$

so that $(\partial f g)^T$ does transform as a vector. This suggests the following definition

Definition 16. The *four-gradient* $\mathrm{Grad}\, f$ of a function on space-time is the four-vector given by

$$\mathrm{Grad}\, f = (\partial f g)^T = \begin{pmatrix} \partial_0 f \\ -\partial_1 f \\ -\partial_2 f \\ -\partial_3 f \end{pmatrix} \tag{5.35}$$

We can then say that the following operator

$$\left(\frac{1}{c}\frac{\partial}{\partial t}, -\frac{\partial}{\partial x}, -\frac{\partial}{\partial y}, -\frac{\partial}{\partial z}\right)$$

transforms as a four-vector. Given a four-vector field X, that is a four-vector that varies from event to event, we can form an invariant scalar $\mathrm{Div}\, X$ by taking the inner product of the four-vector operator with X.

Definition 17. The *four-divergence* $\mathrm{Div}\, X$ of a four-vector field X is the function

$$\mathrm{Div}\, X = \frac{1}{c}\frac{\partial X^0}{\partial t} + \frac{\partial X^1}{\partial x} + \frac{\partial X^2}{\partial y} + \frac{\partial X^3}{\partial z}$$

The four-divergence has the property that it is invariant under change of inertial coordinate system.

Given a function f on space-time, we can form the four gradient $\mathrm{Grad}\, f$ (which is a four-vector) and then take the four-divergence. The result is the d'Alembertian operator

$$\Box f \equiv \mathrm{Div}\ (\mathrm{Grad}\, f) = \frac{1}{c^2}\frac{\partial^2 f}{\partial t^2} - \frac{\partial^2 f}{\partial x^2} - \frac{\partial^2 f}{\partial y^2} - \frac{\partial^2 f}{\partial z^2}$$

It of course follows that the d'Alembertian is an invariant operator.

6 Relativistic Kinematics

Now that we have developed the technology of four-vectors and understood their algebraic properties, we return to the problem of describing the motion of particles and objects in Minkowski space.

6.1 Four-velocity and proper time

We would like to promote the velocity of a particle, a three-vector in Newtonian physics, to a four-vector. That such a promotion should be possible is intuitively clear from the point of view of space-time diagrams, where a particle trajectory is a curve: the vector tangent to the world-line at a given point should be some sort of velocity four-vector.

Naively, we would like to define a four-velocity as the time-derivative of the position in space-time,

$$V^a \stackrel{?}{=} \frac{\mathrm{d}x^a(t)}{\mathrm{d}t}\ . \tag{6.1}$$

The problem is that we need to specify *which time* we should be differentiating with respect to. If it is the time coordinate in the same ICS in which the x^a are defined, then we would have $V^0 = c$, which cannot be true in every Lorentz frame if V transforms as a four-vector.

The key is to identify a canonical parametrisation s of the particle world-line on which all observers will agree. Said differently, at any point on the particle world-line we would like to identify a canonical ICS whose time coordinate we should use in the expression in (6.1). It is then clear that there is only one reasonable option: we should use the time coordinate of the ICS in which the particle is instantaneously at rest.

We have seen in our discussion of time dilation that in a fixed ICS, the time coordinate s on the worldline of a particle moving at instantaneous velocity $\mathbf{v}$ is related to the time coordinate t of the ICS according to

$$\frac{\mathrm{d}s}{\mathrm{d}t} = \frac{1}{\gamma(v)} = \sqrt{1 - \frac{\mathbf{v}^2}{c^2}}\ . \tag{6.2}$$

This gives us the following expression for the infinitesimal change in time in the ICS in which the particle is at rest in terms of the reference coordinates:

$$c\,\mathrm{d}s = \sqrt{c^2\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2}\ . \tag{6.3}$$

This is analogous to the infinitesimal measure of distance in Euclidean space,

$$\mathrm{d}s_{\mathrm{E}} = \sqrt{\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2}\ , \tag{6.4}$$

and just as Euclidean distance in unchanged under a rotation of coordinates, the measure of proper time is unchanged under a Lorentz transformation. Thus we make the following definition.

Definition 18. The *proper time* at any event P on a time-like world-line Γ is given by

$$s(P) = \frac{1}{c}\int_{P_0}^{P} \sqrt{c^2dt^2 - dx^2 - dy^2 - dz^2}\ , \tag{6.5}$$

where (ct, x, y, z) are coordinates in an inertial frame and the integral is taken along the world-line Γ. The freedom to choose the initial event P_0 on Γ is the freedom to re-define proper time by an additive shift. Due to Lorentz invariance of the infinitesimal measure, the proper time is independent of the choice of ICS.

Proper time clearly coincides with coordinate time t for uniform motion in an ICS where the particle is at rest, but differs from t for general motion or in a general frame. Proper time is well-defined for accelerating trajectories, just as Euclidean distance is sensible for curves. The *clock hypothesis* says that it is correct to consider (or perhaps define) an *ideal clock* to measure such proper time, which means that its working must not be affected by acceleration. In practice we may assume that atoms are at least good approximations to ideal clocks. Astronauts in an accelerated rocket will experience time and age according to proper time, since all their atoms may be assumed to be clocks which see proper time. We will return to this below in our discussion of four-acceleration.

Note that for space-like curves a similar definition can be made of proper distance by including an overall minus sign under the square root. This leads to the notion of the *proper length* of an extended body, which can be thought of as the length of the body in its own rest frame. For null lines, on the other hand, no useful analogue exists; null lines can be parametrised as paths in space-time, but there is no canonical parameter.

Our definition of proper time allows us to improve our original naive attempt to make velocity into a four-vector. We can now use s as a natural parametrisation of any time-like world-line,

$$x^a = x^a(s), \quad a = 0, 1, 2, 3 \tag{6.6}$$

and we then define the *velocity four-vector* by

$$V^a = \frac{\mathrm{d}x^a(s)}{\mathrm{d}s} \ . \tag{6.7}$$

This definition has the advantage that the following proposition holds true:

Proposition 9. The V^a are the components of a four-vector.

Proof. As we discussed earlier, the coordinates $x^a(s)$ do not form a four-vector, but the displacement vector $x^a(s + \delta s) - x^a(s)$ does. Taking the limit, we have

$$V^a(s) = \lim_{\delta s \to 0} \frac{x^a(s + \delta s) - x^a(s)}{\delta s} \ .$$

Under a Poincaré transformation, the proper time is invariant, so it is clear that this transforms as a four-vector. □

We can perform a direct computation of the form of the four-velocity in a given ICS:

$$\frac{\mathrm{d}x^a}{\mathrm{d}s} = \frac{\mathrm{d}t}{\mathrm{d}s}\frac{\mathrm{d}x^a}{\mathrm{d}t} = \gamma(v)\,(c, \mathbf{v})$$

where $\mathbf{v}$ is the three-velocity (of the particle whose world-line we are focusing on) in the ICS. Hence we can make the following observation.

Proposition 10. For any four-velocity V, $g(V,V) = c^2$.

Proof. We choose any ICS, and then have

$$g(V,V) = \gamma^2(v)\left(c^2 - \mathbf{v}\cdot\mathbf{v}\right) = c^2 \ .$$

□

The fact that four-velocities have a fixed pseudo-norm makes a certain amount of sense. The three-velocity is determined by three numbers, and to promote it to a four-vector we would naively have to find a fourth parameter somewhere. In fact, we do not, because the four-velocity is a constrained four-vector and is still determined by three parameters.

The use of four-vectors greatly assists in calculations of relativistic effects by eliminating the need to carry out actual transformation of coordinates. Instead, we can make good use of Lorentz invariant quantities whenever possible. (This is exactly as in Prelims Geometry, where many statements about circles and triangles can be simplified by using the dot product instead of writing out coordinates.)

Example 5 (Velocity addition in 3 dimensions.)**.** Relative to some ICS, an observer has velocity $\mathbf{u}$ and a particle has velocity $\mathbf{v}$. Find the speed w of the particle relative to the observer in terms of $\mathbf{u}$ and $\mathbf{v}$.

Solution: This is the velocity addition problem all over again, but now in three dimensions where we have to worry about relative angles between the two velocities $\mathbf{u}$ and $\mathbf{v}$. We will solve it using four-vectors.

Let U be the four-velocity of the observer and let V be the four-velocity of the particle. In the given ICS,

$$U = \gamma(u)(c,\mathbf{u}) \ , \qquad V = \gamma(v)(c,\mathbf{v}) \ ,$$

so

$$g(U,V) = \gamma(u)\gamma(v)(c^2 - \mathbf{u}\cdot\mathbf{v}) \ .$$

Now consider an ICS in which the observer is at rest. In this system, the same four-vectors have components

$$U' = (c,0,0,0) \ , \qquad V' = \gamma(w)(c,\mathbf{w}) \ .$$

so

$$g(U',V') = c^2\gamma(w) \ .$$

But $g(U,V)$ is invariant so that $g(U,V) = g(U',V')$. Therefore

$$c^2\gamma(w) = \gamma(u)\gamma(v)(c^2 - \mathbf{u}\cdot\mathbf{v}) \ . \tag{6.8}$$

Solving for the magnitude w of the three-velocity involves a bit of algebra, but eventually without too much difficulty we find

$$w = \frac{c\sqrt{c^2(\mathbf{u}-\mathbf{v})\cdot(\mathbf{u}-\mathbf{v}) - |\mathbf{u}\wedge\mathbf{v}|^2}}{c^2 - \mathbf{u}\cdot\mathbf{v}} \ . \tag{6.9}$$

This reduces to the one-dimensional addition formula (3.28) when $\mathbf{u}$ and $\mathbf{v}$ are parallel.

Example 6 (Simultaneity and pseudo-orthogonality)**.** An inertial observer O has four-velocity U in some ICS. Let A and B be two events, with displacement four-vector $B - A = X$. Show that O reckons that A and B are simultaneous if and only if $g(U,X) = 0$.

Solution. Pick an ICS where O is at rest, so then U has components $U = (c,0,0,0)$. Suppose that in this ICS, $X = (X^0, X^1, X^2, X^3)$. Now O reckons that A and B are simultaneous if and only if $X^0 = 0$. But $g(U,X) = cX^0$, and the result follows.

Thus the concept of 'simultaneous to an intertial observable' is equivalent to the geometric concept of pseudo-orthogonality with the four-velocity of the observer.

Example 7 (Agreed simultaneity)**.** Two observers O and O' are travelling in straight lines at constant speeds. Show that there is a pair of events A and A', with A on the worldline of O and A' on the worldline of O', which O and O' both reckon are simultaneous.

Solution. Let the four-velocities of O and O' be U and U', respectively. Then the worldline of O is given by $P + sU$, where P is some event on the worldline and s ranges over the real numbers. The worldline of O' is likewise $Q + s'U'$. From the preceding example, the simultaneity conditions can be written as

$$\begin{aligned} g((P+sU) - (Q+s'U'), U\ ) &= 0 \ , \\ g((P+sU) - (Q+s'U'), U') &= 0 \ , \end{aligned} \tag{6.10}$$

which gives two linear equations for s and s',

$$\begin{aligned} g(P-Q,U) + c^2 s - g(U',U)s' &= 0 \ , \\ g(P-Q,U') + g(U,U')s - c^2 s' &= 0 \ . \end{aligned} \tag{6.11}$$

They will have a unique solution provided $g(U,U') \neq c^2$, which is just the condition that U and U' are not equal. (If U and U' are equal then, trivially, the observers are stationary relative to one another and there are infinitely many sets of events satisfying the condition.)

6.2 Four-acceleration

One sometimes reads that Special Relativity is only equipped to describe uniform motion, and that in order to treat acceleration requires the theory of General Relativity. These statements are somewhat misleading. The real restriction in Special Relativity is that the *coordinate systems* we use are inertial coordinate systems, related by Poincaré transformations accounting for uniform relative motion.[6] This restriction does nothing to prevent us from discussing the properties of accelerated bodies. Indeed, Einstein's 1905 paper was titled 'On the Electro*dynamics* of Moving Bodies', and dynamics is all about force and acceleration.

Let us consider a trajectory in spacetime $x^a(s)$, parameterized by $s \in \mathbb{R}$, such that

$$\frac{\mathrm{d}x^a}{\mathrm{d}s} = V^a \ , \qquad g_{ab}V^aV^b = c^2 \ , \tag{6.12}$$

so s is a measure of proper time. Differentiating the right hand equation in (6.12) with respect to proper time gives us

$$g_{ab}V^a\frac{\mathrm{d}V^b}{\mathrm{d}s} + g_{ab}\frac{\mathrm{d}V^a}{\mathrm{d}s}V^b = 2g_{ab}V^a\frac{\mathrm{d}V^b}{\mathrm{d}s} = 0 \ . \tag{6.13}$$

Definition 19. The *acceleration four-vector* for a time-like trajectory with four-velocity V is given by

$$A^a = \frac{\mathrm{d}V^a}{\mathrm{d}s} \ .$$

We see that we have the general relation

$$g(A, V) = 0$$

for *any time-like trajectory*. Again, we see that the four-acceleration is a *constrained four-vector*, so the space of allowed four-accelerations for a trajectory with a fixed four-velocity is three-dimensional.

If we consider the ICS where, instantaneously at some proper time, $V^a = (V, 0, 0, 0)$, then A^a must be of the form $(0, \mathbf{a})$ for some three-vector $\mathbf{a}$. The velocity four-vector takes the general form

$$V = \gamma(v)(c, \mathbf{v}) \ , \tag{6.14}$$

from which we obtain

$$A = \frac{\mathrm{d}V}{\mathrm{d}s} = \gamma(v)\frac{\mathrm{d}V}{\mathrm{d}t} = c^{-2}\gamma(v)^4(c, \mathbf{v})v\frac{\mathrm{d}v}{\mathrm{d}t} + \gamma(v)^2\left(0, \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t}\right) \ . \tag{6.15}$$

[6] There is also a restriction that one must neglect the effects of gravity. It turns out that the key idea for incorporating gravity is exploiting the freedom to use *any coordinates we like*, and this is the origin of the word 'general' in General Relativity.

Since we have $\mathbf{v} = 0$ in the frame where the particle is instantaneously stationary, we obtain $A = (0, \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t})$. Thus in this frame $\mathbf{a}$ is *precisely the three-acceleration.*

Example 8 (Constant acceleration)**.** Find the coordinates in an ICS for a worldline with $y = z = 0$ that experiences constant acceleration with magnitude κ.

Solution: In the ICS, the four-velocity and four-acceleration of the worldline will be given by

$$U = \left(c\frac{\mathrm{d}t}{\mathrm{d}s}\,,\ \frac{\mathrm{d}x}{\mathrm{d}s}\,,\ 0\,,\ 0\right)\ ,\qquad A = \left(c\frac{\mathrm{d}^2t}{\mathrm{d}s^2}\,,\ \frac{\mathrm{d}^2x}{\mathrm{d}s^2}\,,\ 0\,,0\right)\ . \tag{6.16}$$

These four-vectors are required to satisfy the constraints ($g(A, A) = -\kappa^2$ together with the constraint for the four-velocity we found above)

$$c^2\left(\frac{\mathrm{d}t}{\mathrm{d}s}\right)^2 - \left(\frac{\mathrm{d}x}{\mathrm{d}s}\right)^2 = c^2\ ,\quad c^2\left(\frac{\mathrm{d}^2t}{\mathrm{d}s^2}\right)^2 - \left(\frac{\mathrm{d}^2x}{\mathrm{d}s^2}\right)^2 = -\kappa^2\ . \tag{6.17}$$

We differentiate the first equation with respect to s and substitute it into the second to get

$$c\frac{\mathrm{d}^2t}{\mathrm{d}s^2} = \kappa\sqrt{\left(\frac{\mathrm{d}t}{\mathrm{d}s}\right)^2 - 1}\ ,\quad \frac{\mathrm{d}^2x}{\mathrm{d}s^2} = \kappa\frac{\mathrm{d}t}{\mathrm{d}s}\ . \tag{6.18}$$

We can solve the left equation by inspection for the first derivative of t, from which we can simply integrate the right equation to get the first derivative of x,

$$\frac{\mathrm{d}t}{\mathrm{d}s} = \cosh(\kappa s/c)\ ,\qquad \frac{\mathrm{d}x}{\mathrm{d}s} = c\sinh(\kappa s/c)\ . \tag{6.19}$$

Integrating, we find

$$\begin{aligned} ct(s) &= \frac{c^2}{\kappa}\sinh\left(\frac{\kappa s}{c}\right)\ , \\ x(s) &= \frac{c^2}{\kappa}\cosh\left(\frac{\kappa s}{c}\right)\ , \\ y(s) &= 0\ , \\ z(s) &= 0\ , \end{aligned} \tag{6.20}$$

which is a hyperbola in the (x, ct) plane. Note that we have that $\kappa x = c\sqrt{\kappa^2t^2 + c^2}$, which means that for small t, $x \sim c^2/\kappa + \kappa t^2/2$, as in non-relativistic acceleration, but for large t, $x \sim ct$.

A more striking fact is that the growth in x and t are exponential in proper time s. For instance if κ is the acceleration $g \sim 10\,m/s^2$ familiar on the Earth's surface, then from $s = 0$ to $s = 10$ years, we find that t goes from 0 to about $\frac{1}{2}\exp(10)$ years, about 11000 years. (Details left as exercise.) Thus, basically, astronauts in a rocket capable of maintaining g-acceleration for 10 years could travel 11000 years into the future.

This calculation extends the triangle-inequality observation made earlier, by smoothing the non-uniform motion into a continuously accelerating path.

7 Relativistic dynamics and collisions

We have seen that intuitive notions such as velocity and acceleration are upgraded in a relativistic setting to certain constrained four-vectors. More relevant than velocity in the physical setting is *momentum*, so we will need to upgrade momentum to a four-vector as well. This will require revisiting the notion of 'inertial mass' that appears in Newton's second law, but first we recall the relevant Newtonian conservation laws seen at the beginning of the course.

Newtonian conservation laws

In Newtonian theory, conservation laws are of great use in analysing many complicated mechanical systems, especially the collision of particles. Of particular importance were the conservation of momentum and the conservation of mass.

(Newtonian) conservation of mass: If the (inertial) masses of a collection of k incoming particles are m_1, m_2, …, m_k, and those of the $n-k$ outgoing particles are m_{k+1}, m_{k+2}, …, m_n, then

$$\sum_{i=1}^{k} m_i = \sum_{i=k+1}^{n} m_i \ . \tag{7.1}$$

Note that the number of outgoing particles need not be the same as the number of incoming particles — we allow for the particles to break up or coalesce.

(Newtonian) conservation of three-momentum: If the velocities of the same k incoming particles are $\mathbf{v}_1$,$\mathbf{v}_2$, …, $\mathbf{v}_k$ and those of the outgoing $n-k$ particles are $\mathbf{v}_{k+1}$, $\mathbf{v}_{k+2}$, …, $\mathbf{v}_n$, then

$$\sum_{i=1}^{k} m_i\mathbf{v}_i = \sum_{i=k+1}^{n} m_i\mathbf{v}_i \ , \tag{7.2}$$

We have omitted the conservation of energy, though it is applicable and useful in many instances in Newtonian physics. This is because in the particular case of particle collisions, which will be a subject of interest to us in what follows, energy was often lost to 'heat' or 'radiation' in the Newtonian context (in particular when particles collided inelastically, and thus coalesced).

Collectively, the conservations of mass and momentum give four constraint equations that must be obeyed in any Newtonian particle collision, whether elastic or inelastic. We would like to find relativistic generalisations of these constraints that reduce to them in the non-relativistic limit.

7.1 Rest mass, four-momentum, and "inertial mass"

To begin, we will need to revisit the notion of mass in light of relativistic considerations. It was an axiom of Newtonian physics that the mass of a particle is independent of its state of motion. We will not make such an assumption here, but we can still define a quantity that we call mass, or rest mass, by adopting the Newtonian definition *in the special case when a particle is at rest.*

Definition 20. The *rest mass* of a particle is the ratio between an applied force and its acceleration in the ICS where the particle is instantaneously at rest.

This definition is frame-independent, in the sense that the definition requires always going to the rest frame of that object (this is a kind of *operational* definition, like we had in the radar method for defining length and duration). Thus the rest mass is an intrinsic quantity associated to a particle, and we may treat it as a being Lorentz-invariant. We can then define a four-vector version of momentum that is guaranteed to agree with the three-vector version at low speeds,

Definition 21. The *momentum four-vector* of a particle with rest mass m and four-velocity U is the four-vector $P = mU$.

In a given ICS, the momentum four-vector has temporal and spatial parts given by

$$P = (\,\gamma(u)mc\,,\,\gamma(u)m\mathbf{u}\,)\ . \tag{7.3}$$

so that it encodes both, information about the rest mass as well as information about the three-velocity. We can immediately observe that while any four-velocity is constrained to satisfy $g(U,U) = c^2$, the four-momentum of an object has a rest-mass-dependent constraint $g(mU,mU) = m^2c^2$. For positive mass the four-momentum is a future-pointing timelike vector, with the pseudo-norm dictated by the rest mass of the particle.

It is this four-vector version of the momentum that is conserved in a relativistic setting.

(Relativistic) conservation of four-momentum: If the four-momenta of k incoming particles are P_1, P_2, …, P_k, and those of the outgoing $n-k$ particles are P_{k+1}, P_{k+2}, …, P_n, then in the absence of any additional external forces,

$$\sum_{i=1}^{k} P_i = \sum_{j=k+1}^{n} P_j\ , \tag{7.4}$$

We can verify that this is compatible with Newtonian conservation laws in the limit of small velocities. If we expand the relativistic conservation law in powers of u/c, we have

$$\begin{aligned} c\left(\sum_{i=1}^{k} m_i - \sum_{i=k+1}^{n} m_i + O\left(\frac{u^2}{c^2}\right)\right) &= 0 \ , \\ \left(\sum_{i=1}^{k} m_i \mathbf{u}_i - \sum_{i=k+1}^{n} m_i \mathbf{u}_i\right) + O\left(\frac{u^2}{c^2}\right) &= 0 \ . \end{aligned} \tag{7.5}$$

So Newtonian mass and momentum conservation laws are encoded in four-momentum conservation law at low speeds.

At velocities that are not negligible compared to c, the situation is really quite different from the Newtonian limit. The momentum four-vector in a given ICS for the particle i is given by

$$P_i = (\,\gamma(u_i) m_i c\,,\, \gamma(u_i) m_i \mathbf{u}_i\,) \ , \tag{7.6}$$

where m_i is the rest mass of particle i. Let us consider in turns the spatial and temporal parts of our relativistic conservation law.

The spatial part of the law reads as

$$\sum_{i=1}^{k} \gamma(u_i) m_i \mathbf{u}_i = \sum_{i=k+1}^{n} \gamma(u_i) m_i \mathbf{u}_i \ . \tag{7.7}$$

This looks very similar to Newtonian momentum conservation, but the rest mass is replaced by the quantity $\gamma(u_i)m_i$. For this reason, we define the following.

Definition 22. The *relativistic inertial mass* of an object with rest mass m relative to an ICS in which it is moving with three-velocity $\mathbf{u}$ is the quantity $\gamma(u)m$.

Thus in collisions, ordinary conservation of three-momentum still applies if we use the relativistic intertial masses of all involved bodies instead of their invariant rest mass in defining the three-momenta. Note that the relativistic inertial mass of an object increases as the three-velocity increases in magnitude, approaching infinity as $u \to c$, so an object of fixed rest mass moving at approximately the speed of light will seem to carry an infinite amount of momentum.

Turning now to the time component of the four-momentum conservation, we see that this is qualitatively different from its Newtonian limit. The exact expression is

$$\sum_{i=1}^{k} \gamma(u_i) m_i c = \sum_{i=k+1}^{n} \gamma(u_i) m_i c \ . \tag{7.8}$$

We already saw that in the strict Newtonian limit this becomes conservation of mass. Let us now consider an approximation where we keep the first correction to the Newtonian limit.

Then, after multiplying by an overall factor of c, we have

$$\sum_{i=1}^{k}\left(m_i c^2+\frac{1}{2}m_i u_i^2\right)=\sum_{i=k+1}^{n}\left(m_i c^2+\frac{1}{2}m_i u_i^2\right)+O\left(\frac{u^2}{c^4}\right)\ . \tag{7.9}$$

We see here terms that look like Newtonian kinetic energy, and we interpret the conservation equation (in this approximation) as saying that energy can be transferred between rest-mass-energy and kinetic energy. Notice that this is different from the Newtonian treatment of particle collisions, in that it leaves no room for energy to be lost as 'heat' or 'radiation'. Of course, this equation will only hold if the terms on the two sides account for every physical entity engaging in the interaction, so energy could be lost to radiation if we are not keeping track of the particles of radiation, but then momentum could be lost as well. We see that relativity requires that we tie up energy and momentum in a package.

Since the temporal component of the four-momentum encodes, at low energies, the kinetic energy of a particle (along with an additional amount related to the rest mass), we make the following definitions:

Definition 23. The *total energy* of a particle relative to a given ICS is the temporal component of its four-momentum multiplied by the speed of light.

Thus we have

$$P=\left(\frac{E}{c}\ ,\ \mathbf{p}\right)\ , \tag{7.10}$$

where, as we mentioned before, the three-momentum $\mathbf{p}$ is defined using the relativistic inertial mass of the particle, $\mathbf{p}=\gamma(u)m\mathbf{u}$. Using our constraint equation $g(P,P)=m^2c^2$, we get the following expression relating three-momentum, rest mass, and total energy,

$$E^2=m^2c^4+p^2c^2\ . \tag{7.11}$$

We see that in an ICS in which a particle is at rest, so where the three-momentum is zero, the total energy of the particle is not zero, but instead given by its *rest energy.*

Definition 24. The *rest energy* of a particle of rest mass m is $E_{\text{rest}}=mc^2$.

You may have seen this equation before :)

7.2 Relativistic Newton's law of motion

With a solid understanding of four-momentum in place, we can finally write the relativistic analogue of Newton's second law:

$$F^a=\frac{\mathrm{d}P^a}{\mathrm{d}s}=mA^a\ , \tag{7.12}$$

where s is the proper time of the particle under consideration. Note that, unsurprisingly, this equation requires that the force, which used to be a three-vector, be promoted to some kind of four-vector that transforms appropriately under Lorentz transformations and has the property that $g(F, U) = 0$.

We denote the components of the four-force in an ICS in which it is being applied to a particle moving at velocity $\mathbf{u}$ as follows,

$$F = \left(F^0 \, , \, \gamma(u)\mathbf{f} \right) \ . \tag{7.13}$$

We can then see that the spatial components of Newton's equations amount to

$$\gamma(u)\mathbf{f} = \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}s} = \frac{\mathrm{d}t}{\mathrm{d}s}\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = \gamma(u)\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} \ , \tag{7.14}$$

so this is just the non-relativistic Newton's law where $\mathbf{f}$ is the force and where the momentum is the relativistic three-momentum, using the relativistic inertial mass. The temporal component of F does not contain additional information: since F is constrained as mentioned above, F^0 should be adjusted such that $g(F, U) = 0$. Thus the content of Newton's laws in a relativistic setting in a fixed ICS amount to the replacement of rest mass by relativistic inertial mass in the definition of three-momentum.

We can also, for example, derive the relation between work and change of energy in this relativistic setting. Starting with the requirement $E^2 = m^2c^4 + \mathbf{p}\cdot\mathbf{p}\,c^2$, we differentiate both sides and find

$$E\frac{\mathrm{d}E}{\mathrm{d}t} = \mathbf{p}\cdot\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t}c^2 = \mathbf{p}\cdot\mathbf{f}c^2 \ , \tag{7.15}$$

which, after using $E = \gamma mc^2$ and $\mathbf{p} = \gamma m\mathbf{u}$, gives us

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \mathbf{u}\cdot\mathbf{f} \ . \tag{7.16}$$

We see again that the same Newtonian formula emerges, but where the energy E now means the total relativistic energy.

7.3 The four-momentum of a photon

Above we only considered the four-momentum for the very narrow arena of colliding point particles. We also need to know that the identification of energy and inertial mass can be followed through consistently when for example, energy is transferred to an electromagnetic field. In the famous application to nuclear physics, which lies behind the principle of the atomic bomb, the theory of nuclear binding energy must be defined consistently. All this lies far beyond our scope. The main point is that *all physical laws must be Lorentz invariant* for such consistency to be possible. Modern relativistic quantum field theory, as embodied in the Standard Model of forces and particles, achieves this.

One observation that builds confidence in the compatibility of the above picture of momentum conservation with more general electromagnetic phenomena is that it can be extended to include interactions with weak electromagnetic fields in the form of photons, or *light quanta*. The question of how light, which we previously encountered as arising from electromagnetic waves, can be thought of as a particle is beyond the scope of this course. For our purposes here, let us simply take as given that particles called photons exist which are the particulate manifestation of light and thus always travel at speed c. These particles have a characteristic frequency ω that is related to the frequency of the light-wave they embody.

A particle that moves at speed c poses some problems if we try to apply our previous discussion directly. For one, we know that the worldline of such a particle will be moving at the speed of light in every ICS, and so there is no way to define a rest mass for such a particle. Relatedly, the four-velocity of a particle was defined to be normalized so that $g(U,U) = c^2$, but a particle moving at the speed of light will be moving along a null trajectory, so no matter how we normalize the four-vector pointing along its world-line we will have $g(U,U) = 0$.

The connection between velocity and momentum must come in a different way, and here quantum-mechanical physics supplies the link. Planck's constant $\hbar \approx 1.05 \times 10^{-34} J \cdot s$, with the dimensions of [Energy $\times$ Time], relates (angular) frequency to energy. We borrow from quantum mechanics the fact that the energy and momentum carried by a photon are given by

$$E = \hbar\omega \ , \qquad \mathbf{p} = \frac{\hbar\omega}{c}\hat{\mathbf{k}} \ , \tag{7.17}$$

where $\mathbf{k}$ is the unit three vector ($\hat{\mathbf{k}} \cdot \hat{\mathbf{k}} = 1$) in the direction of propagation. We then ensemble these components into the momentum four-vector for a photon

$$P = \left(\frac{\hbar\omega}{c} , \frac{\hbar\omega}{c}\hat{\mathbf{k}} \right) \ . \tag{7.18}$$

So that in particular $g(P,P) = 0$. From (7.17) we see that

$$E^2 = p^2c^2 \ , \tag{7.19}$$

which we can reconcile with (7.11) if we set $m = 0$ in that equation. For this reason, it is sometimes said that a photon is a particle with *zero rest mass*. The expression 'zero rest mass' is, of course, a contradiction in terms: a particle with zero mass is moving at the speed of light and so cannot be at rest. However, it has become established usage. It may be more accurate to call a photon a *massless particle*.

7.4 Collisions

One of the most significant applications of relativistic energy/momentum conservation is in the study of collisions and interactions of elementary particles, also known as *particle scattering*. One immediate consequence of our new picture of energy and momentum is that processes may occur in which the incoming and outgoing particles do not have the same combined rest masses, and this gives rise to a much greater diversity of particle interactions than would otherwise not be allowed.

Example 9 (Particle decay)**.** A simple example of a relativistic particle interaction is the decay of an unstable particle into two stable particles. Consider a particle A of rest mass M that decays (splits) into two new particles of type B, each of rest mass m.

$$A \longrightarrow B + B \ . \tag{7.20}$$

In the inertial frame in which A is at rest, suppose the B particles move with three-velocities $u_1\hat{\mathbf{n}}_1$ and $u_2\hat{\mathbf{n}}_2$, with $\hat{\mathbf{n}}_1, \hat{\mathbf{n}}_2$ unit vectors giving the direction of motion and $u_1, u_2 > 0$. We will show that the masses of the particles and their velocities are related according to

$$\begin{aligned} \hat{\mathbf{n}}_1 &= -\hat{\mathbf{n}}_2 \ , \\ u_1 &= u_2 \equiv u \ , \\ M &= 2m\gamma(u) \ . \end{aligned} \tag{7.21}$$

Solution. By conservation of four-momentum

$$M(c,0,0,0) = m\gamma(u_1)(c, u_1\hat{\mathbf{n}}_1) + m\gamma(u_2)(c, u_2\hat{\mathbf{n}}_2) \ . \tag{7.22}$$

From the spatial components, we have

$$\gamma(u_1)u_1\hat{\mathbf{n}}_1 = -\gamma(u_2)u_2\hat{\mathbf{n}}_2 \ , \tag{7.23}$$

Dotting by $\hat{\mathbf{n}}_1$ and $\hat{\mathbf{n}}_2$ we obtain

$$\gamma(u_1)u_1 = -\gamma(u_2)u_2\hat{\mathbf{n}}_1 \cdot \hat{\mathbf{n}}_2 \ , \qquad \gamma(u_2)u_2 = -\gamma(u_1)u_1\hat{\mathbf{n}}_1 \cdot \hat{\mathbf{n}}_2 \ , \tag{7.24}$$

from which we deduce $(\hat{\mathbf{n}}_1 \cdot \hat{\mathbf{n}}_2)^2 = 1 \to \hat{\mathbf{n}}_1 \cdot \hat{\mathbf{n}}_2 = \pm 1$. Since u_1, u_2 are positive we must have $\hat{\mathbf{n}}_1 \cdot \hat{\mathbf{n}}_2 = -1$. This implies $u_1 = u_2 \equiv u$ and $\hat{\mathbf{n}}_1 = -\hat{\mathbf{n}}_2$.

The temporal component of the conservation equation then becomes $M = 2m\gamma(u)$. For fixed M and m we can solve for the speed of the outgoing particles,

$$\begin{aligned} \frac{1}{\sqrt{1-\frac{u^2}{c^2}}} &= \frac{M}{2m} \ , \\ 1 - \frac{u^2}{c^2} &= \frac{4m^2}{M^2} \ , \\ u &= c\sqrt{1 - \frac{4m^2}{M^2}} \ . \end{aligned} \tag{7.25}$$

We see that for the speed to be well defined, we must have $2m \leqslant M$, with the resulting particles sitting at rest in the case when the inequality is saturated. This is the Newtonian case.

In the opposite extreme, we see that the resulting particles will move closer and closer to the speed of the light as $m \to 0$. In the case when $m = 0$, we recover the result for decay into photons.

Exercise. Determine the frequency of the photons that would be emitted if the particle A decayed into two photons.

Example 10 (Compton scattering)**.** A photon (normally denoted by γ in particle physics) collides with an electron (normally denoted e^-) of rest-mass m_e,

$$\gamma + e^- \longrightarrow \gamma + e^- \ . \tag{7.26}$$

In the ICS in which the electron is initially at rest, the photon has frequency ω. After the collision, the outgoing photon has frequency ω'. Show that

$$\hbar\omega\omega'(1-\cos\theta) = m_e c^2(\omega-\omega') \ , \tag{7.27}$$

where θ is the angle between the initial and final directions of the photon.

Solution. In the ICS where the electron is initially at rest, the four-momenta of the electron before and after collision are

$$P_e = m_e(c,0,0,0) \ , \qquad P_e' = m_e\gamma(u)(c,\mathbf{u}) \ . \tag{7.28}$$

The four-momenta of the photon before and after collision are

$$P_\gamma = \frac{\hbar\omega}{c}(1,\mathbf{e}) \ , \qquad P_\gamma' = \frac{\hbar\omega'}{c}(1,\mathbf{e}') \ . \tag{7.29}$$

Our four-momentum conservation condition reads

$$P_e + P_\gamma = P_e' + P_\gamma' \ . \tag{7.30}$$

We can eliminate $\mathbf{u}$, which does not appear in the required relation, by considering

$$\begin{aligned} m_e^2c^2 &= g(P_e',P_e') \\ &= g(P_e+P_\gamma-P_\gamma', P_e+P_\gamma-P_\gamma') \\ &= g(P_e,P_e) + 2g(P_e,P_\gamma-P_\gamma') - 2g(P_\gamma,P_\gamma') \ , \end{aligned} \tag{7.31}$$

where in going from the second to the third line we have used $g(P_\gamma,P_\gamma) = g(P_\gamma',P_\gamma') = 0$. We also know $g(P_e,P_e) = m_e^2c^2$, which leaves

$$g(P_\gamma,P_\gamma') = g(P_e,P_\gamma-P_\gamma') \ . \tag{7.32}$$

Inserting the components this yields

$$\frac{\hbar^2\omega\omega'}{c^2}(1-\mathbf{e}\cdot\mathbf{e}') = m_e\hbar(\omega-\omega') \ . \tag{7.33}$$

This is equivalent to the desired expression.

Example 11 (Elastic collision of identical particles)**.** We consider the simplest possible collision, when two particles with the same mass collide and emerge as two particles of the same mass (one might consider them as two electrons, with rest mass m_e)

$$e^- + e^- \longrightarrow e^- + e^- \ . \tag{7.34}$$

In the ICS in which one of the initial electrons is at rest, we denote the three-velocity of the other initial electron in that frame by $\mathbf{u}$. After the collision, the electrons have velocities $\mathbf{v}$ and $\mathbf{w}$. We will show that if θ is the angle between $\mathbf{v}$ and $\mathbf{w}$, then

$$\cos\theta = \frac{c^2}{vw}(1-\sqrt{1-v^2/c^2})(1-\sqrt{1-w^2/c^2}) \ . \tag{7.35}$$

Solution. Let the four-momenta of the ingoing particles be

$$P_1 = m(c,0,0,0) \ , \qquad P_2 = m\gamma(u)(c,\mathbf{u}) \ , \tag{7.36}$$

and the four-momenta of the outgoing particles be

$$Q_1 = m\gamma(v)(c,\mathbf{v}) \ , \qquad Q_2 = m\gamma(w)(c,\mathbf{w}) \ . \tag{7.37}$$

The conservation equation $P_1 + P_2 = Q_1 + Q_2$ can be written as

$$\gamma(u) + 1 = \gamma(v) + \gamma(w) \ , \quad \gamma(u)\mathbf{u} = \gamma(v)\mathbf{v} + \gamma(w)\mathbf{w} \ . \tag{7.38}$$

One can proceed by squaring these equations and simplifying the resulting algebraic equations. This is done in Woodhouse (page 130). As an alternative, we can use four-vector calculus as we have in the previous example to eliminate $\mathbf{u}$ directly. We have

$$\begin{aligned} g(P_2,P_2) &= g(Q_1+Q_2-P_1, Q_1+Q_2-P_1) \ , \qquad (7.39)\\ &= g(Q_1,Q_1) + g(Q_2,Q_2) + g(P_1,P_1) + 2g(Q_1,Q_2) - 2g(Q_1,P_1) - 2g(Q_2,P_1) \ . \end{aligned}$$

Now $g(P_i,P_i) = g(Q_i,Q_i) = m^2c^2$, so the last equation is equivalent to

$$m^2c^2 = g(Q_1,P_1) + g(Q_2,P_1) - g(Q_1,Q_2) \ . \tag{7.40}$$

We also have

$$\begin{aligned} g(Q_1,P_1) &= m^2c^2\gamma(v) \ ,\\ g(Q_2,P_1) &= m^2c^2\gamma(w) \ ,\\ g(Q_1,Q_2) &= m^2\gamma(v)\gamma(w)\left(c^2 - \mathbf{v}\cdot\mathbf{w}\right) \ , \end{aligned} \tag{7.41}$$

So substituting back in we have

$$1 = \gamma(v) + \gamma(w) - \gamma(v)\gamma(w)\left(1 - \frac{\mathbf{v}\cdot\mathbf{w}}{c^2}\right),$$

or

$$\frac{\mathbf{v} \cdot \mathbf{w}}{c^2} = \frac{(\gamma(v) - 1)(\gamma(w) - 1)}{\gamma(v)\gamma(w)} \tag{7.42}$$

from which the required result follows.

Exercise. Show that the result can be written as $\cos\theta = \tanh(\alpha/2)\tanh(\beta/2)$, where α, β are the rapidities associated with the velocities v, w.

Exercise. Suppose Newtonian theory applied to such elementary particle collisions. We would have kinetic energy conservation as well as 3-momentum conservation, and so $\mathbf{u} = \mathbf{v} + \mathbf{w}$ and $u^2 = v^2 + w^2$, hence $\mathbf{v} \cdot \mathbf{w} = 0$ and $\theta = \pi/2$ for all allowed values of $\mathbf{v}$ and $\mathbf{w}$. (Chech that the same conclusion follows from taking the limit $c \to \infty$ in (7.42).)

Particle paths in high energy collision experiments, in which θ is observed to obey $\theta < \pi/2$, therefore offer a direct manifestation of relativistic effects in particle scattering.

7.5 Particle creation

Perhaps the most remarkable of all the possibilities suggested by the mass-energy equivalence is the creation of new particles if an adequate amount of energy is made available. A sizeable fraction of all activity in physics today is founded upon this phenomenon, specially for the purpose of creating and studying particles that are too short-lived to be found in nature. To create a particle of rest mass m_0 we need an energy input of at least $m_0 c^2$. In practise, much more is needed. There are two main reasons for this.

- There are fundamental conservation laws, such as electric charge conservation, which prevent us from creating only one new particle in a collision process. An example of this is the creation of an electron-positron pair from the energy of a $\gamma-$ray photon

  $$\gamma \to e^- + e^+$$

  Since the photon is electrically neutral we cannot only create an electron.

- The second reason is practical. Consider for instance the creation of positively charge π-mesons, also known as pions π^+. These can be created by colliding protons P_1, P_2 and the relevant process is

  $$P_1 + P_2 \to P + N + \pi^+$$

  The colliding protons give rise to a proton, a neutron and a pion, as indicated. Since the neutron and proton have almost equal rest masses, the only new rest energy needed is that represented by the pion (about 140 MeV - where the rest energy of an electron, for comparison, is about 0.5 MeV). But! if the target proton P_2 is initially at rest and P_1 has large momentum it is clear a good deal of kinetic energy is locked up in the motion of the system as a whole, and hence unavailable for conversion into the rest mass of new particles. If particles P_1 and P_2 could be made to collide with equal and opposite momenta, the amount of energy associated with the general motion of the system would be zero. However, to produce beams of particles traveling in opposite directions is technically much harder than to have one beam striking a stationary target (*e.g.* very few duels end in a draw with the two bullets hitting each other). Let us consider this creation process in more detail.

Pion production

Let us consider the process $P_1 + P_2 \to P + N + \pi^+$ first in a frame of reference where the total three-momentum is zero. The four-momenta of the incoming protons is then

$$P_1 = m_p \gamma(u)(c, \mathbf{u}), \quad P_2 = m_p \gamma(u)(c, -\mathbf{u})$$

where m_p is the rest mass of the proton. The most economical condition for particle creation is one where all three produced particles are at rest, so that the four-momenta of the outgoing particles is

$$P_P = m_p(c, 0), \quad P_N = m_p(c, 0), \quad P_\pi = m_\pi(c, 0)$$

where we have assumed proton and neutron have the same rest-mass (a very good approximation) and m_π is the rest mass of the pion. Conservation of four-momenta then gives

$$2m_p \gamma(u) = 2m_p + m_\pi$$

What does this mean in terms of energy? the rest energy of a proton is $m_p c^2 = 938Mev$. Since for the Pion $m_\pi c^2 = 140Mev$ then we need $\gamma(u) \sim 1.075$. This means that the speed of every colliding proton in this (zero-momentum) frame is $u \sim 0.37c$ and the 'kinetic' (total minus rest) energy of each incoming proton is then $\gamma(u) m_p c^2 \sim 70Mev$, of course.

Let us look at the same process from a reference frame where the proton P_2 is at rest. This frame will move at speed u with respect to the zero momentum frame (so that P_2 is at rest). But that means, that the velocity of P_1 can be calculated from the addition of velocities formula and is given by

$$u_1 = \frac{u + u}{1 + \frac{u^2}{c^2}} = \frac{2u}{1 + \frac{u^2}{c^2}}$$

So that for the present problem $u_1 = 0.65c$. This in turn implies

$$\gamma(u_1) \sim 1.32$$

This means that the bombarding proton must have an kinetic energy $(\gamma(u_1) - 1) m_p c^2 \sim 300Mev$, a much higher value than the previously obtained. Of these $160Mev$ will be 'lost' in the motion of the system as a whole.

8 Relativistic electrodynamics

8.1 Lorentz transformations of E and B

The requirement that Maxwell's equations should be consistent with the principle of relativity implies that the speed of photons must be independent of the motion of their source and the observer. This led to the conclusion that inertial coordinate systems must be related by the Lorentz transformations. In this section we would like to answer the inverse question: given the Lorentz transformations, are the Maxwell's equations invariant? To answer this question we must find the transformation rules for the components of the electric and magnetic field, and then show that if they solve the Maxwell's equations in an inertial frame, then they do in all inertial frames.

Let us consider a particle with charge e and rest mass m. We are assuming that the charge and rest mass are inherent properties of the particles, they do not depend on the choice of inertial frame and do not change by interaction of the particle with the electromagnetic fields. Suppose this particle is moving, in an inertial frame, with velocity $\mathbf{v}$ and momentum $\mathbf{p}$. We will assume that the electric and magnetic fields are such that as the particle moves it feels the relativistic version of the Lorentz force law

$$\frac{d\mathbf{p}}{dt} = e\left(\mathbf{E} + \mathbf{v}\times\mathbf{B}\right) \quad \text{where} \quad \mathbf{p} = m\gamma(v)\mathbf{v}. \tag{8.1}$$

From this we can derive the following.

Proposition 11. The four acceleration $A = (\alpha, \mathbf{a})$ of a particle of rest mass m and charge e moving with velocity $\mathbf{v}$ in a electromagnetic field is given by

$$mc\alpha = e\gamma(v)\mathbf{E}\cdot\mathbf{v}, \quad m\mathbf{a} = e\gamma(v)\left(\mathbf{E} + \mathbf{v}\times\mathbf{B}\right)$$

Proof. Since m is constant and $\mathbf{p} = m\gamma(v)\mathbf{v}$, the second equation follows from (8.1) together with

$$\mathbf{a} = \frac{d}{ds}\left(\gamma(v)\mathbf{v}\right) = \frac{dt}{ds}\frac{d}{dt}\left(\gamma(v)\mathbf{v}\right) = \gamma(v)\frac{d}{dt}\left(\gamma(v)\mathbf{v}\right).$$

Then the first equation follows from the (pseudo-)orthogonality of the four acceleration $A = (\alpha, \mathbf{a})$ and the four-velocity $V = \gamma(v)(c, \mathbf{v})$, which implies that $c\alpha = \mathbf{a}\cdot\mathbf{v}$. □

Let F denote the 4×4 matrix

$$F = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & -c\,B_3 & c\,B_2 \\ -E_2 & c\,B_3 & 0 & -c\,B_1 \\ -E_3 & -c\,B_2 & c\,B_1 & 0 \end{pmatrix}. \tag{8.2}$$

This matrix is called the electromagnetic field. The entries in F are denoted by F_{ab}, with $a, b = 0, 1, 2, 3$; they are the components of the electromagnetic field. An observer measures the electromagnetic field at an event by measuring the four-acceleration of charged particles, and using the previous proposition. We then have the following

Proposition 12. Suppose that the inertial coordinate systems of two inertial observers are related by (5.10). Then the electromagnetic fields measured by the two observers at an event are related by $F' = L^T F L$. We say that F transforms as a (covariant) tensor.

The general proof can be found in Woodhouse (page 135).

Let us look at a particular example: the standard Lorentz transformation with velocity u in the $x-$direction. In this case the transformation reduces to

$$E_1' = E_1, \quad E_2' = \gamma(u)(E_2 - uB_3), \quad E_3' = \gamma(u)(E_3 + uB_2) \tag{8.3}$$

$$B_1' = B_1, \quad B_2' = \gamma(u)\left(B_2 + \frac{u}{c^2}E_3\right), \quad B_3' = \gamma(u)\left(B_3 - \frac{u}{c^2}E_2\right) \tag{8.4}$$

We will return to this.

8.2 The four current and the four potential

We now turn to the invariance of the Maxwell's equations. We can write them in terms of the potentials as

$$\nabla^2\phi - \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} = -\frac{1}{\epsilon_0}\rho \tag{8.5}$$

$$\nabla^2\mathbf{A} - \frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2} = -\mu_0\mathbf{J} \tag{8.6}$$

with the condition (known as Lorenz gauge)

$$\nabla\cdot\mathbf{A} + \frac{1}{c^2}\frac{\partial\phi}{\partial t} = 0 \tag{8.7}$$

It is customary to introduce the *d'Alembertian operator*, defined by

$$\Box = \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2 = \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial z^2}$$

Which is the $3+1$ analogous of the Laplace operator. We can then write the equations

$$\Box\phi = \frac{\rho}{\epsilon_0}, \quad \Box c\mathbf{A} = \frac{\mathbf{J}}{c\,\epsilon_0} \tag{8.8}$$

Transforming the source terms

Let us first look at the transformation properties, under Lorentz transformations, of the r.h.s. of equations (8.8). We have the following very useful result for a continuous distribution of particles. Suppose that in an inertial coordinate system there are σ particles per unit volume, moving with velocity $\mathbf{v}$, where σ and $\mathbf{v}$ are differentiable functions of position and time. Then the following is true:

Proposition 13. Under change of inertial coordinates $c\sigma$ and $\sigma\mathbf{v}$ transform as the temporal and spatial part of a four-vector.

Proof. Since we know that $\gamma(v)(c, \mathbf{v})$ transforms as a four-vector (the four-velocity of a particle), the task is to show that $\sigma/\gamma(v)$ is the same in all inertial frames. Let us assume $\mathbf{v}$ is constant and the same for all particles in a given small volume. We can always achieve this by choosing a small enough volume. Hence all particles are moving in a straight line, which we take to be in the $x-$direction for simplicity, with the same velocity.

Let $(\tilde{t}, \tilde{x}, \tilde{y}, \tilde{z})$ by a ICS in which the particles are at rest, and suppose there is $\tilde{\sigma}$ particles per unit volume in these coordinates. Consider the particles which at $\tilde{t} = 0$ occupy a small cube of volume L^3 with vertices at

$$A : (0,0,0), \quad B : (L,0,0), \quad C : (0,L,0), \quad D : (0,0,L), \quad \cdots$$

in the coordinate system $(\tilde{t}, \tilde{x}, \tilde{y}, \tilde{z})$. The number of particles in such cube is then $\tilde{\sigma}L^3$. Let us now consider a second inertial frame, with coordinates (t, x, y, z), related to $(\tilde{t}, \tilde{x}, \tilde{y}, \tilde{z})$ by a standard Lorentz transformation with speed v. In these new coordinates A, B, C, D are at

$$\begin{aligned}
A &: \ x = vt, \quad y = z = 0, \\
B &: \ x = vt + L\sqrt{1 - v^2/c^2}, \quad y = z = 0, \\
C &: \ x = vt, \quad y = L, \quad z = 0, \\
D &: \ x = vt, \quad y = 0, \quad z = L.
\end{aligned}$$

Therefore, in the second inertial frame we have a cuboid of volume $L^3/\gamma(v)$ moving with speed v in the positive $x-$direction. Since the number of particles is the same, there is $\sigma = \gamma(v)\tilde{\sigma}$ particles per unit volume in the second inertial coordinate system. We then see that $\sigma/\gamma(v)$ is independent of v. □

Going back to the source terms, suppose now we have a continuous distribution of charge particles, with charge e, with σ particles per unit volume, moving with velocity $\mathbf{v}$. Then

$$c\rho = e\sigma c, \quad \mathbf{J} = e\sigma\mathbf{v}$$

By the proposition we have just proven, and assuming e is invariant, $(c\rho, \mathbf{J})$ transforms as a four-vector. The same should be true for any distribution of charges, since the four-vector transformation rule is linear and we can sum contributions to ρ and $\mathbf{J}$ from different groups of particles. We then introduce the following

Definition 25. $J = (c\rho, \mathbf{J})$ is denoted the *current four-vector.*

The continuity equation (2.11) can be written in an invariant form

$$\mathrm{Div} J = 0,$$

where the four-divergence of a four-vector field X is the function

$$\mathrm{Div} X = \frac{1}{c}\frac{\partial X^0}{\partial t} + \frac{\partial X^1}{\partial x} + \frac{\partial X^2}{\partial y} + \frac{\partial X^3}{\partial z}$$

And, as discussed, it can be shown the four-divergence is invariant under Lorentz transformations.

Transforming the potentials

To deal with the potentials is slightly more complicated, since those are not uniquely defined (due to gauge freedom), but the basic result is that indeed ϕ and $c\mathbf{A}$ fit together into a four-vector. The precise statement is the following

Proposition 14. Let ϕ and $\mathbf{A}$ be the scalar and vector potentials in the Lorenz gauge for an electromagnetic field, in some inertial coordinate system. Suppose that ϕ and $c\,\mathbf{A}$ are transformed as the temporal and spatial parts of a four vector under some Lorentz transformation. Then the results are again scalar and vector potentials for the transformed electromagnetic field, again in the Lorenz gauge.

Proof. Consider $\Phi = (\phi, c\,\mathbf{A})$ and let us assume it transforms as a four-vector. First note that the Lorenz gauge condition is preserved as it can be written in the invariant form

$$\mathrm{Div}\,\Phi = 0.$$

Next, given ϕ and $\mathbf{A}$, define a matrix valued function M by

$$M = \frac{1}{c}\begin{pmatrix} \partial_t \\ \partial_x \\ \partial_y \\ \partial_z \end{pmatrix} \Big(\phi,\ -c\,A_1,\ -c\,A_2,\ -c\,A_3\Big)\,.$$

One can explicitly check that $F = M - M^T$. For instance, consider the second entry in the first row. We obtain

$$M_{12} - M_{21} = -\frac{\partial A_1}{\partial t} - \frac{\partial \phi}{\partial x} = E_1 = F_{12}$$

Check by yourself a few more entries.

Now under a Poincare transformation (5.10) we have

$$\big(\partial_{t'}, c\partial_{x'}, c\partial_{y'}, c\partial_{z'}\big) = \left(\partial_t, c\partial_x, c\partial_y, c\partial_z\right) L$$

Also, since by assumption $\Phi = (\phi, c\,\mathbf{A})$ is a four-vector, ϕ' and $\mathbf{A}'$ in the coordinate system (t', x', y', z') are given by

$$\begin{pmatrix} \phi' \\ cA_1' \\ cA_2' \\ cA_3' \end{pmatrix} = L^{-1}\begin{pmatrix} \phi \\ cA_1 \\ cA_2 \\ cA_3 \end{pmatrix} \tag{8.9}$$

Equivalently, since $L^{-1} = gL^T g$

$$\left(\phi',\, -c\, A_1',\, -c\, A_2',\, -c\, A_3'\right) = \left(\phi,\, -c\, A_1,\, -c\, A_2,\, -c\, A_3\right) L$$

Now consider M' given by

$$M' = \frac{1}{c} \begin{pmatrix} \partial_{t'} \\ \partial_{x'} \\ \partial_{y'} \\ \partial_{z'} \end{pmatrix} \left(\phi',\, -c\, A_1',\, -c\, A_2',\, -c\, A_3'\right).$$

Given the transformation properties of both components we get $M' = L^T M L$, and hence $M' - M'^T = F'$. It then follows that ϕ' and $\mathbf{A}'$ are potentials for the electric and magnetic fields in the primed coordinates. $\square$

Definition 26. The four-vector $\Phi = (\phi, c\, \mathbf{A})$ is called the *four potential*.

Putting all together

In terms of the four-potential in the Lorenz gauge and the current four-vector , Maxwell's equations are written as

$$\Box \Phi = \frac{1}{c\, \epsilon_0} J, \qquad \mathrm{Div}\, \Phi = 0 \tag{8.10}$$

As discussed in section 5, it can be shown that the operator $\Box$ is also invariant. It then follows that, if Maxwell's equations hold for the electromagnetic field F, then they also hold for the transformed field F'.

8.3 Invariants

Recall the transformation properties for the electric and magnetic field fields in two inertial frames O, O' related by the standard Lorentz transformation

$$E_1' = E_1, \quad E_2' = \gamma(u)(E_2 - uB_3), \quad E_3' = \gamma(u)(E_3 + uB_2) \tag{8.11}$$

$$B_1' = B_1, \quad B_2' = \gamma(u)\left(B_2 + \frac{u}{c^2}E_3\right), \quad B_3' = \gamma(u)\left(B_3 - \frac{u}{c^2}E_2\right) \tag{8.12}$$

It follows from this transformations that

$$\mathbf{E}' \cdot \mathbf{B}' = E_1 B_1 + \gamma^2 (E_2 - uB_3)\left(B_2 + \frac{u}{c^2}E_3\right) + \gamma^2 (E_3 + uB_2)\left(B_3 - \frac{u}{c^2}E_2\right) = \mathbf{E} \cdot \mathbf{B}$$

so that $\mathbf{E}\cdot\mathbf{B}$ is actually an invariant. This can be proven also for more general transformations. Another invariant is the combination $\mathbf{E} \cdot \mathbf{E} - c^2 \mathbf{B} \cdot \mathbf{B}$.

Example: Field of a uniformly moving charge

Consider a coordinate system (t', x', y', z') where a charge e is stationary at the origin. Then $\mathbf{B}' = 0$ and

$$E'_1 = k\frac{x'}{r'^3}, \quad E'_2 = k\frac{y'}{r'^3}, \quad E'_3 = k\frac{z'}{r'^3}$$

where $k = \frac{e}{4\pi\epsilon_0}$ and $r'^2 = x'^2 + y'^2 + z'^2$. What is the field in a coordinate system (t, x, y, z) where the charge is moving with velocity $(u, 0, 0)$? We can obtain this from (8.12) replacing $u \to -u$ (which gives the inverse transformation). We then obtain

$$E_1 = \frac{k\gamma(x - ut)}{r'^3}, \quad E_2 = \frac{k\gamma y}{r'^3}, \quad E_3 = \frac{k\gamma z}{r'^3}$$
$$B_1 = 0, \quad B_2 = -\frac{k\gamma uz}{c^2 r'^3}, \quad B_3 = \frac{k\gamma uy}{c^2 r'^3}$$

which can be writing in vector form as

$$\mathbf{E} = \frac{\gamma k \mathbf{R}}{r'^3}, \quad \mathbf{B} = \frac{1}{c^2}\mathbf{u} \times \mathbf{E}$$

where $\mathbf{R} = (x, y, z) - (ut, 0, 0)$, namely the position with respect to the charge at time t. It remains to express r' in terms of t, x, y, z coordinates. A short computation shows

$$r'^2 = \gamma^2(1 - \frac{u^2}{c^2}\sin^2\theta)\mathbf{R} \cdot \mathbf{R}$$

where θ is the angle between $\mathbf{R}$ and $\mathbf{u}$.

Exercise 6. Show that $\mathbf{E} \cdot \mathbf{B}$ and $\mathbf{E} \cdot \mathbf{E} - c^2\mathbf{B} \cdot \mathbf{B}$ are indeed invariant for this example.

8.4 Monochromatic Plane waves

Consider the equations for the four-potential in the Lorenz gauge in the absence of sources

$$\Box\,\Phi = 0, \quad \operatorname{Div}\Phi = 0. \tag{8.13}$$

We propose solutions of the form

$$\Phi = P\cos\left(\frac{\omega}{c}\left(c\,t - \hat{\mathbf{k}} \cdot \mathbf{x}\right)\right),$$

where P is a constant four-vector, w is the frequency, and $\hat{\mathbf{k}} \cdot \hat{\mathbf{k}} = 1$. This corresponds to a monochromatic plane wave propagating at speed c in the direction $\hat{\mathbf{k}}$. Without loss of generality we take the frequency w to be positive.

It is convenient to introduce the *frequency four-vector* given by

$$K = \omega(1, \hat{\mathbf{k}})\ . \tag{8.14}$$

Note in particular that $K = c\,\mathrm{Grad}\frac{\omega}{c}\left(c\,t - \hat{\mathbf{k}}\cdot\mathbf{x}\right)$. The source-free equations (8.13) then imply

$$g(K,K) = 0, \qquad g(K,P) = 0$$

so that K must be a future-pointing null vector, and P must be a space-like four-vector orthogonal to K. Note furthermore that the solution can be written as

$$\Phi(X) = P\cos\left(g(K,X)/c\right) \ . \tag{8.15}$$

In the previous section we assigned a four-vector P_{photon} to a photon particle. Note that this four-vector is nothing but the frequency four-vector, rescaled in such as way as to obtain to correct quantum mechanical energy

$$P_{photon} = \frac{\hbar}{c}K$$

This relation connects the two views of light, as a plane wave, and as a particle, the photon. With the help of the frequency four-vector we can express the frequency ω in an invariant form as

$$\omega = g(U,K)/c$$

where U is the four-velocity of the observer. This gives us an elegant four-vector-based method for solving the relativistic Doppler shift problem (already solved the poor man's way)

Example 12 (Doppler shift)**.** In some ICS, let the frequency of a photon traveling in the x-direction be ω, so its frequency four-vector is $\omega(1,1,0,0)$. What frequency is observed for the photon by an observer with velocity four-vector U relative to the ICS?

Solution. The observer will see a frequency of $g(U,K)/c$. In particular, for an observer moving in the same direction as the photon, with velocity four-vector $U = (\gamma(u)c, \gamma(u)u, 0, 0)$, the observed frequency is

$$\omega_{obs} = \omega\gamma(u)\left(1 - \frac{u}{c}\right) = \omega\sqrt{\frac{c-u}{c+u}} \ , \tag{8.16}$$

while for an observer moving in the opposite direction from the photon, with velocity four-vector $U = (\gamma(u)c, -\gamma(u)u, 0, 0)$, the observed frequency will be

$$\omega_{obs} = \omega\gamma(u)\left(1 + \frac{u}{c}\right) = \omega\sqrt{\frac{c+u}{c-u}} \ . \tag{8.17}$$

Which agrees with our previous result of course. On the other hand, for an observer moving at an angle θ in the $y-z$ plane, with velocity four vector $U = (\gamma(u)c, 0, \gamma(u)u\cos\theta, \gamma(u)u\sin\theta)$, the observed frequency is simply given by

$$\omega_{bs} = \omega\gamma(u) \ . \tag{8.18}$$

8.5 Electromagnetic Energy

Consider a gang of charged particles, each of rest mass m and charge e. Each particle generates an electromagnetic field that influences all other particles, each of which wins or losses energy due to the Lorentz force. There is hence a constant exchange of energy between the particles and the fields. In the following we will calculate the energy density of the electromagnetic field.

Suppose that in the inertial coordinate system t, x, y, z there are $\sigma(t, x, y, z)$ particles per unit volume moving at velocity $\mathbf{u}(t, x, y, z)$. Then

$$\rho = e\sigma, \quad \mathbf{J} = e\sigma\mathbf{u}$$

Following Proposition 11, the motion of a particle satisfies

$$\frac{d}{dt}\left(m\gamma(u)c^2\right) = e\,\mathbf{E}\cdot\mathbf{u}$$

Recall that $m\gamma(u)c^2$ is the total energy of a particle. Hence, it follows that between t and $t + \delta t$ the energy of a particle changes $e\,\mathbf{E}\cdot\mathbf{u}\delta t$. If T is the total energy of the particles and V is a fixed volume containing all the particles then

$$\frac{dT}{dt} = \int_V e\sigma\mathbf{E}\cdot\mathbf{u}dV = \int_V \mathbf{E}\cdot\mathbf{J}dV = \frac{1}{\mu_0}\int_V \mathbf{E}\cdot\left(\nabla\times\mathbf{B} - \frac{1}{c^2}\frac{\partial\mathbf{E}}{\partial t}\right),$$

where in the last equality we used Maxwell's equations. Now

$$\mathbf{E}\cdot(\nabla\times\mathbf{B}) \overset{\text{vector identity}}{=} \nabla\cdot(\mathbf{B}\times\mathbf{E}) + \mathbf{B}\cdot(\nabla\times\mathbf{E}) \overset{\text{Maxwell's eqn.}}{=} \nabla\cdot(\mathbf{B}\times\mathbf{E}) - \mathbf{B}\cdot\frac{\partial\mathbf{B}}{\partial t}$$

Hence

$$\begin{aligned}\frac{dT}{dt} &= \frac{1}{\mu_0}\int_V\left(\nabla\cdot(\mathbf{B}\times\mathbf{E}) - \frac{1}{2}\frac{\partial}{\partial t}\left(\frac{1}{c^2}\mathbf{E}\cdot\mathbf{E} + \mathbf{B}\cdot\mathbf{B}\right)\right)dV\\ &= -\int_{\partial V}\frac{1}{\mu_0}(\mathbf{E}\times\mathbf{B})\cdot dS - \frac{d}{dt}\int_V\left(\frac{\epsilon_0}{2}\mathbf{E}\cdot\mathbf{E} + \frac{1}{2\mu_0}\mathbf{B}\cdot\mathbf{B}\right)dV\end{aligned}$$

We therefore make the following identifications

- Energy density
$$\frac{\epsilon_0}{2}\mathbf{E}\cdot\mathbf{E} + \frac{1}{2\mu_0}\mathbf{B}\cdot\mathbf{B}.$$
- Energy flux (Poynting vector)
$$\frac{\mathbf{E}\times\mathbf{B}}{\mu_0}.$$

We can then interpret the formula for dT/dt as the statement of energy conservation: the increase/decrease in the energy of the particles is matched by the flow of energy across the surface of V.

Exercise 7. Check that for the monochromatic plane wave the Poynting vector points in the direction of propagation.