PREPARED WITH THE JHEP LaTeX STYLE Trinity Term 2018

Special Relativity

Oxford University Mathematics, Part A Short Option

Christopher Beem

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

christopher.beem@maths.ox.ac.uk

Course Information and Disclaimers

These notes accompany an eight hour lecture course given by the author at the University of Oxford during Trinity Term of the 2017-2018 academic year.

The official course overview reads as follows:

> The unification of space and time into a four-dimensional space-time is essential to the modern understanding of physics. This course will build on first-year algebra, geometry, and applied mathematics to show how this unification is achieved. The results will be illustrated throughout by reference to the observed physical properties of light and elementary particles.

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-year work.

The emphasis of this course is on building up a picture of the geometry of Minkowski space and of the dynamics that play out upon it. There will be no time devoted to the historically important interplay between Maxwell's equations for electromagnetism and relativity; students in the course are not assumed to be well versed in Maxwell's equations. There will also be no time to discuss applications of relativity to observations in the real world, especially in the context of astronomy and cosmology. Students are encouraged to look into this in their spare time.

Resources

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. In addition to the references below, Wikipedia is your friend.

The primary textbook reference for this course is

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

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. In order to provide some diversity in presentation, my lectures will not follow the text book on the nose, though I will do my best to make all notation and terminology consistent.

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).

I have borrowed heavily from the lecture notes for the previous incarnation of this course, which was taught by Andrew Hodges until Trinity Term of 2015. Another excellent set of lectures notes are those by David Tong at the University of Cambridge. I have borrowed from these in a number of places, but they are definitely worth looking at in their own right.

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 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:

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

- 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 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 From Galileo and Newton to Einstein

We begin by revisiting the world of Newton and Galileo (and prelims Dynamics). We will identify assumptions of the Newtonian worldview that will need to be modified or abandoned in the relativistic setting. Motivated by the experimental fact that the speed of light is the same in all inertial frames, we will lay out the postulates that undergird Einstein's special theory of relativity.

1.1 Galilean space-time

We must first introduce some elementary notions of space and time. We start with the concept of an event.

Definition 1. An *event* is a specific point in space at a specific moment in time.

The quintessential example of an event is *right here* and *right now*. *Thirsty Meeples in Gloucester Green at 16:00 next Sunday* is another example. In a given coordinate system, an event is determined by a choice of time t and a choice of spatial coordinates, often denoted (x, y, z). Under changes of coordinates, the representation of an event may change, but the event itself is an invariant notion.

Taken together, the set of all events constitutes the entirety of space and time, which we give the name *space-time*.

Definition 2. *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, $\mathbb{R}^3 \times \mathbb{R}_t \cong \mathbb{R}^4$. This gives us, for example, a notion of continuous paths in space-time.

Definition 3. The (continuous) path through spacetime traced out by an object/particle/observer is its *world-line*.

We should be careful to not think of Galilean space-time being equipped with the natural Euclidean inner product on $\mathbb{R}^4$; Euclidean distances in $\mathbb{R}^4$ are not invariant under Galilean transformations.

With space-time in place, let us recall the Prelims notion of a reference frame:

Definition 4. In Newtonian mechanics, a *reference frame* is a choice of spatial origin, together with a set of perpendicular (right handed) Cartesian coordinate axes for space, which vary continuously with time.

Newtonian mechanics entails a number of assumptions about the reference frames we can choose for space-time:

- 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).*
- Any two ICSs are related by a Galilean transformation.
- The inertial mass of an object (that appearing in Newton's second law) is independent of motion, and so in particular is the same in any inertial frame.
- The laws of physics (Newton's laws and gravity) hold in any inertial frame (and in particular take the same form in any ICS).

The final statement implies that there is no preferred inertial frame – only motion *relative to* some inertial frame is observable and meaningful.[2] This is the *principle of relativity.* This particular version of the principle of relativity, where we demand that ICSs be related by Galilean transformations, is known as *Galilean relativity.*

You will show in the problems that requiring invariance under Galilean transformations means that to two events we can always assign a meaningful a time difference, but we can assign a spatial distance *only* if those two events are simultaneous. This apparent asymmetry between the treatment of spatial and temporal displacements is overcome in Einstein's relativity.

1.2 The speed of light

Though Newtonian mechanics gives an excellent description of Nature in most "reasonable" circumstances, it is not universally valid. In particular, it needs improving when discussing

[2] This, of course, is a statement about the *laws of physics.* One might very well find themself in a local environment where one inertial reference frame is natural.

the physics of the very small, the very heavy, or the very fast-moving. It is the latter case that will be of interest to us here.[3]

The relevant benchmark for what counts as "very fast" is the speed of light, which is

$$c = 299792458 \ m/s \ . \tag{1.1}$$

This is actually an *exact* value; the length of one meter is defined to be the distance traveled by light in 1/299792458 seconds, where a second is defined in terms of the period of radiative transitions of a Cesium atom.

The first thing that one notices is that this is pretty fast. The earth moves about the sun at a speed of about 30000 m/s (still pretty fast!), which is only about one one-hundredth of one percent of the speed of light. It is not surprising that Newtonian physics is a good approximation if its failure occurs only when objects move at speeds close to c.

The second thing that one notices (as anticipated in the introduction) is that this discussion doesn't seem like it should make any sense. Indeed, in our earlier discussion we stated the very plausible principle of relativity, according to which only relative velocities are meaningful. So the speed given above is the speed of light *relative to what?*

In the time before Einstein, wonderful theories were developed to answer this question. These theories required abandoning the principle of relativity by postulating the existence of a *luminiferous aether* that permeates all of space and whose rest frame defines a preferred reference frame in which the light propagates at its "correct" speed. However, these attempts turned out to be misguided. Indeed, the idea of a luminiferous aether has been banished to the dustbin of history, and instead the following remarkable observation has been confirmed many times over in a variety of ingenious experiments,[4]

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

1.3 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 electro-*

[3]When discussing the very small, one must pass to the world of *quantum mechanics*, while the very heavy will require *general relativity*, and eventually (at extremely high densities) a theory of *quantum gravity*.

[4]Everyone should spend some time online learning about the various tests of this fact, perhaps using Wikipedia as a jumping off point.

dynamics of moving bodies, Albert Einstein explained how to 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.

2 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 the formula for Lorentz transformations in two different ways. Our first derivation will be pragmatic and utilitarian. 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.

2.1 Coordinate transformations and the speed of light

We again consider two inertial observers 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

$$x' = f(x,t) \ , \qquad t' = g(x,t) \ . \tag{2.1}$$

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

$$f(0,0) = g(0,0) = 0 \ . \tag{2.2}$$

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,

$$x' = \alpha_1 x + \alpha_2 t \ , \qquad t' = \beta_1 x + \beta_2 t \ , \tag{2.3}$$

for some real numbers $\alpha_{1,2}$ and $\beta_{1,2}$. 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

$$x' = \gamma(x - vt) \ , \qquad t' = \beta_1 x + \beta_2 t \ . \tag{2.4}$$

For some coefficient γ that can in principle depend on the relative velocity v — let us write $\gamma = \gamma_v$.

We have further specified that a light beam will have to be observed by both O and O' as moving with speed c, meaning that the line $x = ct$ must map to the line $x' = ct'$. Imposing this requirements gives us

$$t' = \gamma_v \left(1 - \frac{v}{c}\right) t \quad \text{for} \quad x = ct \ . \tag{2.5}$$

Now note that if we performed the same argument changing coordinates from (x', t') to (x, t), so going from observer O' to observer O, then we would have gotten the equation

$$t = \gamma_{-v} \left(1 + \frac{v}{c}\right) t' \quad \text{for} \quad x' = ct' \ . \tag{2.6}$$

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

$$\tilde{x}' = \gamma_{-v}(\tilde{x} + vt) \quad \Longrightarrow \quad x' = \gamma_{-v}(x - vt) \ . \tag{2.7}$$

Combined with (2.4), this tells us that $\gamma_v = \gamma_{-v}$.

At last we have the the two equations

$$t' = \gamma_v \left(1 - \frac{v}{c}\right) t \ , \qquad t = \gamma_v \left(1 + \frac{v}{c}\right) t' \ , \tag{2.8}$$

that must hold at every event where $x = ct$. A bit of algebra gives us an expression for γ purely in terms of c and the relative velocity v,

$$\gamma = \sqrt{\frac{1}{1 - \frac{v^2}{c^2}}} \ . \tag{2.9}$$

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 now have gathered enough equations to go back and completely determine the form for our Lorentz transformations. Indeed, from the two equations

$$x' = \gamma(x - vt) \ , \qquad x = \gamma(x' + vt') \ . \tag{2.10}$$

We can solve for t in terms of t' and x', and for t' in terms of x and t. After a bit of algebra, we find our final results

$$\begin{aligned} x' &= \gamma(x - vt) \ , & x &= \gamma(x' + vt') \ , \\ ct' &= \gamma\left(ct - \frac{v}{c}x\right) \ , & ct &= \gamma\left(ct' + \frac{v}{c}x'\right) \ . \end{aligned} \tag{2.11}$$

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

$$\begin{pmatrix} ct' \\ x' \end{pmatrix} = \gamma \begin{pmatrix} 1 & -v/c \\ -v/c & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \end{pmatrix} \ . \tag{2.12}$$

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

$$\begin{pmatrix} ct \\ x \end{pmatrix} = \gamma \begin{pmatrix} 1 & v/c \\ v/c & 1 \end{pmatrix} \begin{pmatrix} ct' \\ x' \end{pmatrix} \ . \tag{2.13}$$

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

2.1.1 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. 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,

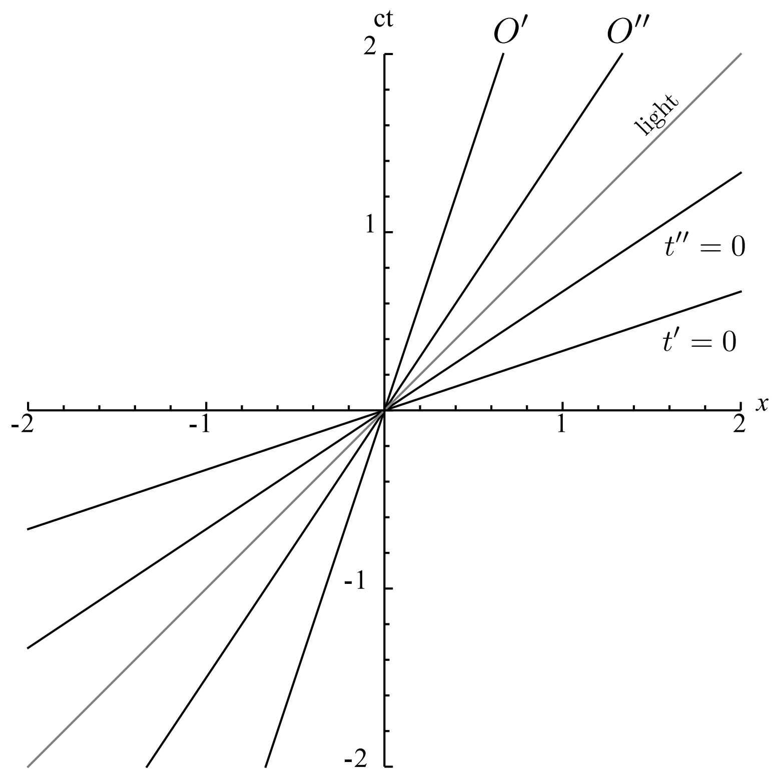

Figure 1. A space-time diagram showing lines of constant position and time coordinates for frames O and O' that are boosted relative to the the stationary frame of the diagram.

- We 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 are 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.*
- We will sometimes draw lines indicating simultaneity, *i.e.*, $t =$ constant or $t' =$ constant, but must always remembering 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.)

2.2 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 (2.11).

2.2.1 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 realiable clocks exist!), and take readings from the clock as they progress through time, thus assigning coordinates $(t, 0, 0, 0)$ for $t \in \mathbb{R}$.

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

This (hypothetical) operation allows any inertial observer to, in principal, 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.

2.2.2 Lorentz transforms from the radar method

We again consider two observers O and O' travelling in one spatial dimension with (unequal) constant speeds. 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.

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.)

Thus O reckons the (relative) speed of O' (as distance/time) to be

$$v = \frac{c(k^2 - 1)}{k^2 + 1} \ . \tag{2.14}$$

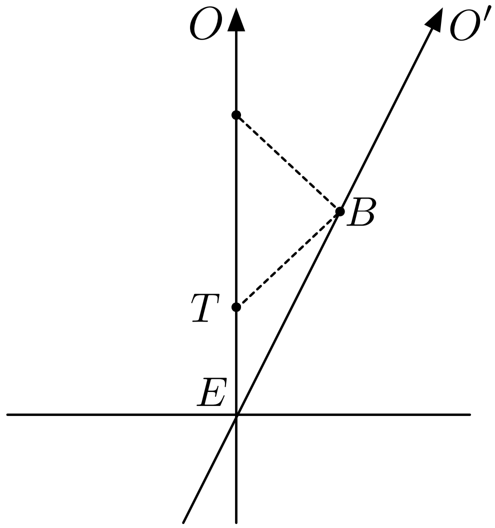

Figure 2. Comparing clocks with the radar method. The dotted lines are light rays.

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

$$k = \sqrt{\frac{c+v}{c-v}}\ . \tag{2.15}$$

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

$$T' = kT = \sqrt{\frac{c+v}{c-v}}\,T\ . \tag{2.16}$$

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 that the amount of time that has passed from E to B is $\sqrt{\frac{c+v}{c-v}}\,T$. We recognize the ratio of these two times, which is independent of T, as the Lorentz factor we found before,

$$\frac{\text{Time measured by } O}{\text{Time measured by } O'} = \frac{1}{\sqrt{1-v^2/c^2}} \equiv \gamma\ . \tag{2.17}$$

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

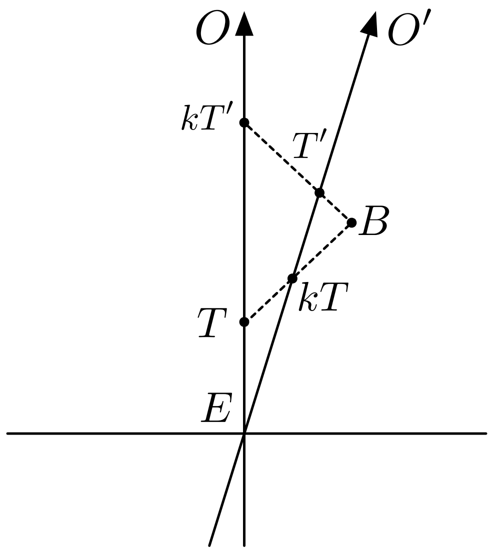

Figure 3. Two inertial observers coordinatize the same space-time event B using the radar method. Times displayed on the worldlines of the observers correspond to those observers' local measurements.

respective ICSs. We can immediately see that the coordinates assigned to B by observer O will be

$$t = \frac{1}{2}(kT' + T) \ , \qquad x = \frac{1}{2}c(kT' - T) \ . \tag{2.18}$$

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

$$t' = \frac{1}{2}(T' + kT) \ , \qquad x' = \frac{1}{2}c(T' - kT) \ . \tag{2.19}$$

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 (2.11). See Woodhouse, pages 66 – 68 for the details.

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 space-time 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 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. We will not pursue these more involved questions in this course, but we 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.

3.1 Length contraction

A famous consequences of special relativity is the phenomenon of 'length contraction'. We consider two inertial observers O and O' whose inertial coordinate systems are related by (2.11). 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

$$(1) \qquad \begin{pmatrix} ct \\ x \end{pmatrix} = \gamma \begin{pmatrix} 1 & v/c \\ v/c & 1 \end{pmatrix} \begin{pmatrix} ct' \\ 0 \end{pmatrix} = \gamma \begin{pmatrix} ct' \\ vt' \end{pmatrix} , \tag{3.1}$$

$$(2) \qquad \begin{pmatrix} ct \\ x \end{pmatrix} = \gamma \begin{pmatrix} 1 & v/c \\ v/c & 1 \end{pmatrix} \begin{pmatrix} ct' \\ L \end{pmatrix} = \gamma \begin{pmatrix} ct' + Lv/c \\ vt' + L \end{pmatrix} . \tag{3.2}$$

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

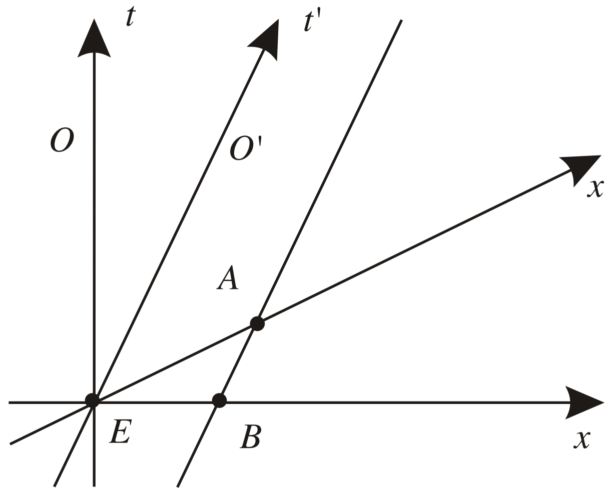

Figure 4. Space-time diagram demonstrating the phenomenon of length contraction.

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. What O measures is the separation between E and B; but when O' measures the length, he is assigning a distance to the separation between E and A.

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 very large 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 $\gamma(nc\delta,\ n\delta v)$. 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.

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

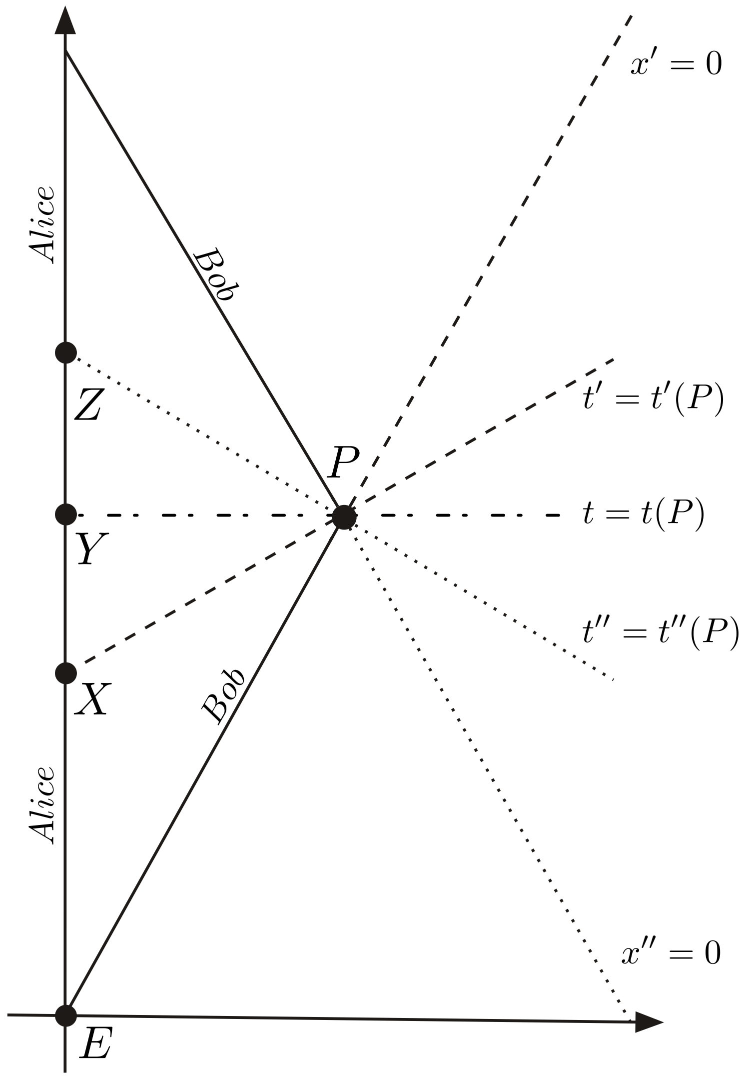

Figure 5. Space-time diagrams showing the twin paradox phenomenon. Dotted lines are lines of constant O' coordinates, dashed lines are lines of constant O'' coordinates. The dot-dashed line is a line of constant time in the original O frame of Alice.

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. Weve 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 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

$$t'(E) = x'(E) = 0 \ , \qquad t'(A) = T' \ , \ x'(A) = X' = vT' \ . \tag{3.3}$$

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,

$$w = \frac{x(A)}{t(A)} = \frac{u+v}{1 + \frac{uv}{c^2}} \ . \tag{3.4}$$

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

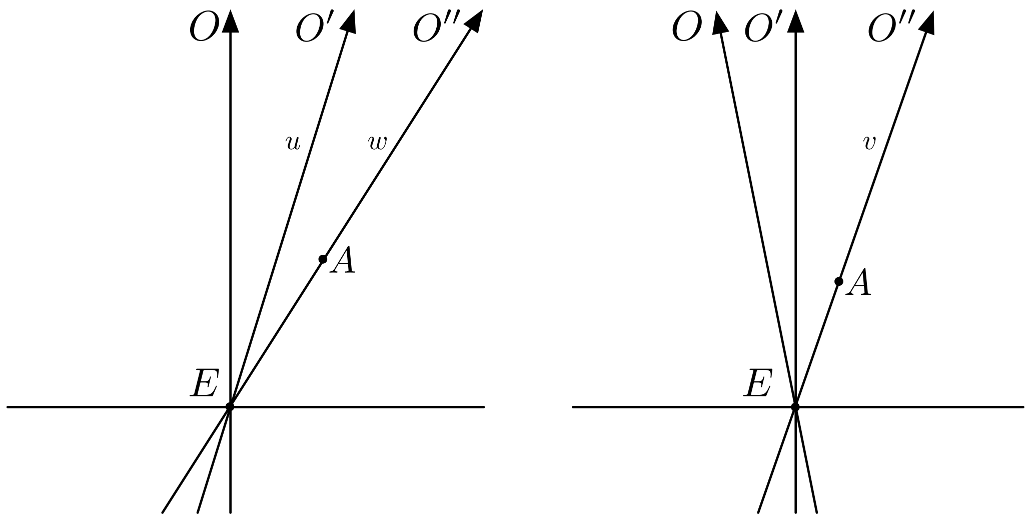

Figure 6. Space-time diagrams showing the configuration of observers relevant for the velocity addition problem. The velocity addition formula determines the relative velocity w in terms of u and v.

have $w < c$, so a change of frame of reference will never result in an object traveling faster than the speed of light. We will see a simpler way to derive this formula in the next section when we discuss the group structure of Lorentz transformations.

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 (x, t)-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.

4.1 Rapidity and group structure

We start with a definition,

Definition 5. The *rapidity* associated with a relative velocity u with $|u| < c$ is given by

$$\phi(u) = \mathrm{Tanh}^{-1}(u/c) \ . \tag{4.1}$$

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

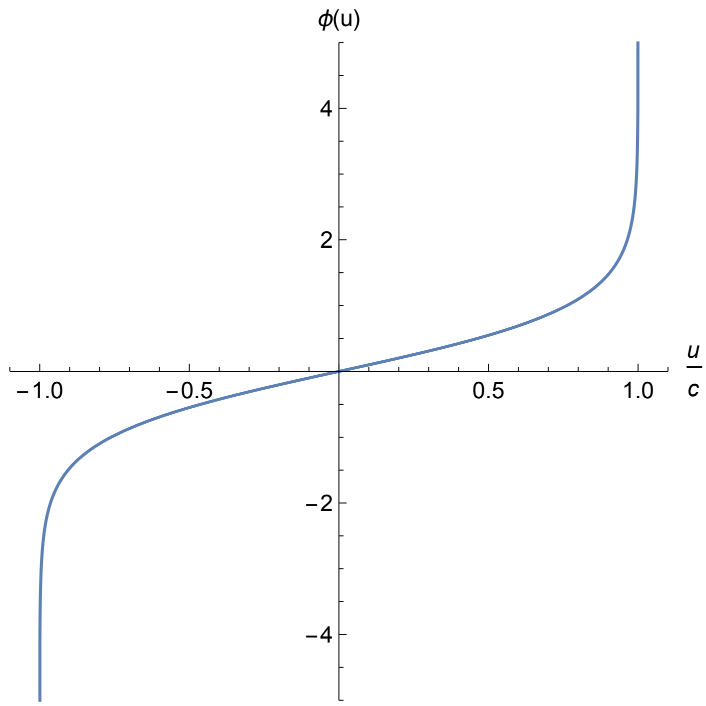

Figure 7. Rapidity as a function of velocity. The map to rapidity maps the set of physical velocities (with magnitude less than the speed of light) to the entire real line.

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,

$$\begin{aligned} \gamma(u) &= \cosh\phi \ , \\ u\gamma(u)/c &= \sinh\phi \ , \\ k(u) &= \exp\phi \ , \end{aligned} \tag{4.2}$$

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

$$\gamma(u)\begin{pmatrix} 1 & u/c \\ u/c & 1 \end{pmatrix} = \begin{pmatrix} \cosh\phi & \sinh\phi \\ \sinh\phi & \cosh\phi \end{pmatrix} \ . \tag{4.3}$$

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

$$\begin{pmatrix} \cosh\phi & \sinh\phi \\ \sinh\phi & \cosh\phi \end{pmatrix}\begin{pmatrix} \cosh\psi & \sinh\psi \\ \sinh\psi & \cosh\psi \end{pmatrix} = \begin{pmatrix} \cosh(\phi+\psi) & \sinh(\phi+\psi) \\ \sinh(\phi+\psi) & \cosh(\phi+\psi) \end{pmatrix} \ . \tag{4.4}$$

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. 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

$$\phi(w) = \phi(u) + \phi(v) \ , \tag{4.5}$$

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

$$w = \frac{u+v}{1+uv/c^2} \ . \tag{4.6}$$

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

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \ , \qquad \theta \in [0, 2\pi) \ . \tag{4.7}$$

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,

$$(H\mathbf{x})^{\mathrm{T}} \cdot (H\mathbf{x}) = \mathbf{x}^{\mathrm{T}} \cdot \mathbf{x} \ , \qquad \forall \mathbf{x} \in \mathbb{R}^2 \ . \tag{4.8}$$

Equivalently, these are the matrices that obey

$$H^{\mathrm{T}} H = I \ , \tag{4.9}$$

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

$$H^{\mathrm{T}} I H = I \ . \tag{4.10}$$

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 (4.10).

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.

4.3 Pseudo-orthogonal transformations in two dimensions

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

$$g = \begin{pmatrix} +1 & 0 \\ 0 & -1 \end{pmatrix} . \tag{4.11}$$

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

$$\begin{aligned} L(\phi)^{\mathrm{T}} g L(\phi) &= \begin{pmatrix} \cosh\phi & \sinh\phi \\ \sinh\phi & \cosh\phi \end{pmatrix} \begin{pmatrix} \cosh\phi & \sinh\phi \\ -\sinh\phi & -\cosh\phi \end{pmatrix} \\ &= \begin{pmatrix} \cosh^2\phi - \sinh^2\phi & 0 \\ 0 & \sinh^2\phi - \cosh^2\phi \end{pmatrix} \\ &= g \ . \end{aligned} \tag{4.12}$$

We make the following definition:

Definition 6. 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 (4.3). 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 7. 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 (4.3) are clearly proper.

Definition 8. 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 (4.3) 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 9. 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

$$L(\phi) = \begin{pmatrix} \cosh\phi & \sinh\phi \\ \sinh\phi & \cosh\phi \end{pmatrix} , \tag{4.13}$$

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 $rmO(1,1)$, 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

$$\begin{aligned}(ct)^2 - x^2 &= (ct, x)\ g \begin{pmatrix} ct \\ x \end{pmatrix} \\ &= (ct, x)\, L^{\mathrm{T}} g L \begin{pmatrix} ct \\ x \end{pmatrix} \\ &= \left(ct', x'\right)\ g \begin{pmatrix} ct' \\ x' \end{pmatrix} \\ &= (ct')^2 - (x')^2 \ .\end{aligned} \tag{4.14}$$

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),

$$c^2(t_2 - t_1)^2 - (x_2 - x_1)^2 = c^2(t_2' - t_1')^2 - (x_2' - x_1')^2 \ . \tag{4.15}$$

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.

5 The Lorentz group in 1+3 dimensions

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 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 the matrix g, or equivalently, that preserve the four-dimensional invariant interval,

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

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 still 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 multiplet 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 10. 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

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

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.

5.2 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. It turns out that the additional y and z coordinates are unchanged.

Exercise 3. Show that under a standard Lorentz transformation in $1+3$ dimensions, the y and z coordinates must 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 stanard 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 u and for some $H, H' \in \mathrm{SO}(3)$. The proof is left as an exercise.[5]

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 4. 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?

5.3 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,

$$\begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} = L \begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} + \begin{pmatrix} C^0 \\ C^1 \\ C^2 \\ C^3 \end{pmatrix} , \tag{5.10}$$

where the C^a are constants. This transformation relates the origin in frame O' to the point C' in frame O. Similarly, the transformation that relates the point C in frame O' to the origin in frame O is given by

$$\begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} = L \begin{pmatrix} ct' - C'^0 \\ x' - C'^1 \\ y' - C'^2 \\ z' - C'^3 \end{pmatrix} , \tag{5.11}$$

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

Poincaré group (not examinable). The Poincaré transformations also form a group, known (unsurprisingly) as the *Poincaré group.* This group turns out to be the *semi-direct*

[5] 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.

product of the Lorentz group $SO^+(1,3)$ extended by the translation group of four-dimensional spacetime $\mathbb{R}^{1,3}$,

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

This is the (identity component of) the *isometry group* of the four-dimensional Minkowski spacetime that we will meet in the next section.

6 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.

6.1 Preliminaries about four-vectors

The general idea of a four-dimensional vector space is nothing new here. 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.

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

$$X = LX' \ , \tag{6.1}$$

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

$$\overrightarrow{PQ} = (\, ct_2 - ct_1 \ , \ x_2 - x_1 \ , \ y_2 - y_1 \ , \ z_2 - z_1 \,) \tag{6.2}$$

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 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.
- The same, 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. We use lower indices for dual vectors. 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.

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 3-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.

6.2 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,

$$x^a = \sum_{b=0}^{3} L^a{}_b x'^b + T^a \ , \qquad a = 0, 1, 2, 3 \ . \tag{6.3}$$

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

$$X^a = \sum_{b=0}^{3} L^a{}_b X'^b \ . \tag{6.4}$$

We immediately see that the coordinates x^a themselves are not four-vectors due to the inhomogeneous term in (6.3), but displacement vectors of 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:

$$g(X, Y) = \sum_{a,b=0}^{3} g_{ab} X^a Y^b \ , \tag{6.5}$$

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. □

Definition 11. $g(X, Y)$ is said to define a *pseudo-inner product* on the vector space of four-vectors. It also defines a notion of *pseudo-orthonormal basis*, (X^0, X^1, X^2, X^3) with the property $g(X^0, X^0) = 1, g(X^1, X^1) = g(X^2, X^2) = g(X^3, X^3) = -1, g(X^0, X^1) = g(X^0, X^2) = g(X^0, X^3) = g(X^1, X^2) = g(X^1, X^3) = g(X^2, X^3) = 0$.

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

$$X_a = \sum_b g_{ab} X^b \tag{6.6}$$

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 (6.3) 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

$$x^a = L^a{}_b x'^b + T^a \,, \tag{6.7}$$

and the relation between components of vectors by

$$X^a = L^a{}_b X'^b \,. \tag{6.8}$$

The condition for L to be Lorentz is

$$L^a{}_c g_{ab} L^b{}_d = g_{cd} \,. \tag{6.9}$$

The pseudo-inner-product is defined as

$$g(X,Y) = g_{ab}X^aY^b = X^0Y^0 - X^1Y^1 - X^2Y^2 - X^3Y^3\,, \tag{6.10}$$

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

Lowering the index is written simply as

$$X_a = g_{ab}X^b\,, \tag{6.11}$$

and now $g(X,Y) = X_aY^a = X^aY_a$.

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, take care of everything.

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

$$g^{ab} = \mathrm{diag}(1,-1,-1,-1)\,. \tag{6.12}$$

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}.

With this definition,

$$g^{ab}X_b = X^a\ , \tag{6.13}$$

and now we can freely raise and lower indices.

Exercise 5. Check that this definition of lowering and raising indices is consistent, i.e that $g^{ab}g_{bc}X^c = X^a$.

Exercise 6. 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. □

6.3 Classification of four-vectors

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

Definition 12. 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 four-vector whose spatial part vanishes in some ICS must be timelike; and a 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:

$$\frac{1}{\sqrt{P^2 - p^2}}(P, p\,\mathbf{e}), \quad \frac{1}{\sqrt{P^2 - p^2}}(p, P\,\mathbf{e}), \quad (0, \mathbf{q}), \quad (0, \mathbf{r}) \;, \tag{6.14}$$

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. □

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 13. 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 8, where the x^0 axis is vertical and one spatial dimension is suppressed. The null vectors lie on the cone

$$(x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2 = 0 \ , \tag{6.15}$$

which has its vertex at the origin.

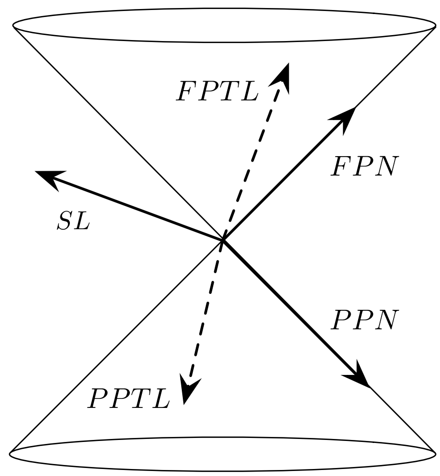

Figure 8. *The space of four-vectors, illustrating future-pointing timelike, future-pointing null, spacelike, past-pointing null, past-pointing timelike.*

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

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.

6.4 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

$$|\mathbf{x}+\mathbf{y}| \leqslant |\mathbf{x}| + |\mathbf{y}| \ . \tag{6.16}$$

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

$$\sqrt{g(U+V,U+V)} \geqslant \sqrt{g(U,U)} + \sqrt{g(V,V)} \ . \tag{6.17}$$

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

$$\sqrt{(U^0+V^0)^2-(V^1)^2} \geqslant U^0 + \sqrt{(V^0)^2-(V^1)^2} \tag{6.18}$$

Squaring both sides, this is equivalent to

$$(U^0+V^0)^2-(V^1)^2 \geqslant (U^0)^2 + 2U^0\sqrt{(V^0)^2-(V^1)^2} + (V^0)^2-(V^1)^2 \ , \tag{6.19}$$

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

$$2U^0V^0 \geqslant 2U^0\sqrt{(V^0)^2-(V^1)^2} \ , \tag{6.20}$$

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.

7 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.

7.1 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,

$$V^a \stackrel{?}{=} \frac{\mathrm{d}x^a(t)}{\mathrm{d}t} \ . \tag{7.1}$$

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 parameterization s of the particle world-line on which all observers will agree. Said differently, at for 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 (7.1). It is then clear that there is only one reasonable option: we should use the time coordinate of the 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

$$\frac{\mathrm{d}s}{\mathrm{d}t} = \frac{1}{\gamma(v)} = \sqrt{1 - \frac{\mathbf{v}^2}{c^2}} \ . \tag{7.2}$$

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

$$c\,\mathrm{d}s = \sqrt{c^2\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2} \ . \tag{7.3}$$

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

$$\mathrm{d}s_{\mathrm{E}} = \sqrt{\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2} \ , \tag{7.4}$$

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 14. The *proper time* at any event P on a time-like world-line γ is given by

$$s(P) = \frac{1}{c} \int_{P_0}^{P} \sqrt{c^2 dt^2 - dx^2 - dy^2 - dz^2} \ , \tag{7.5}$$

where (ct, x, y, z) are coordinates in an inertial frame and the integral is taken along 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 infintesimal measure, the proper time is independent 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 is 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 parameterized 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 use s as a natural parameterization of any time-like world-line,

$$x^a = x^a(s) \ , \tag{7.6}$$

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

$$V^a = \frac{\mathrm{d}x^a(s)}{\mathrm{d}s} \ . \tag{7.7}$$

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 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 1 (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. Therefore

$$c^2\gamma(w) = \gamma(u)\gamma(v)(c^2 - \mathbf{u}\cdot\mathbf{v})\ . \tag{7.8}$$

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

$$w = \frac{c\sqrt{c^2(\mathbf{u}-\mathbf{v})\cdot(\mathbf{u}-\mathbf{v}) - |\mathbf{u}\wedge\mathbf{v}|^2}}{c^2 - \mathbf{u}\cdot\mathbf{v}}\ . \tag{7.9}$$

This reduces to the one-dimensional addition formula (4.6) when $\mathbf{u}$ and $\mathbf{v}$ are parallel.

Example 2 (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 3 (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

$$\begin{aligned} g((P+sU) - (Q+s'U'), U\ ) &= 0\ ,\\ g((P+sU) - (Q+s'U'), U') &= 0\ , \end{aligned} \tag{7.10}$$

which gives two linear equations for s and s',

$$\begin{aligned} g(P-Q,U) + c^2 s - g(U',U)s' &= 0\ ,\\ g(P-Q,U') + g(U,U')s - c^2 s' &= 0\ . \end{aligned} \tag{7.11}$$

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.)

7.2 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 *co-ordinate 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

$$\frac{\mathrm{d}x^a}{\mathrm{d}s} = U^a \ , \qquad g_{ab}U^aU^b = c^2 \ , \tag{7.12}$$

so s is a measure of proper time. Differentiating the right hand equation in (7.12) with respect to proper time gives us

$$g_{ab}U^a\frac{\mathrm{d}U^b}{\mathrm{d}s} + g_{ab}\frac{\mathrm{d}U^a}{\mathrm{d}s}U^b = 2g_{ab}U^a\frac{\mathrm{d}U^b}{\mathrm{d}s} = 0 \ . \tag{7.13}$$

Definition 15. The *acceleration four-vector* for a time-like trajectory with four-velocity U is given by

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

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

$$V = \gamma(v)(c, \mathbf{v}) \ , \tag{7.14}$$

from which we obtain

$$A = \frac{\mathrm{d}V}{\mathrm{d}s} = \gamma(v)\frac{\mathrm{d}V}{\mathrm{d}t} = c^{-2}\gamma(v)^4(c, \mathbf{v})v\frac{\mathrm{d}v}{\mathrm{d}t} + \gamma(v)^2\left(0, \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t}\right) \ . \tag{7.15}$$

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

Example 4 (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

$$U = \left(c\frac{\mathrm{d}t}{\mathrm{d}s} \, , \; \frac{\mathrm{d}x}{\mathrm{d}s} \, , \; 0 \, , \; 0 \right) \; , \quad A = \left(c\frac{\mathrm{d}^2t}{\mathrm{d}s^2} \cdot \frac{\mathrm{d}^2x}{\mathrm{d}s}^2 \, . \, 0 \, , 0 \right) \; . \tag{7.16}$$

These four-vectors are required to satisfy the constraints

$$c^2 \left(\frac{\mathrm{d}t}{\mathrm{d}s} \right)^2 - \left(\frac{\mathrm{d}x}{\mathrm{d}s} \right)^2 = c^2 \; , \quad c^2 \left(\frac{\mathrm{d}^2t}{\mathrm{d}s^2} \right)^2 - \left(\frac{\mathrm{d}^2x}{\mathrm{d}s^2} \right)^2 = -\kappa^2 \; . \tag{7.17}$$

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

$$c\frac{\mathrm{d}^2t}{\mathrm{d}s^2} = \kappa\sqrt{\left(\frac{\mathrm{d}t}{\mathrm{d}s} \right)^2 - 1} \; , \quad \frac{\mathrm{d}^2x}{\mathrm{d}s^2} = \kappa\frac{\mathrm{d}t}{\mathrm{d}s} \; . \tag{7.18}$$

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,

$$\frac{\mathrm{d}t}{\mathrm{d}s} = \cosh(\kappa s/c) \; , \qquad \frac{\mathrm{d}x}{\mathrm{d}s} = c \sinh(\kappa s/c) \; . \tag{7.19}$$

Integrating, we find

$$\begin{aligned} ct(s) &= \frac{c^2}{\kappa} \sinh\left(\frac{\kappa s}{c} \right) \; , \\ x(s) &= \frac{c^2}{\kappa} \cosh\left(\frac{\kappa s}{c} \right) \; , \\ y(s) &= 0 \; , \\ z(s) &= 0 \; , \end{aligned} \tag{7.20}$$

which is a hyperbola in the (x, ct) plane. Note that we have that $\kappa x = c\sqrt{\kappa^2 t^2 + c^2}$, 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/sec/sec 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.

7.3 Frequency four-vector

We introduce one final four-vector, this one associated with waves that propagate at the speed of light, such as light waves. Recall from the problems that in one dimension, a solution of the wave equation moving with velocity $+c$, is given by a function ϕ of the form

$$\phi(t, x) = \phi(ct - x) \ . \tag{7.21}$$

A typical solution with frequency ω is then given by $\cos(\omega(ct - x)/c)$. The generalisation of this to a wave travelling in three dimensions with direction $\mathbf{n}$ and with frequency ω is given by

$$\phi(t, x) = \cos\left(\frac{\omega}{c}\left(ct - \mathbf{n}\cdot\mathbf{x}\right)\right) \ . \tag{7.22}$$

We can put this in a four-vector form in terms of the *frequency four-vector* which is defined by

$$K = \omega(1, \mathbf{n}) \ . \tag{7.23}$$

The wave is then given by

$$\phi(X) = \cos\left(g(K, X)/c\right) \ . \tag{7.24}$$

The frequency ω can now be expressed in an invariant form as $g(U, K)/c$, where U is the four-velocity of the observer. This gives us an elegant four-vector-based method for solvig the relativistic Doppler shift problem.

Example 5 (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

$$\omega_{obs} = \omega\gamma(u)\left(1 - \frac{u}{c}\right) = \sqrt{\frac{c-u}{c+u}} \ , \tag{7.25}$$

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

$$\omega_{obs} = \omega\gamma(u)\left(1 + \frac{u}{c}\right) = \sqrt{\frac{c+u}{c-u}} \ . \tag{7.26}$$

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

$$\omega_{bs} = \omega\gamma(u) \ . \tag{7.27}$$

8 Relativistic dynamics

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.

8.1 Newtonian conservation laws

In Newtonian theory, conservation laws were of great use in analyzing many complicated mechanical systems, especially the collision of particles. Of particular import 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

$$\sum_{i=1}^{k} m_i = \sum_{i=k+1}^{n} m_i \ . \tag{8.1}$$

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

$$\sum_{i=1}^{k} m_i\mathbf{v}_i = \sum_{i=k+1}^{n} m_i\mathbf{v}_i \ , \tag{8.2}$$

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 the next section, energy was often lost to 'heat' or 'radiation' in the Newtonian context (in particular when particles collided inelastically, and thus coalesced).

Collectively, the conservation 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 generalizations of these constraints that reduce to them in the non-relativistic limit.

8.2 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 16. 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 17. The *momentum four-vector* of a particle with rest mass m is and four-velocity U is the four-vector $P = mU$.

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 $g(mU, mU) = m^2c^2$. Thus the four-momentum can be 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,

$$\sum_{i=1}^{k} P_i = \sum_{j=k+1}^{n} P_j \ , \tag{8.3}$$

where the m_i are the rest masses of the particles.

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 c, we have

$$\begin{gathered} c\left(\sum_{i=1}^{k} m_i - \sum_{i=k+1}^{n} m_i + O\left(\frac{u^2}{c^2}\right)\right) = 0 \ , \\ \left(\sum_{i=1}^{k} m_i\mathbf{u}_i - \sum_{i=k+1}^{n} m_i\mathbf{u}_i\right) + O\left(\frac{u^2}{c^2}\right) = 0 \ . \end{gathered} \tag{8.4}$$

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 is given by

$$P_i = (\, \gamma(u_i) m_i c_i \, , \, \gamma(u_i) m_i \mathbf{u}_i \,) \ . \tag{8.5}$$

Let us consider in turns the spatial and temporal parts of our relativistic conservation law.

The spatial part of the law reads as

$$\sum_{i=1}^{k} \gamma(u_i) m_i \mathbf{u}_i = \sum_{i=k+1}^{n} \gamma(u_i) m_i \mathbf{u}_i \ . \tag{8.6}$$

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 18. 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 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

$$\sum_{i=1}^{k} \gamma(u_i) m_i c = \sum_{i=k+1}^{n} \gamma(u_i) m_i c \ . \tag{8.7}$$

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 power of c, we have

$$\sum_{i=1}^{k} \left(m_i c^2 + \frac{1}{2} m_i u_i^2 \right) = \sum_{i=k+1}^{n} \left(m_i c^2 + \frac{1}{2} m_i u_i^2 \right) + O\left(\frac{u^2}{c^4} \right) \ . \tag{8.8}$$

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 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 19. 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

$$P = \left(\frac{E}{c} \; , \; \mathbf{p} \right) \; , \tag{8.9}$$

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,

$$E^2 = m^2c^4 + p^2c^2 \; . \tag{8.10}$$

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

8.3 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:

$$F^a = \frac{\mathrm{d}P^a}{\mathrm{d}s} = mA^a \; . \tag{8.11}$$

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,

$$F = \left(F^0 \, , \, \gamma(u)\mathbf{f} \right) \; . \tag{8.12}$$

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

$$\gamma(u)\mathbf{f} = \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}s} = \frac{\mathrm{d}t}{\mathrm{d}s}\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = \gamma(u)\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} \; , \tag{8.13}$$

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. 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 $E^2 = m^2c^2 + \mathbf{p}\cdot\mathbf{p}\,c^4$, we differentiate both sides and find

$$E\frac{\mathrm{d}E}{\mathrm{d}t} = \mathbf{p}\cdot\mathbf{f}c^2 \ , \tag{8.14}$$

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

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \mathbf{u}\cdot\mathbf{f} \ . \tag{8.15}$$

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

8.4 The four-momentum of a photon

Above we only considered 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, and therefore also have a frequency four-vector K.

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

$$E = \hbar\omega \ , \qquad \mathbf{p} = \frac{\hbar\omega}{c}\hat{\mathbf{n}} \ , \tag{8.16}$$

Recalling our expression for the frequency four-vector, this means that the momentum four-vector for a photon is just given by

$$P = \frac{\hbar}{c}K = \left(\frac{\hbar\omega}{c}, \frac{\hbar\omega}{c}\hat{\mathbf{n}}\right) \ . \tag{8.17}$$

We see that the analogue of (8.10) in this case is given by

$$E^2 = p^2c^2 \ , \tag{8.18}$$

which is the same as (8.10) 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*.

9 Particle physics

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 particle interactions than would otherwise be allowed.

Example 6 (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 mass m_A that decays (splits) into two new particles of type B, each of rest mass m_B.

$$A \longrightarrow B + B \ . \tag{9.1}$$

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 $u_2\hat{\mathbf{n}}_2$. We will show that the masses of the particles and their velocities are related according to

$$\begin{aligned} \hat{\mathbf{n}}_1 &= -\hat{\mathbf{n}}_2 \ , \\ u_1 &= u_2 =: u \ , \\ M &= 2m\gamma(u) \ . \end{aligned} \tag{9.2}$$

Solution. By conservation of four-momentum

$$M(c,0,0,0) = m\gamma(u_1)(c, u_1\hat{\mathbf{n}}_1) + m\gamma(u_2)(c, u_2\hat{\mathbf{n}}_1) \ . \tag{9.3}$$

From the spatial components, we have

$$\gamma(u_1)u_1\hat{\mathbf{n}}_1 = \gamma(u_2)u_2\hat{\mathbf{n}}_2 \ , \tag{9.4}$$

which is solved by $u_1 = u_2 =: u$ 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,

$$\begin{aligned} \frac{1}{\sqrt{1-\frac{u^2}{c^2}}} &= \frac{M}{2m} \ , \\ 1 - \frac{u^2}{c^2} &= \frac{4m^2}{M^2} \ , \\ u &= \sqrt{1 - \frac{4m^2}{M^2}} \ . \end{aligned} \tag{9.5}$$

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

$$\gamma + e^- \longrightarrow \gamma + e^- \ . \tag{9.6}$$

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

$$\hbar\omega\omega'(1-\cos\theta) = mc^2(\omega - \omega') \ , \tag{9.7}$$

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

$$P_e = m_e(c,0,0,0) \ , \qquad P'_e = m_e\gamma(u)(c, \mathbf{u}) \ . \tag{9.8}$$

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

$$P_\gamma = \frac{\hbar\omega}{c}(1, \mathbf{e}) \ , \qquad P'_\gamma = \frac{\hbar\omega'}{c}(1, \mathbf{e}') \ . \tag{9.9}$$

Our four-momentum conservation condition reads

$$P_e + P_\gamma = P'_e + P'_\gamma \ . \tag{9.10}$$

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

$$\begin{aligned} m_e^2 c^2 &= g(P'_e, P'_e) \\ &= g(P_e + P_\gamma - P'_\gamma, P_e + P_\gamma - P'_\gamma) \\ &= g(P_e, P_e) + 2g(P_e, P_\gamma - P'_\gamma) - 2g(P_\gamma, P'_\gamma) \ , \end{aligned} \tag{9.11}$$

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 $g(P_e, P_e) = m_e^2 c^2$, which leaves

$$g(P_e, P_\gamma - P'_\gamma) = g(P_\gamma, P'_\gamma) \ . \tag{9.12}$$

Inserting the components this yields

$$\frac{\hbar^2 \omega\omega'}{c^2}(1 - \mathbf{e}\cdot\mathbf{e}') = m\hbar(\omega - \omega') \ . \tag{9.13}$$

This is equivalent to the desired expression.

Example 8 (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)

$$e^- + e^- \longrightarrow e^- + e^- \ . \tag{9.14}$$

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

$$\cos\theta = \frac{c^2}{vw}(1 - \sqrt{1 - v^2/c^2})(1 - \sqrt{1 - w^2/c^2}) \ . \tag{9.15}$$

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

$$P_1 = m(c, 0, 0, 0) \ , \qquad P_2 = m\gamma(u)(c, \mathbf{u}) \ , \tag{9.16}$$

and the four-momenta of the outgoing elections be

$$Q_1 = m\gamma(v)(c, \mathbf{v}) \ , \qquad Q_2 = m\gamma(w)(c, \mathbf{w}) \ . \tag{9.17}$$

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

$$\gamma(u) + 1 = \gamma(v) + \gamma(w) \ , \quad \gamma(u)\mathbf{u} = \gamma(v)\mathbf{v} + \gamma(w)\mathbf{w} \ . \tag{9.18}$$

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

$$\begin{aligned} g(P_2, P_2) &= g(Q_1 + Q_2 - P_1, Q_1 + Q_2 - P_1) \ , \\ &= g(Q_1, Q_1) + g(Q_2, Q_2) + g(P_1, P_1) + 2g(Q_1, Q_2) - 2g(Q_1, P_1) - 2g(Q_2, P_1) \ . \end{aligned} \tag{9.19}$$

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

$$m_e^2 c^2 = g(Q_1, P_1) + g(Q_2, P_1) - g(Q_1, Q_2) \ . \tag{9.20}$$

We also have

$$\begin{aligned} g(Q_1, P_1) &= m_e^2 c^2 \gamma(v) \ , \\ g(Q_2, P_1) &= m_e^2 c^2 \gamma(w) \ , \\ g(Q_1, Q_2) &= m_e^2 \gamma(v)\gamma(w) \left(c^2 - \mathbf{v} \cdot \mathbf{w}\right) \ , \end{aligned} \tag{9.21}$$

So substituting back in we have

$$1 - \frac{\mathbf{v} \cdot \mathbf{w}}{c^2} = \frac{1}{\gamma(v)} + \frac{1}{\gamma(w)} - \frac{1}{\gamma(v)} \frac{1}{\gamma(w)} \ , \tag{9.22}$$

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 $\mathbf{w}$. (The same conclusion follows from taking the limit $c \to \infty$ in (9.22).)

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.