THE GEOMETRY OF SPECIAL RELATIVITY

Tevian Dray
Department of Mathematics, Oregon State University
tevian@math.orst.edu

8 July 2003

Lorentz transformations are just hyperbolic rotations.

Preface

The unification of space and time introduced by Einstein's special theory of relativity is one of the cornerstones of the modern scientific description of the universe. Yet the unification is counterintuitive, since we perceive time very differently from space. And, even in relativity, time is not just another dimension, it is one with different properties. Some authors have tried to "unify" the treatment of time and space, typically by replacing t by it, thus hiding some annoying minus signs. But these signs carry important information: Our universe, as described by relativity, is *not* Euclidean.

This short book treats the geometry of hyperbolas as the key to understanding special relativity. This approach can be summarized succinctly as the replacement of the ubiquitous γ symbol of most standard treatments with the appropriate hyperbolic trigonometric functions. In most cases, this not only simplifies the appearance of the formulas, but emphasizes their geometric content in such a way as to make them almost obvious. Furthermore, many important relations, including but not limited to the famous relativistic addition formula for velocities, follow directly from the appropriate trigonometric addition formulas.

I am unaware of any other introductory book on special relativity which adopts this approach as fundamental. Many books point out the relationship between Lorentz transformations and hyperbolic rotations, but few actually make use of it. A pleasant exception was the original edition of Taylor and Wheeler's marvelous book [1], but much of this material was removed from the second edition [2].

At the same time, this book is not intended as a replacement for that or any of the other excellent textbooks on special relativity. Rather, it is intended as an introduction to a particularly beautiful way of looking at special relativity, in hopes of encouraging students to see beyond the formulas to the deeper structure. Enough applications are included to get the basic idea, but these would probably need to be supplemented for a full course.

While much of the material presented can be understood by those familiar with the ordinary trigonometric functions, occasional use is made of elementary differential calculus. In addition, the chapter on electricity and magnetism assumes the reader has seen Maxwell's equations, and has at least a passing acquaintance with vector calculus. A prior course in calculus-based physics, up to and including electricity and magnetism, should provide the necessary background.

After a general introduction in Chapter 1, the basic physics of special relativity is described in Chapter 2. This is a quick, intuitive introduction to special relativity, which sets the stage for the geometric treatment which follows. Chapter 3 summarizes some standard (and some not so standard) properties of ordinary 2-dimensional Euclidean space, expressed in terms of the usual circular trigonometric functions; this geometry will be referred to as *circle geometry*. This material has deliberately been arranged so that it closely parallels the treatment of 2-dimensional Minkowski space in Chapter 4 in terms of hyperbolic trigonometric functions, which we call *hyperbola geometry*. [1] Special relativity is covered again from the geometric point of view in Chapter 5, which is followed by a discussion of some of the standard "paradoxes" in Chapter 8, applications to relativistic mechanics in Chapter 9, and the relativistic unification of electricity and magnetism in Chapter 11. Finally, Chapter 13 contains a brief discussion of the further steps leading to Einstein's general theory of relativity.

[1]Not to be confused with *hyperbolic geometry*, the curved geometry of the 2-dimensional unit hyperboloid. See Chapter 13.

Acknowledgments

This book grew out of class notes for a course on *Reference Frames*, which in turn forms part of a major upper-division curriculum reform effort, entitled *Paradigms in Physics*, which was begun in the Department of Physics at Oregon State University in 1997. I am grateful to all of the faculty involved in this effort, but especially to the leader of the project, Corinne Manogue, for support and encouragement at every stage. The *Paradigms in Physics* project was supported in part by NSF grant DUE–965320, supplemented with funds from Oregon State University; my own participation was made possible thanks to the (sometimes reluctant!) support of my department chair, John Lee. I was fortunate in having excellent teaching assistants, Jason Janesky and Emily Townsend, the first times I taught the course. A course based on an early draft of this book was taught at Mount Holyoke College in 2002, giving me an opportunity to make further revisions; my stay at Mount Holyoke was partially supported by their Hutchcroft Fund. I am grateful to Greg Quenell for having carefully read the manuscript at that time, and for suggesting improvements. Last but not least, I thank the many students who struggled to learn physics from a mathematician, enriching all of us.

Contents

List of Figures

Chapter 1

Introduction

1.1 Newton's Relativity

Our daily experience leads us to believe in Newton's laws. When you drop a ball, it falls straight down. When you throw a ball, it travels in a uniform (compass) direction — and falls down. We appear to be in a constant gravitational field, but apart from that there are no forces acting on the ball. This isn't the full story, of course, as we are ignoring things like air resistance and the spin of the ball. Nevertheless, it seems to give a pretty good description of what we observe, and so we base our intuitive understanding of physics on it.

But it's wrong.

Yes, gravity is more complicated than this simple picture. The gravitational field of the Earth isn't really constant. And there are other nearby objects, notably the Moon, whose gravity acts on us. As discussed in the final chapter, this causes tides.

A bigger problem is that the Earth is round. Due East is *not* a straight line, defined in this case as the shortest distance between 2 points, as anyone who flies from San Francisco to New York is aware. In fact, if you travel in a straight line (initially) due East from my home in Oregon, you will eventually pass to the south of the southern tip of Africa! [1]

So East is not East.

But the real problem is that the Earth is rotating. Try playing catch

[1] You can check this by stretching a string on a globe so that it goes all the way around, is as tight as you can make it, and goes through Oregon in an East/West direction.

on a merry-go-round! Balls certainly don't seem to travel in a straight line! Newton's laws don't work here, and strictly speaking they don't work on (that is, in the reference frame of) the Earth's surface. The motion of a Foucault pendulum can be thought of as a Coriolis effect, caused by an external *pseudoforce*. And a plumb bob doesn't actually point towards the center of the Earth!

So down is not down.

1.2 Einstein's Relativity

All of the above problems come from the fact that, even without worrying about gravity, the surface of the Earth is *not* an inertial frame. An inertial frame is, roughly speaking, one in which Newton's laws do hold. Playing catch on a train is little different from on the ground — at least in principle, and so long as the train is not speeding up or slowing down. Furthermore, an observer on the ground would see nothing out of the ordinary, it merely being necessary to combine the train's velocity with that of the ball.

However, shining a flashlight on a moving train, and especially the description of this from the ground, turns out to be another story, which we will study in more detail below. Light doesn't behave the way balls do, and this difference forces a profound change in our description of the world around us. As we will see, this forces moving objects to change in unexpected ways: their clocks slow down, they change size, and, in a certain sense, they get heavier.

So time is not time.

Of course, these effects are not very noticeable in our daily lives, any more than Coriolis forces affect a game of catch. But some modern conveniences, notably global positioning technology, are affected by relativistic corrections.

The bottom line is that the reality is quite different from what our intuition says it ought to be. The world is neither Euclidean nor Newtonian. Special relativity isn't just some bizarre theory, it is a correct description of nature (ignoring gravity). It is also a beautiful theory, as I hope you will agree. Let's begin.

Chapter 2

The Physics of Special Relativity

In which it is shown that time is not the same for all observers.

2.1 Observers and Measurement

Special relativity involves comparing what different observers see. But we need to be careful about what these words mean.

A *reference frame* is a way of labeling each event with its location in space and the time at which it occurs. Making a measurement corresponds to recording these labels for a particular event. When we say that an *observer* "sees" something, what we really mean is that a particular event is recorded in a the reference frame associated with the observer. This has nothing to do with actually *seeing* anything, a much more complicated process which would involve keeping track of the light reflected into the observer's eyes! Rather, an "observer" is really an entire army of observers, who record any interesting events; an "observation" consists of reconstructing from their journals what took place.

2.2 The Postulates of Special Relativity

The most fundamental postulate of relativity is

Postulate I: The laws of physics apply in all inertial reference frames.

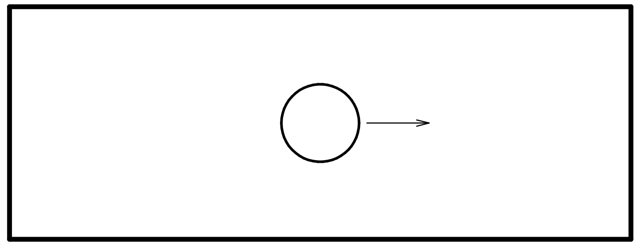


Figure 2.1: A passenger on a train throws a ball to the right. On an ideal train, it makes no difference whether the train is moving.

The first ingredient here is a class of preferred reference frames. Simply put, an inertial (reference) frame is one without external forces. More precisely, an inertial frame is one in which an object initially at rest will remain at rest. Because of gravity, inertial frames must be in free fall — a spaceship with its drive turned off, or a falling elevator. Gravity causes additional complications, such as tidal effects, which force such freely falling frames to be small (compared to, say, the Earth); we will revisit this in the final chapter. But special relativity describes a world without gravity, so in practice we describe inertial frames in terms of relative motion at constant velocity, typically in the form of an idealized train.

Applied to mechanics, Postulate I is the principle of *Galilean relativity.* For instance, consider a ball thrown to the right with speed u. Ignoring things like gravity and air friction, since there are no forces acting on the ball, it keeps moving at the same speed forever. Try the same thing on a train, which is itself moving to the right with speed v. Then Galilean relativity leads to the same conclusion: As seen from the train, the ball moves to the right with speed u forever. An observer on the ground, of course, sees the ball move with speed $u + v$; Galilean relativity insists only that both observers observe the same physics, namely the lack of acceleration due to the absence of any forces, but not necessarily the same speed. This situation is shown in Figure 2.1.

Einstein generalized Postulate I by applying it not just to mechanics, but also to electrodynamics. However, Maxwell's equations make explicit reference to the speed of light! In MKS units, Gauss' Law (Equation (11.76) below) involves the permittivity constant ϵ_0, and Ampère's Law (Equation (11.79) below) involves the permeability constant μ_0; both of these can be measured experimentally. But Maxwell's equations predict electromagnetic

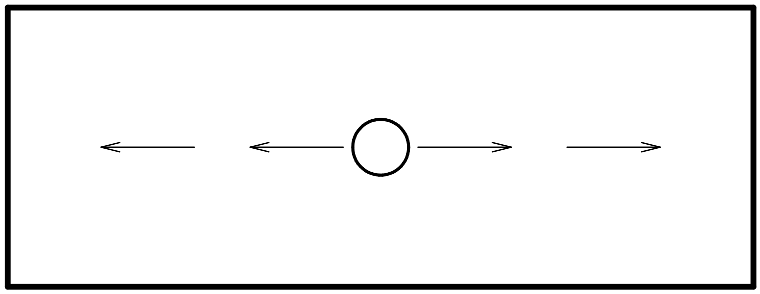


Figure 2.2: The same situation as the previous example, with the ball replaced by a lamp in the exact middle of the train. The light from the lamp reaches both ends of the train at the same time, regardless of whether the train is moving.

waves — including light — with a speed (in vacuum) of

$$c = \frac{1}{\sqrt{\epsilon_0\mu_0}} \tag{2.1}$$

Thus, from some relatively simple experimental data, Maxwell's equations *predict* that the speed of light in vacuum is

$$c = 3 \times 10^8 \ \frac{\mathrm{m}}{\mathrm{s}} \tag{2.2}$$

The famous Michelson/Morley experiment set out to show that this speed is relative to the *ether*, so that we should be able to measure our own motion relative to the ether by measuring direction-dependent variations in c. Instead, the experiment showed that there were no such variations; Einstein argued that there is therefore no ether! Postulate I together with Maxwell's equations therefore lead to

Postulate II: The speed of light is the same for all inertial observers.

As we will now show, an immediate consequence of this is that two inertial observers disagree about whether two events are simultaneous!

Consider a train at rest, with a lamp in the middle, as shown in Figure 2.2. After the light is turned on, light reaches both ends of the train at the same time, having traveled in both directions at constant speed c. Now try the same experiment on a moving train. This is still an inertial frame, and so, just as with the ball in the previous example, one obtains the same result,

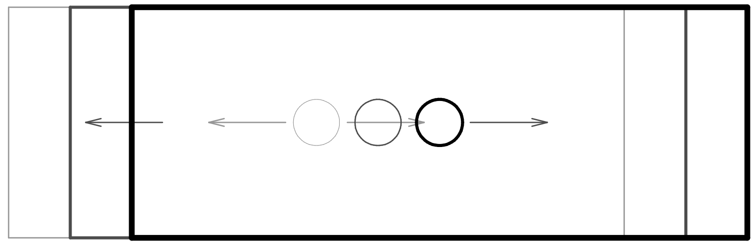


Figure 2.3: An observer on the ground sees the train go past with its lamp, as in the previous example. Since this observer must also see the light travel with speed c, and since the ends of the train are moving while the light is traveling, this observer concludes that the light reaches the rear of the train before the front.

namely that the light reaches both ends of the train at the same time *as seen by an observer on the train*. However, the second postulate leads to a very different result for the observer on the ground. According to this postulate, the light travels at speed c as seen from the ground, *not* the expected $c \pm v$. But, as seen from the ground, the ends of the train also move while the light is getting from the middle of the train to the ends! The rear wall "catches up" with the approaching light beam, while the front wall "runs away"! As shown in Figure 2.3, the net result of this is that the ground-based observer sees the light reach the rear of the train before it reaches the front; these two observers disagree about whether the light does or does not reach both ends of the train simultaneously.

2.3 Time Dilation and Length Contraction

We have seen that the postulates of relativity force the surprising conclusion that time is observer-dependent. We now examine this phenomenon in more detail.

Consider again a train, of height h, with a beam of light bouncing up and down between mirrors on the floor and ceiling, as shown in the first sketch in Figure 2.4. The time between bounces can be interpreted as the "ticks" of a clock, and of course this interval, *as measured on the train*, is independent of whether the train is moving. However, as shown in the second sketch, a stationary observer sees something quite different.

From the ground, the light appears to move diagonally, and hence travels

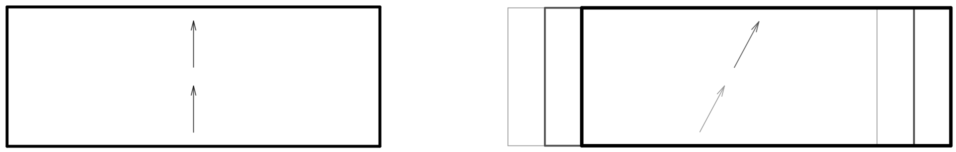


Figure 2.4: A beam of light bounces up and down between mirrors on the floor and ceiling of a moving train. The time between bounces can be used as a unit of time, but a moving observer and a stationary observer obtain different results.

a longer path than the vertical path seen on the train. But since the light must move at the same speed for both observers, each "tick" takes longer according to the ground-based observer than for the observer on the train. Thus, the observer at rest sees the "clock" of the moving observer run slow!

Work through each step of this argument carefully; the key assumption is Postulate II, namely that the light must travel at the same speed for both observers. This is not the behavior we expect from our daily experience!

To compute how the times are related, we must first introduce some notation. Let t denote time as measured on the ground, and t' denote time as measured on the train; we will similarly use x' and x to measure length. One tick of the clock as seen on the train takes time $\Delta t'$, where

$$h = c\,\Delta t' \tag{2.3}$$

Suppose the same tick takes time Δt as seen from the ground. In this time, the light travels a distance $c\,\Delta t$, which is the hypotenuse of a right triangle with legs h and $v\,\Delta t$. The Pythagorean Theorem now leads to

$$(c\,\Delta t)^2 = (v\,\Delta t)^2 + h^2 \tag{2.4}$$

which can be solved for h. Comparing the result with (2.3), we obtain

$$\Delta t = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}\,\Delta t' \tag{2.5}$$

which indeed shows that the moving clock runs slower than the stationary one ($\Delta t > \Delta t'$), at least for speeds $v < c$. The factor relating these times shows up so often that we give it a special name, defining

$$\gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \tag{2.6}$$

Figure 2.5: A beam of light bounces back and forth between mirrors at the front and back of a moving train. The time of the roundtrip journey can be used to measure the length of the train, but a moving observer and a stationary observer obtain different results.

This effect, called *time dilation*, has important consequences for objects traveling at a significant fraction of the speed of light, but has virtually no effect for objects at everyday speeds ($v \ll c$). It is this effect which allows cosmic rays to reach the earth — the particles' lifetimes, as measured by their own clocks, is many orders of magnitude shorter than the time we observe them traveling through the atmosphere.

Time is not the only thing on which observers cannot agree. Consider now a beam of light bouncing horizontally between the front and back of the train, as shown in the first sketch in Figure 2.5. As seen on the train, if it takes time $\Delta t'$ to make a round trip, and the length of the train is $\Delta x'$, then we must have

$$c\,\Delta t' = 2\,\Delta x' \tag{2.7}$$

What does the observer on the ground see? Don't forget that the train is moving, so that, as shown in the second sketch, the distance traveled in one direction is different from that in the other. More precisely, light starting from the back of the train must "chase" the front; if it takes time Δt_1 to catch up, then the distance traveled is $\Delta x + v\,\Delta t_1$, the sum of the length of the train (as seen from the ground!) and the distance the front of the train traveled while the light was under way. Similarly, if the time taken on the return journey is Δt_2, then the distance traveled is $\Delta x - v\,\Delta t_2$. Thus,

$$\begin{align} c\,\Delta t_1 &= \Delta x + v\,\Delta t_1 \tag{2.8} \\ c\,\Delta t_2 &= \Delta x - v\,\Delta t_2 \tag{2.9} \end{align}$$

or equivalently

$$\Delta t_1 = \frac{\Delta x}{c - v} \tag{2.10}$$

$$\Delta t_2 = \frac{\Delta x}{c+v} \tag{2.11}$$

Combining these results leads to

$$c\,\Delta t = c\,\Delta t_1 + c\,\Delta t_2 \tag{2.12}$$
$$= \frac{2\,\Delta x}{1-\frac{v^2}{c^2}} \tag{2.13}$$

and comparing with (2.5) and (2.7) leads to

$$\Delta x = \sqrt{1-\frac{v^2}{c^2}}\,\Delta x' = \frac{1}{\gamma}\,\Delta x' \tag{2.14}$$

Thus, a moving object appears to be shorter in the direction of motion than it would be at rest; this effect is known as *length contraction*.

2.4 Lorentz Transformations

Suppose the frame O' is moving to the right with speed v. If x denotes the distance of an object from O, [1] then the distance between the object and O' *as measured by* O will be $x - vt$. But using the formula for length contraction derived above, namely $\Delta x' = \gamma\,\Delta x$, we see that

$$x' = \gamma\,(x - vt) \tag{2.15}$$

By Postulate I, the framework used by each observer to describe the other must be the same. In particular, if we interchange the roles of O and O', nothing else should change — except for the fact that the relative velocity (of O with respect to O') is now $-v$ instead of v. By symmetry, we therefore have immediately that

$$x = \gamma\,(x' + vt') \tag{2.16}$$

where we have been careful not to assume that $t' = t$. In fact, comparing these equations quickly yields

$$t' = \gamma\left(t - \frac{v}{c^2}x\right) \tag{2.17}$$

and a similar expression for t in terms of x' and t'.

[1] This distance need not be constant; but could be a function of time.

2.5 Addition of Velocities

Suppose that, as seen from O, O' is moving to the right with speed v and that an object is moving to the right with speed u. According to Galileo we would simply add velocities to determine the velocity of the object as seen from O':

$$u = u' + v \tag{2.18}$$

This equation can be derived by differentiating the Galilean transformation

$$x = x' + vt \tag{2.19}$$

thus obtaining

$$\frac{dx}{dt} = \frac{dx'}{dt} + v \tag{2.20}$$

To derive the relativistic formula for the addition of velocities, we proceed similarly. However, since *both* x and t transform, it is useful to use the differential form of the Lorentz transformations, namely

$$dx = d\Big(\gamma\,(x' + vt')\Big) = \gamma\,(dx' + v\,dt') \tag{2.21}$$

$$dt = d\Big(\gamma\,(t' + \frac{v}{c^2}x')\Big) = \gamma\left(dt' + \frac{v}{c^2}dx'\right) \tag{2.22}$$

Dividing these expressions leads to

$$\frac{dx}{dt} = \frac{\frac{dx'}{dt'} + v}{1 + \frac{v}{c^2}\frac{dx'}{dt'}} \tag{2.23}$$

or equivalently

$$u = \frac{u' + v}{1 + \frac{u'v}{c^2}} \tag{2.24}$$

2.6 The Interval

Direct computation using the Lorentz transformations shows that

$$x'^2 - c^2t'^2 = \gamma^2\,(x - vt)^2 - \gamma^2\left(ct - \frac{v}{c}x\right)^2 \tag{2.25}$$

$$= x^2 - c^2t^2 \tag{2.26}$$

so that the quantity $x^2 - c^2t^2$, known as the *interval*, does not depend on the observer who computes it. We will explore this further in later chapters.

Chapter 3

Circle Geometry

In which some standard properties of 2-dimensional Euclidean geometry are reviewed, and some more subtle properties are pointed out.

3.1 Distance

The key concept in Euclidean geometry is the *distance function* that measures the distance between two points. In two dimensions, the (squared!) distance between a point $B = (x, y)$ and the origin is given by

$$d^2 = x^2 + y^2 \tag{3.1}$$

which is of course just the Pythagorean Theorem; see Figure 3.1.

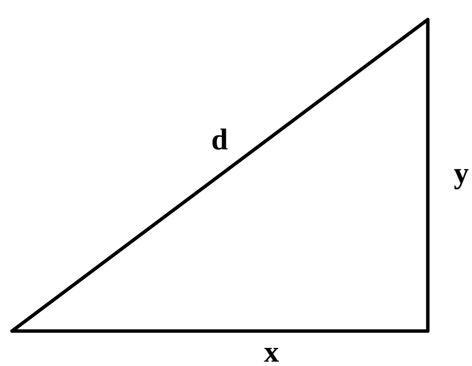


Figure 3.1: Measuring distance in Euclidean geometry using the Pythagorean Theorem.

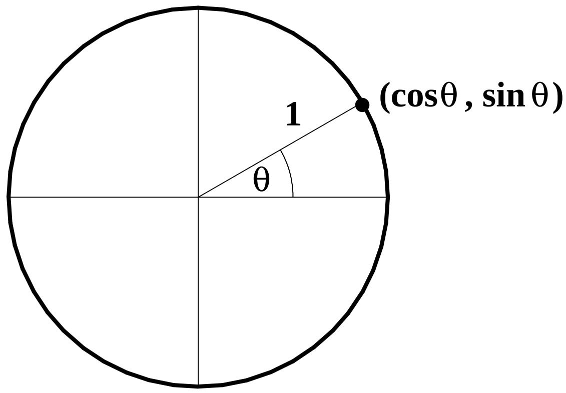


Figure 3.2: Defining the (circular) trig functions via the unit circle.

It is natural to study the set of points that are a constant distance from a given point, which of course form a *circle*.

3.2 Trigonometry

Consider a point P on the unit circle, as shown in Figure 3.2. The angle between the line from the origin to P and the (positive) x-axis is *defined* to be the length of the arc of the unit circle between P and the point $(1,0)$. Denoting the coordinates of P by (x,y), the basic (circular) trig functions are then *defined* by

$$\cos\theta = x \tag{3.2}$$
$$\sin\theta = y \tag{3.3}$$
$$\tan\theta = \frac{\sin\theta}{\cos\theta} \tag{3.4}$$

and the fundamental identity

$$\cos^2\theta + \sin^2\theta = 1 \tag{3.5}$$

then follows from the definition of the unit circle. It is straightforward to verify the addition formulas

$$\sin(\theta+\phi) = \sin\theta\cos\phi + \cos\theta\sin\phi \tag{3.6}$$

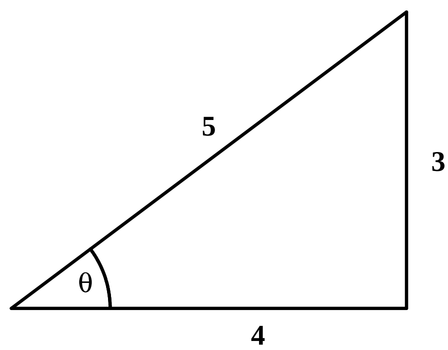


Figure 3.3: A triangle with $\tan\theta = \frac{3}{4}$.

$$\cos(\theta+\phi) = \cos\theta\cos\phi - \sin\theta\sin\phi \tag{3.7}$$

$$\tan(\theta+\phi) = \frac{\tan\theta + \tan\phi}{1 - \tan\theta\tan\phi} \tag{3.8}$$

as well as the derivative formulas

$$\frac{d}{d\theta}\sin\theta = \cos\theta \tag{3.9}$$

$$\frac{d}{d\theta}\cos\theta = -\sin\theta \tag{3.10}$$

3.3 Triangle Trig

An important class of trig problems involve determining, say, $\cos\theta$ if $\tan\theta$ is known. One can of course do this algebraically, using the identity

$$\cos^2\theta = \frac{1}{1+\tan^2\theta} \tag{3.11}$$

But it is often easier to do this geometrically, as illustrated with the following example.

Suppose you know $\tan\theta = \frac{3}{4}$, and you wish to determine $\cos\theta$. Draw *any* triangle containing an angle whose tangent is $\frac{3}{4}$. In this case, the obvious choice would be the triangle shown in Figure 3.3, with sides of 3 and 4. What is $\cos\theta$? The hypotenuse clearly has length 5, so that $\cos\theta = \frac{4}{5}$.

Trigonometry is not merely about ratios of sides, it is also about projections. Another common use of triangle trig is to determine the sides of a triangle given the hypotenuse d and one angle θ. The answer, of course, is that the sides are $d\cos\theta$ and $d\sin\theta$, as shown in in Figure 3.4.

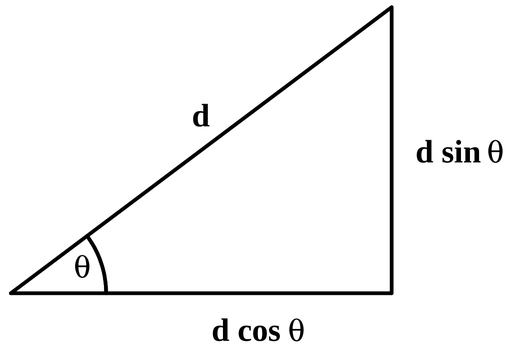


Figure 3.4: A triangle in which the hypotenuse and one angle are known.

3.4 Rotations

Now consider a new set of coordinates, call them (x', y'), based on axes rotated clockwise through an angle θ from the original ones, as shown in Figure 3.5. For instance, the y-axis could point towards true north, whereas the y'-axis might point towards magnetic north. In the primed coordinate system, the point B clearly has coordinates $(1, 0)$, while the point A has coordinates $(0, 1)$, while the unprimed coordinates of B and A are $(\cos\theta, \sin\theta)$ and $(-\sin\theta, \cos\theta)$, respectively. A little work shows that the coordinates are related by a rotation matrix

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} \tag{3.12}$$

(where we are of course assuming that all coordinates are measured in the same units!). Furthermore, both coordinate systems lead to the same result for the (squared) distance, since

$$x^2 + y^2 = x'^2 + y'^2 \tag{3.13}$$

3.5 Projections

Consider the rectangular object of width 1 meter shown in the first sketch in Figure 3.6, which has been rotated so as to be parallel to the *primed* axes. How wide is it? As worded, this question is poorly posed. If width means “extent in the x' direction”, then of course the answer is 1 meter. If, however,

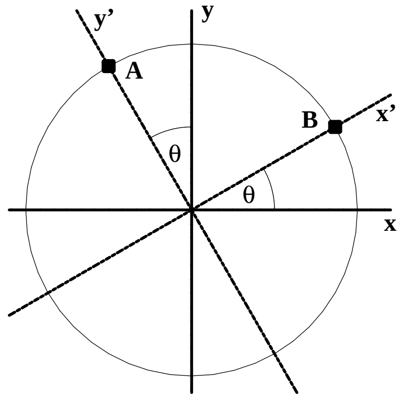


Figure 3.5: A rotated coordinate system.

width means "extent in the x direction", then the answer is obtained by measuring the *horizontal* distance between the sides of the rectangle, which results in a value *larger* than 1. (The exact value is easily seen to be $1/\cos\theta$.)

Repeat this exercise in the opposite direction. Take the same rectangular object, but orient it parallel to the *unprimed* axes, as shown in the second sketch. How wide is it? Clearly the "unprimed" width is 1 meter, and the "primed" width is larger (and again given by $1/\cos\theta$).

In one of the cases above, the "primed" width is smaller, yet in the other the "unprimed" width is smaller. What is happening here? If you turn your suitcase at an angle, it is *harder* to fit under your seat! It has, in effect, become "longer"! But which orientation is best depends, of course, on the orientation of the seat!

Remember this discussion when we address the corresponding questions in relativity in subsequent chapters.

3.6 Addition Formulas

Consider the line through the origin which makes an angle ϕ with the (positive) x-axis, as shown in the first sketch in Figure 3.7. What is its slope? The equation of the line is

$$y = x\ \tan\phi \tag{3.14}$$

so that slope is $\tan\phi$, at least in "unprimed" coordinates. Consider now the line through the origin shown in the second sketch, which makes an angle ϕ

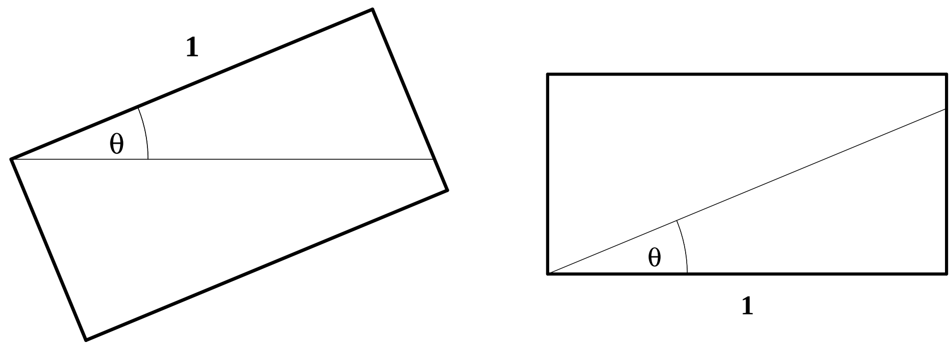


Figure 3.6: Width is coordinate-dependent.

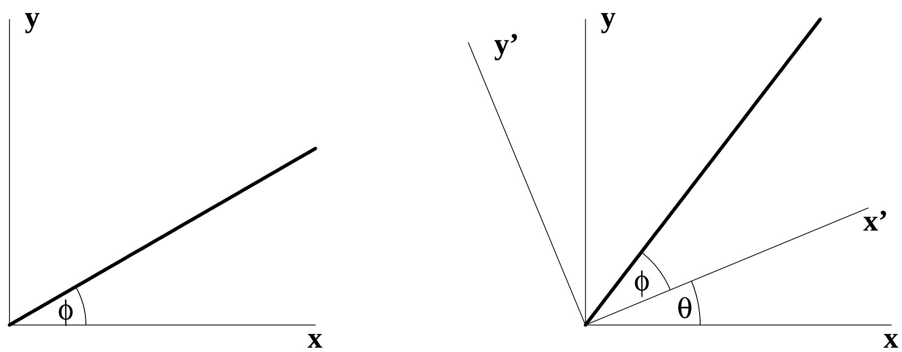


Figure 3.7: The addition formula for slopes.

with the (positive) x'-axis. What is its slope? In "primed" coordinates, the equation of the line is just

$$y' = x' \tan\phi \tag{3.15}$$

so that, in these coordinates, the slope is again $\tan\phi$. But what about in "unprimed" coordinates? The x' axis itself makes an angle θ with the x-axis. It is tempting to simply add these slopes, obtaining $\tan\phi + \tan\theta$, but this not correct. Slopes don't add; angles do! The correct answer is that

$$y = x \tan(\theta + \phi) \tag{3.16}$$

so that the slope is given by (3.8).

Remember this discussion when we discuss the Einstein addition law.

Chapter 4

Hyperbola Geometry

In which a 2-dimensional non-Euclidean geometry is constructed, which will turn out to be identical with special relativity.

4.1 Trigonometry

The hyperbolic trig functions are usually defined using the formulas

$$\cosh\beta = \frac{e^\beta + e^{-\beta}}{2} \tag{4.1}$$

$$\sinh\beta = \frac{e^\beta - e^{-\beta}}{2} \tag{4.2}$$

and then

$$\tanh\beta = \frac{\sinh\beta}{\cosh\beta} \tag{4.3}$$

and so on. We will discuss an alternative definition below. The graphs of these functions are shown in Figure 4.1.

It is straightforward to verify from these definitions that

$$\cosh^2\beta - \sinh^2\beta = 1 \tag{4.4}$$

$$\sinh(\alpha+\beta) = \sinh\alpha\cosh\beta + \cosh\alpha\sinh\beta \tag{4.5}$$

$$\cosh(\alpha+\beta) = \cosh\alpha\cosh\beta + \sinh\alpha\sinh\beta \tag{4.6}$$

$$\tanh(\alpha+\beta) = \frac{\tanh\alpha + \tanh\beta}{1 + \tanh\alpha\tanh\beta} \tag{4.7}$$

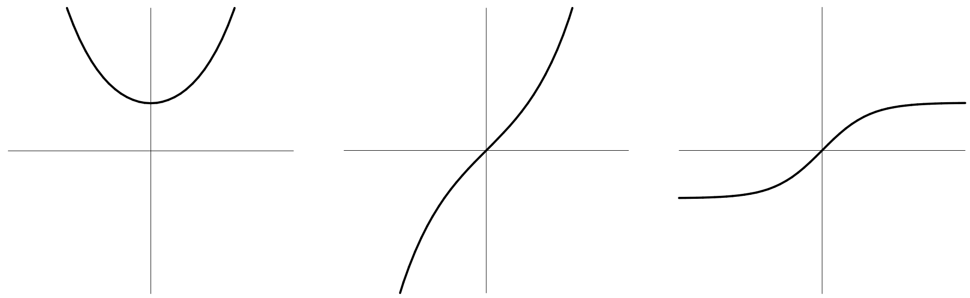


Figure 4.1: The graphs of $\cosh\beta$, $\sinh\beta$, and $\tanh\beta$, respectively.

$$\frac{d}{d\beta}\sinh\beta = \cosh\beta \tag{4.8}$$

$$\frac{d}{d\beta}\cosh\beta = \sinh\beta \tag{4.9}$$

These hyperbolic trig identities look very much like their ordinary trig counterparts (except for signs). This similarity derives from the fact that

$$\cosh\beta \equiv \cos(i\beta) \tag{4.10}$$

$$\sinh\beta \equiv -i\sin(i\beta) \tag{4.11}$$

4.2 Distance

We saw in the last chapter that Euclidean distance is based on the *unit circle*, the set of points which are unit distance from the origin. Hyperbola geometry is obtained simply by using a different distance function! Measure the "squared distance" of a point $B = (x, y)$ from the origin using the definition

$$\delta^2 = x^2 - y^2 \tag{4.12}$$

Then the unit "circle" becomes the unit hyperbola

$$x^2 - y^2 = 1 \tag{4.13}$$

and we further restrict ourselves to the branch with $x > 0$. If B is a point on this hyperbola, then we can *define* the hyperbolic angle β between the line from the origin to B and the (positive) x-axis to be the *Lorentzian* length [1]

[1] No, we haven't defined this. In Euclidean geometry, the the length of a curve is obtained by integrating ds along the curve, where $ds^2 = dx^2 + dy^2$. In a similar way, the Lorentzian length is obtained by integrating $d\sigma$, where $d\sigma^2 = dx^2 - dy^2$.

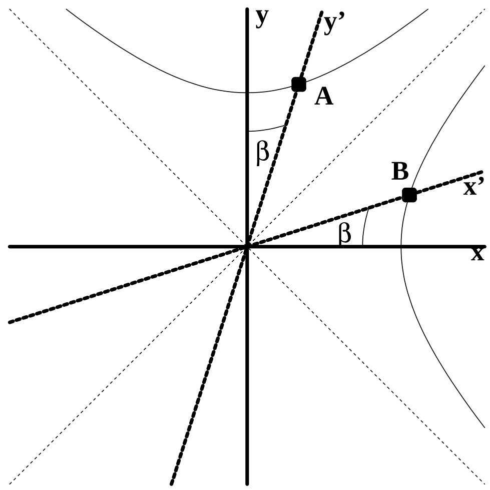


Figure 4.2: The unit hyperbola. The point A has coordinates $(\sinh\beta, \cosh\beta)$, and $B = (\cosh\beta, \sinh\beta)$.

of the arc of the unit hyperbola between B and the point $(1, 0)$. We could then *define* the hyperbolic trig functions to be the coordinates (x, y) of B, that is

$$\cosh\beta = x \tag{4.14}$$
$$\sinh\beta = y \tag{4.15}$$

and a little work shows that this definition is exactly the same as the one above. [2] This construction is shown in Figure 4.2, which also shows another "unit" hyperbola, given by $x^2 - y^2 = -1$. By symmetry, the point A on this hyperbola has coordinates $(x, y) = (\sinh\beta, \cosh\beta)$. We will discuss the importance of this hyerbola later.

Many of the features of the graphs shown in Figure 4.1 follow immediately from this definition of the hyperbolic trig functions in terms of coordinates along the unit hyperbola. Since the minimum value of x on this hyperbola

[2] Use $x^2 - y^2 = 1$ to compute

$$d\beta^2 \equiv d\sigma^2 = dy^2 - dx^2 = \frac{dx^2}{x^2 - 1} = \frac{dy^2}{y^2 + 1}$$

then take the square root of either expression and integrate. (The integrals are hard.) Finally, solve for x or y in terms of β, yielding (4.1) or (4.2), respectively.

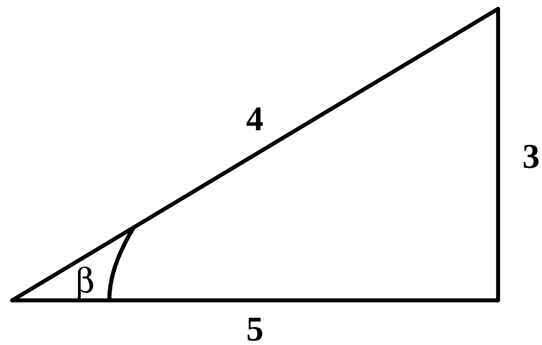


Figure 4.3: A hyperbolic triangle with $\tanh\beta = \frac{3}{5}$.

is 1, we must have $\cosh\beta \geq 1$. As β approaches $\pm\infty$, x approaches ∞ and y approaches $\pm\infty$, which agrees with the asymptotic behavior of the graphs of $\cosh\beta$ and $\sinh\beta$, respectively. Finally, since the hyperbola has asymptotes $y = \pm x$, we see that $|\tanh\beta| < 1$, and that $\tanh\beta$ must approach ± 1 as β approaches $\pm\infty$.

So how do we measure the distance between two points? The "squared distance" was defined in (4.12), and can be positive, negative, or zero! We adopt the following convention: *Take the square root of the absolute value of the "squared distance"*. As we will see in the next chapter, it will also be important to remember whether the "squared distance" was positive or negative, but this corresponds directly to whether the distance is "mostly horizontal" or "mostly vertical".

4.3 Triangle Trig

We now recast ordinary triangle trig into hyperbola geometry.

Suppose you know $\tanh\beta = \frac{3}{5}$, and you wish to determine $\cosh\beta$. One can of course do this algebraically, using the identity

$$\cosh^2\beta = \frac{1}{1 - \tanh^2\beta} \tag{4.16}$$

But it is easier to draw *any* triangle containing an angle whose hyperbolic tangent is $\frac{3}{5}$. In this case, the obvious choice would be the triangle shown in Figure 4.3, with sides of 3 and 5.

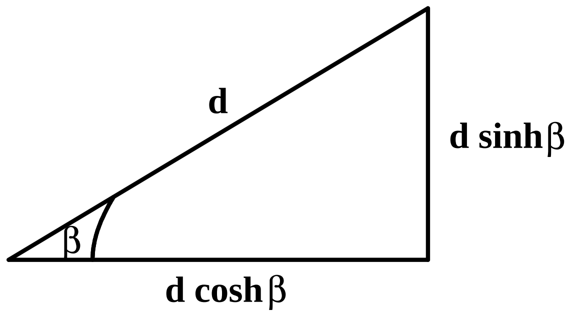


Figure 4.4: A hyperbolic triangle in which the hypotenuse and one angle are known.

What is $\cosh\beta$? Well, we first need to work out the length δ of the hypotenuse. The (hyperbolic) Pythagorean Theorem tells us that

$$5^2 - 3^2 = \delta^2 \tag{4.17}$$

so δ is clearly 4. Take a good look at this 3-4-5 triangle of hyperbola geometry, which is shown in Figure 4.3! But now that we know all the sides of the triangle, it is easy to see that $\cosh\beta = \frac{5}{4}$.

Trigonometry is not merely about ratios of sides, it is also about projections. Another common use of triangle trig is to determine the sides of a triangle given the hypotenuse d and one angle β. The answer, of course, is that the sides are $d\cosh\beta$ and $d\sinh\beta$, as shown in in Figure 4.4.

4.4 Rotations

By analogy with the Euclidean case, we *define* a hyperbolic rotation through the relations

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cosh\beta & \sinh\beta \\ \sinh\beta & \cosh\beta \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} \tag{4.18}$$

This corresponds to "rotating" both the x and y axes into the first quadrant, as shown in Figure 4.2. While this may seem peculiar, it is easily verified that the "distance" is invariant, that is,

$$x^2 - y^2 \equiv x'^2 - y'^2 \tag{4.19}$$

which follows immediately from the hyperbolic trig identity (4.4).

4.5 Projections

We can ask the same question as we did for Euclidean geometry. Consider a rectangle of width 1 whose sides are parallel to the unprimed axes. How wide is it when measured in the primed coordinates? It turns out that the width of the box in the primed coordinate system is *less than* 1. This is length contraction, to which we will return in the next chapter, along with time dilation.

4.6 Addition Formulas

What is the slope of the line from the origin to the point A in Figure 4.2? The equation of this line, the y'-axis, is

$$x = y \tanh \beta \tag{4.20}$$

Consider now a line with equation

$$x' = y' \tanh \alpha \tag{4.21}$$

What is its (unprimed) slope? Again, slopes don't add, but (hyperbolic) angles do; the answer is that

$$x = y \tanh(\alpha + \beta) \tag{4.22}$$

which can be expressed in terms of the slopes $\tanh \alpha$ and $\tanh \beta$ using (4.7). As discussed in more detail in the next chapter, this is the Einstein addition formula!

Chapter 5

The Geometry of Special Relativity

In which it is shown that special relativity is just hyperbolic geometry.

5.1 Spacetime Diagrams

A brilliant aid in understanding special relativity is the *Surveyor's parable* introduced by Taylor and Wheeler [1, 2]. Suppose a town has daytime surveyors, who determine North and East with a compass, nighttime surveyors, who use the North Star. These notions of course differ, since magnetic north is not the direction to the North Pole. Suppose further that both groups measure north/south distances in miles and east/west distances in meters, with both being measured from the town center. How does one go about comparing the measurements of the two groups?

With our knowledge of Euclidean geometry, we see how to do this: Convert miles to meters (or vice versa). Furthermore, distances computed with the Pythagorean theorem do not depend on which group does the surveying. Finally, it is easily seen that "daytime coordinates" can be obtained from "nighttime coordinates" by a simple rotation. The moral of this parable is therefore:

1. *Use the same units.*
2. *The (squared) distance is invariant.*
3. *Different frames are related by rotations.*

Applying that lesson to relativity, the first thing to do is to measure both time and space in the same units. How does one measure distance in seconds? that's easy: simply multiply by c. Thus, since $c = 3 \times 10^8 \ \frac{\mathrm{m}}{\mathrm{s}}$, 1 second of distance is just 3×10^8 m. [1] Note that this has the effect of setting $c = 1$, since the number of seconds (of distance) traveled by light in 1 second (of time) is precisely 1.

Of course, it is also possible to measure time in meters: simply divide by c. Thus, 1 meter of time is the time it takes for light (in vacuum) to travel 1 meter. Again, this has the effect of setting $c = 1$.

5.2 Lorentz Transformations

The Lorentz transformation between a frame (x, t) at rest and a frame (x', t') moving to the right at speed v was derived in Chapter 2. The transformation from the moving frame to the frame at rest is given by

$$x = \gamma\,(x' + vt') \tag{5.1}$$

$$t = \gamma\left(t' + \frac{v}{c^2}x'\right) \tag{5.2}$$

where

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{5.3}$$

The key to converting this to hyperbola geometry is to measure space and time in the same units by replacing t by ct. The transformation from the moving frame, which we now denote (x', ct'), to the frame at rest, now denoted (x, ct), is given by

$$x = \gamma\,(x' + \frac{v}{c}\,ct') \tag{5.4}$$

$$ct = \gamma\left(ct' + \frac{v}{c}\,x'\right) \tag{5.5}$$

which makes the symmetry between these equations much more obvious.

We can simplify things still further. Introduce the *rapidity* β via [2]

$$\frac{v}{c} = \tanh\beta \tag{5.6}$$

[1] A similar unit of distance is the *lightyear*, namely the distance traveled by light in 1 year, which would here be called simply a *year* of distance.

[2] WARNING: Some authors use β for $\frac{v}{c}$, not the rapidity.

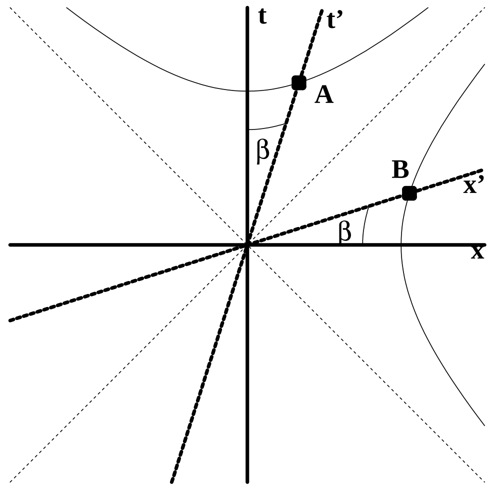


Figure 5.1: The Lorentz transformation between an observer at rest and an observer moving at speed $\frac{v}{c} = c \tanh \beta$ is shown as a hyperbolic rotation. The point A has coordinates $(\sinh \beta, \cosh \beta)$, and $B = (\cosh \beta, \sinh \beta)$. (Units have been chosen such that $c = 1$.)

Inserting this into the expression for γ we obtain

$$\gamma = \frac{1}{\sqrt{1 - \tanh^2 \beta}} = \sqrt{\frac{\cosh^2 \beta}{\cosh^2 \beta - \sinh^2 \beta}} = \cosh \beta \tag{5.7}$$

and

$$\frac{v}{c} \gamma = \tanh \beta \cosh \beta = \sinh \beta \tag{5.8}$$

Inserting these identities into the Lorentz transformations above brings them to the remarkably simple form

$$x = x' \cosh \beta + ct' \sinh \beta \tag{5.9}$$
$$ct = x' \sinh \beta + ct' \cosh \beta \tag{5.10}$$

which in matrix form are just

$$\begin{pmatrix} x \\ ct \end{pmatrix} = \begin{pmatrix} \cosh \beta & \sinh \beta \\ \sinh \beta & \cosh \beta \end{pmatrix} \begin{pmatrix} x' \\ ct' \end{pmatrix} \tag{5.11}$$

But (5.11) is just (4.18), with $y = ct$!

Thus, Lorentz transformations are just hyperbolic rotations! As noted in the previous chapter, the invariance of the interval follows immediately from the fundamental hyperbolic trig identity (4.4). This invariance now takes the form

$$x^2 - c^2t^2 \equiv x'^2 - c^2t'^2 \tag{5.12}$$

We thus have precisely the situation described in Figure 4.2, but with y replaced by ct; this is shown in Figure 5.1.

5.3 Space and Time

We now return to the peculiar fact that the "squared distance" between two points can be positive, negative, or zero. This sign is positive for horizontal distances and negative for vertical distances. But these directions correspond to the coordinates x and t, and measure space and time, respectively — as seen by the given observer. But *any* observer's space axis must intersect the unit hyperbola somewhere, and hence corresponds to positive "squared distance". Such directions have more space than time, and will be called *spacelike*. Similarly, any observer's time axis intersects the hyperbola $x^2 - c^2t^2 = -1$, corresponding to negative "squared distance"; such directions are *timelike*.

What about diagonal lines at a (Euclidean!) angle of 45°? These correspond to a "squared distance" of zero — and to moving at the speed of light. All observers agree about these directions, which will be called *lightlike*. In hyperbola geometry, there are thus preferred directions of "length zero". Indeed, this is the geometric realization of the idea that the speed of light is the same for all observers!

It is important to realize that *every* spacelike direction corresponds to the space axis for *some* observer. Events separated by a spacelike line occur at the simultaneously for that observer — and the (square root of the) "squared distance" is just the distance between the events as seen by that observer. Similarly, events separated by a timelike line occur at the same place for some observer, and the (square root of -1 times the) "squared distance" is just the time which elapses between the events as seen by that observer.

On the other hand, events separated by a timelike line do not occur simultaneously for *any* observer! We can thus divide the spacetime diagram into causal regions as follows: Those points connected to the origin by spacelike lines occur "now" for some observer, whereas those points connected to the

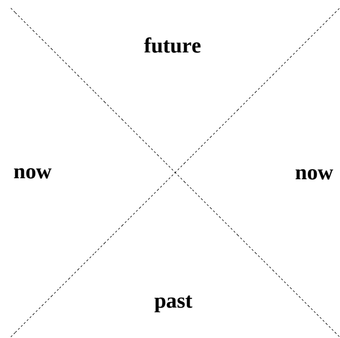


Figure 5.2: The causal relationship between points in spacetime and the origin.

origin by timelike lines occur unambiguously in the future or the past. This is shown in Figure 5.2. [3]

In order to be able to make sense of cause and effect, only events in our past can influence us, and we can only influence events in our future. Put differently, if information could travel faster than the speed of light, then different observers would no longer be able to agree on cause and effect.

5.4 Dot Product

In Euclidean geometry, distances can be described by taking the (squared!) length of a vector using the dot product. Denoting the unit vectors in the x and y directions by $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$, respectively, then the vector from the origin to the point (x, y) is just

$$\vec{\boldsymbol{r}} = x\,\hat{\boldsymbol{x}} + y\,\hat{\boldsymbol{y}} \tag{5.13}$$

whose (squared) length is just

$$|\vec{\boldsymbol{r}}|^2 = \vec{\boldsymbol{r}} \cdot \vec{\boldsymbol{r}} = x^2 + y^2 \tag{5.14}$$

It is straightforward to generalize this to hyperbola geometry. Denote the unit vectors in the t and x directions by $\hat{\boldsymbol{t}}$ and $\hat{\boldsymbol{x}}$. [4] Then the (Lorentzian)

[3] With two or more spatial dimensions, the lightlike directions would form a surface called the *light cone*, and the regions labeled "now" would be connected.

[4] Unit vectors are dimensionless! It is neither necessary nor desirable to include a factor of c in the definition of $\hat{\boldsymbol{t}}$.

dot product can be defined by the requirement that this be an orthonormal basis, in the sense that

$$\begin{aligned} \hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{x}} &= 1 && (5.15)\\ \hat{\boldsymbol{t}} \cdot \hat{\boldsymbol{t}} &= -1 && (5.16)\\ \hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{t}} &= 0 && (5.17) \end{aligned}$$

Any point (x, ct) in spacetime can thus be identified with the vector

$$\vec{\boldsymbol{r}} = x\,\hat{\boldsymbol{x}} + ct\,\hat{\boldsymbol{t}} \tag{5.18}$$

from the origin to that point, whose "squared length" is just the "squared distance" from the origin, namely

$$|\vec{\boldsymbol{r}}|^2 = \vec{\boldsymbol{r}} \cdot \vec{\boldsymbol{r}} = x^2 - c^2t^2 \tag{5.19}$$

One of the fundamental properties of the Euclidean dot product is that

$$\vec{\boldsymbol{u}} \cdot \vec{\boldsymbol{v}} = |\vec{\boldsymbol{u}}||\vec{\boldsymbol{v}}| \cos\theta \tag{5.20}$$

where θ is the (smallest) angle between the directions of $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$. This relationship between the dot product and projections of one vector along another can in fact be used to *define* the dot product. What happens in hyperbola geometry?

First of all, the dot product can be used to define right angles: Two vectors $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$ are said to be *orthogonal* (or *perpendicular*) precisely when their dot product is zero, that is

$$\vec{\boldsymbol{u}} \perp \vec{\boldsymbol{v}} \Longleftrightarrow \vec{\boldsymbol{u}} \cdot \vec{\boldsymbol{v}} = 0 \tag{5.21}$$

We will adopt this definition unchanged in hyperbola geometry.

When are $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$ perpendicular? Assume first that $\vec{\boldsymbol{u}}$ is spacelike. We can assume without loss of generality that $\vec{\boldsymbol{u}}$ is a unit vector, in which case it takes the form

$$\vec{\boldsymbol{u}} = \cosh\alpha\,\hat{\boldsymbol{x}} + \sinh\alpha\,\hat{\boldsymbol{t}} \tag{5.22}$$

What vectors are perpendicular to $\vec{\boldsymbol{u}}$? One such vector is

$$\vec{\boldsymbol{v}} = \sinh\alpha\,\hat{\boldsymbol{x}} + \cosh\alpha\,\hat{\boldsymbol{t}} \tag{5.23}$$

and it is easy to check that all other solutions are multiples of this one. Note that $\vec{\boldsymbol{v}}$ is timelike! Had we assumed instead that $\vec{\boldsymbol{v}}$ were timelike, we would merely have interchanged the roles of $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$.

Furthermore, $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$ are just the space and time axes, respectively, of an observer moving with speed $\frac{v}{c} = \tanh\alpha$. So orthogonal directions correspond precisely to the coordinate axes of some observer.

What if $\vec{\boldsymbol{u}}$ is lightlike? It is a peculiarity of Lorentzian (hyperbola) geometry that there are nonzero vectors of length zero. But since the dot product gives the length, having length zero means that lightlike vectors are perpendicular to themselves!

We can finally define the *length* of a vector $\vec{\boldsymbol{v}}$ by

$$|\vec{\boldsymbol{v}}| = \sqrt{|\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{v}}|} \tag{5.24}$$

If $\vec{\boldsymbol{v}}$ is spacelike we can write

$$\vec{\boldsymbol{v}} = |\vec{\boldsymbol{v}}|(\cosh\alpha\,\hat{\boldsymbol{x}} + \sinh\alpha\,\hat{\boldsymbol{t}}) \tag{5.25}$$

while if $\vec{\boldsymbol{v}}$ is timelike we can write

$$\vec{\boldsymbol{v}} = |\vec{\boldsymbol{v}}|(\sinh\alpha\,\hat{\boldsymbol{x}} + \cosh\alpha\,\hat{\boldsymbol{t}}) \tag{5.26}$$

(If $\vec{\boldsymbol{v}}$ is lightlike, $|\vec{\boldsymbol{v}}| = 0$, so no such expression exists.)

The above argument shows that timelike vectors can only be perpendicular to spacelike vectors, and *vice versa*. We will also say in this case that the vectors form a *right angle*. Recall that hyperbolic angles were defined along the unit hyperbola, hence only exist (as originally defined) between spacelike directions! It is straightforward to extend this to timelike directions using the hyperbola $x^2 - ct^2 = -1$; this was implicitly done when drawing Figure 4.2. But there is no hyperbola relating timelike directions to spacelike ones. Thus, a “right angle” isn’t an angle at all!

A *right triangle* is one which contains a right angle. By the above discussion, one of the legs of such a triangle must be spacelike, and the other timelike. Consider first the case where the hypotenuse is either spacelike or timelike. The only hyperbolic angle in such a triangle is the one between the hypotenuse and the leg of the same type, that is between the two timelike sides if the hypotenuse is timelike, and between the two spacelike sides if the hypotenuse is spacelike. Several such hyperbolic right triangles are shown in Figures 5.3. It is also possible for the hypotenuse to be null, as shown in Figure 5.4. Such triangles do not have any hyperbolic angles!

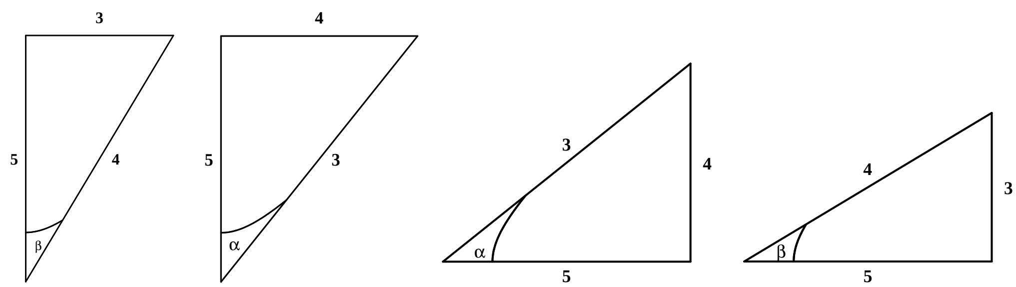


Figure 5.3: Some hyperbolic right triangles.

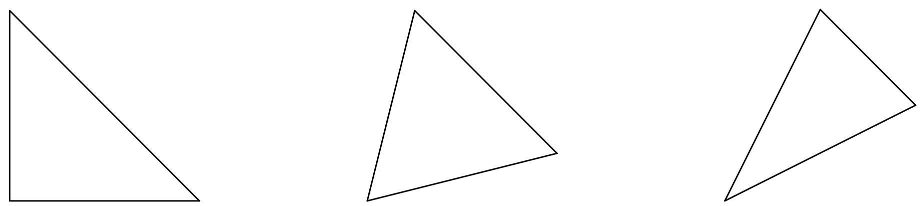


Figure 5.4: More hyperbolic right triangles. The right angle is on the left!

What happens if we take the dot product between two spacelike vectors? We can assume without loss of generality that one vector is parallel to the x axis, in which case we have

$$\vec{\boldsymbol{u}} = |\vec{\boldsymbol{u}}|\,\hat{\boldsymbol{x}} \tag{5.27}$$
$$\vec{\boldsymbol{v}} = |\vec{\boldsymbol{v}}|(\cosh\alpha\,\hat{\boldsymbol{x}} + \sinh\alpha\,\hat{\boldsymbol{t}}) \tag{5.28}$$

so that the dot product satisfies

$$\vec{\boldsymbol{u}}\cdot\vec{\boldsymbol{v}} = |\vec{\boldsymbol{u}}||\vec{\boldsymbol{v}}|\cosh\alpha \tag{5.29}$$

What happens if both vectors are timelike? The above argument still works, except that the roles of $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{t}}$ must be interchanged, resulting in

$$\vec{\boldsymbol{u}}\cdot\vec{\boldsymbol{v}} = -|\vec{\boldsymbol{u}}||\vec{\boldsymbol{v}}|\cosh\alpha \tag{5.30}$$

In both cases, note that $|\vec{\boldsymbol{v}}|\cosh\alpha$ is the projection of $\vec{\boldsymbol{v}}$ along $\vec{\boldsymbol{u}}$; see Figure 5.5.

But what happens if we take the dot product between a timelike vector and a spacelike vector? We can again assume without loss of generality that the spacelike vector is parallel to the x axis, so that

$$\vec{\boldsymbol{u}} = |\vec{\boldsymbol{u}}|\,\hat{\boldsymbol{x}} \tag{5.31}$$
$$\vec{\boldsymbol{v}} = |\vec{\boldsymbol{v}}|(\sinh\alpha\,\hat{\boldsymbol{x}} + \cosh\alpha\,\hat{\boldsymbol{t}}) \tag{5.32}$$

Figure 5.5: Hyperbolic projections between two spacelike vectors, or between two timelike vectors.

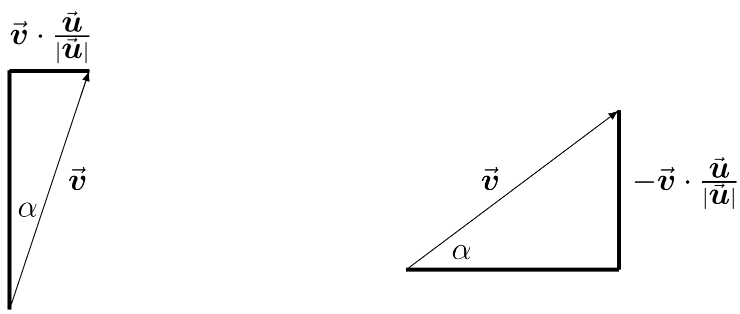


Figure 5.6: Hyperbolic projections between timelike and spacelike vectors.

The dot product now satisfies

$$\vec{\boldsymbol{u}} \cdot \vec{\boldsymbol{v}} = |\vec{\boldsymbol{u}}||\vec{\boldsymbol{v}}| \sinh \alpha \tag{5.33}$$

At first sight, this is something new. But note from the first drawing in Figure 5.6 that $\vec{\boldsymbol{v}} \sinh \alpha$ is just the projection of $\vec{\boldsymbol{v}}$ along $\vec{\boldsymbol{u}}$! The new feature here is that we can't define the angle between a timelike direction and a spacelike direction. The only angle in the triangle which is defined is the one shown! [5]

[5] Alternatively, we could have assumed that the timelike vector was parallel to the t axis, resulting in the second drawing in Figure 5.6. The conclusion is the same, although now it represents the projection of $\vec{\boldsymbol{u}}$ along $\vec{\boldsymbol{v}}$.

Chapter 6

Applications

6.1 Addition of Velocities

What is the rapidity β? Consider an observer moving at speed v to the right. This observer's world line intersects the unit hyperbola

$$c^2t^2 - x^2 = 1 \qquad (ct > 0) \tag{6.1}$$

at the point $A = (\sinh\beta, \cosh\beta)$; this line has "slope" [1]

$$\frac{v}{c} = \tanh\beta \tag{6.2}$$

as required. Thus, β can be thought of as the *hyperbolic angle* between the ct-axis and the worldline of a moving object. As discussed in the preceding chapter, β turns out to be precisely the distance from the axis as measured along the hyperbola (in hyperbola geometry!). This was illustrated in Figure 5.1.

Consider therefore an object moving at speed u relative to an observer moving at speed v. Their rapidities are given by

$$\frac{u}{c} = \tanh\alpha \tag{6.3}$$

$$\frac{v}{c} = \tanh\beta \tag{6.4}$$

[1] It is not obvious whether "slope" should be defined by $\frac{\Delta x}{c\,\Delta t}$ or by the reciprocal of this expression. This is further complicated by the fact that both (x, ct) and (ct, x) are commonly used to denote the coordinates of the point A!

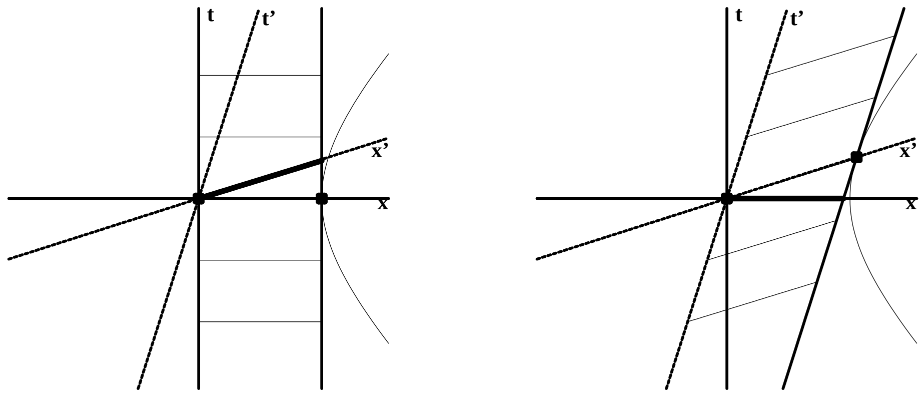


Figure 6.1: Length contraction as a hyperbolic projection.

To determine the resulting speed with respect to an observer at rest, simply add the *rapidities*! One way to think of this is that you are adding the arc lengths along the hyperbola. Another is that you are following a (hyperbolic) rotation through a (hyperbolic) angle β (to get to the moving observer's frame) with a rotation through an angle α. In any case, the resulting speed w is given by

$$\frac{w}{c} = \tanh(\alpha + \beta) = \frac{\tanh\alpha + \tanh\beta}{1 + \tanh\alpha\tanh\beta} = \frac{\frac{u}{c} + \frac{v}{c}}{1 + \frac{uv}{c^2}} \tag{6.5}$$

which is — finally — precisely the Einstein addition formula!

6.2 Length Contraction

We now return to the question of how "wide" things are.

Consider first a meter stick at rest. In spacetime, the stick "moves" vertically, that is, it ages. This situation is shown in the first sketch in Figure 6.1, where the horizontal lines show the meter stick at various times (according to an observer at rest). How "wide" is the *worldsheet* of the stick? The observer at rest of course measures the length of the stick by locating both ends *at the same time*, and measuring the distance between them. At $t = 0$, this corresponds to the 2 heavy dots in the sketch, one at the origin and the other on the unit hyperbola. But *all* points on the unit hyperbola are at an interval of 1 meter from the origin. The observer at rest therefore concludes, unsurprisingly, that the meter stick is 1 meter long.

How long does a moving observer think the stick is? This is just the "width" of the worldsheet *as measured by the moving observer*. This observer follows the same procedure, by locating both ends of the stick *at the same time*, and measuring the distance between them. But time now corresponds to t', not t. At $t' = 0$, this measurement corresponds to the heavy line in the sketch. Since this line fails to reach the unit hyperbola, it is clear that the moving observer measures the length of a stationary meter stick to be less than 1 meter. This is length contraction.

To determine the exact value measured by the moving observer, compute the intersection of the line $x = 1$ (the right-hand edge of the meter stick) with the line $t' = 0$ (the x'-axis), or equivalently $ct = x \tanh\beta$, to find that

$$ct = \tanh\beta \tag{6.6}$$

so that x' is just the interval from this point to the origin, which is

$$x' = \sqrt{x^2 - c^2t^2} = \sqrt{1 - \tanh^2\beta} = \frac{1}{\cosh\beta} \tag{6.7}$$

What if the stick is moving and the observer is at rest? This situation is shown in the second sketch in Figure 6.1. The worldsheet now corresponds to a "rotated rectangle", indicated by the parallelograms in the sketch. The fact that the meter stick is 1 meter long in the moving frame is shown by the distance between the 2 heavy dots (along $t' = 0$), and the measurement by the observer at rest is indicated by the heavy line (along $t = 0$). Again, it is clear that the stick appears to have shrunk, since the heavy line fails to reach the unit hyperbola.

Thus, a moving object appears shorter by a factor $1/\cosh\beta$. It doesn't matter whether the stick is moving, or the observer; all that matters is their relative motion.

6.3 Time Dilation

We now investigate moving clocks. Consider first the smaller dot in Figure 6.2. This corresponds to $ct = 1$ (and $x = 0$), as evidenced by the fact that this point is on the (other) unit hyperbola, as shown. Similarly, the larger dot, lying on the same hyperbola, corresponds to $ct' = 1$ (and $x' = 0$). The horizontal line emanating from this dot gives the value of ct there, which

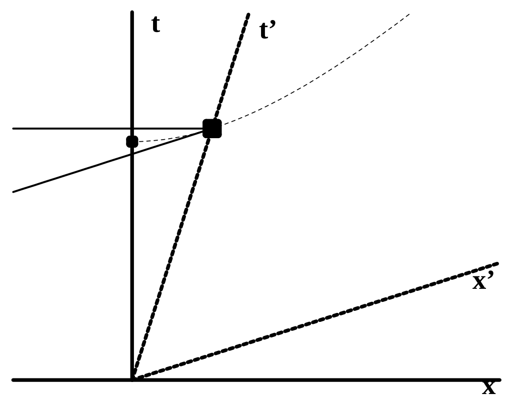


Figure 6.2: Time dilation as a hyperbolic projection.

is clearly greater than 1. This is the time measured by the observer at rest when the moving clock says 1; the moving clock therefore runs slow. But now consider the diagonal line emanating from the larger dot. At all points along this line, $ct' = 1$. In particular, at the smaller dot we must have $ct' > 1$. This is the time measured by the moving observer when the clock at rest says 1; the moving observer therefore concludes the clock at rest runs slow!

There is no contradiction here; one must simply be careful to ask the right question. In each case, observing a clock in another frame of reference corresponds to a projection. In each case, a clock in relative motion to the observer appears to run slow.

6.4 Doppler Shift

The frequency f of a beam of light is related to its wavelength λ by the formula

$$f\lambda = c \tag{6.8}$$

How do these quantities depend on the observer?

Consider an inertial observer moving to the right in the laboratory frame who is carrying a flashlight that is pointing to the left; see Figure 6.3. Then the moving observer is traveling along a path of the form $x' = x'_1 = \text{const}$. Suppose the moving observer turns on the flashlight (at time t'_1) just long enough to emit 1 complete wavelength of light, and that this takes time dt'.

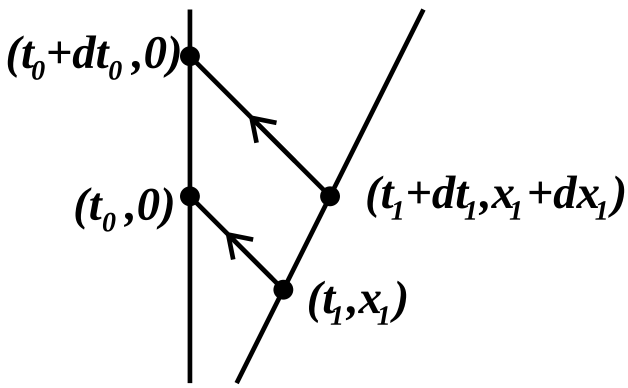


Figure 6.3: The Doppler effect: An observer moving to the right emits a pulse of light to the left, which is later seen by a stationary observer. The wavelengths measured by the two observers differ, causing a *Doppler shift* in the frequency.

Then the moving observer "sees" a wavelength

$$\lambda' = c\,dt' \tag{6.9}$$

According to the lab, the flashlight was turned on at the event (t_1, x_1), and turned off dt_1 seconds later, during which time the moving observer moved a distance dx_1 meters to the right. But when was the light received, at $x = 0$, say?

Let $(t_0, 0)$ denote the first reception of light by a lab observer at $x = 0$, and suppose this observer sees the light stay on for dt_0 seconds. Since light travels at the speed of light, we have the equations

$$\begin{aligned} c(t_0 - t_1) &= x_1 && (6.10) \\ c[(t_0 + dt_0) - (t_1 + dt_1)] &= x_1 + dx_1 && (6.11) \end{aligned}$$

from which it follows that

$$c(dt_0 - dt_1) = dx_1 \tag{6.12}$$

so that

$$\begin{aligned} c\,dt_0 &= dx_1 + c\,dt_1 && (6.13) \\ &= (dx_1' \cosh\beta + c\,dt_1' \sinh\beta) + (c\,dt_1' \cosh\beta + dx_1' \sinh\beta) && (6.14) \\ &= (\cosh\beta + \sinh\beta)\,c\,dt_1' && (6.15) \end{aligned}$$

since $dx_1' = 0$. But the wavelength as seen in the lab is

$$\lambda = c\,dt_0 \tag{6.16}$$

so that

$$\begin{aligned}\frac{\lambda}{\lambda'} &= \frac{dt_0}{dt_1'} = \cosh\beta + \sinh\beta \\ &= \cosh\beta\,(1 + \tanh\beta) = \gamma\left(1 + \frac{v}{c}\right) = \sqrt{\frac{1 + \frac{v}{c}}{1 - \frac{v}{c}}}\end{aligned} \tag{6.17}$$

The frequencies transform inversely, that is

$$\frac{f'}{f} = \sqrt{\frac{1 + \frac{v}{c}}{1 - \frac{v}{c}}} \tag{6.18}$$

Chapter 7

Problems I

7.1 Cosmic Rays

Consider μ-mesons produced by the collision of cosmic rays with gas nuclei in the atmosphere 60 kilometers above the surface of the earth, which then move vertically downward at nearly the speed of light. The half-life before μ-mesons decay into other particles is 1.5 microseconds (1.5×10^{-6} s).

1. *Assuming it doesn't decay, how long would it take a μ-meson to reach the surface of the earth?*
2. *Assuming there were no time dilation, approximately what fraction of the mesons would reach the earth without decaying?*
3. *In actual fact, roughly $\frac{1}{8}$ of the mesons would reach the earth! How fast are they going?*

1. Without much loss of accuracy, assume the mesons travel at the speed of light. Then it takes them

$$\frac{60 \text{ km}}{3 \times 10^{8} \ \frac{\text{m}}{\text{s}}} = 200 \ \mu\text{s} \tag{7.1}$$

2. 200 μs is $\dfrac{200}{1.5} = \dfrac{400}{3}$ half-lives, so only $2^{-\frac{400}{3}}$ of the mesons reach the earth!

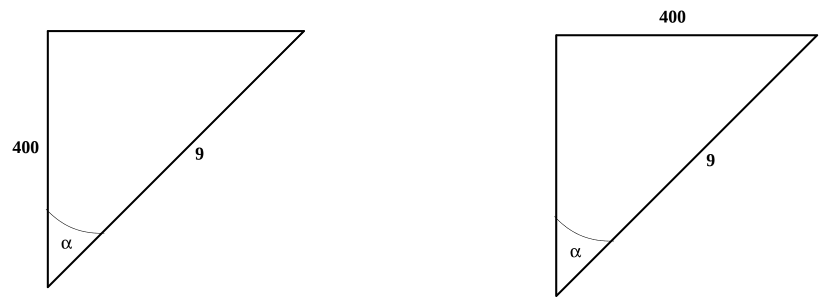


Figure 7.1: Hyperbolic triangles for the cosmic ray example.

3. *First solution:* This corresponds to 3 half-lives. Thus, the time is dilated by a factor of $\frac{400/3}{3}$, so that

$$\cosh\alpha = \frac{400}{9} \tag{7.2}$$

But

$$\frac{v}{c} = \tanh\alpha = \frac{\sqrt{400^2 - 9^2}}{400} \approx .99974684 \tag{7.3}$$

See the first drawing in Figure 7.1.

Second solution: The above argument assumes $v \approx c$! (This was used to obtain the figure 200 μs in part (a), which is hence only an approximation.)

A more accurate argument would use the fact that the mesons travel 60 km in 4.5×10^{-6} s (of proper time). Thus,

$$\sinh\alpha = \frac{(60\text{ km})(1000\ \frac{\text{m}}{\text{km}})}{(4.5\times 10^{-6}\text{ s})(3\times 10^{8}\ \frac{\text{m}}{\text{s}})} = \frac{400}{9} \tag{7.4}$$

so that

$$\frac{v}{c} = \tanh\alpha = \frac{400}{\sqrt{400^2 + 9^2}} \approx .99974697 \tag{7.5}$$

See the second drawing in Figure 7.1.

It is important to realize not only that the second answer is more accurate than the first (assuming sufficient accuracy in the original data!), but also that that the "shortcut" used in the first answer is justified! [1]

[1] This can be made rigorous using a power series expansion. Equivalently, by simply redoing the computation using the approximation obtained in part (c) to recalculate part (a), the correct answer can be obtained to any desired accuracy. The speed at which this iterative procedure converges to the exact answer justifies having made the approximation in the first place.

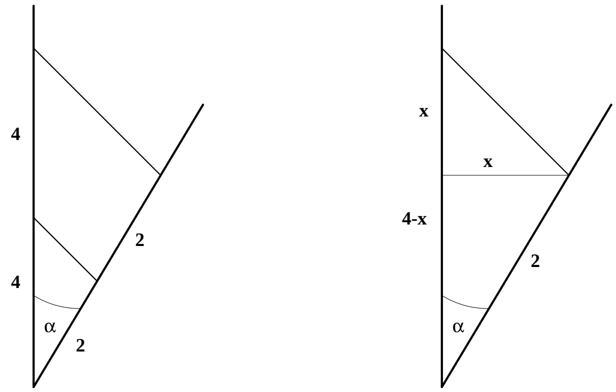


Figure 7.2: Computing Doppler shift.

7.2 Doppler Effect

1. *A rocket sends out flashes of light every 2 seconds in its own rest frame, which you receive every 4 seconds. How fast is the rocket going?*

1. *First solution:* This situation is shown in the first drawing in Figure 7.2. In order to find the hyperbolic angle α, draw a horizontal line as shown in the enlarged second drawing, resulting in the system of equations [2]

$$\tanh\alpha = \frac{x}{4-x} \tag{7.7}$$

$$(4-x)^2 - x^2 = 2^2 \tag{7.8}$$

which is easily solved for $x = \frac{3}{2}$, so that $\frac{v}{c} = \tanh\alpha = \frac{3}{5}$.

Second solution: Insert $\lambda = 4$ and $\lambda' = 2$ into (6.17), and solve for $\frac{v}{c}$.

[2] This method can be be used to derive the Doppler shift formula in general, yielding

$$\frac{v}{c} = \frac{\lambda^2 - \lambda'^2}{\lambda^2 + \lambda'^2} \tag{7.6}$$

which is equivalent to (6.17); in this example, $\lambda = 4$ and $\lambda' = 2$.

Chapter 8

Paradoxes

In which impossible things are shown to be possible.

8.1 Special Relativity Paradoxes

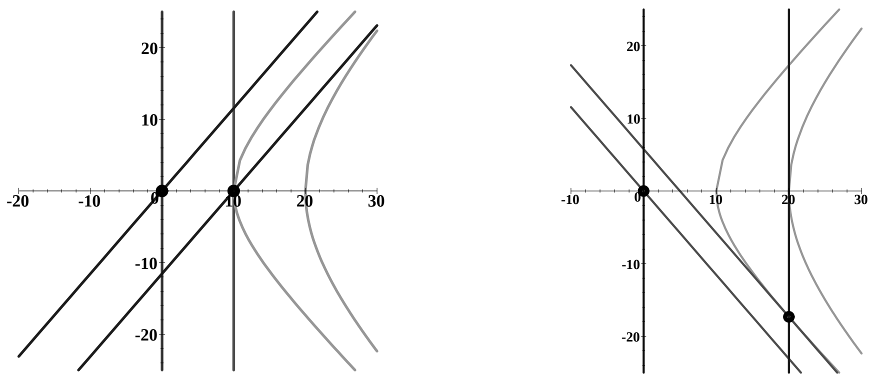
It is easy to create seemingly impossible scenarios in special relativity by playing on the counterintuitive nature of observer-dependent time. These scenarios are usually called paradoxes, because they seem to be impossible. Yet there is nothing paradoxical about them!

The best way to resolve these so-called paradoxes is to draw a good spacetime diagram. This requires careful reading of the problem, making sure always to associate the given information with a particular reference frame. A single spacetime diagram suffices to determine what *all* observers see. It is nevertheless instructive to draw separate spacetime diagrams for each observer, making sure that they all agree.

In this chapter, we discuss the two most famous paradoxes in special relativity, the Pole and Barn Paradox and the Twin Paradox. We also briefly discuss the more subtle aspects of flying manhole covers.

8.2 The Pole and Barn Paradox

A 20 foot pole is moving towards a 10 foot barn fast enough that the pole appears to be only 10 feet long. As soon as both ends of the pole are in the barn, slam the doors. How can a 20 foot pole fit into a 10 foot barn?



BARN **POLE**

Figure 8.1: In each diagram, the heavy straight lines represent the ends of the pole and the lighter straight lines represent the front and back of the barn. The hyperbolas of "radius" 10 and 20 are also shown. The dot at the origin labels the event where the *back* of the pole *enters* the barn, and the other dot labels the event where the *front* of the pole *leaves* the barn.

This is the beginning of the Pole and Barn Paradox. It's bad enough trying to imagine what happens to the pole when it suddenly stops and finds itself in a barn which is too small. But what does the pole see?

Since length contraction is symmetric, if the barn sees the pole shortened by a factor of 2, then the pole sees the barn shortened by the same factor of 2. This factor is just $\cosh\beta$, where $|\beta|$ is the same in both cases. So the 20 foot pole sees this 5 foot barn approaching. No way is the pole going to fit in the barn!

So what does the pole see?

The spacetime diagrams for this situation are shown in Figure 8.1, as seen first from the barn's reference frame and then from the pole's reference frame. The 2 dots represent the events of closing the barn doors when the ends of the pole are even with the corresponding door.

In the *barn's* frame, these events happen simultaneously, so it is possible in principle to trap the pole in the barn by shutting the doors "at the same time". (We omit speculation about what happens when the pole hits the closed door!)

In the *pole's* frame, the exit door is closed long before the rear of the pole enters the barn. Assuming the pole keeps going, for instance by virtue of the door opening again, then the entrance door is closed much later, when the rear of the pole finally gets there. The pole thinks it silly to try to catch it by waiting to close the entrance door until most of the pole has already escaped through the exit door!

8.3 The Twin Paradox

One twin travels 24 light-years to star X at speed $\frac{24}{25}c$; her twin brother stays home. When the traveling twin gets to star X, she immediately turns around, and returns at the same speed. How long does each twin think the trip took?

Star X is 24 light-years away, so, according to the twin at home, it takes her 25 years to get there, and 25 more to return, for a total of 50 years away from earth. But the traveling twin's clock runs slow by a actor of

$$\cosh\beta = \frac{1}{1-\tanh^2\beta} = \frac{25}{7}$$

This means that, according to the traveling twin, it only takes her 7 years each way. Thus, she has only aged 14 years while her brother has aged 50!

This is in fact correct, and represents a sort of time travel into the future: It only takes the traveling twin 14 years to get 50 years into earth's future. (Unfortunately, there's no way to get back!)

But wait a minute. The traveling twin should see her brother's clock run slow, by the same factor of $\frac{25}{7}$. So when 7 years of her time elapse, she thinks her brother has only aged $\frac{49}{25} \approx 2$ years! Her brother should therefore only have aged 4 years when she returns!

This can't be right. Either her brother is 4 years older, or he is 50 years older. Both siblings must surely agree on that!

The easiest way to resolve this is to draw a single spacetime diagram showing the entire trip in the reference frame of the stay-at-home twin, as shown in Figure 8.2. The lower half of the figure is a hyperbolic triangle with $\tanh\beta = \frac{24}{25}$, and hypotenuse 7 years. The remaining diagonal lines are lines of constant time for the traveling twin, at the point of turnaround, while going and returning, respectively. There is another right trangle containing the hyperbolic angle β; the right angle is at the point of turnaround, and the hypotenuse is $\frac{7}{\cosh\beta} = \frac{49}{25}$ years, the age of the stay-at-home twin "when" the traveling twin turns around.

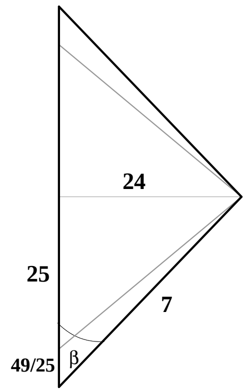


Figure 8.2: The spacetime diagram for the Twin Paradox. The vertical line represents the world line of the stay-at-home twin, while the heavy diagonal lines show the world line of the traveling twin.

How much does each twin age? Simply measure the length of their world lines! Intervals are invariant; it doesn't matter how you compute them. The clear answer is that the brother has indeed aged 50 years, while the sister has only aged 14 years.

So what was wrong with the argument that the brother should have only aged 4 years? There are really 3 reference frames here: earth, going, and returning. The "going" and "returning" frames yield different times on earth for the turnaround — and these times differ by precisely $\frac{1152}{25} \approx 46$ years! From this point of view, it takes the traveling twin 46 years to turn around!

One difference between the twins is that the traveling twin is *not* in an inertial frame — she is in 2 inertial frames, but must accelerate in order to switch from one to the other. This breaks the symmetry between the 2 twins.

However, this is not really the best way to explain this paradox. It is possible to remove this particular asymmetry by assuming the universe is closed, so that the traveling twin doesn't need to turn around! A simplified version of this is to put the problem on a cylinder [4]. It turns out that this introduces another sort of asymmetry, but there is a simpler way to look at it.

The amount an observer ages is just the timelike interval measured along his or her worldline. We used this argument above. This approach also works for curved worldlines, corresponding to noninertial observers — except that one must integrate the infinitesimal timelike interval $d\tau = \sqrt{dt^2 - ds^2}$ along the worldline.

A little thought leads to the following remarkable result: *The timelike line connecting 2 events (assuming there is one) is the* ***longest*** *path from one to the other.* (Think about it. Use this line as the t axis. Then any other path has a nonzero contribution from the change in x — which *decreases* the hyperbolic length of the path, and hence the time taken.)

8.4 Manhole Covers

There are 2 well-known paradoxes involving manhole covers, which illustrate some unexpected implications of special relativity. In the first, a 2 foot manhole cover approaches a 2 foot manhole at relativistic velocity. Since the hole sees the cover as much smaller than 2 feet long, the cover must fall into the manhole. It does.

But what does the cover see? It sees a very small hole rushing at it. No way is this enormous manhole cover going to fit into this small hole!

The resolution of this paradox requires careful consideration of what it means for the something to begin to fall, and is left to the reader.

There is also a higher-dimensional version of this problem, without the complication of falling, that is, without gravity. Suppose the manhole cover is flying to the right as before, but now the hole is in a metal sheet which is rising up to meet it. Again, from the point of view of the hole, the cover is very small and so — if the timing is right — the cover will pass through the hole. It does.

But what does the cover see? It again sees a very small hole rushing at it. How do you get a big object through a small hole? This time one must consider what it means for the cover to "pass through" the whole.

If you have successfully resolved these 2 paradoxes, you will realize that some properties of materials which we take for granted will are quite impossible in special relativity!

Chapter 9

Relativistic Mechanics

In which it is shown that mass is energy.

9.1 Proper Time

In the rest frame, position doesn't change. Let τ denote "wristwatch time" [2], that is, time as measured by a clock carried by an observer moving at constant speed u with respect to the given frame. In the moving observer's rest frame, position doesn't change. We therefore have

$$(\Delta x)^2 - c^2(\Delta t)^2 = 0 - c^2(\Delta\tau)^2 \tag{9.1}$$

so that

$$(\Delta\tau)^2 = \left(1 - \frac{1}{c^2}\left(\frac{\Delta x}{\Delta t}\right)^2\right)(\Delta t)^2 \tag{9.2}$$

or equivalently

$$d\tau = \sqrt{1 - \frac{u}{c^2}}\,dt = \frac{1}{\gamma}\,dt = \frac{1}{\cosh\alpha}\,dt \tag{9.3}$$

Note that proper time is independent of reference frame!

9.2 Energy and Momentum

Consider the *ordinary velocity* of a moving object, defined by

$$u = \frac{d}{dt}\,x \tag{9.4}$$

This transforms in a complicated way, since

$$\frac{1}{c}\frac{dx'}{dt'} = \frac{\frac{1}{c}\frac{dx}{dt} - \frac{v}{c}}{1 - \frac{v}{c^2}\frac{dx}{dt}} \tag{9.5}$$

The reason for this is that both the numerator and the denominator need to be transformed. The invariance of proper time suggests that we should instead differentiate with respect to proper time, since of course

$$\frac{d}{d\tau}\, x' = \frac{dx}{d\tau} \tag{9.6}$$

or in other words since the operator $\frac{d}{d\tau}$ pulls through the Lorentz transformation, so that only the numerator is transformed when changing reference frames.

Furthermore, the same argument can be applied to t, which suggests that there are (in 2 dimensions) 2 components to the velocity. We therefore consider the "2-velocity"

$$\boldsymbol{u} = \frac{d}{d\tau}\begin{pmatrix} ct \\ x \end{pmatrix} = \begin{pmatrix} c\frac{dt}{d\tau} \\ \frac{dx}{d\tau} \end{pmatrix} \tag{9.7}$$

But since

$$dt = \cosh\alpha\, d\tau \tag{9.8}$$

and

$$dx^2 - c^2dt^2 = -c^2d\tau^2 \tag{9.9}$$

we also have

$$dx = c\,\sinh\alpha\, d\tau \tag{9.10}$$

so that

$$\boldsymbol{u} = c\begin{pmatrix} \cosh\alpha \\ \sinh\alpha \end{pmatrix} \tag{9.11}$$

Note that $\frac{1}{c}\,\boldsymbol{u}$ is a *unit* vector, that is

$$\frac{1}{c^2}\,\boldsymbol{u}\cdot\boldsymbol{u} = 1 \tag{9.12}$$

and further that

$$\frac{u}{c} = \frac{dx}{dt} = \tanh\alpha \tag{9.13}$$

as expected.

9.3 Conservation Laws

Suppose that (Newtonian) momentum is conserved in a given frame, that is

$$\sum m_i v_i = \sum \hat{m}_j \hat{v}_j \tag{9.14}$$

(Both of these would be 0 in the center-of-mass frame.) Changing to another frame moving with respect to the first at speed v, we have

$$v_i = v_i' + v \tag{9.15}$$
$$\hat{v}_j = \hat{v}_j' + v \tag{9.16}$$

so that

$$\sum m_i(v_i' + v) = \sum \hat{m}_j(\hat{v}_j' + v) \tag{9.17}$$

We therefore see that

$$\sum m_i v_i' = \sum \hat{m}_j \hat{v}_j' \iff \sum m_i = \sum \hat{m}_j \tag{9.18}$$

that is, momentum is conserved in *all* inertial frames provided it is conserved on one frame *and* mass is conserved.

Repeating the computation for the kinetic energy, we obtain starting from

$$\frac{1}{2}\sum m_i v_i^2 = \frac{1}{2}\sum \hat{m}_j \hat{v}_j^2 \tag{9.19}$$

that

$$\frac{1}{2}\sum m_i(v_i' + v)^2 = \frac{1}{2}\sum \hat{m}_j(\hat{v}_j' + v)^2 \tag{9.20}$$

Expanding this out, we discover that (kinetic) energy is conserved in all frames provided it is conserved in one frame *and* both mass and momentum are conserved.

The situation in special relativity is quite different.

Consider first the momentum defined by the ordinary velocity, namely

$$p = mu = m\frac{dx}{dt} \tag{9.21}$$

This momentum is *not* conserved!

We use instead the momentum defined by the 4-velocity, which is given by

$$p = m\frac{dx}{d\tau} = mc\sinh\alpha \tag{9.22}$$

Suppose now that, as seen in a particular inertial frame, the total momentum of a collection of particles is the same before and after some interaction, that is

$$\sum m_i c \sinh \alpha_i = \sum \hat{m}_j c \sinh \hat{\alpha}_j \tag{9.23}$$

Consider now the same situation as seen by another inertial reference frame, moving with respect to the first with speed

$$v = c \tanh \beta \tag{9.24}$$

We therefore have

$$\alpha_i = \alpha_i' + \beta \tag{9.25}$$

$$\hat{\alpha}_j = \hat{\alpha}_j' + \beta \tag{9.26}$$

Inserting this into the conservation rule (9.23) leads to

$$\sum m_i c \sinh \alpha_i' = \sum m_i c \sinh(\alpha_i - \beta) \tag{9.27}$$

$$= \left(\sum m_i c \sinh \alpha_i\right) \cosh \beta - \left(\sum m_i c \cosh \alpha_i\right) \sinh \beta \tag{9.28}$$

and similarly

$$\sum \hat{m}_j c \sinh \hat{\alpha}_j' = \left(\sum \hat{m}_j c \sinh \hat{\alpha}_j\right) \cosh \beta - \left(\sum \hat{m}_j c \cosh \hat{\alpha}_j\right) \sinh \beta \tag{9.29}$$

The coefficients of $\cosh \beta$ in these 2 expressions are equal due to the assumed conservation of momentum in the original frame. We therefore see that conservation of momentum will hold in the new frame if and only if we have in addition that the coefficients of $\sinh \beta$ agree, namely

$$\sum m_i c \cosh \alpha_i = \sum \hat{m}_j c \cosh \hat{\alpha}_j \tag{9.30}$$

But what is this?

9.4 Energy

This mystery is resolved by recalling that momentum is mass times velocity, and that there is also a "t-component" to the velocity. In analogy with the 2-velocity, we therefore define the "2-momentum" to be

$$\boldsymbol{p} = m \begin{pmatrix} c\frac{dt}{d\tau} \\ \frac{dx}{d\tau} \end{pmatrix} = mc \begin{pmatrix} \cosh \alpha \\ \sinh \alpha \end{pmatrix} \tag{9.31}$$

The second term is clearly the momentum, which we denote by p, but what is the first term? If the object is at rest, $\alpha = 0$, and the first term is therefore just mc. But Einstein's famous equation

$$E = mc^2 \tag{9.32}$$

leads us to suspect that this is some sort of energy. In fact, mc^2 is called the *rest energy* or *rest mass*.

In general, we *define* the energy of an object moving at speed $u = c \tanh \alpha$ to be the first component of $\boldsymbol{p}$, that is we define

$$E \;:=\; mc^2 \cosh \alpha \tag{9.33}$$

$$p \;:=\; mc \sinh \alpha \tag{9.34}$$

or equivalently

$$\boldsymbol{p} = \begin{pmatrix} \frac{1}{c}E \\ p \end{pmatrix} \tag{9.35}$$

Is this definition reasonable? Consider the case $\frac{u}{c} \ll 1$. Then

$$E = mc^2 \cosh \alpha = mc^2 \gamma \tag{9.36}$$

$$= \frac{mc^2}{\sqrt{1 - \frac{u^2}{c^2}}} \tag{9.37}$$

$$\approx mc^2 + \frac{1}{2} mu^2 + \frac{3}{8} m \frac{u^4}{c^2} + ... \tag{9.38}$$

The first term is the rest energy, the next term is the Newtonian kinetic energy, and the remaining terms are relativistic corrections to the kinetic energy.

The moral is that conservation of 2-momentum is equivalent to both conservation of momentum and conservation of energy, but that there is no requirement that the total mass be conserved.

Taking the (squared) norm of the 2-momentum, we obtain

$$-c^2 \boldsymbol{p} \cdot \boldsymbol{p} = E^2 - p^2 c^2 = m^2 c^4 \tag{9.39}$$

Note that this equation continues to makes sense if $m = 0$, although the expressions for E and p separately in terms of α or γ do not. In fact, γ must approach ∞, or equivalently $\frac{u^2}{c^2} = 1$, so that $|u| = c$; such particles *always* move at the speed of light!

We therefore conclude that there can be massless particles, which move at the speed of light, and which satisfy ($m = 0$ and)

$$E = |p|c \neq 0 \tag{9.40}$$

Photons are examples of such particles; quantum mechanically, one has $E = \hbar\nu$, where ν is the frequency of the light (and $\hbar = \frac{h}{2\pi}$ where h is Planck's constant.)

9.5 Useful Formulas

The key formulas for analyzing the collision of relativistic articles can all be derived from (9.33) and (9.34). Taking the difference of squares leads to the key formula (9.39) relating energy, momentum, and (rest) mass, which holds also for massless particles. Rewriting (9.33) and (9.34) leads directly to

$$\gamma = \cosh\alpha = \frac{E}{mc^2} = \frac{1}{\sqrt{1 - \frac{u^2}{c^2}}} \tag{9.41}$$

and

$$\sinh\alpha = \frac{p}{mc} = \frac{u}{c}\gamma \tag{9.42}$$

and dividing these formulas yields

$$\tanh\alpha = \frac{pc}{E} = \frac{u}{c} \tag{9.43}$$

Finally, another useful formula is

$$\frac{m^2c^4}{E^2} = 1 - \frac{u^2}{c^2} = \left(1 + \frac{u}{c}\right)\left(1 - \frac{u}{c}\right) \approx 2\left(1 - \frac{u}{c}\right) \tag{9.44}$$

where the final approximation holds if $u \approx c$.

Chapter 10

Problems II

10.1 Mass isn't Conserved

Two identical lumps of clay of (rest) mass m collide head on, with each moving at $\frac{3}{5}c$. What is the mass of the resulting lump of clay?

We assume this is an elastic collision, that is, we do not worry about the details of the actual collision. Conservation of momentum doesn't help here — there is no momentum either before or afterwards. So we need to use conservation of energy. After the collision, there is no kinetic energy, so we have

$$E' = Mc^2 \tag{10.1}$$

Before the collision, we know that the energy of each lump is

$$E = mc^2 \cosh\alpha \tag{10.2}$$

but how do we find α? We are given that each lump is moving at $\frac{3}{5}c$. But this means we know

$$\tanh\alpha = \frac{3}{5} \tag{10.3}$$

Yes, we could now use the formula $\cosh\alpha = \frac{1}{\sqrt{1-\tanh^2\alpha}}$, but it is easier to use a triangle. Since $\tanh\alpha = \frac{3}{5}$, we can scale things so that the legs have "length" 3 and 5. Using the (hyperbolic!) Pythagorean Theorem, the hypotenuse has "length" $\sqrt{5^2 - 3^2} = 4$. This is just the triangle in Figure 4.3! Thus,

$$\cosh\alpha = \frac{5}{4} \tag{10.4}$$

and so

$$Mc^2 = E' = 2E = 2mc^2 \cosh\alpha = \frac{5}{2}\,mc^2 \tag{10.5}$$

so that finally

$$M = \frac{5}{2}\,m \tag{10.6}$$

10.2 Colliding particles

Consider the head on collision of 2 identical particles each of mass m and energy E.

1. *In Newtonian mechanics, what multiple of E is the energy E' of one particle as observed in the reference frame of the other?*
2. *In special relativity, what is the energy E' of one particle as observed in the reference frame of the other?*
3. *Suppose we collide 2 protons ($mc^2 = 1$ GeV) with energy $E = 30$ GeV. Roughly what multiple of E is E' in this case?*

1. In the center-of-mass frame, each particle has (Newtonian) kinetic energy

$$E = \frac{1}{2}mv^2 \tag{10.7}$$

In the reference frame of one of the particles, the other is moving twice as fast, so that

$$E' = \frac{1}{2}m(2v)^2 = 4E \tag{10.8}$$

2. Now we must use the relativistic energy

$$E = mc^2 \cosh\alpha \tag{10.9}$$

In the reference frame of one of the particles, the other is *not* moving twice as fast. Rather, the hyperbolic angle has doubled. Thus

$$E' = mc^2 \cosh(2\alpha) = mc^2(2\cosh^2\alpha - 1) \tag{10.10}$$

so that

$$\frac{E'}{E} = \frac{2\cosh^2\alpha - 1}{\cosh\alpha} \tag{10.11}$$

3. We are given that

$$\cosh\alpha = \frac{E}{mc^2} = 30 \tag{10.12}$$

so that

$$\frac{E'}{E} = \frac{2\cosh^2\alpha - 1}{\cosh\alpha} \approx 60 \tag{10.13}$$

Chapter 11

Relativistic Electromagnetism

In which it is shown that electricity and magnetism can no more be separated than space and time.

11.1 Magnetism from Electricity

Our starting point is the electric and magnetic fields of an infinite straight wire, which are derived in most introductory textbooks on electrodynamics, such as Griffiths [3], and which we state here without proof.

The electric field of an infinite straight wire with charge density λ points away from the wire with magnitude

$$E = \frac{\lambda}{2\pi\epsilon_0 r} \tag{11.1}$$

where r is the perpendicular distance from the wire and ϵ_0 is the permittivity constant. The magnetic field of such a wire with current density I has magnitude

$$B = \frac{\mu_0 I}{2\pi r} \tag{11.2}$$

with r as above, and where μ_0 is the permeability constant, which is related to ϵ_0 by

$$\epsilon_0\mu_0 = \frac{1}{c^2} \tag{11.3}$$

(The direction of the magnetic field is obtained as the cross product of the direction of the current and the position vector from the wire to the point in question.)

We will also need the *Lorentz force law*, which says that the force $\vec{\boldsymbol{F}}$ on a test particle of charge q and velocity $\vec{\boldsymbol{v}}$ is given by

$$\vec{\boldsymbol{F}} = q(\vec{\boldsymbol{E}} + \vec{\boldsymbol{v}} \times \vec{\boldsymbol{B}}) \tag{11.4}$$

where $\vec{\boldsymbol{E}}$ and $\vec{\boldsymbol{B}}$ denote the electric and magnetic fields (with magnitudes E and B, respectively).

Consider an infinite line charge, consisting of identical particles of charge ρ, separated by a distance ℓ. This gives an infinite wire with (average) charge density

$$\lambda_0 = \frac{\rho}{\ell} \tag{11.5}$$

Suppose now that the charges are moving to the right with speed

$$u = c \tanh \alpha \tag{11.6}$$

Due to length contraction, the charge density seen by an observer at rest *increases* to

$$\lambda = \frac{\rho}{\frac{\ell}{\cosh \alpha}} = \lambda_0 \cosh \alpha \tag{11.7}$$

Suppose now that there are positively charged particles moving to the right, and equally but negatively charged particles moving to the left, each with speed u. Consider further a test particle of charge q situated a distance r from the wire and moving with speed

$$v = c \tanh \beta \tag{11.8}$$

to the right. Then the net charge density in the laboratory frame is 0, so that there is no electrical force on the test particle in this frame. There is of course a net current density, however, namely

$$I = \lambda u + (-\lambda)(-u) = 2\lambda u \tag{11.9}$$

What does the test particle see? Switch to the rest frame of the test particle; this makes the negative charges appear to move faster, with speed $u_- > u$, and the positive charges move slower, with speed $u_+ < u$. The relative speeds satisfy

$$\frac{u_+}{c} = \tanh(\alpha - \beta) \tag{11.10}$$

$$\frac{u_-}{c} = \tanh(\alpha + \beta) \tag{11.11}$$

resulting in current densities

$$\lambda_\pm = \lambda \cosh(\alpha \mp \beta) = \lambda(\cosh\alpha\cosh\beta \mp \sinh\alpha\sinh\beta) \tag{11.12}$$

resulting in a total charge density of

$$\begin{aligned} \lambda' &= \lambda_+ - \lambda_- && (11.13)\\ &= -2\lambda_0 \sinh\alpha\sinh\beta && (11.14)\\ &= -2\lambda\tanh\alpha\sinh\beta && (11.15)\end{aligned}$$

According to (11.1), this results in an electric field of magnitude

$$E' = \frac{\lambda'}{2\pi\epsilon} \tag{11.16}$$

which in turn leads to an electric force of magnitude

$$\begin{aligned} F' &= qE' = -\frac{\lambda}{\pi\epsilon_0 r}\, q\tanh\alpha\sinh\beta && (11.17)\\ &= -\frac{\lambda u}{\pi\epsilon_0 c^2 r}\, qv\cosh\beta && (11.18)\\ &= -\frac{\mu_0 I}{2\pi r}\, qv\cosh\beta && (11.19)\end{aligned}$$

To relate this to the force observed in the laboratory frame, we must consider how force transforms under a Lorentz transformation. We have [1]

$$\vec{\boldsymbol{F}}' = \frac{d\vec{\boldsymbol{p}}'}{dt'} \tag{11.20}$$

and of course also

$$\vec{\boldsymbol{F}} = \frac{d\vec{\boldsymbol{p}}}{dt} \tag{11.21}$$

But since in this case the force is perpendicular to the direction of motion, we have

$$d\vec{\boldsymbol{p}} = d\vec{\boldsymbol{p}}' \tag{11.22}$$

and since $dx' = 0$ in the comoving frame we also have

$$dt = dt'\cosh\beta \tag{11.23}$$

[1] This is the traditional notion of force, which does not transform simply between frames. As discussed briefly below, a possibly more useful notion of force is obtained by differentiating with respect to proper time.

Thus, in this case, the magnitudes are related by

$$F = \frac{F'}{\cosh\beta} = -\frac{\mu_0 I}{2\pi r}\,qv \tag{11.24}$$

But this is just the Lorentz force law

$$\vec{\boldsymbol{F}} = q\,\vec{\boldsymbol{v}} \times \vec{\boldsymbol{B}} \tag{11.25}$$

with $B = |\vec{\boldsymbol{B}}|$ given by (11.2)!

We conclude that in the laboratory frame there is a *magnetic* force on the test particle, which is just the *electric* force observed in the comoving frame!

11.2 Lorentz Transformations

We now investigate more general transformations of electric and magnetic fields between different inertial frames. Our starting point is the electromagnetic field of an infinite flat metal sheet, which is derived in most introductory textbooks on electrodynamics, such as Griffiths [3], and which we state here without proof.

The electric field of an infinite metal sheet with charge density σ points away from the sheet and has the constant magnitude

$$|E| = \frac{\sigma}{2\epsilon_0} \tag{11.26}$$

The magnetic field of such a sheet with current density $\vec{\boldsymbol{\kappa}}$ has constant magnitude

$$|B| = \frac{\mu}{2}|\vec{\boldsymbol{\kappa}}| \tag{11.27}$$

and direction determined by the right-hand-rule.

Consider a capacitor consisting of 2 horizontal (z = constant) parallel plates, with equal and opposite charge densities. For definiteness, take the charge density on the bottom plate to be σ_0, and suppose that the charges are at rest, that is, that the current density of each plate is zero. Then the electric field is given by

$$\vec{\boldsymbol{E}}_0 = E_0\,\vec{\boldsymbol{\jmath}} = \frac{\sigma_0}{\epsilon_0}\,\vec{\boldsymbol{\jmath}} \tag{11.28}$$

between the plates and vanishes elsewhere. Now let the capacitor move to the left with velocity

$$\vec{\boldsymbol{u}} = -u\,\vec{\boldsymbol{\imath}} = -c\tanh\alpha\ \vec{\boldsymbol{\imath}} \tag{11.29}$$

Then the *width* of the plate is unchanged, but, just as for the line charge (11.7), the *length* is Lorentz contracted, which *decreases* the area, and hence *increases* the charge density. The charge density (on the bottom plate) is therefore

$$\sigma = \sigma_0 \cosh \alpha \tag{11.30}$$

But there is now also a current density, which is given by

$$\vec{\boldsymbol{\kappa}} = \sigma \, \vec{\boldsymbol{u}} \tag{11.31}$$

on the lower plate. The top plate has charge density $-\sigma$, so its current density is $-\vec{\boldsymbol{\kappa}}$. Then both the electric and magnetic fields vanish outside the plates, whereas inside the plates one has

$$\begin{aligned} \vec{\boldsymbol{E}} &= E^y \, \vec{\boldsymbol{\jmath}} = \frac{\sigma}{\epsilon_0} \, \vec{\boldsymbol{\jmath}} && (11.32) \\ \vec{\boldsymbol{B}} &= B^z \, \vec{\boldsymbol{k}} = -\mu_0 \, \sigma u \, \vec{\boldsymbol{k}} && (11.33) \end{aligned}$$

which can be rewritten using (11.29) and (11.30) in the form

$$\begin{aligned} E^y &= E_0 \cosh \alpha && (11.34) \\ B^z &= cB_0 \sinh \alpha && (11.35) \end{aligned}$$

For later convenience, we have introduced in the last equation the quantity

$$B_0 = -\mu_0 \sigma_0 = -\mu_0 \epsilon_0 E_0 = -\frac{1}{c^2} \, E_0 \tag{11.36}$$

which does *not* correspond to the magnetic field when the plate is at rest — which of course vanishes since $\vec{\boldsymbol{u}} = 0$.

The above discussion gives the electric and magnetic fields seen by an observer at rest. What is seen by an observer moving to the right with speed $v = c \tanh \beta$? To compute this, first use the velocity addition law to compute the correct rapidity to insert in (11.35), which is simply the sum of the rapidities α and β!

The moving observer therefore sees an electric field $\vec{\boldsymbol{E}}'$ and a magnetic field $\vec{\boldsymbol{B}}'$. From (11.35), (11.36), and the hyperbolic trig formulas (4.5) and (4.6), we have

$$\begin{aligned} E'^y &= E_0 \cosh(\alpha + \beta) \\ &= E_0 \cosh \alpha \cosh \beta + E_0 \sinh \alpha \sinh \beta \\ &= E_0 \cosh \alpha \cosh \beta - c^2 B_0 \sinh \alpha \sinh \beta \\ &= E^y \cosh \beta - cB^z \sinh \beta && (11.37) \end{aligned}$$

and similarly

$$\begin{aligned} B'^z &= cB_0 \sinh(\alpha + \beta) \\ &= cB_0 \sinh\alpha \cosh\beta + cB_0 \cosh\alpha \sinh\beta \\ &= B^z \cosh\beta - \frac{1}{c} E^y \sinh\beta \end{aligned} \tag{11.38}$$

Repeating the argument with the y and z axes interchanged (and being careful about the orientation), we obtain the analogous formulas

$$E'^z = E^z \cosh\beta + cB^y \sinh\beta \tag{11.39}$$

$$B'^x = B^y \cosh\beta + \frac{1}{c} E^z \sinh\beta \tag{11.40}$$

Finally, by considering motion perpendicular to the plates one can show [3]

$$E'^x = E^x \tag{11.41}$$

and by considering a solenoid one obtains [3]

$$B'^x = B^x \tag{11.42}$$

Equations (11.37)–(11.42) describe the behavior of the electric and magnetic fields under Lorentz transformations. These equations can be nicely rewritten in vector language by introducing the projections parallel and perpendicular to the direction of motion of the observer, namely

$$\vec{\boldsymbol{E}}_\parallel = \frac{\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{E}}}{\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{v}}} \vec{\boldsymbol{v}} \tag{11.43}$$

$$\vec{\boldsymbol{B}}_\parallel = \frac{\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{B}}}{\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{v}}} \vec{\boldsymbol{v}} \tag{11.44}$$

and

$$\vec{\boldsymbol{E}}_\perp = \vec{\boldsymbol{E}} - \vec{\boldsymbol{E}}_\parallel \tag{11.45}$$

$$\vec{\boldsymbol{B}}_\perp = \vec{\boldsymbol{B}} - \vec{\boldsymbol{B}}_\parallel \tag{11.46}$$

We then have

$$\vec{\boldsymbol{E}}'_\parallel = \vec{\boldsymbol{E}}_\parallel \tag{11.47}$$

$$\vec{\boldsymbol{B}}'_\parallel = \vec{\boldsymbol{B}}_\parallel \tag{11.48}$$

and

$$\vec{\boldsymbol{B}}'_\perp = \left(\vec{\boldsymbol{B}}_\perp - \frac{1}{c^2}\,\vec{\boldsymbol{v}} \times \vec{\boldsymbol{E}}_\perp\right)\cosh\beta \tag{11.49}$$

$$\vec{\boldsymbol{E}}'_\perp = \left(\vec{\boldsymbol{E}}_\perp + \vec{\boldsymbol{v}} \times \vec{\boldsymbol{B}}_\perp\right)\cosh\beta \tag{11.50}$$

11.3 Vectors

In the previous chapter, we used 2-component vectors to describe spacetime, with one component for time and the other for space. In the case of 3 spatial dimensions, we use 4-component vectors, namely

$$x^\nu = \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} = \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \tag{11.51}$$

These are called *contravariant* vectors, and their indices are written "upstairs", that is, as superscripts.

Just as before, Lorentz transformations are hyperbolic rotations, which must now be written as 4×4 matrices. For instance, a "boost" in the x direction now takes the form

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \cosh\beta & -\sinh\beta & 0 & 0 \\ -\sinh\beta & \cosh\beta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \tag{11.52}$$

A general Lorentz transformation can be written in the form

$$x'^\mu = \Lambda^\mu{}_\nu\, x^\nu \tag{11.53}$$

where $\Lambda^\mu{}_\nu$ are (the components of) the appropriate 4×4 matrix, and where we have adopted the *Einstein summation convention* that repeated indices, in this case ν, are to be summed from 0 to 3. In matrix notation, this can be written as

$$\boldsymbol{x'} = \boldsymbol{\Lambda x} \tag{11.54}$$

Why are some indices up and others down? In relativity, both special and general, it is essential to distinguish between 2 types of vectors. In addition to contravariant vectors, there are also *covariant* vectors, often referred to as dual vectors. The dual vector associated with x^μ is [2]

$$x_\mu = \begin{pmatrix} -x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -ct \\ x \\ y \\ z \end{pmatrix} \tag{11.55}$$

We won't have much need for covariant vectors, but note that the invariance of the interval can be nicely written as

$$\begin{aligned} x_\mu x^\mu &= -c^2t^2 + x^2 + y^2 + z^2 \\ &= x'^\mu x'_\mu \end{aligned} \tag{11.56}$$

(Don't forget the summation convention!) In fact, this equation can be taken as the *definition* of Lorentz transformations, and it is straightforward to determine which matrices $\Lambda^\mu{}_\nu$ are allowed.

Taking the derivative with respect to proper time leads to the *4-velocity*

$$u^\mu = \frac{dx^\mu}{d\tau} = \frac{dx^\mu}{dt}\frac{dt}{d\tau} \tag{11.57}$$

It is often useful to divide these into space and time in the form

$$u = \begin{pmatrix} c\gamma \\ \vec{\boldsymbol{v}}\gamma \end{pmatrix} = \begin{pmatrix} c\cosh\beta \\ \hat{\boldsymbol{v}}\, c\sinh\beta \end{pmatrix} \tag{11.58}$$

where $\hat{\boldsymbol{v}}$ is the unit vector in the direction of $\vec{\boldsymbol{v}}$. Note that the 4-velocity is a unit vector in the sense that

$$\frac{1}{c^2}\, u_\mu u^\mu = -1 \tag{11.59}$$

The 4-momentum is is simply the 4-velocity times the rest mess, that is

$$p^\mu = mu^\mu = \begin{pmatrix} \frac{1}{c}E \\ \vec{\boldsymbol{p}} \end{pmatrix} = \begin{pmatrix} mc\gamma \\ m\vec{\boldsymbol{v}}\gamma \end{pmatrix} = \begin{pmatrix} mc\cosh\beta \\ \hat{\boldsymbol{v}}\, mc\sinh\beta \end{pmatrix} \tag{11.60}$$

and note that

$$p_\mu p^\mu = -m^2c^2 \tag{11.61}$$

which is equivalent to our earlier result

$$E^2 - p^2c^2 = m^2c^4 \tag{11.62}$$

[2] Some authors use different conventions.

11.4 Tensors

Roughly speaking, tensors are like vectors, but with more components, and hence more indices. We will only consider one particular case, namely *rank 2 contravariant tensors*, which have 2 "upstairs" indices. Such a tensor has components in a particular reference frame which make up a 4×4 matrix,

$$T^{\mu\nu} = \begin{pmatrix} T^{00} & T^{01} & T^{02} & T^{03} \\ T^{10} & T^{11} & T^{12} & T^{13} \\ T^{20} & T^{21} & T^{22} & T^{23} \\ T^{30} & T^{31} & T^{32} & T^{33} \end{pmatrix} \tag{11.63}$$

How does the tensor $\boldsymbol{T}$ transform under Lorentz transformations? Well, it has *two* indices, *each* of which must be transformed. This leads to a transformation of the form

$$T'^{\mu\nu} = \Lambda^{\mu}{}_{\rho}\Lambda^{\nu}{}_{\sigma}T^{\rho\sigma} = \Lambda^{\mu}{}_{\rho}T^{\rho\sigma}\Lambda^{\nu}{}_{\sigma} \tag{11.64}$$

where the second form (and the summation convention!) leads naturally to the matrix equation

$$\boldsymbol{T'} = \boldsymbol{\Lambda T \Lambda}^t \tag{11.65}$$

where t denotes matrix transpose.

Further simplification occurs in the special case where $\boldsymbol{T}$ is antisymmetric, that is

$$T^{\nu\mu} = -T^{\mu\nu} \tag{11.66}$$

so that the components of $\boldsymbol{T}$ take the form

$$T^{\mu\nu} = \begin{pmatrix} 0 & a & b & c \\ -a & 0 & f & -e \\ -b & -f & 0 & d \\ -c & e & -d & 0 \end{pmatrix} \tag{11.67}$$

11.5 The Electromagnetic Field

Why have we done all this? Well, first of all, note that, due to antisymmetry, $\boldsymbol{T}$ has precisely 6 independent components. Next, compute $\boldsymbol{T'}$, using matrix multiplication and the fundamental hyperbolic trig identity (4.4). As you should check for yourself, the result is

$$T'^{\mu\nu} = \begin{pmatrix} 0 & a' & b' & c' \\ -a' & 0 & f' & -e' \\ -b' & -f' & 0 & d' \\ -c' & e' & -d' & 0 \end{pmatrix} \tag{11.68}$$

where

$$a' = a \tag{11.69}$$
$$b' = b\cosh\beta - f\sinh\beta \tag{11.70}$$
$$c' = c\cosh\beta + e\sinh\beta \tag{11.71}$$
$$d' = d \tag{11.72}$$
$$e' = e\cosh\beta + c\sinh\beta \tag{11.73}$$
$$f' = f\cosh\beta - b\sinh\beta \tag{11.74}$$

The first 3 of these are the transformation rule for the electric field, and the remaining 3 are the transformation rule for the magnetic field!

We are thus led to introduce the *electromagnetic field tensor*, namely

$$F^{uv} = \begin{pmatrix} 0 & \frac{1}{c}E^x & \frac{1}{c}E^y & \frac{1}{c}E^z \\ -\frac{1}{c}E^x & 0 & B^z & -B^y \\ -\frac{1}{c}E^y & -B^z & 0 & B^x \\ -\frac{1}{c}E^z & B^y & -B^x & 0 \end{pmatrix} \tag{11.75}$$

11.6 Maxwell's equations

Maxwell's equations in vacuum (and in MKS units) are

$$\vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{E}} = \frac{1}{\epsilon_0}\rho \tag{11.76}$$
$$\vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{B}} = 0 \tag{11.77}$$
$$\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{E}} = -\frac{\partial \vec{\boldsymbol{B}}}{\partial t} \tag{11.78}$$
$$\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{B}} = \mu_0 \vec{\boldsymbol{J}} + \mu_0\epsilon_0 \frac{\partial \vec{\boldsymbol{E}}}{\partial t} \tag{11.79}$$

where ρ is the charge density, $\vec{\boldsymbol{J}}$ is the current density, and the constants μ_0 and ϵ_0 satisfy (11.3). Equation (11.76) is just Gauss' Law, (11.78) is

Faraday's equation, and (11.79) is Ampère's Law corrected for the case of a time-dependent electric field. We also have the charge conservation equation

$$\vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{J}} = -\frac{\partial \rho}{\partial t} \tag{11.80}$$

and the Lorentz force law

$$\vec{\boldsymbol{F}} = q(\vec{\boldsymbol{E}} + \vec{\boldsymbol{v}} \times \vec{\boldsymbol{B}}) \tag{11.81}$$

The middle two of Maxwell equations are automatically solved by introducing the scalar potential Φ and the vector potential $\vec{\boldsymbol{A}}$ and defining

$$\vec{\boldsymbol{B}} = \vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{A}} \tag{11.82}$$

$$\vec{\boldsymbol{E}} = -\frac{\partial \vec{\boldsymbol{A}}}{\partial t} - \vec{\boldsymbol{\nabla}}\Phi \tag{11.83}$$

Consider the following derivatives of $\boldsymbol{F}$:

$$\frac{\partial F^{\mu\nu}}{\partial x^{\nu}} = \frac{\partial F^{\mu 0}}{\partial t} + \frac{\partial F^{\mu 1}}{\partial x} + \frac{\partial F^{\mu 2}}{\partial y} + \frac{\partial F^{\mu 3}}{\partial z} \tag{11.84}$$

This corresponds to four different expressions, one for each value of μ. For $\mu = 0$, we get

$$0 + \frac{1}{c}\frac{\partial E^x}{\partial x} + \frac{1}{c}\frac{\partial E^y}{\partial y} + \frac{1}{c}\frac{\partial E^z}{\partial z} = \frac{1}{c}\vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{E}} = \frac{\rho}{c\epsilon_0} = c\mu_0\rho \tag{11.85}$$

where Gauss' Law was used to get the final two equalities. Similarly, for $\mu = 1$ we have

$$-\frac{1}{c^2}\frac{\partial E^x}{\partial t} + 0 + \frac{\partial B^z}{\partial y} - \frac{\partial B^y}{\partial z} \tag{11.86}$$

and combining this with the expressions for $\mu = 2$ and $\mu = 3$ yields the left-hand-side of

$$-\frac{1}{c^2}\frac{\partial \vec{\boldsymbol{E}}}{\partial t} + \nabla \times \vec{\boldsymbol{B}} = \mu_0 \vec{\boldsymbol{J}} \tag{11.87}$$

where the right-hand-side follows from Ampère's Law. Combining these equations, and introducing the *4-current density*

$$J^{\mu} = \begin{pmatrix} c\rho \\ \vec{\boldsymbol{J}} \end{pmatrix} \tag{11.88}$$

leads to

$$\frac{\partial F^{\mu\nu}}{\partial x^\nu} = \mu_0 \vec{\boldsymbol{J}}^\mu \tag{11.89}$$

which is equivalent to the two Maxwell equations with a physical source, namely Gauss' Law and Ampère's Law.

Furthermore, taking the (4-dimensional!) divergence of the 4-current density leads to

$$\mu_0 \frac{\partial J^\mu}{\partial x^\mu} = \frac{\partial}{\partial x^\mu}\frac{\partial F^{\mu\nu}}{\partial x^\nu} = 0 \tag{11.90}$$

since there is an implicit double sum over both μ and ν, and the derivatives commute but $F^{\mu\nu}$ is antisymmetric. (Check this by interchanging the order of summation.) Working out the components of this equation, we have

$$\frac{1}{c}\frac{\partial J^0}{\partial t} + \frac{\partial J^1}{\partial x} + \frac{\partial J^2}{\partial y} + \frac{\partial J^3}{\partial z} = 0 \tag{11.91}$$

which is just the charge conservation equation (11.80).

What about the remaining equations? Introduce the *dual tensor* $G^{\mu\nu}$ obtained from $F^{\mu\nu}$ by replacing $\frac{1}{c}\vec{\boldsymbol{E}}$ by $\vec{\boldsymbol{B}}$ and $\vec{\boldsymbol{B}}$ by $-\frac{1}{c}\vec{\boldsymbol{E}}$, resulting in

$$G^{uv} = \begin{pmatrix} 0 & B^x & B^y & B^z \\ -B^x & 0 & -\frac{1}{c}E^z & \frac{1}{c}E^y \\ -B^y & \frac{1}{c}E^z & 0 & -\frac{1}{c}E^x \\ -B^z & -\frac{1}{c}E^y & \frac{1}{c}E^x & 0 \end{pmatrix} \tag{11.92}$$

Then the four equations

$$\frac{\partial G^{\mu\nu}}{\partial x^\nu} = 0 \tag{11.93}$$

correspond to

$$\vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{B}} = 0 \tag{11.94}$$

$$-\frac{1}{c}\frac{\partial \vec{\boldsymbol{B}}}{\partial t} - \frac{1}{c}\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{E}} = 0 \tag{11.95}$$

which are precisely the two remaining Maxwell equations.

Some further properties of these tensors are

$$\frac{1}{2}F_{\mu\nu}F^{\mu\nu} = -\frac{1}{c^2}\left|\vec{\boldsymbol{E}}\right|^2 + |\vec{\boldsymbol{B}}|^2 = -\frac{1}{2}G_{\mu\nu}G_{\mu\nu} \tag{11.96}$$

$$\frac{1}{4}G_{\mu\nu}F^{\mu\nu} = -\frac{1}{c}\,\vec{\boldsymbol{E}} \cdot \vec{\boldsymbol{B}} \tag{11.97}$$

where care must be taken with the signs of the components of the *covariant* tensors $F_{\mu\nu}$ and $G_{\mu\nu}$. You may recognize these equations as corresponding to important scalar invariants of the electromagnetic field.

Finally, it is possible to solve the sourcefree Maxwell equations by introducing a *4-potential*

$$A^\mu = \begin{pmatrix} \frac{1}{c}\Phi \\ \vec{\boldsymbol{A}} \end{pmatrix} \tag{11.98}$$

and defining

$$F^{\mu\nu} = \frac{\partial A^\nu}{\partial x_\mu} - \frac{\partial A^\mu}{\partial x_\nu} \tag{11.99}$$

where again care must be taken with the signs of the components with "downstairs" indices. Furthermore, the Lorentz force law can be rewritten in the form

$$m\frac{\partial p^\mu}{\partial \tau} = qu_\nu F^{\mu\nu} \tag{11.100}$$

Note the appearance of the proper time τ in this equation. Just as in the previous chapter, this is because differentiation with respect to τ pulls through a Lorentz transformation, which makes this a valid tensor equation, valid in any inertial frame.

Chapter 12

Problems III

12.1 Electricity vs. Magnetism I

Suppose you know that in a particular inertial frame neither the electric field $\vec{\boldsymbol{E}}$ *nor the magnetic field* $\vec{\boldsymbol{B}}$ *has an* x *component, but neither* $\vec{\boldsymbol{E}}$ *nor* $\vec{\boldsymbol{B}}$ *is zero. Consider another inertial frame moving with respect to the first one with velocity* v *in the* x*-direction, and denote the electric and magnetic fields in this frame by* $\vec{\boldsymbol{E}}'$ *and* $\vec{\boldsymbol{B}}'$*, respectively.*

1. *What are the conditions on* $\vec{\boldsymbol{E}}$ *and* $\vec{\boldsymbol{B}}$*, if any, and the value(s) of* v*, if any, such that* $\vec{\boldsymbol{E}}'$ *vanishes for some value of* v*?*

2. *What are the conditions on* $\vec{\boldsymbol{E}}$ *and* $\vec{\boldsymbol{B}}$*, if any, and the value(s) of* v*, if any, such that* $\vec{\boldsymbol{B}}'$ *vanishes for some value of* v*?*

3. *What are the conditions on* $\vec{\boldsymbol{E}}$ *and* $\vec{\boldsymbol{B}}$*, if any, and the value(s) of* v*, if any, such that* **both** $\vec{\boldsymbol{E}}'$ *and* $\vec{\boldsymbol{B}}'$ *vanish for the same value of* v*.*

4. *Is it possible that* $\vec{\boldsymbol{E}}'$ *and* $\vec{\boldsymbol{B}}''$ *vanish for different values of* v*?*
   (We write $\vec{\boldsymbol{B}}''$ *rather than* $\vec{\boldsymbol{B}}'$ *to emphasize that* $\vec{\boldsymbol{E}}'$ *and* $\vec{\boldsymbol{B}}''$ *are with respect to different reference frames.)*

1. Inserting $\vec{\boldsymbol{E}}' = 0$ into (11.50), we get

$$v = |\vec{\boldsymbol{v}}| = \frac{|\vec{\boldsymbol{E}}|}{|\vec{\boldsymbol{B}}|} \tag{12.1}$$

since $\vec{\boldsymbol{v}}$ is perpendicular to $\vec{\boldsymbol{B}}$. This is only possible if $|\vec{\boldsymbol{E}}| < c|\vec{\boldsymbol{B}}|$, which also follows immediately from the invariance of (11.96).
2. Inserting $\vec{\boldsymbol{B}}' = 0$ into (11.49), we get

$$v = |\vec{\boldsymbol{v}}| = \frac{c^2|\vec{\boldsymbol{B}}|}{|\vec{\boldsymbol{E}}|} \tag{12.2}$$

which is only possible if $c|\vec{\boldsymbol{B}}| < |\vec{\boldsymbol{E}}|$. This condition also follows immediately from the invariance of (11.96).
3. This is not possible; if the electric and magnetic fields are both zero in any frame, they are zero in all frames.
4. No; the conditions in the first two problems can not both be satisfied.

12.2 Electricity vs. Magnetism II

Suppose that in a particular inertial frame the electric field $\vec{\boldsymbol{E}}$ and magnetic field $\vec{\boldsymbol{B}}$ are neither perpendicular nor parallel to each other.

1. *Is there another inertial frame in which the fields $\vec{\boldsymbol{E}}''$ and $\vec{\boldsymbol{B}}''$ are parallel to each other?*

2. *Is there another inertial frame in which the fields $\vec{\boldsymbol{E}}''$ and $\vec{\boldsymbol{B}}''$ are perpendicular to each other?*

   (You may assume without loss of generality that the inertial frames are in relative motion parallel to the x-axis, and that neither $\vec{\boldsymbol{B}}$ nor $\vec{\boldsymbol{E}}$ has an x-component. Why? Briefly justify your assumptions!)

1. Yes! Setting the cross product of (11.50) and (11.49) equal to zero, using the vector identity $\vec{\boldsymbol{u}} \times (\vec{\boldsymbol{v}} \times \vec{\boldsymbol{w}}) = (\vec{\boldsymbol{u}} \cdot \vec{\boldsymbol{w}})\vec{\boldsymbol{v}} - (\vec{\boldsymbol{u}} \cdot \vec{\boldsymbol{v}})\vec{\boldsymbol{w}}$ and the fact that $\vec{\boldsymbol{v}}$ is perpendicular to both $\vec{\boldsymbol{E}}$ and $\vec{\boldsymbol{B}}$, leads to

$$\frac{\frac{\vec{\boldsymbol{v}}}{c}}{1 + \frac{|\vec{\boldsymbol{v}}|^2}{c^2}} = \frac{\vec{\boldsymbol{E}} \times c\vec{\boldsymbol{B}}}{|\vec{\boldsymbol{E}}|^2 + c^2|\vec{\boldsymbol{B}}|^2} \tag{12.3}$$

2. No; $\vec{\boldsymbol{E}} \cdot \vec{\boldsymbol{B}}$ is invariant according to (11.97).

Chapter 13

Beyond Special Relativity

NEXT STOP: General Relativity!

13.1 Problems with Special Relativity

We began in Chapter 2 by using moving trains to model inertial reference frames. But we made an implicit assumption beyond assuming an ideal train, with no friction and a perfectly straight track. We also assumed that there was no gravity. Einstein's famous thought experiment for discussing gravity is to consider a "freely falling" reference frame, typically a falling elevator. Objects thrown horizontally in such an elevator will not be seen to fall — there is no gravity (for a little while, at least!). But even here there is an implicit assumption, namely that the elevator is small compared to the Earth.

Return to our ideal train moving to the right, but now assume there is gravity. A ball thrown straight up will, according to an observer on the train, eventually turn around and fall back down, moving along a straight line, as shown in the first drawing in Figure 13.2. Since the acceleration due to gravity at the surface of the Earth is taken to be constant, the exact motion is described by a quadratic equation in t. From the ground, the same effect is seen, but combined with a constant motion (linear in t) to the right. The motion therefore takes place along a parabola; see Figure 13.1.

Now suppose that that train is also accelerating to the right with constant acceleration. Then a ball thrown straight up on the train still moves along the same parabolic trajectory as before (as seen from the ground), but the rear wall of the train might now catch up with it before it lands! This shows,

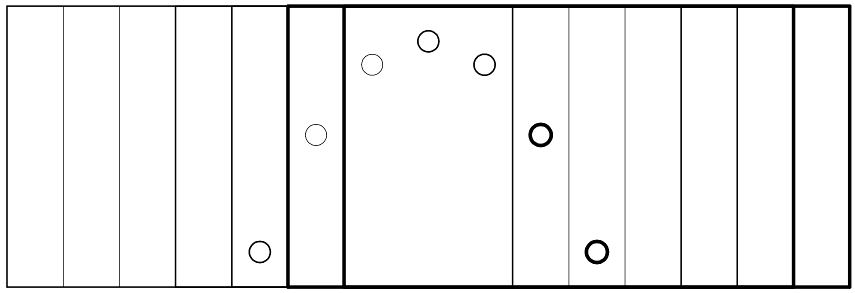


Figure 13.1: Throwing a ball in a moving train, as seen from the ground.

first of all, that Newton's laws fail in a noninertial frame.

But let's analyze the situation more carefully. Try to compensate by giving the ball a horizontal component of velocity. If the ball's initial horizontal speed is chosen appropriately, the resulting trajectory could look like the "boomerang" in the second drawing in Figure 13.2, in which (as seen from the train) the ball is thrown up at an angle, and returns along the same path!

What's going here? Acceleration and gravity produce the same kind of effect, and what the "boomerang" is telling us is that the *effective* force of gravity is no longer straight down, but rather at an angle. Throw the ball up at that angle, and it goes "up" and "down" in a straight line!

13.2 Tidal Effects

Consider now two objects falling towards the Earth, but far from it, as shown in the first sketch in Figure 13.3. Both objects fall towards the center of the Earth — which is not quite the same direction for each object. Assuming the objects start at the same distance from the Earth, their paths will converge.

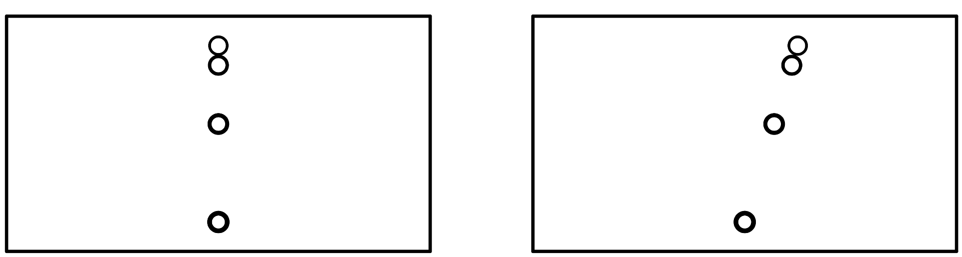


Figure 13.2: Throwing a ball in a moving train, as seen from the train.

Figure 13.3: Two objects falling towards the Earth from far away either move closer together or further apart depending on their initial configuration.

Now, if they don't realize they are falling — by virtue of being in a large falling elevator, say — they will nevertheless notice that they are approaching each other. This is gravity!

Similarly, if, as shown in the second sketch, one object starts out above the other, it is slightly further from the Earth, and hence experiences a slightly weaker gravitational attraction. Thus, the lower object will always accelerate more then the upper one, and so the distance between the objects will increase.

This effect causes tides! To see this, imagine that the Earth is falling towards the Moon, as shown in Figure 13.4. The motion of the center of the Earth is not relevant; subtract it out. What is left is a relative attraction at *both* the nearest and furthest points from the Moon, and a relative repulsion on either side. This explains why there are two, not one, high (and low) tides every day! A similar effect, but roughly half as strong, is caused by the Sun.

13.3 Differential Geometry

In 2-dimensions, Euclidean geometry is the geometry of a *flat* piece of paper. But there are also of course *curved* 2-dimensional surfaces. The simplest of these is the sphere, which has constant positive curvature and is a model of (double) *elliptic geometry*. Another important example is the hyperboloid, which also has constant curvature, and is a model of *hyperbolic geometry*. Hyperbolic and elliptic geometry form the 2 main categories of non-Euclidean

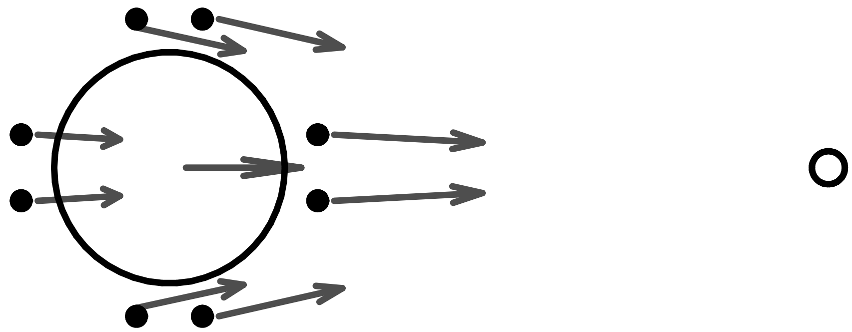


Figure 13.4: Tides are caused by the Earth falling towards the Moon!

geometries.

In fact, *any* 2-dimensional surface in Euclidean 3-space provides a possible geometry, most of which are curved. However, it is important to realize that distances are always positive in all such geometries. One measures the distance between 2 points on such a surface by stretching a string between them *along the surface*. This does not measure the (3-dimensional) Euclidean distance between the points, but instead corresponds to integrating the arclength along the shortest path between them.

In hyperbola geometry, we instead made a fundamental change to the distance function, allowing it to become negative or zero. If there is precisely one (basis) direction in which distances turn out to be negative, such geometries are said to have *Lorentzian signature*, as opposed to the *Euclidean signature* of ordinary surfaces. As implied by the way we have drawn it, hyperbola geometry turns out to be flat in a well-defined sense, which immediately raises the question of whether there are *curved* geometries with Lorentzian signature.

The mathematical study of curved surfaces forms a central part of *differential geometry*, and the further restriction to surfaces on which distances are positive is known as *Riemannian geometry*. The much harder case of Lorentzian signature is known, not surprisingly, as *Lorentzian geometry*, and the important special case where the curvature vanishes is *Minkowskian geometry*. This classification of geometries by signature and curvature is summarized in Table 13.5.

What does this have to do with physics? We have seen that hyperbola

signature	**flat**	**curved**
$(+ + ... +)$	Euclidean	Riemannian
$(- + ... +)$	Minkowskian	Lorentzian

Table 13.5: Classification of geometries.

geometry, more correctly called Minkowski space, is the geometry of special relativity. Lorentzian geometry turns out to be the geometry of general relativity. In short, according to Einstein, gravity is curvature!

13.4 General Relativity

Just as one studies flat (2-dimensional) Euclidean geometry before studying curved surfaces, one studies special relativity before general relativity. But the analogy goes much further.

The basic notion in Euclidean geometry is the distance between two points, which is given by the Pythagorean theorem. The basic notion on a curved surface is still the distance "function", but this is now an statement about infinitesimal distances. Euclidean geometry is characterized by the line element

$$ds^2 = dx^2 + dy^2 \tag{13.1}$$

which can be used to find the distance along any curve. The simplest curved surface, a sphere of radius r, can be characterized by the line element

$$ds^2 = r^2\,d\theta^2 + r^2 \sin^2\theta\,d\phi^2 \tag{13.2}$$

Remarkably, it is possible to calculate the curvature of the sphere from this line element alone; it is *not* necessary to use any 3-dimensional geometry. It's also possible to calculate the "straight lines", that is, the shortest path between two given points. These are, of course, straight lines in the plane, but great circles (diameters) on the sphere.

Similarly, special relativity is characterized by the line element [1]

$$ds^2 = -dt^2 + dx^2 \tag{13.3}$$

[1]Note that we have set $c = 1$!

which is every bit as flat as a piece of paper. You get general relativity simply by considering more general line elements!

Of course, it's not quite that simple. The line element must have a minus sign. And the curvature must correspond to a physical source of gravity — that's where Einstein's field equations come in. But, given a line element, one can again calculate the "straight lines", which now correspond to freely falling observers! Matter curves the universe, and the curvature tells objects which paths are "straight" — that's gravity!

Here are two examples to whet your appetite further.

The (3-dimensional) line element

$$ds^2 = -dt^2 + \sin^2(t) \left(d\theta^2 + \sin^2\theta \, d\phi^2 \right) \tag{13.4}$$

describes the 2-dimensional surface of a spherical balloon, whose radius changes with time. This roughly corresponds to a cosmological model for an expanding universe produced by a Big Bang. By studying the properties of this model more carefully, you will be led to ask good questions about relativistic cosmology.

And finally, the line element

$$ds^2 = -\left(1 - \frac{2m}{r}\right) dt^2 + \frac{dr^2}{1 - \dfrac{2m}{r}} \tag{13.5}$$

describes a simplified model of a black hole, with an apparent singularity at $r = 2m$, which however is just due to a poor choice of coordinates. Trying to understand what actually happens at $r = 2m$ will give you some understanding of what a black hole really is.

If you wish to pursue these ideas further, you may wish to take a look at an introductory textbook on general relativity, such as d'Inverno [5] or Taylor & Wheeler [6].

Bibliography

[1] Edwin F. Taylor and John Archibald Wheeler, **Spacetime Physics**, W. H. Freeman, San Francisco, 1963.

[2] Edwin F. Taylor and John Archibald Wheeler, **Spacetime Physics**, second edition, W. H. Freeman, New York, 1992.

[3] David J. Griffiths, **Introduction to Electrodynamics**, 3rd edition, Prentice Hall, Upper Saddle River, 1999.

[4] Tevian Dray, *The Twin Paradox Revisited*, Am. J. Phys. **58**, 822–825 (1990).

[5] Ray A. d'Inverno, **Introducing Einstein's Relativity**, Oxford University Press, Oxford, 1992.

[6] Edwin F. Taylor, John Archibald Wheeler, **Exploring Black Holes: Introduction to General Relativity**, Addison-Wesley, San Francisco, 2000.