An Introduction to Special Relativity, version 0.2

Kenan Diab, Massachusetts Institute of Technology '11, kdiab@mit.edu, web.mit.edu/kdiab/www

May 15, 2008

1 What are we trying to accomplish? Why?

Overzealous physics teachers often summarize the world of mechanics in the following table:

	Thingie going slow	Thingie going fast
Thingie very massive	Newtonian mechanics	Relativity
Thingie not massive	Quantum mechanics	Relativistic quantum field theory

While this is not an incorrect description, it is pretty imprecise. Special relativity specifically aims to answer the following question: Suppose we have two observers, one moving at some constant speed relative to the other. If something happens from the standpoint of one of the observers, what does the other see? Before we attempt to answer this question, it is necessary to define a couple terms so that we can state this question even more precisely:

Definition 1.1. *An **event** is a vector that specifies a point in space along with a time.*

Definition 1.2. *A **frame of reference** is a coordinate system that specifies the location of an event. A frame of reference S' is called **inertial** with respect to another frame of reference S if the velocity of S' with respect to S is constant.*

Thus, we can reframe the question as follows: Given the coordinates of an event $E = (x, y, z, t)$ in some frame of reference S, find the coordinates $E' = (x', y', z', t')$ of this event in a frame of reference S' inertial to S (As a side note, the inertial part of this question is what makes special relativity "special." Behavior between reference frames that are accelerating with respect to one another is in the domain of "general relativity.") For convenience, we choose the axes of S and S' so that the velocity of S' lies along the positive x-axis in S, and we force the origins of S and S' to coincide at $t = 0$. Common sense seems to dictate a very simple set of transformations: the Galilean transformations. They are given below. Try to convince yourself that they make sense intuitively:

$$\begin{aligned} x' &= x - vt \\ y' &= y \\ x' &= z \\ t' &= t \end{aligned}$$

Scientists in the 1800's realized these equations, although generally pretty good, were inadequate. In particular, they noticed that the Galilean transformations don't play nicely with electrodynamics. The details of this are not important right now, but the basic idea is that the explanations of phenomena in different inertial frames of reference are inconsistent, even though electrodynamics predicts the same final result in every frame. There were two possible explanations: either the Galilean transformations were wrong, or there must exist some universal, "true" frame of reference that lets you determine what is "really" happening. Given the seeming obviousness of the Galilean transformations, scientists postulated that the second explanation must be correct by appealing to an "ether", an invisible, undetectable substance that permeated all of space. All events would have to be measured with respect to the ether. The Michelson-Morley experiment, however, shattered this theory by doing precision measurements of the speed of light. If the ether was moving in some direction, it should "drag" light anisotropically, but this was shown to be false (Michelson and Morley subsequently won the Nobel prize for this). Moreover, Hendrik Lorentz, a Dutch physicist, noted that all the laws of electrodynamics remained unchanged under a set of transformations different from the Galilean transformation (which, shockingly, are called the Lorentz transformations). The stage was set: in 1905, German physicist Albert Einstein published the theory of special relativity.

2 The mathematics of special relativity

2.1 A reminder of basic geometry and algebra

You should (hopefully) already know something about vectors in Euclidean space. To remind you of some basic properties, consider the xy-plane. We can specify the location of a point in the xy-plane by two real numbers: x and y. These two numbers also define a vector, a mathematical object that has a length and a direction. As you (hopefully) know, we can specify a vector in the xy-plane as a 1 by 2 matrix: the vector extending from the origin to the point (x, y) is written $\begin{pmatrix} x \\ y \end{pmatrix}$.

We can take the dot product of two vectors by multiplying them component-wise. The dot product of $\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$ and $\begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$ is $x_1 y_1 + x_2 y_2$. When we take the dot product of a vector with itself, we get the length of the vector squared: $d^2 = x^2 + y^2$. This is an immediate consequence of the Pythagorean Theorem, which you (hopefully) know. For reasons that will become clear later, the dot product is often referred to as the *Euclidean inner product*, and the square root of the Euclidean inner product is called the *Euclidean metric*. When we combine the set of all vectors with real components with the Euclidean inner product, we have what is known as *Euclidean space*.

Of course, we don't live in a two-dimensional world. We have three spatial dimensions and one time dimension for a total of four dimensions. Thus, we need a total of four numbers to specify an event: (x, y, z, t). This isn't the only important difference between 2-D space and 4-D space-time, however. As you'll see in the next section, it turns out the world we live in isn't Euclidean. That is, the length of a vector is NOT given by the Euclidean metric. We will, however, state the mathematical formulation for the old, incorrect, Galilean theory so that you can compare it to the correct, relativistic theory later.

For convenience, we rewrite the Galilean transformations as a matrix equation. If we have an event (x, y, z, t) in a frame S, the coordinates of the event (x', y', z', t') in a frame S' moving at a velocity v relative to S is given by:

$$\begin{pmatrix} x' \\ y' \\ z' \\ t' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & -v \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix}$$

Although the Galilean transformations are not quite right, they have two essential properties that must be preserved by the Lorentz transformations. The first is left as an exercise:

Exercise 2.1. *Show that if we multiply the matrices of two different Galilean transformations, we get another Galilean transformation. Also, show that for every Galilean transformation, there exists another Galilean transformation such that the product of the two transformations is the identity – i.e. every Galilean transformation has an inverse.*

Mathematically, this means the Galilean transformations form a "group". We also need the Lorentz transformations to be a group. The second property is pretty obvious: the Galilean transformations are *linear*: that is, if we plot x' versus x, we get a line. Ditto for the other variables. This implies the intuitive assumption that it doesn't matter where you set the origin in your coordinate system: space-time looks the same no matter where and when you are. As you can verify, this implies that the Lorentz transformations also have a matrix representation. Now we are ready to develop the core mathematical tools of special relativity:

2.2 The postulates of special relativity and the derivation of the Lorentz transformations

The Michelson-Morley experiment motivated Einstein to assert two postulates as the starting point of his theory. First, the laws of physics are the same in all inertial reference frames. There is no universal frame of reference. Second, the speed of light is the same to all inertial observers, no matter how the source moves. Think about that second one for a bit; this is very counter-intuitive, but the experimental data unequivocally confirms it. We're going to use these postulates, along with the discussion in the previous section to derive the Lorentz transformations.

First, we recall what we're trying to do: Given the coordinates of an event $E = (x, y, z, t)$ in some frame of reference S, find the coordinates $E' = (x', y', z', t')$ of this event in a frame of reference S' inertial to S. To simplify the math a little bit, we make the usual choice of axes so that the velocity of S' is along the x-axis in S and the origins of S and S' coincide at $t = 0$. Thus, we have $y' = y$ and $z' = z$ since there is no relative velocity in those directions. Hence, we can simply omit these two terms when we calculate how x and t should change. As we noted before, the Lorentz transformations are linear, so we want to find a matrix with entries A, B, C, D so that the following equation is true:

$$\begin{pmatrix} x' \\ t' \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} x \\ t \end{pmatrix} \tag{1}$$

In linear form:

$$x' = Ax + Bt \tag{2}$$
$$t' = Cx + Dt \tag{3}$$

Note that the origin of S' has $x' = 0$. Remember that we specified that S' had velocity v relative to S, so the x-coordinate of the origin of S' in S is $x = vt$. Substituting these into (2), we have $0 = Avt + Bt$, so that $B = -Av$. Thus, (2) becomes

$$x' = A(x - vt) \tag{4}$$

Then, we do this in the opposite direction. The origin of S has $x = 0$. Substituting into (3) and (4), we have $t' = Dt$ and $x' = -Avt$. Since S moves with velocity $-v$ relative to S', we have $x' = -vt'$. Combining this equation with $x' = -Avt$, we have $A = D$. Let us pause to see what we have accomplished so far. Defining $E = C/A$, our original equations (2) and (3) are now:

$$x' = A(x - vt) \tag{5}$$
$$t' = A(Ex + t) \tag{6}$$

In matrix form:

$$\begin{pmatrix} x' \\ t' \end{pmatrix} = A \begin{pmatrix} 1 & -v \\ E & 1 \end{pmatrix} \begin{pmatrix} x \\ t \end{pmatrix} \tag{7}$$

Now comes the ugly part. The only information we are given that distinguishes two inertial reference frames is their relative velocity. Thus, A and E must be functions of v only. We will write A_v and E_v to denote the numbers

A and E that we get if we assume the relative velocity between two inertial frames is v. Now, we use the fact that the Lorentz transformations form a group - that is, if we do two Lorentz transformations back-to-back, we get another Lorentz transformation. Suppose we introduce another reference frame S'' that moves with velocity w relative to S' in the direction of the x-axis. Recall that S' moves with velocity v relative to S. Thus, we have two matrix equations:

$$\begin{pmatrix} x' \\ t' \end{pmatrix} = A_v \begin{pmatrix} 1 & -v \\ E_v & 1 \end{pmatrix} \begin{pmatrix} x \\ t \end{pmatrix} \tag{8}$$

$$\begin{pmatrix} x'' \\ t'' \end{pmatrix} = A_w \begin{pmatrix} 1 & -w \\ E_w & 1 \end{pmatrix} \begin{pmatrix} x' \\ t' \end{pmatrix} \tag{9}$$

Now, we substitute (8) into (9) as below:

$$\begin{aligned} \begin{pmatrix} x'' \\ t'' \end{pmatrix} &= A_w \begin{pmatrix} 1 & -w \\ E_w & 1 \end{pmatrix} A_v \begin{pmatrix} 1 & -v \\ E_v & 1 \end{pmatrix} \begin{pmatrix} x \\ t \end{pmatrix} & (10) \\ &= A_w A_v \begin{pmatrix} 1 & -w \\ E_w & 1 \end{pmatrix} \begin{pmatrix} 1 & -v \\ E_v & 1 \end{pmatrix} \begin{pmatrix} x \\ t \end{pmatrix} & (11) \\ &= A_w A_v \begin{pmatrix} 1 - E_v w & -v - w \\ E_w + E_v & 1 - E_w v \end{pmatrix} \begin{pmatrix} x \\ t \end{pmatrix} & (12) \end{aligned}$$

Now, we want the matrix in (12) to satisfy the properties that we verified earlier. In particular, we showed $A = D$, so the two diagonal elements of this matrix must be equal. Thus, we have:

$$\begin{aligned} 1 - E_v w &= 1 - E_w v & (13) \\ \implies \frac{v}{E_v} &= \frac{w}{E_w} & (14) \end{aligned}$$

Since (14) must be true for all choices of v and w, we have $\frac{v}{E_v} = k$ for some constant k, so that $E_v = v/k$. Substituting into (7):

$$\begin{pmatrix} x' \\ t' \end{pmatrix} = A \begin{pmatrix} 1 & -v \\ v/k & 1 \end{pmatrix} \begin{pmatrix} x \\ t \end{pmatrix} \tag{15}$$

All that's left is to find A. We now use the last piece of information: Lorentz transformations have an inverse. In particular, the inverse should just be the Lorentz transformation with velocity $-v$. Thus, substituting $w = -v$ in equation (9), we have $x = x''$ and $t = t''$. Using equation (15), we have two matrix equations again:

$$\begin{pmatrix} x' \\ t' \end{pmatrix} = A_v \begin{pmatrix} 1 & -v \\ v/k & 1 \end{pmatrix} \begin{pmatrix} x \\ t \end{pmatrix} \tag{16}$$

$$\begin{pmatrix} x \\ t \end{pmatrix} = A_{-v} \begin{pmatrix} 1 & v \\ -v/k & 1 \end{pmatrix} \begin{pmatrix} x' \\ t' \end{pmatrix} \tag{17}$$

Combining as before:

$$\begin{aligned} \begin{pmatrix} x \\ t \end{pmatrix} &= A_{-v} \begin{pmatrix} 1 & v \\ -v/k & 1 \end{pmatrix} A_v \begin{pmatrix} 1 & -v \\ v/k & 1 \end{pmatrix} \begin{pmatrix} x \\ t \end{pmatrix} & (18) \\ &= A_{-v} A_v \begin{pmatrix} 1 + v^2/k & 0 \\ 0 & 1 + v^2/k \end{pmatrix} \begin{pmatrix} x \\ t \end{pmatrix} & (19) \end{aligned}$$

Since we want this to be the identity transformation, we need the following equation to be true:

$$A_{-v}A_v \begin{pmatrix} 1+v^2/k & 0 \\ 0 & 1+v^2/k \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{20}$$

This implies that

$$A_{-v}A_v = \frac{1}{1+v^2/k} \tag{21}$$

Since our choice of the direction of v was arbitrary, we must have $A_{-v} = A_v$. Thus,

$$A_v^2 = \frac{1}{1+v^2/k} \tag{22}$$

$$\implies A_v = \sqrt{\frac{1}{1+v^2/k}} \tag{23}$$

We then make a judicious choice of k. We write $k = -c^2$. Hence, our final Lorentz transformation is:

$$\begin{pmatrix} x' \\ t' \end{pmatrix} = \frac{1}{\sqrt{1-v^2/c^2}} \begin{pmatrix} 1 & -v \\ -v/c^2 & 1 \end{pmatrix} \begin{pmatrix} x \\ t \end{pmatrix} \tag{24}$$

If we expand it out, add y and z, and write it in separate equations, we have:

$$\begin{aligned} x' &= \frac{x-vt}{\sqrt{1-v^2/c^2}} \\ y' &= y \\ z' &= z \\ t' &= \frac{-xv/c^2+t}{\sqrt{1-v^2/c^2}} \end{aligned}$$

I leave it to you to prove that c is indeed constant under the Lorentz transformations, and so we interpret it as the speed of light:

Exercise 2.2. *Show that if $x = ct$, then $x' = ct'$.*

This finishes the derivation of the Lorentz transformations. Phew!

Now we make a couple simplifications so that we can rewrite the Lorentz transformations in a nicer way. First, we would like all the components of a four-vector to have the same units. Thus, instead of writing t for time, we will write ct so that it has units of distance, just like x, y, and z. Second, we define $\beta = v/c$ and $\gamma = \frac{1}{\sqrt{1-v^2/c^2}} = \frac{1}{\sqrt{1-\beta^2}}$. Third, we rearrange the elements of our event so that time comes first. Thus, the Lorentz transformations are now as follows:

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \tag{25}$$

This is our final form. Burn it permanently in your mind. You will apply it over and over and over again.

2.3 More geometry: invariants, intervals, four-vectors

There are a couple immediate mathematical implications that are crucial to note. First, note that c is a kind of speed limit of the universe. If the velocity of a frame is larger than c, then γ becomes imaginary, which is unphysical. Thus, nothing can go faster than the speed of light. Second, note that the Galilean transformations actually have the same form as the Lorentz transformations – indeed, the Lorentz transformations reduce to the Galilean transformations if we assume light is infinitely fast (verify this for yourself). Third, and perhaps most crucially, the "distance" between events is conserved by the Lorentz transformations, but we have to define the notion of "distance" in space-time a little differently than in normal Euclidean space:

Exercise 2.3. *If an event has coordinates (ct, x, y, z), show that the quantity $-(ct)^2 + x^2 + y^2 + z^2$ doesn't change when you apply the Lorentz transformations.*

Unsurprisingly, this quantity is often referred to as the "invariant" of an event. Pay close attention to that negative sign! That negative sign is the reason for all the weirdness in special relativity, as you will soon see. One useful property of the invariant is that if we add or subtract the coordinates of any two events, we get another vector which has an invariant of the same form:

Exercise 2.4. *Suppose we have two events $E_1 = (ct_1, x_1, y_1, z_1,)$ and $E_2 = (ct_2, x_2, y_2, z_2,)$. Show that the quantity $I = -(\Delta ct)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2$ is invariant under the Lorentz transformations.*

We call I the *interval* betwen two events in spacetime. This result suggests the intuitive conclusion that if we add or subtract events, we get another event back. Of course, this property should hold for not only events, but also things like velocity and momentum as well. Thus, we can generalize the notion of an "event" to cover all sorts of physical quantities in relativity by defining a mathematical object called a "four-vector." Note the superscript indexes instead of the usual subscript indexes!

Definition 2.1. *A **four-vector** is a vector (x^0, x^1, x^2, x^3) that has the following two properties:*
***Linearity:** If v and w are four vectors, then $av + bw$ is a four-vector for any choice of real numbers a and b.*
***Invariant:** $-(x^0)^2 + (x^1)^2 + (x^2)^2 + (x^3)^2$ is invariant under the Lorentz transformations.*

Of course, don't just take my word for it:

Exercise 2.5. *Prove that all vectors that obey the Lorentz transformations have the two four-vector properties.*

Thus, the interval between two events has the same form as the invariant of a single event since linearity implies that $E_1 - E_2$ is also a four-vector. For obvious reasons, we call the difference $E_1 - E_2$ between two events the *displacement four-vector.*

A mathematical aside: To keep track of the negative sign in the invariant, we can introduce some notation. We call a vector indexed with superscript numbers instead of subscript numbers *contravariant*, and vectors with subscript indexes are called *covariant*. We call the first component x^0 of a four-vector the *time-like* part of the four-vector, and the other three components the *spacelike* part. We define $x^0 = -x_0$, $x^1 = x_1$, $x^2 = x_2$, $x^3 = x_3$. Note that we only introduce a minus sign for the time-like part. Using this mathematical notation, we can define the analog of the Euclidean inner product for four-vectors and write the invariant compactly. We define the *Minkowski inner product* of two four-vectors $x = (x^0, x^1, x^2, x^3)$ and $y = (y^0, y^1, y^2, y^3)$ to be $\sum_{\mu=0}^{3} x_\mu y^\mu$. Using the Einstein summation notation, which specifies that repeated indices are to be summed over all possible values, we can rewrite this as simply $x_\mu y^\mu$. Thus, the invariant of a four vector x is simply $x_\mu x^\mu$. Compare with the Euclidean inner product $x_\mu x_\mu$: it's different! When we use the Minkowski inner product instead of the Euclidean inner product, we get *Minkowski space* instead of Euclidean space: this is the geometry of space-time. You should note that, unlike the dot product, the Minkowski inner product has no simple geometric meaning. For example, vectors which lie along the line $x = ct$ have invariant zero, but we can't really say those vectors are "perpendicular" to themselves. Moreover, observe that Minkowski space can be treated just like Euclidean space if we multiply the timelike part of a four-vector by the imaginary unit i. We will not, however, use this convention.

Finally, at long last, we have all the mathematical tools we need to address some real physical phenomena! The next section is an optional note highlighting the parallels between rotations in Euclidean space and rotations in Minkowski space.

2.4 Optional note: rotations and rapidity

First, we will recall some facts about rotations in Euclidean space. In standard two-dimensional Euclidean space (i.e. the xy-plane), we can rotate vectors by multiplying them by a matrix. Indeed, if we want to rotate a vector v in the xy-plane through an angle θ in the counter-clockwise direction, we use the following transformation:

$$\begin{aligned} v_\theta &= \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\ &= \begin{pmatrix} x\cos\theta + y\sin\theta \\ -x\sin\theta + y\cos\theta \end{pmatrix} \end{aligned}$$

This transformation preserves the length of the vector. To show this, we use the well-known trig identity $\sin^2\theta + \cos^2\theta = 1$:

$$\begin{aligned} (x\cos\theta + y\sin\theta)^2 + (-x\sin\theta + y\cos\theta)^2 &= x^2\cos^2\theta + 2xy\cos\theta\sin\theta + y^2\sin^2\theta \\ &\quad +x^2\sin^2\theta - 2xy\cos\theta\sin\theta + y^2\cos^2\theta \\ &= x^2(\cos^2\theta + \sin^2\theta) + y^2(\cos^2\theta + \sin^2\theta) \\ &= x^2 + y^2 \\ &= d^2 \end{aligned}$$

If we take the fundamental operation on Euclidean space to be rotations and define the invariant to be the length of a vector, vectors in Euclidean space have the same properties as four-vectors in Minkowski space! (I leave it to you to show linearity.) The fact that the Lorentz transformation matrix vaguely looks like a rotation matrix is no mistake. Just as regular Euclidean rotations preserve the Euclidean inner product, the Lorentz transformations preserve the Minkowski inner product. In this sense, the Lorentz transformations are the analog of rotations in Minkowski space. The extra negative sign in the form of the Lorentz transformation matrix motivates us to consider using hyperbolic trigonometric functions to make the matrix look more like a rotation. Indeed, we define the "rapidity" of the particle to be B so that $\tanh B = \beta$. Then, the Lorentz transformations can be written as:

$$\begin{pmatrix} ct' \\ x' \end{pmatrix} = \begin{pmatrix} \cosh B & -\sinh B \\ -\sinh B & \cosh B \end{pmatrix} \begin{pmatrix} ct \\ x \end{pmatrix} \tag{26}$$

This might look kind of contrived, but it is very natural in a couple ways. First, B ranges from $-\infty$ to ∞ instead of just $-c$ to c, which is more in line with our intuition that the domain of velocities shouldn't be artificially restricted. Second, if you apply consecutive Lorentz transformations, there's the nice property that the composite Lorentz transformation has rapidity equal to the sum of the rapidities of the two consitutent transformations. Verify this fact for yourself using the usual hyperbolic trigonometric identities, and compare this with the velocity addition formula in section 3.2 – this formulation is much, much nicer.

Remember that rapidity isn't any sort of physical quantity. It is an imaginary mathematical construct, but it highlights the reason why the Lorentz transformations are deeply related to rotations in Euclidean space: it's because the Lorentz transformations are rotations in Minkowski space. They form a group, just like SO_2 in two-dimensional Euclidean space.

3 Relativistic kinematics

From now on, we really don't care about the y or z directions, so we're going to omit those parts and only write four-vectors with one time-like and space-like component.

3.1 Visualizing Minkowski space: space-time diagrams, non-simultaneity, causality

So far, our discussion of relativity has been pretty abstract. We can make our equations a little more concrete by drawing diagrams of space-time. Under our choice of axes, we have only one relevant spacelike component and one timelike component, so we can draw out a particle's path in the xy-plane. By convention, we take the y-axis to be ct and the x-axis to be x. It is also useful to represent light pulses in the diagram. Since light pulses move at the speed of light, we have $x = \pm ct$, so they travel through space-time on lines of slope 1 and -1.

The first, and most fundamental question, that we must answer is: what events happen in the past (that is, $t < 0$ **in all frames**), and what events happen in the future? ($t > 0$ **in all frames**). This is actually not so obvious: consider the two events $E_1 = (0, 0)$ and $E_2 = (0, x)$ in some frame - in this frame, they appear to happen at the same time. An observer with velocity v relative to this frame, however, will observer these events at $E_1' = (0, 0)$ and $E_2' = (-\gamma\beta x, \gamma x)$ – verify this for yourself by applying the Lorentz transformations. Since γ is always positive, and since β ranges from -1 to 1, this means that E_2' could seem to happen before E_1' in some frame, and after E_1' in others! This seems to raise troubling implications about the nature of cause and effect. If the ordering of events can change, how can we say that any event can *cause* any other event. An event can't be influenced by some other event that happens after it! Here, our space-time diagram is very useful in resolving the paradox. We will classify all events by their invariants.

Suppose an event has invariant $-(ct)^2 + x^2 < 0$. Then, time must be nonzero, since x^2 is always a positive number. Thus, events with negative invariant happen in the definite past or the definite future. If we plot $-(ct)^2 + x^2 = k$ for any negative number k, this is confirmed. We get a hyperbola with two branches: one below the x-axis, and one above the x-axis. These represent all the possible space-time coordinates for an event with that invariant. Note that the asymptotes of both branches are the two lines we drew for light, and the hyperbola lies entirely inside the light cone. This is good, since it confirms that any valid inertial frame of reference must have $v < c$. Furthermore, since the two branches of the hyperbola are disconnected, the timelike part of the displacement four-vector between any two events with negative interval always has the same sign – if the displacement four-vector between two events lies on the one branch of the hyperbola, no Lorentz transformation can move it to the other branch. In plain language: one of the events is observed to happen before the other in all reference frames. Thus, we call events which have negative interval *causally connected*. It is possible for the first event to influence the second since the order of the events never changes.

Now, suppose the invariant is greater than zero. Plotting $-(ct)^2 + x^2 = k$ for some positive k, we find that the hyperbola is not oriented up and down anymore – the two branches of our hyperbola are oriented left and right. One branch is to the right of the ct axis, the other to the left, and both branches intersect points on all possible values of time. As we discussed earlier, any event here can happen in the past or in the future, depending on who you ask. Since these events don't occur in the definite past or definite future, we say that they happen "elsewhere". Note that the hyperbola never enters the light cone, confirming our intuition that events that happen "elsewhere" can never cross into the definite past or future. In terms of intervals, since the time between two events with positive interval can take on any value, the ordering of these events is not fixed. Either can happen before the other one, depending on who you ask. The sign of the timelike part of the displacement four-vector between two events with positive interval can change.

Finally, if the invariant is zero, we have $-(ct)^2 + x^2 = 0$, which implies $ct = \pm x$. That is, these events lie along the boundary of the light cone. Two events with this interval are only connected by signals that travel at the speed of light. Although the physical interpretation of this case is difficult to imagine, as you will see later on when we discuss mass and energy, this severely limits what kind of information can be transmitted between events with zero interval.

For these reasons, we refer to four-vectors with negative invariant *timelike*, four-vectors with positive invariant *spacelike*, and four-vectors with zero invariant *lightlike*. Similarly, if the displacement four-vector of two events has negative invariant, we say those two events are separated by a timelike interval, and so on.

3.2 Time dilation, length contraction, velocity addition

Finally, some physical applications! First, we need to define some terminology:

Definition 3.1. *The **proper time** τ between two events is the length of time measured by an observer who sees these two events happen in the same place in space.*

Definition 3.2. *The **proper length** L of some object is the length measured by an observer who sees the object at rest.*

Our first strange phenomenon is time dilation. Suppose the proper time between two events is τ. Then, the displacement four-vector between these events is $(c\tau, 0)$. Lorentz transform this four-vector into some other frame moving at a velocity v. The four-vector becomes $(\gamma c\tau, \gamma\beta c\tau)$. Thus, we have $t = \gamma\tau$. Since γ is always greater than 1, this means that an observer in some frame inertial to the rest frame will always observe that it takes more time for processes to occur in the rest frame. A popular way of stating this result in plain language is "moving clocks run slow."

Exercise 3.1. *The usual way less hardcore treatments of the subject motivate time dilation is by using a geometric argument with a light-clock. Suppose we mount two mirrors on the ceiling and the floor of a train moving with velocity v with respect to the ground, separated by a distance d. Observers in both the rest frame and the ground frame see a pulse of light start traveling up from the floor mirror at time $t = 0$. To the observer in the rest frame, the light travels straight up and down, but in the ground frame, the light does not. Using only the constancy of the speed of light and some geometry (i.e. don't use the Lorentz transformations), show that the time dilation formula works.*

We can see the effects of time dilation through spacetime diagrams, too. Suppose an observer sees some object moving with velocity $v < c$. Plot this line on a spacetime diagram. Then suppose we label an event at regularly spaced intervals in time on this line. Using the hyperbolas from the previous section, find the space-time coordinates of these points in the frame of the object – i.e. the frame where they all happen at $x = 0$. Note that the events are more closely spaced. The observer measures a longer time between events than the object does.

Also, note that if the velocity of the object is not close to c, γ is close to 1, so t is close to τ. This is why we don't notice the effects of time dilation in our everyday lives. It simply doesn't become important until objects start moving very, very fast. Even if an object is moving at a tenth the speed of light, we only see a correction on the order of about a half-percent.

The next phenomenon we're going to study is length contraction. Suppose an observer sees an object with proper length L whizzing past him at a velocity v. To find the length of this object in his frame, suppose we make a measurement at each end at the same time. The displacement four-vector between these events is $(0, l)$, where lowercase l is the measured length. Using the Lorentz transformation, the displacement four-vector becomes $(\gamma\beta l, \gamma l)$ in the rest frame. Thus, we have $L = \gamma l$, or $l = \dfrac{L}{\gamma}$. Since $\gamma > 1$, we conclude that the object looks shorter to the observer.

Exercise 3.2. *Extend the thought experiment given in the previous exercise to motivate the length contraction formula without use of the Lorentz transformations.*

As before, it's worth noting that for speeds that are not close to the speed of light, length contraction is negligible.

3.3 Velocity addition

Now, we deal with how velocities add in relativity. To clarify what this actually means, we present an example. Suppose we observe two particles moving at speed $0.99c$ in opposite directions. In the frame of one particle, what is the velocity of the other particle? Clearly, it cannot be $1.98c$, because that would violate the cosmic speed limit we derived earlier. The Lorentz transformations should apparently tell us what the answer is, but how? The answer: we need a four-vector to represent the particle's velocity. We call it the *four-velocity.* It is not so clear, however, how we should define it. To motivate our definition, we appeal to another example:

Suppose we are standing on the ground and we see an airplane flying overhead. What exactly do we mean when we say it has some velocity v? Recall from classical physics that velocity is $v = \frac{dx}{dt}$. This definition still works in relativity, but we should note that both x and t are measured in our frame, not the plane's frame. Neither of these are proper quantities. We call this velocity the *ordinary velocity*, which we denote u. Of course, in relativity, the displacement between events has a timelike part ct, but we know that ct divided by t is just c, so u has components (c, u_x). Although it's not always the case that the velocity of the particle has no y or z component, we're ignoring them for now. It'll be an exercise for you to figure out what happens in that case!

This meaning of velocity, however, is not the only one. We can also define how much distance *in our frame* x the plane covers per unit *proper time* τ. This is weird – we're measuring distance in some external frame, and time in the rest frame! We call this interpretation of velocity the *proper velocity*, which we denote η. To make the difference between these two definitions of velocity concrete, think about what each velocity means to a traveler on the plane. The ordinary velocity will tell him how long his flight will seem to an observer on the ground, but the proper velocity tells him how long the flight will seem to him.

When we define four-velocity, we want to use proper velocity. The reason for this is immediate: proper time never changes, but the time an observer measures depends on who you ask! Using proper velocity instead of four-velocity therefore simplifies calculations a great deal, because you only have to Lorentz transform the displacement, not the time. Indeed, since we know from the previous section that $t = \gamma\tau$, we have the relationship $\eta = \gamma u$. We thus have $\eta = (\gamma c, \gamma u_x)$. Since we often deal with multiple velocities in relativity, we can easily confuse our γ's if we're not careful. Thus, we write γ_u and β_u to mean that we plug in the value u to the equations for γ and β. That is $\gamma_u = \dfrac{1}{\sqrt{1 - u^2/c^2}}$ and $\beta_u = u/c$. Thus, our final form for the four-velocity is $\eta = (\gamma_u c, \gamma_u u)$. As you can confirm, this obeys all the Lorentz transformations, just like events.

Exercise 3.3. *What is the invariant of four-velocity? The easiest way to do this is to consider the frame where the particle is at rest, i.e.* $\gamma = 1$*.*

We can use η to solve our original question about the two particles. To be general, suppose the particles are moving with velocities u and v. The four-velocity of one particle is $\eta = (\gamma_u c, \gamma_u u)$ to an external observer. Lorentz transforming into a frame with velocity v (i.e. the frame of the other particle), we have:

$$\eta' = (\gamma_{u'} c, \gamma_{u'} u') = (\gamma_v(\gamma_u c - \beta_v \gamma_u u), \gamma_v(\gamma_u u - \beta_v \gamma_u c))$$

Solving for u':

$$u' = c\frac{\gamma_{u'} u'}{\gamma_{u'} c} = c\frac{\gamma_v(\gamma_u u - \beta_v \gamma_u c)}{\gamma_v(\gamma_u c - \beta_v \gamma_u u)} = \frac{u - v}{1 - \frac{uv}{c^2}}$$

Taking $u = 0.99c$ and $v = -0.99c$, we have $u' = 0.999949498c$. Less than the speed of light!

Exercise 3.4. *So far, we've been assuming that velocity is always going to be parallel to the relative motion of the reference frames. This will not always be the case. Suppose the velocity of one of the particles has a* y *component and a* z *component. Find how these transform into the frame of the other particle. Here's the answers, just so you can check for yourself:*

$$u'_x = \frac{u_x - v}{(1 - \frac{u_x v}{c^2})}$$
$$u'_y = \frac{u_y}{\gamma_v(1 - \frac{u_x v}{c^2})}$$
$$u'_z = \frac{u_z}{\gamma_v(1 - \frac{u_x v}{c^2})}$$

Not to sound like a broken record, but you should verify that all these formulas reduce to the Galilean formulas when the velocities studied are not close to c.

3.4 Famous paradoxes

With just these results, we can state a couple famous relativity paradoxes. I hope the discussion so far is thorough enough so that you can produce the solutions to these on your own (not to mention it's a good idea to think through relativity problems yourself to get used to them), so they aren't given here:

Exercise 3.5. *(Laser beam nuking an ant) - Suppose an ant is constrained to run along the x-axis only. Its initial position is* $(0, x)$*. At time* $t = 0$ *a powerful laser is turned on at* $x = 0$*. The laser beam moves at the speed of light c, but of course, the ant can never be that fast. However, the ant can still manage to live for an arbitrarily long time (assuming it has infinite energy). How?*

Exercise 3.6. *(Barn-pole paradox) - Suppose a farmer running at velocity* v *close to* c *is carrying a ten meter pole. He is running towards a barn that has length five meters To the farmer, the barn has length less than five meters, by length contraction, so he concludes he can't fit the pole into the barn. To an observer in the barn, the pole is length contracted, so the pole can fit into the barn. Who is right?*

Exercise 3.7. *(Doomsday button) - Suppose we have a t-shaped piece of metal, where the vertical bar of the T has length two meters. Also, suppose that we have a u-shaped piece of metal with height one meter. The width of the u-shaped piece is small enough so that only the vertical part of the t-shaped piece can fit inside the u-shaped piece. At the bottom of the u-shaped piece is a button that will cause a bomb to explode. The t-shaped piece moves toward the u-shaped piece at a velocity* v *close to c. To the u-shaped piece, the t-shaped piece is length contracted, so the bomb will not go off. To the t-shaped piece, the u-shaped piece is length contracted, so the bomb will go off. Who is right?*

Exercise 3.8. *(Twin paradox) - Suppose we have two twins. One leaves Earth in a rocketship going at a velocity* v *close to c, and the other stays behind. The traveling twin travels for some distance and then turns around and comes home. To the twin on Earth, the traveling twin experienced time dilation, and should thus be older than the twin on Earth. To the twin in the rocket, his twin on Earth seemed to experience time dilation, so the twin on Earth should be older. Who is right?*

4 Relativistic dynamics

This section discusses only the most important concepts and results in relativistic dynamics (i.e. the topics which useful at the IPhO). Things like Minkowski forces and relativistic electrodynamics are cool, but I won't touch on them. See the reading suggestions at the end if you're interested in more.

4.1 Mass, energy, momentum: conservation and non-conservation

First, we define four-momentum. The definition is very simple. The four-momentum is ηmc, where m is the mass of the particle. You might notice that these have units of energy, not momentum. This is indeed correct. The old-fashioned convention is to define it to be ηm so that it actually has units of momentum, but as you'll see, adding the extra factor of c makes the results of our discussion a bit nicer. Indeed, consider the timelike part of ηmc. We call it the relativistic energy $E = \gamma mc^2$. Consider a frame where the particle is at rest, i.e. $\gamma = 1$.

Then, the relativistic energy becomes the *rest energy* $E_{rest} = mc^2$ (As a simple exercise, verify that this implies that $-(mc^2)^2$ is the invariant of four-momentum). This equation really ought to look familiar to all of you! This is extremely important because it establishes the equivalence of mass and energy; the difference between the two quantities is simply multiplication by a constant. Since this is the energy we have when the particle is at rest, we can define the kinetic energy of a particle as $E - E_{rest} = mc^2(\gamma - 1)$. As usual, verify that this reduces to the classical equation in the low-velocity limit blah blah blah.

The form of these equations should give you some pause. They are saying that as an object's velocity approaches the speed of light, its energy asymptotically approaches infinity. Thus, it is impossible for a particle with nonzero mass not only to exceed the speed of light, but also to reach it. To do so would require an infinite amount of energy, which is absurd. To reference an earlier section, this means that events separated by a lightlike interval can only communicate information via massless particles moving at c. The only object that satisfies this property is the photon, the particle of light.

Another important equation comes from the equation for the invariant of the four-momentum. Defining the *relativistic momentum* to be $p = \gamma mu$ (i.e. classical momentum times γ), we have $-(mc^2)^2 = -E^2 + p^2c^2$. Rearranging, we have $E^2 = m^2c^4 + p^2c^2$. This result is very useful, because it relates the momentum and energy of a particle without information about the velocity. Indeed, writing $\eta mc = (E, pc)$, it becomes obvious why the four-momentum is often called the *energy-momentum four-vector*.

Of course, the notions of mass, energy, and momentum are only useful since they have very nice properties with respect to invariance and conservation. As a reminder, we call a quantity *invariant* if it has the same value in all reference frames, and we call a quantity *conserved* if its value doesn't change over time in some given frame. Of course, we also stipulate that the system in question is closed when we consider questions of conservation (i.e. there are no external forces). The invariance/conservation of various quantities are given in the table below:

Quantity	Classically invariant?	Classically conserved?	Relativistically invariant?	Relativistically conserved?
Mass	Yes	Yes	Yes	**No**
Energy	No	Yes	No	Yes
Velocity	No	No	No	No
Momentum	No	Yes	No	Yes
Electric Charge	Yes	Yes	Yes	Yes

The non-conservation of mass should seem weird to you, but you just need to wrap your head around the concept that mass and energy are, in a very literal sense, equivalent. It is relativistic energy, not mass, that must be conserved. To see this in action, we give an example. Suppose that some object with mass 20 kg moving at $0.6c$ collides with another object of mass 30 kg at rest and sticks to it. Assuming no energy is released, what is the final mass and velocity of the pair of objects? To do this, we simply impose the conservation of four-momentum. In the first component:

$$\frac{20c^2}{\sqrt{1-0.6^2}} + 30c^2 = \frac{Mc^2}{\sqrt{1-v^2/c^2}}$$

In the second component:

$$\frac{(20)(0.6c)c}{\sqrt{1-0.6^2}} + 0 = \frac{Mvc}{\sqrt{1-v^2/c^2}}$$

Two equations, two unknowns. We know how to do this. After a bit of algebraic manipulation, we find $M = 52.9$ kg, and $v = 0.273c$. Observe that the combined mass is larger than the sum of the two original masses.

Some more oddness comes into the mix when we try to find the four-momentum of light itself. For a photon, γ is infinite, but the mass of a photon is 0, so it is at least plausible that the energy of a photon is finite. Mathematically, this makes no sense, but since the photon exists anyway and has finite energy, it's true. The four momentum of a photon is simply (E, E). Using Planck's well known quantum-mechanical relation $E = hf$ for a photon, the four-momentum of a photon is simply (hf, hf). More oddness still:

Exercise 4.1. *Suppose a proton and an antiproton annihilate (i.e., they collide with each other and are converted to pure energy in the form of photons). Why isn't it okay for only one photon to be emitted in this interaction?*

As a final note for this section, be aware that in your own reading, you may encounter a quantity called relativistic mass given by γm, where m is the "rest mass" (the mass of the particle measured in a frame where it is stationary). This is a very old-fashioned construct which dates way back to Einstein's original 1905 paper, and it has fallen out of use in modern texts because frankly, it stinks. The notion of relativistic energy is not only equivalent, but also more useful and less confusing. When we use relativistic energy instead of relativistic mass, we don't have to distinguish two different kinds of mass (the only mass there is is rest mass), and the intrinsic relationship between energy and mass becomes obvious.

4.2 Relativistic Doppler shift, Compton scattering

To continue our discussion of photons, we derive a couple more phenomena related to light. First is the relativistic Doppler shift. Recall the classical Doppler shift:

$$\frac{f_{observed}}{f_{source}} = \frac{v_{source}}{v_{observed}}$$

To derive the relativistic shift, we merely apply the Lorentz transformations to the four-momentum of a photon: $hf' = \gamma(hf - \beta hf)$. Cancelling and rearranging, we have $f'/f = \gamma(1-\beta) = \sqrt{\frac{1-\beta}{1+\beta}}$. That was easy.

The Compton effect is a little more difficult. Suppose we have a photon with energy E_0 that collides with some particle of mass m initially at rest. It gets absorbed and re-emitted at an angle θ to the original direction. We want to find the energy E_1 of the re-emitted photon. For simplicity, we choose our axes so that the photon initially travels along the x-axis. Then, we use the conservation of four-momentum:

$$\begin{aligned}
(\eta mc)_{\text{photon, before}} + (\eta mc)_{\text{particle, before}} &= (\eta mc)_{\text{photon, after}} + (\eta mc)_{\text{particle, after}} \\
\implies (\eta mc)_{\text{particle, after}} &= (\eta mc)_{\text{photon, before}} + (\eta mc)_{\text{particle, before}} - (\eta mc)_{\text{photon, before}} \\
&= (E_0, E_0, 0, 0) + (mc^2, 0, 0, 0) - (E_1, E_1\cos\theta, E_1\sin\theta, 0) \\
&= (E_0 + mc^2 - E_1, E_0 - E_1\cos\theta, -E_1\sin\theta, 0)
\end{aligned}$$

We then take a deep breath and use the fact that the four-momentum of the particle has invariant $-(mc^2)^2$. I will not show all the steps of the algebra. I trust that you know how to multiply large polynomials and know some basic trig identities. Of course, you should verify all of this for yourself.

$$\begin{aligned}
-(mc^2)^2 &= -(E_0 + mc^2 - E_1)^2 + (E_0 - E_1\cos\theta)^2 + (-E_1\sin\theta)^2 \\
&= \cdots \\
&= -(mc^2)^2 - 2E_0 mc^2 + 2E_0E_1 + 2E_1 mc^2 - 2E_0E_1\cos\theta \\
\implies \frac{1}{E_1} - \frac{1}{E_0} &= \frac{1}{mc^2}(1-\cos\theta)
\end{aligned}$$

Using the well known equation for the wavelength of a photon $\lambda = \frac{hc}{E}$, we have

$$\lambda_1 - \lambda_0 = \frac{h}{mc}(1 - \cos\theta)$$

This is the Compton scattering result. Note that the right hand side has dimensions of wavelength. Thus, we call $\frac{h}{mc}$ the *Compton wavelength* of the particle.

These are all the results of special relativity you'll need for contests (and basically everything you'd learn in a one-semester intro treatment of the subject). Of course, one can introduce relativistic corrections to virtually every other area of physics, but now, you should have a deep understanding of the core concepts and results in special relativity. To close, we'll tie up some odds and ends and give a tricky question to see our principles at work.

4.3 Natural units, an example

You may have noticed that we seem to write down the speed of light c a lot in our equations. We physicists, however, can't deal with that kind of redundancy. We always want to make our equations as dense and concise as possible. To this end, the $c = 1$ convention is ubiquitous in relativity. Essentially, we redefine our system of units so that the speed of light is equal to 1. Velocities are measured in fractions of the speed of light. Distances are measured in light-seconds, and so on. This convention not only makes things less tedious to write out, but it also makes our equations look a lot nicer. For example, the energy-mass equivalence statement becomes $E = m$. It's hard to make equivalence any more obvious. This convention often confuses students who have never learned relativity before, but it is important to get used to it. Suppress the voice in your head that complains about the seemingly inconsistent units that we're using. You can always multiply by c at the end!

To illustrate, we will walk you through a difficult problem from the 2003 International Physics Olympiad:

Exercise 4.2. *(IPhO 2003, 3A) A free neutron of mass m_n decays at rest in the laboratory frame of reference into three non-interacting particles: a proton, an electron, and an anti-neutrino. The rest mass of the proton is m_p, while the rest mass of the anti-neutrino m_ν is assumed to be nonzero and much smaller than the rest mass of the electron m_e. Denote the speed of light in vacuum by c. The measured values of mass are as follows: $m_n = 939.56563 MeV/c^2$, $m_p = 938.27231 MeV/c^2$, $m_e = 0.5109907 MeV/c^2$*

In the following, all energies and velocities are referred to the laboratory frame. Let E be the total energy of the electron coming out of the decay. Find the maximum possible value E_{max} of E and the speed ν_m of the anti-neutrino when $E = E_{max}$. Both answers must be expressed in terms of the rest masses of the particles and the speed of light. Given that $m_v < 7.3 eV/c^2$, compute E_{max} and the ratio ν_m/c to 3 significant digits.

First, we make an observation. By conservation of momentum, the velocity (and therefore the energy) of the electron is at a maximum when the proton and antineutrino are headed directly away from the electron. Then, we use a trick. Since the four-momentum is a four-vector, we can add the proton and neutron four-momenta to get another four-momentum. Call this combined four-momentum (E_x, p_x). Note that by doing this, we're essentially considering the center of mass of the proton and anti-neutrino. We also have the electron four-momentum (E_e, p_e). Since the initial velocity of the neutron was 0, we can apply conservation of four-momentum in the space-like part to get $p_x = p_e = p$. Writing the invariants of these two four-momenta with our new convention, we have:

$$\begin{aligned} -m_x^2 &= E_x^2 + p^2 \\ -m_e^2 &= E_e^2 + p^2 \end{aligned}$$

Subtracting, we have $m_e^2 - m_x^2 = E_e^2 - E_x^2$. Conservation of energy tells us $E_x = E_n - E_e$ (where n stands for neutron), and since $m_\nu \ll m_P$ by 9 orders of magnitude, we make the approximation that $m_x \approx m_p$. Thus, combining all these equations, we have:

$$\begin{aligned}
m_e^2 - m_x^2 &= E_e^2 - E_x^2 \\
\implies m_e^2 - m_p^2 &= E_e^2 - (E_n - E_e)^2 \\
m_e^2 - m_p^2 &= E_e^2 - E_n^2 + 2E_nE_e - E_e^2 \\
\implies E_e &= \frac{m_e^2 - m_p^2 - E_n^2}{2E_n}
\end{aligned}$$

Now, we just plug in all of our values, noting that $E_n = m_n$ since the neutron was at reast in the beginning:

$$\begin{aligned}
E_e &= \frac{m_e^2 - m_p^2 - E_n^2}{2E_n} \\
&= 1.29257MeV
\end{aligned}$$

To get the other part of the problem, we first find the electron velocity:

$$\begin{aligned}
E_e = \gamma m_e &= 1.29257MeV \\
\implies \gamma &= 2.52953
\end{aligned}$$

Now, we note that, from the form of our earlier equations, E_e was maximized by minimizing E_x. Since E_x is defined relative to the center of mass of the proton and antineutrino, the "rest mass" of our imaginary x particle is simply the total energy of the proton and antineutrino in the center of mass frame. To minimize this rest mass, we need the proton and anti-neutrino to travel at the same velocity so that $\gamma = 1$ for both particles in this frame. Thus, we can directly apply conservation of four-momentum, this time considering the time-like part. Using our approximation $m_x \approx m_p$ again, we find:

$$\begin{aligned}
m_n &= (2.52953)m_e + \gamma_{\nu_m} m_x \\
\implies \gamma_{\nu_m} &= 1.0000008 \\
\implies \nu_m &= 0.00126c
\end{aligned}$$

And we're done! I hope you've seen how keeping track of all those c's might have become really annoying.

5 Topics for further reading, book suggestions

If you'd like some books that cover special relativity in more detail than these notes, consider the following: Spacetime Physics: Introduction to Special Relativity by Edwin F. Taylor and John Archibald Wheeler is a battle-tested textbook used by many universities across the country. It's pretty clear, and the problems are good. One criticism often levied on Taylor and Wheeler is that they're not extremely rigorous in most places, but this is forgivable in an introductory treatment. For those of you who really want to see the mathematical tools of special relativity derived from first principles, The Geometry of Minkowski Spacetime by Gregory L. Naber makes a great companion text. Caveat emptor, though: it assumes some familiarity with some abstract algebra and vector calculus. Finally, to plug my university's didactic strategy, MIT uses Special Relativity by A.P. French. I've never read it before, though, since French's book on waves and oscillations sucks hardcore. Beyond this, there are myriad applications of relativity that I haven't covered because of lack of time, or because I couldn't assume the level of mathematical knowledge necessary to discuss them, etc. Two of the most accessible applications are relativistic electrodynamics and general relativity:

Relativistic electrodynamics is particularly natural because the inadequacy of classical physics in explaning electrodynamical phenomena was the impetus for developing special relativity in the first place. David Griffiths'

Introduction to Electrodynamics is great for learning about this. The vast majority of the book is about classical electrodynamics, but knowledge of classical electrodynamics is necessary to fully appreciate the results of relativistic electrodynamics, which is discussed at the end. To reiterate, I strongly recommend you buy this book. Griffiths is a canonical, standard textbook in physics education because it's so good: the exposition is casual, but always rigorous, the problems are challenging, and the field is covered comprehensively. As a nice bonus, it covers some details of relativistic dynamics that I left out earlier, and it gives a slightly different perspective on the derivation and application of the Lorentz transformations than these lecture notes.

Then, if you are bold and fearless, you can try learning about general relativity (GR). Before you undertake this, you need to know about classical mechanics at an advanced undergraduate level. Classical Mechanics by Goldstein is the canonical textbook for this. As for actual GR books, Gravitation by Misner, Thorne, and Wheeler (universally referred to as MTW) is a behemoth of a textbook, but it's considered *the* authoritative, comprehensive treatment of the subject. Be aware that it's intended to be a graduate level textbook and the math is very intense, so unless you know a lot of analysis and topology, you'll probably be blindsided. General Relativity by Robert Wald is considered one of the best references on the topic. It is also very mathematical, but it also covers some neat applications in thermodynamics and cosmology. For actually learning GR for the first time, I've heard great things about Sean Carroll's An Introduction to General Relativity: Spacetime and Geometry. It's what MIT's first GR class uses.

Then, there's things like relativistic quantum field theory, quantum electrodynamics, quantum chromodynamics, and various other applications in high-energy physics, but you should probably know a great deal of quantum mechanics (and actually understand the Lagrangian and Hamiltonian formulations of classical mechanics) before you tackle those.

6 Acknowledgements

These notes are based on the U.S. Physics Team's horribly pun-laden relativity lecture notes, courtesy of Andrew Lin. Some ideas and explanations were taken from David Griffiths' classic book Introduction to Electrodynamics. The idea for using group structure in the derivation of the Lorentz transformations came from Victor Yakovenko.