

SUPPLEMENTAL LECTURE II: SPECIAL RELATIVITY IN TENSOR NOTATION

© Joel C. Corbo, 2005

This set of notes accompanied the second in a series of “fun” lectures about relativity given during the Fall 2005 Physics H7C course at UC Berkeley. It is an introduction to the tensor formulation of special relativity and is meant for undergraduates who have had an introduction to the basic concepts of special relativity.

1 Index Notation in Three Dimensions

Index notation is a powerful tool that greatly simplifies the math involved in dealing with vector quantities. Before introducing the full machinery of index notation in four-dimensional spacetime, it is worth taking the time to look at it in three dimensions where things are more familiar.

1.1 Three-Vectors and Spatial Rotations

Before discussing index notation, we will take the time to recall what we mean when we call a quantity a vector; this material should be familiar, but the concepts involved will come in handy when trying to understand what we mean by a four-vector.

Given three-dimensional Euclidean space, we can define a coordinate system, known as *Cartesian coordinates*, involving three mutually-perpendicular axes commonly referred to as the x -axis, y -axis, and z -axis. Any point P in this space can then be represented by a set of three numbers x , y , and z , the *coordinates* of the point, which are interpreted as giving the position of the point in space. The *position vector* \mathbf{x}^* is a directed line segment, starting at the origin and ending at P . Because there is one unique vector defined for each point in space, we can interpret this vector as giving the position of the point, just as its coordinates do.

*In these notes, I will denote three-vectors in boldface instead of with an arrow.

Suppose that we rotate our coordinate system by an angle θ about the z -axis. Our vector \mathbf{x} will have new components x' , y' , and z' related to the old components by

$$\begin{aligned}x' &= x \cos \theta + y \sin \theta \\y' &= -x \sin \theta + y \cos \theta \\z' &= z.\end{aligned}\tag{1}$$

However, even though the components of our vector have changed, the length L of the vector, given by the square root of the sum of the squares of the components, remains unchanged. That this is true should be pretty obvious, but let's verify it with a calculation anyway:

$$\begin{aligned}L^2 &= (x')^2 + (y')^2 + (z')^2 \\&= (x \cos \theta + y \sin \theta)^2 + (-x \sin \theta + y \cos \theta)^2 + z^2 \\&= x^2 \cos^2 \theta + y^2 \sin^2 \theta + 2xy \sin \theta \cos \theta \\&\quad + x^2 \sin^2 \theta + y^2 \cos^2 \theta - 2xy \sin \theta \cos \theta + z^2 \\&= x^2 + y^2 + z^2.\end{aligned}\tag{2}$$

Thus, we see explicitly that rotations leave the length of this vector invariant[†].

Technically, we have not yet talked about arbitrary three-vectors, but only about the position three-vector. However, the generalization is straightforward: a three-vector \mathbf{A} is a three-component object whose components transform in the manner of a position three-vector, that is, in the manner described by Eq. (1). Its length, defined in the same way as that of the position three vector, is invariant under spatial rotations. Thus, what we define as a vector is a direct consequence of the properties of the space in which we define it, because it is those properties that determine the transformation law given in Eq. (1). Again, this statement is totally trivial in Euclidean space, but it is useful to keep in mind when we step out of Euclidean space.

Before heading down that road, we will make note of a convenient way to write a vector in terms of its components. As we indicated above, Cartesian space is built up out of three coordinate axes. Each of these axes as associated with is a *unit vector*;

[†]You may be worried that we have only shown that vectors are invariant under rotations about the z -axis. However, for rotation about an arbitrary axis, we can always transform our coordinate system such that the rotation axis becomes the z -axis.

we will write these as \mathbf{e}_x , \mathbf{e}_y , and \mathbf{e}_z . With this notation, we can write an arbitrary vector \mathbf{A} as

$$\mathbf{A} = A_x \mathbf{e}_x + A_y \mathbf{e}_y + A_z \mathbf{e}_z. \quad (3)$$

This notation makes distinct the unit vectors and the components of the vector, which are scalars and so are not written in boldface.

In addition to this notation, we can also write a vector as a column matrix. If we all agree in advance which coordinate system we are using (Cartesian versus spherical, for example), then all of the information in a vector is given by its components. Therefore, we can write

$$\mathbf{A} = \begin{pmatrix} A_x \\ A_y \\ A_z \end{pmatrix}. \quad (4)$$

This allows us a convenient way to write the transformation in Eq. (1) is as a matrix equation:

$$\begin{pmatrix} A'_x \\ A'_y \\ A'_z \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} A_x \\ A_y \\ A_z \end{pmatrix}. \quad (5)$$

We see that if we expand out the right-hand side of this equation, we reproduce the desired transformation laws. In even nicer notation, we can write the 3x3 matrix representing the rotation about the z -axis as \mathcal{R}_z . Then, we can rewrite the matrix equation as a vector equation:

$$\mathbf{x}' = \mathcal{R}_z \mathbf{x}. \quad (6)$$

Now our transformation law contains just three objects: our original and transformed vectors and a rotation matrix that contains all of the information about the transformation. This equation also illustrates a simple but important fact about vector equations: if there is a vector quantity on one side of an equation, there must be a vector quantity on the other side of the equation. This will be useful to keep in mind when we start representing vectors and matrices as objects with indices.

1.2 Index Notation

In Eq. (3), we wrote the components of the vector \mathbf{A} as A_x , A_y , and A_z . However, it will be to our advantage to write them instead as A_1 , A_2 , and A_3 , where 1 stands for x , 2 for y , and 3 for z . We can use this exact same convention to rewrite the Cartesian

unit vectors. With this convention, Eq. (3) becomes

$$\mathbf{A} = A_1\mathbf{e}_1 + A_2\mathbf{e}_2 + A_3\mathbf{e}_3. \quad (7)$$

This may seem like a trivial change, but it's actually very powerful. To see this, let's further rewrite this expression as

$$\mathbf{A} = \sum_{i=1}^3 A_i\mathbf{e}_i. \quad (8)$$

If we expand out the right-hand side of the above equation using the usual rules for sums, we will reproduce Eq. (3). We have simplified our notation by introducing an *index* i . By introducing this index, we are denoting the i th component of \mathbf{A} as A_i , with the understanding that i can equal only 1, 2, or 3.

Let's look at how this works when applied to the dot product of two vectors \mathbf{A} and \mathbf{B} . Using our 1, 2, 3 notation we can write

$$\begin{aligned} \mathbf{A} \cdot \mathbf{B} = & A_1B_1(\mathbf{e}_1 \cdot \mathbf{e}_1) + A_1B_2(\mathbf{e}_1 \cdot \mathbf{e}_2) + A_1B_3(\mathbf{e}_1 \cdot \mathbf{e}_3) \\ & + A_2B_1(\mathbf{e}_2 \cdot \mathbf{e}_1) + A_2B_2(\mathbf{e}_2 \cdot \mathbf{e}_2) + A_2B_3(\mathbf{e}_2 \cdot \mathbf{e}_3) \\ & + A_3B_1(\mathbf{e}_3 \cdot \mathbf{e}_1) + A_3B_2(\mathbf{e}_3 \cdot \mathbf{e}_2) + A_3B_3(\mathbf{e}_3 \cdot \mathbf{e}_3). \end{aligned} \quad (9)$$

Of course, we know that writing everything on the right-hand side of that expression is a bit silly, because a unit vector dotted with itself equals 1 and a unit vector dotted with a different unit vector, because they are orthogonal by construction, equals 0. However, instead of making that simplification right now, let's instead introduce two indices i and j :

$$\mathbf{A} \cdot \mathbf{B} = \sum_{i=1}^3 \sum_{j=1}^3 A_iB_j(\mathbf{e}_i \cdot \mathbf{e}_j). \quad (10)$$

Once again, if we expand out the right-hand side of the above equation, we will reproduce Eq. (9).

The summation notation is better from the point of view of saving writing, but writing the summation symbol is still a bit annoying. Therefore, let's drop them entirely. Then, our expression for a vector and our expression for a dot product become

$$\mathbf{A} = A_i\mathbf{e}_i, \quad (11)$$

and

$$\mathbf{A} \cdot \mathbf{B} = A_iB_j(\mathbf{e}_i \cdot \mathbf{e}_j). \quad (12)$$

This way of writing a vector expression is called the *Einstein summation convention*. The rule is that every time you see a term in which the same index is repeated twice, you sum over that index from 1 to 3, even though the summation symbol is not explicitly written. Note that it does not matter what letter or symbol is given to the index; all that matters is that if it is repeated twice, it should be summed over[‡].

While we have greatly simplified the expression for our dot product, we have still not made use of the fact that $\mathbf{e}_i \cdot \mathbf{e}_j$ equals either 0 or 1, depending on the values of i and j . In particular, we know that if i equals j , then this expression equals 1, while if i and j are not equal, then this expression equals zero. Given this, we can define a new symbol to replace $\mathbf{e}_i \cdot \mathbf{e}_j$, the *Kronecker delta* δ_{ij} , such that

$$\delta_{ij} \equiv \mathbf{e}_i \cdot \mathbf{e}_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} . \quad (13)$$

Using the Kronecker delta, we can write the dot product that we have been working with as

$$\mathbf{A} \cdot \mathbf{B} = \delta_{ij} A_i B_j = A_i B_i, \quad (14)$$

where in the last step we used the fact that δ_{ij} is only nonzero when i equals j . We see that this answer is reasonable, because if we expand out the right-hand side we get

$$\mathbf{A} \cdot \mathbf{B} = A_i B_i = A_1 B_1 + A_2 B_2 + A_3 B_3, \quad (15)$$

which is, of course, the right expression for the dot product.

There is one last useful point to make about the Kronecker delta. From Eq. (14), we see that we can write

$$\delta_{ij} B_j = B_i, \quad (16)$$

by dropping the A_i . Thus, we have used the Kronecker delta to change the index on a vector component. This may seem like an abuse of notation, but it really is a valid and correct use of the summation convention.

1.2.1 Cross Products with Index Notation

Since we have talked about dot products, we will now move on to cross products. We can write the cross product of the vectors \mathbf{A} and \mathbf{B} as

$$\mathbf{A} \times \mathbf{B} = A_i B_j (\mathbf{e}_i \times \mathbf{e}_j), \quad (17)$$

[‡]In other words, $A_i B_i = A_j B_j = A_{\text{♣}} B_{\text{♣}}$. The symbol used as the index is irrelevant.

where the cross product between the unit vectors can equal either 0, 1, or -1 times the unit vector orthogonal to both of the original ones, which we will call \mathbf{e}_k .

We will play a similar trick to that of introducing the Kronecker delta by introducing the *Levi-Civita symbol*, ϵ_{ijk} . The Levi-Civita symbol can equal 1, -1, or 0 under the following conditions:

$$\epsilon_{ijk} = \begin{cases} 1, & ijk = \{123, 231, 312\} \\ -1, & ijk = \{321, 213, 132\} \\ 0, & \text{otherwise} \end{cases} . \quad (18)$$

So, this object is nonzero only when its indices are all different, and it changes sign when any two of its indices are interchanged. Given this, we can write the cross product above as

$$\mathbf{A} \times \mathbf{B} = \epsilon_{ijk} A_i B_j \mathbf{e}_k. \quad (19)$$

Expanding out the right-hand side, we get

$$\mathbf{A} \times \mathbf{B} = (A_2 B_3 - A_3 B_2) \mathbf{e}_1 + (A_3 B_1 - A_1 B_3) \mathbf{e}_2 + (A_1 B_2 - A_2 B_1) \mathbf{e}_3, \quad (20)$$

which is the correct expression for a cross product. We can also write only a single component of the cross product as

$$\mathbf{A} \times \mathbf{B}|_k = \epsilon_{ijk} A_i B_j, \quad (21)$$

which is sometimes useful in expressions involving a number of cross products at once. We will see an example of this in the next section.

It is often useful to write the product of two Levi-Civita symbols. This can be written as the product of Kronecker deltas:

$$\epsilon_{ijk} \epsilon_{ilm} = \delta_{jl} \delta_{km} - \delta_{jm} \delta_{kl}. \quad (22)$$

It is left as an exercise to the reader to show that this is true.

1.2.2 Vector Calculus with Index Notation

Suppose that we wanted to take the derivative of a vector, that is, the divergence or the curl. We can define the symbol ∂_i to help us. Let's take the divergence as an example:

$$\begin{aligned} \nabla \cdot \mathbf{A} &= \frac{\partial A_1}{\partial x_1} + \frac{\partial A_2}{\partial x_2} + \frac{\partial A_3}{\partial x_3} \\ &= \frac{\partial A_i}{\partial x_i} \\ &\equiv \partial_i A_i. \end{aligned} \quad (23)$$

Here, we have defined a new object, ∂_i , which conveniently acts like a vector whose components are partial derivatives, but which is not an actual vector. We can use it to find any combination of divergences and curls we want simply by treating it as we would any other vector in a dot or cross product. For example, we would write the curl of a vector as

$$\nabla \times \mathbf{A} = \epsilon_{ijk} \partial_i A_j \mathbf{e}_k. \quad (24)$$

All of this notation may seem more trouble than it's worth, but it's actually much easier to use than regular vector calculus once you get used to it. To illustrate this point, let's simplify a particularly nasty quantity, $\nabla \times \nabla \times \mathbf{A}$:

$$\begin{aligned} \nabla \times \nabla \times \mathbf{A} &= \epsilon_{ijk} \partial_i (\nabla \times \mathbf{A})_j \mathbf{e}_k \\ &= \epsilon_{ijk} \partial_i \epsilon_{lmj} \partial_l A_m \mathbf{e}_k \\ &= \epsilon_{ijk} \epsilon_{lmj} \partial_i \partial_l A_m \mathbf{e}_k \\ &= -\epsilon_{jik} \epsilon_{lmj} \partial_i \partial_l A_m \mathbf{e}_k \\ &= \epsilon_{jik} \epsilon_{jml} \partial_i \partial_l A_m \mathbf{e}_k \\ &= (\delta_{im} \delta_{kl} - \delta_{il} \delta_{km}) \partial_i \partial_l A_m \mathbf{e}_k \\ &= \delta_{im} \delta_{kl} \partial_i \partial_l A_m \mathbf{e}_k - \delta_{il} \delta_{km} \partial_i \partial_l A_m \mathbf{e}_k \\ &= \partial_i \partial_l A_i \mathbf{e}_l - \partial_i \partial_i A_m \mathbf{e}_m \\ &= \partial_l \partial_i A_i \mathbf{e}_l - \partial_i \partial_i A_m \mathbf{e}_m \\ &= \nabla (\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}. \end{aligned} \quad (25)$$

This may seem like a lot of steps, but we have been very careful to write out each and every step in detail to help convey how a calculation like this is supposed to proceed[§]. If you've ever calculated this out the long way, you can appreciate how much more simply we arrived at this answer using index notation.

1.3 Index Notation and Matrices

There is a very strong connection between index notation and matrices. Above, we noted that a standard representation of a vector is as a column matrix, and a standard

[§]For example, note that we pick up a minus sign every time we interchange two indices on a Levi-Civita symbol. Note also that we are always careful to *never* interchange derivatives with anything that is not a constant. The reason for this is that derivatives always act to their right, so if we interchange the order of derivatives with other symbols in the equation, we are changing what the derivatives are acting on. It is OK, however, to interchange derivatives with each other, because derivatives commute.

representation of a vector operation, like a rotation, is as a square matrix.

Objects with indices can always be interpreted as matrices. As we have seen, an object with one index is associated with the components of a vector. In a small, but standard, abuse of notation, we can write

$$A_i = \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix}. \quad (26)$$

In other words, we say that A_i represents the *entire* vector \mathbf{A} ; it only represents a particular component of that vector when i is set equal to 1, 2, or 3. Thus we see that the index acts as a placeholder; it picks out particular entries in the column matrix that represents our vector.

If a vector is represented by a single-index object, how is a square matrix, like the rotation matrix, represented? A 3x3 matrix has nine entries, each of which is specified by its row and column. The natural way to represent it is as a two-index object; the first index gives the row and the second index gives the column of a given entry:

$$M_{ij} = \begin{pmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{pmatrix}. \quad (27)$$

Recall that we have already come across an object with two indices: the Kronecker delta! Given this discussion, we should be able to write it as a 3x3 matrix. Here it is:

$$\delta_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (28)$$

It's the identity matrix!

Let's stop briefly to look at a practical example of the equivalence of indexed objects and matrices. Suppose we wanted to calculate the work done on an object by a force. Newtonian mechanics tells us that

$$W = \mathbf{F} \cdot \mathbf{d} = F_1d_1 + F_2d_2 + F_3d_3. \quad (29)$$

We can rewrite this equation using index notation as

$$W = \delta_{ij}F_id_j = F_1d_1 + F_2d_2 + F_3d_3. \quad (30)$$

We can also rewrite this in matrix form as

$$W = \begin{pmatrix} F_1 & F_2 & F_3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} = F_1 d_1 + F_2 d_2 + F_3 d_3. \quad (31)$$

We see that no matter which representation we choose, we always get the same result[¶].

We now have all of the conceptual pieces of this new index notation in place. An object with one index is a vector, which is a set of three numbers that transforms in a particular way under rotations. An object with two indices is a matrix, which is a set of nine numbers that describes how a vector transforms under a particular transformation such as a rotation or translation. Though we haven't explicitly mentioned it, an object with no indices is just a scalar, a quantity that does not transform at all under a rotation or any other vector transformation. In this context all of these objects, and also those with more than two indices, are given a new name: they are called *tensors*.

Tensors are distinguished by their *rank*, which is just the number of indices they have. Hence, a vector is a first-rank tensor, a matrix is a second-rank tensor, an object with three indices (like the Levi-Civita symbol) is a third-rank tensor, and so on. Because both sides of an equation must be the same type of quantity (that is, we can only equate scalars with scalars, vector with vectors, and so on), the terms on both sides of an equation must have the same set of unsummed, or *free*, indices.

We now turn to the task of extending this framework from three-dimensional Euclidean space to four-dimensional Minkowski spacetime.

2 Four-vectors and the Spacetime Interval

Special relativity relates space and time through the Lorentz transformations, which tell us how to transform the spacetime coordinates of an event from one inertial frame to another. For a frame S' moving with a speed v in the x direction relative to a

[¶]Of course, we picked a somewhat trivial example in that the Kronecker delta is the identity matrix. However, when the matrix in question is more complicated, the mathematical power of this notation becomes more apparent.

frame S , the Lorentz transformations take the form

$$ct' = \gamma(ct - \beta x) \quad (32a)$$

$$x' = \gamma(x - \beta ct) \quad (32b)$$

$$y' = y \quad (32c)$$

$$z' = z, \quad (32d)$$

where we have replaced the velocity of the frame by $\beta = \frac{v}{c}$. In the spirit of Eq. (5), we can rewrite this in matrix form as

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix}. \quad (33)$$

This is very suggestive: we have an object made of four numbers which transforms through multiplication by a matrix in a well-defined way, just like a vector. However, is it really a vector? Above, we defined a vector as a quantity which transforms in a particular way under rotations such that its length remains the same. If we are to call this four-component object a vector, we should be able to define a length for it that remains invariant under *some* transformation. Note that we already know of a quantity that remains invariant under a Lorentz transformation: the *spacetime interval*. It is given by

$$s^2 = (ct)^2 - x^2 - y^2 - z^2. \quad (34)$$

We can show that it is invariant with a bit of algebra:

$$\begin{aligned} (s')^2 &= (ct')^2 - (x')^2 - (y')^2 - (z')^2 \\ &= (\gamma(ct - \beta x))^2 - (\gamma(x - \beta ct))^2 - y^2 - z^2 \\ &= \gamma^2((ct)^2 + (\beta x)^2 - 2\beta xct - x^2 - (\beta ct)^2 + 2\beta xct) \\ &\quad - y^2 - z^2 \\ &= \gamma^2(1 - \beta^2)((ct)^2 - x^2) - y^2 - z^2 \\ &= (ct)^2 - x^2 - y^2 - z^2 \\ &= s^2. \end{aligned} \quad (35)$$

Thinking about this quantity as a length, we will define a new object, called the *position four-vector*, by

$$x^\mu = \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} \equiv \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix}. \quad (36)$$

The 0th component of the four-vector is called the *timelike component*, and others are called the *spacelike components*. Notice that we have replaced the Latin index by a Greek index; from now on Greek indices will denote four-vectors while Latin indices will denote three-vectors. Note also that the index is a superscript rather than a subscript; the importance of this will become clear soon.

The length of a three-vector is given by the dot product of the vector with itself. In the same way, we would like to define a dot product between four-vectors such that the dot product of the position four-vector with itself gives the spacetime interval. If we use the definition of the dot product given in Eq. (14), we will not get the minus signs we want. Therefore, we define a new dot product

$$A \cdot B = \eta_{\mu\nu} A^\mu B^\nu, \quad (37)$$

where $\eta_{\mu\nu}$ is called the *metric* and is given by

$$\eta_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \quad (38)$$

With this definition, we see that dotting x^μ with itself will produce the desired interval:

$$\begin{aligned} \eta_{\mu\nu} x^\mu x^\nu &= \eta_{00} x^0 x^0 + \eta_{11} x^1 x^1 + \eta_{22} x^2 x^2 + \eta_{33} x^3 x^3 \\ &= (ct)^2 - x^2 - y^2 - z^2. \end{aligned} \quad (39)$$

Note that we have not written down the twelve terms for which $\eta_{\mu\nu}$ is identically zero.

Because the metric is cumbersome to work with, we can define a new quantity, x_μ , by

$$x_\mu \equiv \eta_{\mu\nu} x^\nu, \quad (40)$$

so that

$$x_\mu = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} \equiv \begin{pmatrix} x^0 \\ -x^1 \\ -x^2 \\ -x^3 \end{pmatrix}. \quad (41)$$

We call x^μ a *contravariant four-vector* and x_μ a *covariant four-vector*. The metric is used to raise and lower indices, to convert between covariant and contravariant four-vectors; it can also be used to raise and lower the indices on tensors. The dot product between any two four-vectors can now be written as

$$A \cdot B = A_\mu B^\mu = A^\mu B_\mu. \quad (42)$$

We can write Eq. (33) with this new notation as

$$x^{\mu'} = \Lambda^{\mu'}_{\nu} x^\nu, \quad (43)$$

where $\Lambda^{\mu'}_{\nu}$ represents a general Lorentz transformation (note that the information about whether a vector is being written in the primed or the unprimed frame is carried by the indices). Eq. (33) is but one example of such a transformation because the form of the matrix depends on the direction of relative motion of the two frames (just as the form of a rotation matrix depends on which axis we rotate around)^{||}. We can write an inverse Lorentz transformation as $\Lambda^{\mu}_{\nu'}$; this takes a vector from the primed to the unprimed frame. In matrix form, $\Lambda^{\mu}_{\nu'}$ is the same as $\Lambda^{\mu'}_{\nu}$, except that β is replaced by $-\beta$.

In analogy to our generalization from the position vector to general vectors, we can now generalize from the position four-vector to any sort of four-vector. A four-vector is any four-component quantity that transforms under a Lorentz transformation such

^{||}Notice that the spatial rotations are *also* Lorentz transformations. For example,

$$\Lambda^{\mu'}_{\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & \sin \theta & 0 \\ 0 & -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (44)$$

is a perfectly good Lorentz transformation as defined by its action on a four-vector, because it leaves the spacetime interval unchanged. This matrix is just a rotation about the z-axis by an angle θ . Therefore, there are really six fundamental Lorentz transformations: the three spatial rotations and the three transformations that correspond to changing inertial frames, which are called *boosts*. All more generic Lorentz transformations can be constructed by sequential applications of these six.

that its length, defined by the dot product of Eq. (37), remains unchanged. The position four-vector is a canonical example of a four-vector, and will help us construct more of them.

2.1 The Lorentz Transformations as Rotations

Before moving on to constructing more four-vectors, we will take one more look at the Lorentz transformations. Above we made an analogy between the Lorentz transformations and three-dimensional spatial rotations in order to motivate the introduction of four-vectors. In fact, this connection is much stronger than it initially appears.

In order to see this, we first note that

$$\gamma^2 - (\gamma\beta)^2 = 1. \quad (45)$$

Remembering our hyperbolic trig identities, we define a quantity θ such that $\cosh \theta = \gamma$ and $\sinh \theta = \gamma\beta$, because $\cosh^2 \theta - \sinh^2 \theta$ equals 1. Dividing one by the other, we find

$$\frac{\gamma\beta}{\gamma} = \frac{\sinh \theta}{\cosh \theta} \Rightarrow \beta = \tanh \theta. \quad (46)$$

The quantity θ is called the *rapidity* and is just another way of writing the frame velocity. It is often a convenient quantity to work with because rapidities corresponding to parallel velocities add linearly whereas the velocities add via the messy relativistic velocity addition law. Substituting rapidity for velocity in that law, we find:

$$u' = \frac{u + v}{1 + \frac{uv}{c^2}} \Rightarrow \tanh \theta_{u'} = \frac{\tanh \theta_u + \tanh \theta_v}{1 + \tanh \theta_u \tanh \theta_v} = \tanh(\theta_u + \theta_v), \quad (47)$$

where the equality of the hyperbolic tangents implies

$$\theta_{u'} = \theta_u + \theta_v. \quad (48)$$

If we rewrite Eq. (33) in terms of the rapidity, we find

$$\Lambda_{\mu}^{\nu} = \begin{pmatrix} \cosh \theta & -\sinh \theta & 0 & 0 \\ -\sinh \theta & \cosh \theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (49)$$

This transformation matrix looks very much like an ordinary spatial rotation matrix. The fact that rapidities add, just like angles in a plane, is also suggestive of rotations.

In fact, a Lorentz transformation *is a rotation in spacetime* that mixes x and t . The fact that the functions that come up in the rotation matrix are hyperbolic trig functions rather than the usual trig functions is a consequence of the fact that spacetime is Minkowski, not Euclidean, in nature**.

3 More Four-vectors and Invariants

Now that we have the position four-vector, let's try to build some more four-vectors and calculate their related invariants.

3.1 The Velocity Four-vector

The next most obvious quantity to try to construct is the velocity four-vector U^μ , which should just be the derivative of the position four-vector x^μ with respect to time. However, we quickly run into a problem: with respect to *which time* do we differentiate? To make this question more specific, imagine that we are in an inertial frame S , and we observe a particle fly by. The particle's trajectory is given in our coordinates by $x(t)$, $y(t)$, and $z(t)$. To find its velocity in our frame, we differentiate these functions with respect to our coordinate time t , producing $u_x(t)$, $u_y(t)$, and $u_z(t)$; this is the standard Newtonian approach. If we use this idea to construct the four-velocity, we get $U^\mu \equiv \frac{dx^\mu}{dx^0}c$. Is this the correct expression for the four-velocity?

The answer is no. Recall that in order for a set of four numbers to constitute a four-vector, they must transform under a Lorentz transformation like the components of the position four-vector. This is not how the quantity $\frac{dx^\mu}{dx^0}$ transforms. The “numerator” dx^μ basically is the four-position, so it transforms correctly alone. However, the “denominator” dx^0 spoils the transformation properties of $\frac{dx^\mu}{dx^0}$ because it also transforms^{††}. This suggests that if we want to construct a sensible four-velocity, we need to find a time that *does not transform under a Lorentz transformation*.

There is another time associated with the particle with respect to which we could differentiate: the particle's *proper time* τ , which is the time that passes for the particle

**For a more detailed discussion of what this means, see the first installment of these notes.

††To connect this to the more familiar world of vectors in 3D Euclidean space, let's suppose we have a vector \mathbf{A} with components A_x , A_y , and A_z . We can construct the derivative $\frac{d\mathbf{A}}{dA_x}$. Is this quantity a vector? It's not, because it does not transform correctly under a spatial rotation. We can see this by noting that the first “component” of this quantity, $\frac{dA_x}{dA_x}$ equals 1 no matter how we rotate the coordinate system; the other two “components” have similar problems.

as measured by a clock that moves with the particle. One very nice property of the proper time is that it is Lorentz invariant; by definition, observers in any inertial frame will observe the same amount of proper time passing for a given particle traversing a particular section of its worldline. This suggests that we could define the four-velocity as the derivative of four-position with respect to proper time.

Before we do this, let's say a few more words about the proper time. Mathematically, proper time is given by

$$d\tau = \frac{dt}{\gamma_u(t)}, \quad (50)$$

where by $\gamma_u(t)$ we mean

$$\gamma_u(t) = \frac{1}{\sqrt{1 - \frac{u^2(t)}{c^2}}}. \quad (51)$$

This is *not* the same as the γ related to the frame velocity v : γ is a number related to the constant relative velocity between two Lorentz frames whereas $\gamma_u(t)$ is a function of coordinate time t related to the velocity $\mathbf{u}(t)$ of a particle as observed in that frame. The fact that $\gamma_u(t)$ is a function of t is the reason why Eq. (50) is an equation involving differentials. To make a finite proper time, we integrate:

$$\tau = \int_{t_1}^{t_2} \frac{dt}{\gamma_u(t)}. \quad (52)$$

Of course, if the particle travels at constant velocity, then we could define a Lorentz frame that corresponds to that velocity, and in that situation, $\gamma = \gamma_u(t)$. Also note that for $u(t) \ll c$, $\gamma_u(t) \approx 1$, which means that $t \approx \tau$. This means that for small particle velocity (the *Newtonian limit*), coordinate and proper time are almost equal, which is what we expect in Newtonian mechanics.

With this understanding, we can return to our original question by calculating the derivative of the four-position with respect to $\tau^{\ddagger\ddagger}$:

$$\frac{dx^\mu}{d\tau} = \frac{dt}{d\tau} \frac{dx^\mu}{dt} = \gamma_u \frac{dx^\mu}{dt} = \gamma_u \begin{pmatrix} c \\ u_x \\ u_y \\ u_z \end{pmatrix}. \quad (53)$$

^{‡‡}To save typing, let's drop the explicit time dependence on $\gamma_u(t)$ and $u(t)$.

Therefore[✎],

$$U^\mu \equiv \frac{dx^\mu}{d\tau} \quad \Rightarrow \quad \begin{pmatrix} U^0 \\ U^1 \\ U^2 \\ U^3 \end{pmatrix} \equiv \gamma_u \begin{pmatrix} c \\ u_x \\ u_y \\ u_z \end{pmatrix}. \quad (54)$$

We see that the spatial components of the four-velocity, also known as the *proper velocity*, are related to the Newtonian velocity by a factor of γ_u . This result is reasonable because in the Newtonian limit, the spatial components of U^μ reduce to the usual Newtonian notion of velocity. In other words, spatial components of U^μ are approximately equal to the components of \mathbf{u} for small particle velocity, which is why the distinction between the two doesn't come up in the day-to-day life of a human being.

Since U^μ is a four-vector, its length must be an invariant. Squaring it, we find

$$U^\mu U_\mu = \gamma_u^2 (c^2 - u^2) = \gamma_u^2 c^2 (1 - \beta_u^2) = c^2, \quad (55)$$

This is indeed an invariant: it is the speed of light!

Since U^μ is a four-vector, it must also transform via the Lorentz transformations:

$$U^{\mu'} = \Lambda^{\mu'}{}_\nu U^\nu. \quad (56)$$

Expanding this out, we find:

$$\gamma_{u'} c = \gamma (\gamma_u c - \beta \gamma_u u_x) \quad (57a)$$

$$\gamma_{u'} u'_x = \gamma (\gamma_u u_x - \beta \gamma_u c) \quad (57b)$$

$$\gamma_{u'} u'_y = \gamma_u u_y \quad (57c)$$

$$\gamma_{u'} u'_z = \gamma_u u_z, \quad (57d)$$

or, cleaning up the notation,

$$\gamma_{u'} = \gamma \gamma_u (1 - \beta \beta_{u_x}) \quad (58a)$$

$$\gamma_{u'} u'_x = \gamma \gamma_u (u_x - v) \quad (58b)$$

$$\gamma_{u'} u'_y = \gamma_u u_y \quad (58c)$$

$$\gamma_{u'} u'_z = \gamma_u u_z. \quad (58d)$$

[✎]You might be worried that a particle traveling at the speed of light has no proper time (because for $u = c$, γ goes to infinity), and that this would cause problems in defining the four-velocity in this case. This is true; we cannot use proper time to parameterize a path in spacetime that is lightlike. However, one can show that there is always *some* parameter, called an *affine parameter*, that can be used to sensibly define a four-velocity even for lightlike worldlines. We will not discuss this point further in these notes.

If we divide the second, third, and fourth equations by the first in turn, we find

$$u'_x = \frac{u_x - v}{1 - \frac{u_x v}{c^2}} \quad (59a)$$

$$u'_y = \frac{u_y}{\gamma(1 - \frac{u_x v}{c^2})} \quad (59b)$$

$$u'_z = \frac{u_z}{\gamma(1 - \frac{u_x v}{c^2})}. \quad (59c)$$

These are nothing but the velocity transformation laws.

3.2 The Energy-Momentum Four-vector

In Newtonian mechanics, we construct an object's momentum by multiplying its velocity by its mass. Let's construct the four-momentum p^μ in a similar fashion. We get

$$p^\mu \equiv mU^\mu \quad \Rightarrow \quad \begin{pmatrix} p^0 \\ p^1 \\ p^2 \\ p^3 \end{pmatrix} \equiv m\gamma_u \begin{pmatrix} c \\ u_x \\ u_y \\ u_z \end{pmatrix}. \quad (60)$$

We see that the spacelike part of the relativistic momentum p^i is given by

$$p^i = \gamma_u m u_i, \quad (61)$$

which is the Newtonian three-momentum multiplied by γ_u . This reduces to the Newtonian result, $m\mathbf{u}$, in the low velocity limit. However, what does the timelike component p^0 reduce to? Taylor expanding it for $\beta_u \ll 1$ produces

$$\begin{aligned} cp^0 = \gamma_u mc^2 &= \frac{mc^2}{\sqrt{1 - \beta_u^2}} = mc^2 \left(1 + \frac{\beta_u^2}{2} + \frac{3\beta_u^4}{8} + \dots \right) \\ &= mc^2 + \frac{1}{2}mu^2 + \mathcal{O}(\beta_u^4), \end{aligned} \quad (62)$$

where we introduced an extra factor of c for convenience. We see that the first term in this expansion is independent of velocity while the rest of the terms are velocity-dependent. Considering only the velocity-dependent terms, we note that the first of these is just the Newtonian result for the kinetic energy of a particle of mass m and speed u , while the rest are corrections suppressed by factors of β_u^2 . Therefore, we can interpret these terms as the relativistic generalization of the kinetic energy of a

particle:

$$\begin{aligned}
 K &= cp^0 - mc^2 \\
 &= \gamma_u mc^2 - mc^2 \\
 &= (\gamma_u - 1)mc^2.
 \end{aligned} \tag{63}$$

The velocity-independent term (mc^2) also has the dimensions of an energy. It describes the energy a particle has simply by virtue of the fact that it has mass, the particle's *rest energy*. This is a concept unique to special relativity, with no Newtonian counterpart.

Thus, we interpret the quantity cp^0 as the total energy E of a particle, which is a sum of the particle's kinetic energy and mc^2 , its rest energy. Hence,

$$p^0 = \gamma_u mc \equiv \frac{E}{c}. \tag{64}$$

There is a point buried in this discussion that has the potential to be conceptually confusing: the fact that special relativity fundamentally changes the definitions of both energy and momentum as compared to the Newtonian definitions. Of course, the relativistic definitions are the correct ones. However, we noted above that for small frame velocity β_u , the relativistic results reduce to the familiar Newtonian results, so that we incur very little error in calculating with the Newtonian formulas when figuring out the dynamics of, for example, a baseball. Keeping this in mind, when we use the symbols E and p in these notes, we of course mean the relativistic energy and momentum, not the Newtonian energy and momentum.

3.2.1 Invariants and Transformations

Now that we understand the components of the momentum four-vector, we can conclude a number of useful things about it. First of all, the momentum four-vector is a four-vector, which means that it should transform like the position four-vector via the Lorentz transformations. Multiplying by our standard Lorentz transformation matrix produces

$$\frac{E'}{c} = \gamma \left(\frac{E}{c} - \beta p_x \right) \tag{65a}$$

$$p'_x = \gamma \left(p_x - \frac{\beta E}{c} \right) \tag{65b}$$

$$p'_y = p_y \tag{65c}$$

$$p'_z = p_z. \tag{65d}$$

We see that energy and momentum get “mixed up” in a Lorentz transformation just like position and time. This is a very strange result, but it follows directly from the Lorentz transformations.

We can also calculate the length of the momentum four-vector, which must be invariant under Lorentz transformations. We find that

$$p_\mu p^\mu = m^2 U_\mu U^\mu = m^2 c^2, \quad (66)$$

which is indeed invariant. Writing $p_\mu p^\mu$ explicitly in terms of energy and three-momentum, we find[♣]

$$\left(\frac{E}{c}\right)^2 - p^2 = m^2 c^2, \quad (67)$$

or, multiplying out the c 's,

$$E^2 - c^2 p^2 = m^2 c^4. \quad (68)$$

This quantity is invariant under Lorentz transformations, and is just as useful as the spacetime interval in solving problems in special relativity.

Suppose we wanted to describe a collision between two particles, A and B . Both the total momentum of the particles and the total energy of the particles are conserved independently in the collision; this is an experimental fact. It would be nice to construct a quantity involving both energy and momentum that is also conserved. Because each particle has a four-momentum associated with it, and four-vectors add together to form new four-vectors in the same way as three-vectors add to form new three-vectors, we can define the total four-momentum of the system before the collision to be

$$\mathcal{P}^\mu \equiv p_{total}^\mu = p_A^\mu + p_B^\mu = \begin{pmatrix} E_{total}/c \\ p_{x,total} \\ p_{y,total} \\ p_{z,total} \end{pmatrix}. \quad (69)$$

The length of this total four-momentum is

$$\mathcal{P}^\mu \mathcal{P}_\mu = \left(\frac{E_{total}}{c}\right)^2 - p_{total}^2. \quad (70)$$

[♣]Note that the “ p ” in this equation is the momentum three-vector. Since we are squaring it, we only care about its magnitude, so we do not write it in boldface. However, in cases where we are summing the momenta of several particles, as we will in the next paragraph, we must take the *vector* sum of the momenta before finding the magnitude.

Since both total momentum and total energy are conserved during a collision, this quantity is also conserved. However, the power in writing this combination comes from the fact that because it is the length of a four-vector, it is also *invariant under Lorentz transformations*. This means that if we are analyzing a collision which is complicated in one Lorentz frame, we can transform it into a more convenient frame, like the center of momentum frame[◇], and this quantity will have the same numerical value. This is, of course, completely analogous to the fact that the spacetime interval has the same value in all inertial reference frames. Because this quantity is both conserved during the collision and invariant when transformed between different inertial frames, it is clearly important and should be given a name: it is the *total mass-energy* of the system.

3.2.2 Massless Particles

One last item to discuss before we move on is the concept of a *massless particle*[♣]. In Newtonian mechanics, a massless particle makes no sense because it has neither momentum nor kinetic energy and it cannot have a net force applied on it. However, if we look at Eq. (68), we see that a particle with no mass makes sense in special relativity. Setting $m = 0$, we find

$$E^2 - p^2c^2 = 0 \quad \Rightarrow \quad E = pc. \quad (71)$$

So, a massless particle is one whose energy and momentum are equal, up to a factor of c . Therefore, if it is moving in the x direction, its four-momentum can be written as

$$p^\mu = \frac{E}{c} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}. \quad (72)$$

[◇]The center of momentum frame is the frame in which the total momentum of the system under consideration is zero.

[♣]These are not just theoretical constructs. Photons are massless particles, as are gluons (the carriers of the strong nuclear force) and gravitons (the carriers of the gravitational force). For many years it was believed that neutrinos are also massless particles, but in fact they have a tiny, but nonzero, mass.

Compare this to Eq. (60). For a particle moving in the x direction, we have

$$p^\mu = mc\gamma_u \begin{pmatrix} 1 \\ u_x/c \\ 0 \\ 0 \end{pmatrix}. \quad (73)$$

These two results should be consistent in the case where $m = 0$. However, we see that if we set $m = 0$ in Eq. (73), we seem to get zero for all components of the four-momentum. The way to make these results consistent is to simultaneously assert that the velocity of the massless particle is c . By doing that, we force the γ_u in Eq. (73) to go to infinity while the mass multiplying it goes to zero; they do this in such a way that the product is finite. Also, the u_x/c factor in Eq. (73) becomes 1, as it should to be consistent with Eq. (72). Therefore, we find that *massless particles must always travel at the speed of light*. This is the only way to make our concepts of energy and momentum consistent for massless particles.

3.3 The Force Four-vector

The last concept from classical mechanics we need to talk about is that of forces. Given our trick for constructing a four-velocity, it should be clear how to construct the four-force K^μ (also known as the *Minkowski force*): differentiate the four-momentum with respect to proper time, producing

$$\frac{dp^\mu}{d\tau} = \gamma_u \frac{dp^\mu}{dt} = \gamma_u \frac{d}{dt} \begin{pmatrix} p^0 \\ p^1 \\ p^2 \\ p^3 \end{pmatrix} = \gamma_u \frac{d}{dt} \begin{pmatrix} E/c \\ p_x \\ p_y \\ p_z \end{pmatrix}. \quad (74)$$

Recall that the time derivative of the momentum of a particle is just the force F acting on it, and the time derivative of its energy is the power P it is absorbing or emitting. Therefore, we find that

$$K^\mu \equiv \frac{dp^\mu}{d\tau} \quad \Rightarrow \quad \begin{pmatrix} K^0 \\ K^1 \\ K^2 \\ K^3 \end{pmatrix} \equiv \gamma_u \begin{pmatrix} P/c \\ F_x \\ F_y \\ F_z \end{pmatrix}. \quad (75)$$

As in the case of the four-velocity and the four-momentum, we find that the components of the four-force reduce to the familiar Newtonian expressions for force and

power in the low velocity limit. The components of the four-force, naturally, transform via the Lorentz transformations, while the Newtonian force has a relatively nasty transformation law given by

$$F'_x = \frac{F_x - \frac{\beta \mathbf{u} \cdot \mathbf{F}}{c}}{1 - \frac{\beta u_x}{c}} \quad (76a)$$

$$F'_y = \frac{F_y}{\gamma \left(1 - \frac{\beta u_x}{c}\right)} \quad (76b)$$

$$F'_z = \frac{F_z}{\gamma \left(1 - \frac{\beta u_x}{c}\right)}. \quad (76c)$$

We will not derive this result here.

Before moving on, let's take some time to explicitly show that the relevant time derivatives above actually equal F and P in the low velocity limit, that is, that the relativistic and Newtonian conceptions of force and power agree at low velocity. First, we can calculate the time derivative of γ_u :

$$\frac{d\gamma_u}{dt} = \frac{d}{dt} \left(1 - \frac{\mathbf{u}^2}{c^2}\right)^{-1/2} = \left(1 - \frac{\mathbf{u}^2}{c^2}\right)^{-3/2} \frac{\mathbf{u} \cdot \dot{\mathbf{u}}}{c^2} = \gamma_u^3 \frac{\mathbf{u} \cdot \dot{\mathbf{u}}}{c^2}, \quad (77)$$

where the overdot on the \mathbf{u} indicates a time derivative. Using this, let's calculate the time derivative of p^0 :

$$\frac{dp^0}{dt} = \frac{d}{dt}[mc\gamma_u] = mc \frac{d\gamma_u}{dt} = mc\gamma_u^3 \frac{\mathbf{u} \cdot \dot{\mathbf{u}}}{c^2} = \frac{\gamma_u^3}{c} \mathbf{u} \cdot (m\dot{\mathbf{u}}). \quad (78)$$

Expanding the factor of γ_u^3 for small β_u , we get

$$\frac{dp^0}{dt} = \frac{1}{c} \mathbf{u} \cdot (m\dot{\mathbf{u}}) \left(1 + \frac{3}{2}\beta_u^2 + \frac{15}{8}\beta_u^4 + \dots\right). \quad (79)$$

Now, we know from Newtonian mechanics that m times $\dot{\mathbf{u}}$ is equal to force, and we know that force dotted into velocity equals power. Therefore, this quantity, to lowest order, is simply the Newtonian power (divided by c), which justifies our interpretation of it, with all terms included, as the relativistic generalization of power (divided by c).

We can use similar reasoning to evaluate the time derivative of the spacelike components of p^μ . To simplify our notation, let's do all three components in one by defining the vector \mathbf{p} as having components p^1 , p^2 , and p^3 . Then, we find that

$$\frac{d\mathbf{p}}{dt} = \frac{d}{dt}[m\gamma_u \mathbf{u}] = m\gamma_u \dot{\mathbf{u}} + m\dot{\gamma}_u \mathbf{u}. \quad (80)$$

Substituting in for $\dot{\gamma}_u$ (from above) and expanding for small β_u produces

$$\begin{aligned} \frac{d\mathbf{p}}{dt} &= m\dot{\mathbf{u}}\gamma_u + m\mathbf{u}\gamma_u^3 \frac{\mathbf{u} \cdot \dot{\mathbf{u}}}{c^2} \\ &= m\dot{\mathbf{u}} \left(1 + \frac{1}{2}\beta_u^2 + \frac{3}{8}\beta_u^4 + \dots \right) + m\beta_u(\beta_u \cdot \dot{\mathbf{u}}) \left(1 + \frac{3}{2}\beta_u^2 + \frac{15}{8}\beta_u^4 + \dots \right) \end{aligned} \quad (81)$$

The lowest order term in this expansion is equal to m times $\dot{\mathbf{u}}$, which is the Newtonian force[∇]. Therefore, we are justified in identifying this entire expression with the relativistic generalization of force.

4 Relativistic Electrodynamics

Special relativity and electrodynamics are intimately related to each other. One of the driving forces behind Einstein's discovery of relativity was the desire to understand what would happen if one ran as fast as a beam of light, and since his time it has been shown that magnetism is nothing but a relativistic consequence of electricity in moving frames. It should be no surprise, then, that the most simple and elegant formulation of Maxwell's theory of electromagnetism is in terms of relativistic four-vectors and tensors. In this section, we will turn electromagnetism into a manifestly relativistic theory.

4.1 The Current Density Four-vector

The theory of classical electromagnetism introduces many new quantities into the arena of physics, including electric and magnetic fields, scalar and vector potentials, charges, currents, and many others. Our first task will be to see what relativistically invariant quantities we can construct out of this list.

First, let's consider charges and currents. Imagine an infinitesimally small lump of charge Q moving by a stationary observer with velocity \mathbf{u} . Classically, that charge would have an associated current density given by

$$\mathbf{J} = \rho\mathbf{u}, \quad (82)$$

where ρ is the charge density in the infinitesimal region. Even though charge density is an invariant quantity classically, it is not a relativistically invariant quantity because

[∇]Note that the lowest order term in the part of the expansion corresponding to the $\dot{\gamma}$ is actually proportional to β^2 , whereas the lowest order term in the other part of the expansion ($m\dot{\mathbf{u}}$) contains no factors of β . Hence, only the $m\dot{\mathbf{u}}$ survives to lowest order in the entire expression.

it is the amount of charge per unit volume, and volumes are Lorentz contracted^b. However, we can define a *proper charge density* ρ_0 as the charge density seen in the rest frame of the lump of charge.

With this definition, we can postulate a form for the current density four-vector J^μ in terms of the four-velocity as

$$J^\mu \equiv \rho_0 U^\mu \quad \Rightarrow \quad \begin{pmatrix} J^0 \\ J^1 \\ J^2 \\ J^3 \end{pmatrix} \equiv \rho_0 \gamma_u \begin{pmatrix} c \\ u_x \\ u_y \\ u_z \end{pmatrix}. \quad (83)$$

We see that the spacelike part of the four-current J^i is given by

$$J^i \equiv \gamma_u \rho_0 u_i, \quad (84)$$

which reproduces the classical result at low velocity. Therefore, we call this quantity the relativistic current density. Similarly, we interpret the timelike component using our usual Taylor expansion trick as

$$\begin{aligned} \frac{J^0}{c} = \gamma_u \rho_0 &= \rho_0 \left(1 + \frac{\beta_u^2}{2} + \frac{3\beta_u^4}{8} + \dots \right) \\ &= \rho_0 + \mathcal{O}(\beta_u^2). \end{aligned} \quad (85)$$

Evidently, it is the relativistic generalization of the charge density, which we will call ρ . We find

$$J^0 = \gamma_u c \rho_0 = c \rho. \quad (86)$$

Now that we have the four-current, we can use it to find the relativistic equivalent of the equation of continuity, which is the mathematical expression of local conservation of charge. The equation of continuity tells us that an outward flux of current from a region leads to a reduction in the charge density in that region over time, and vice versa. In equation form, we have

$$\nabla \cdot \mathbf{J} = -\frac{d\rho}{dt}. \quad (87)$$

^bCharge, however, *is* a relativistically invariant quantity. First of all, it is a scalar quantity, so it should be unchanged under a Lorentz transformation. Also, it is experimentally true that, for example, electrons are always seen to have the same charge regardless of their state of motion.

Now let's take the divergence of the four-current[‡]

$$\partial_\mu J^\mu = \partial_0 J^0 + \partial_i J^i = \frac{d(c\rho)}{d(ct)} + \nabla \cdot \mathbf{J} = \frac{d\rho}{dt} + \nabla \cdot \mathbf{J}. \quad (89)$$

By the equation of continuity, the right hand side of this equation is zero. Hence, we find that

$$\partial_\mu J^\mu = 0, \quad (90)$$

that is, the four-current is divergenceless. This statement is equivalent to Eq. (87), but is written in the tidier notation of four-vectors.

4.2 The Potential Four-vector

The last four-vector we will construct is the potential four-vector A^μ , which is constructed out of the scalar potential ϕ and the vector potential \mathbf{A} . These potentials are related to the electric and magnetic fields via

$$\mathbf{E} = -\nabla\phi - \frac{\partial\mathbf{A}}{\partial t} \quad (91a)$$

$$\mathbf{B} = \nabla \times \mathbf{A}. \quad (91b)$$

Motivating the definition of the four-potential requires a greater knowledge of electromagnetism than is expected of the reader of these notes, but we can at least sketch the argument. Starting with Maxwell's equations in differential form, it is possible to use Eqs. (91) to replace \mathbf{E} and \mathbf{B} by ϕ and \mathbf{A} , yielding Maxwell's equations in terms of the vector and scalar potentials. These equations are not very pleasant. However, because of the way the potentials are defined, we have some freedom, known as *gauge freedom*, in exactly how we choose them[‡]. In particular, we are free to set the

[‡]Here we are using the generalization of the partial derivative vector defined in the section on three-vector calculus. The generalization is

$$\partial_\mu = \begin{pmatrix} \partial_0 \\ \partial_1 \\ \partial_2 \\ \partial_3 \end{pmatrix} = \begin{pmatrix} \partial/\partial x^0 \\ \partial/\partial x^1 \\ \partial/\partial x^2 \\ \partial/\partial x^3 \end{pmatrix}. \quad (88)$$

Notice that this object is *naturally lowered*, which means that its components are all positive when its index is lowered. This is the opposite convention from the one for the four-position and related four-vectors, but it is the right choice to make our notation internally consistent.

[‡]A simple but familiar example of this is the fact that in electrostatics, we are always free to add a constant term to our scalar potential without changing the electric field it generates.

divergence of \mathbf{A} to be whatever we like. If we make the choice that

$$\nabla \cdot \mathbf{A} = -\frac{1}{c^2} \frac{\partial \phi}{\partial t}, \quad (92)$$

then we can write Maxwell's equations (in SI units) as

$$\square^2 \frac{\phi}{c} = -\mu_0(c\rho) \quad (93a)$$

$$\square^2 \mathbf{A} = -\mu_0 \mathbf{J}, \quad (93b)$$

where \square^2 is called the d'Alembertian operator and is given by

$$\square^2 \equiv \nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} = -\partial_\mu \partial^\mu. \quad (94)$$

This choice of divergence is called the *Lorentz gauge*. From the form of Eqs. (93), we see that the two right-hand sides (up to some constants) are the components of the four-current. We also see, from Eq. (94), that the d'Alembertian is a Lorentz scalar (that is, it is invariant under a Lorentz transformation, much like the three-dimensional Laplacian ∇^2 is invariant under a coordinate rotation). Therefore, $\frac{\phi}{c}$ and \mathbf{A} must be the components of some other four-vector: the four-potential A^μ defined by

$$A^\mu = \begin{pmatrix} A^0 \\ A^1 \\ A^2 \\ A^3 \end{pmatrix} \equiv \begin{pmatrix} \phi/c \\ A_x \\ A_y \\ A_z \end{pmatrix}. \quad (95)$$

With this result, we can rewrite Eqs. (93) as one equation in tensor notation:

$$\partial_\mu \partial^\mu A^\nu = \mu_0 J^\nu. \quad (96)$$

4.3 The Field Strength Tensor

Now that we have the four-potential, we would like to learn how to derive the electric and magnetic fields from it. Given the proliferation of four-vectors in these notes, it would be reasonable to assume that the components of the electric and magnetic field also form some sort of four-vector. However, this assumption is wrong. The components of these two fields are actually entries in a second-rank tensor $F_{\mu\nu}$ known as the *electromagnetic field strength tensor*.

In order to see this, let's look more closely at Eqs. (91). Notice that any given component of \mathbf{E} or \mathbf{B} involves sums of derivatives of the potentials. We can therefore write the components of \mathbf{E} and \mathbf{B} in terms of the components of ∂_μ and A_μ as

$$\frac{E_x}{c} = \partial_0 A_1 - \partial_1 A_0 \quad (97a)$$

$$\frac{E_y}{c} = \partial_0 A_2 - \partial_2 A_0 \quad (97b)$$

$$\frac{E_z}{c} = \partial_0 A_3 - \partial_3 A_0 \quad (97c)$$

$$B_x = \partial_3 A_2 - \partial_2 A_3 \quad (97d)$$

$$B_y = \partial_1 A_3 - \partial_3 A_1 \quad (97e)$$

$$B_z = \partial_2 A_1 - \partial_1 A_2. \quad (97f)$$

For example, the first of these says

$$\frac{E_x}{c} = \partial_0 A_1 - \partial_1 A_0 = \frac{1}{c} \frac{\partial}{\partial t}(-A_x) - \frac{\partial}{\partial x} \frac{\phi}{c}, \quad (98)$$

or, eliminating the factor of c ,

$$E_x = -\frac{\partial \phi}{\partial t} - \frac{\partial A_x}{\partial x}, \quad (99)$$

which reproduces one component of Eqs. (91). It is left to the reader to verify the other five components.

The form of Eqs. (97) leads us to postulate an expression for $F_{\mu\nu}$:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \quad (100)$$

We know what six of the components of $F_{\mu\nu}$ are from Eqs. (97). However, a general two-index tensor should have sixteen independent components, so there are ten components unaccounted for. Might there be quantities other than the components of \mathbf{E} and \mathbf{B} in $F_{\mu\nu}$?

To answer this question, notice that if we set μ equal to ν in Eq. (100), $F_{\mu\mu}$ equals zero, leaving us with only twelve nonzero components. Also notice that $F_{\mu\nu}$ is *antisymmetric*, which means that $F_{\mu\nu} = -F_{\nu\mu}$. This cuts the number of independent components in half. Therefore, $F_{\mu\nu}$ has only six independent components, the components of \mathbf{E} and \mathbf{B} .

To finish up our discussion of the structure of $F_{\mu\nu}$, we can write it down in matrix form:

$$F_{\mu\nu} = \begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ -E_x/c & 0 & -B_z & B_y \\ -E_y/c & B_z & 0 & -B_x \\ -E_z/c & -B_y & B_x & 0 \end{pmatrix}. \quad (101)$$

So, as advertised, the field tensor is nothing but the relativistic form of the electric and magnetic fields.

There is a tensor related to the field tensor which is useful in the formulation of Maxwell's equations. It is called the *dual tensor* and is given the symbol $\mathcal{F}^{\mu\nu}$. It is given by

$$\mathcal{F}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma}F_{\rho\sigma}, \quad (102)$$

where $\epsilon^{\mu\nu\rho\sigma}$ is a generalization of the Levi-Civita symbol such that

$$\epsilon^{\mu\nu\rho\sigma} = \begin{cases} 1, & \mu\nu\rho\sigma = 0123 \text{ and even permutations} \\ -1, & \mu\nu\rho\sigma = \text{odd permutations} \\ 0, & \text{otherwise} \end{cases} \quad (103)$$

For practice, let's calculate one entry in the dual tensor, for example \mathcal{F}^{10} :

$$\begin{aligned} \mathcal{F}^{10} &= \frac{1}{2}\epsilon^{10\mu\nu}F_{\mu\nu} \\ &= \frac{1}{2}(\epsilon^{1023}F_{23} + \epsilon^{1032}F_{32}) \\ &= \frac{1}{2}(-(-B_x) + B_x) \\ &= B_x. \end{aligned} \quad (104)$$

Working through the rest of the components, we find that

$$\mathcal{F}^{\mu\nu} = \begin{pmatrix} 0 & -B_x & -B_y & -B_z \\ B_x & 0 & E_z/c & -E_y/c \\ B_y & -E_z/c & 0 & E_x/c \\ B_z & E_y/c & -E_x/c & 0 \end{pmatrix}. \quad (105)$$

4.3.1 Transforming the Field Tensor

The field tensor is the first non-trivial tensor we have seen so far in these notes, so let's illustrate a few points about how to manipulate one. When we defined the field

tensor, we made the choice to write it with lowered indices, but if you look back at the definition, we could have just as well written it with upper indices as $F^{\mu\nu}$. Fortunately, we know how to convert between upper and lower indices: just multiply by the metric! For the field tensor, we find

$$F^{\mu\nu} = \eta^{\mu\rho}\eta^{\sigma\nu}F_{\rho\sigma}, \quad (106)$$

or, in matrix notation,

$$F^{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ -E_x/c & 0 & -B_z & B_y \\ -E_y/c & B_z & 0 & -B_x \\ -E_z/c & -B_y & B_x & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \quad (107)$$

Multiplying all this out, we find

$$F^{\mu\nu} = \begin{pmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ E_x/c & 0 & -B_z & B_y \\ E_y/c & B_z & 0 & -B_x \\ E_z/c & -B_y & B_x & 0 \end{pmatrix}. \quad (108)$$

This is useful to remember, because it is often the case that we might need to raise or lower all or some of the indices on a tensor in order to use it in an equation. All it involves is multiplication by an appropriate number of copies of the metric.

Those of us who have taken H7B (or an equivalent class) will remember that electric and magnetic fields transform between different inertial frames (in other words, observers in different frames will observe different values for the components of \mathbf{E} and \mathbf{B} in the same region of spacetime). The transformations for a boost along x are given by

$$E'_x/c = E_x/c \quad (109a)$$

$$E'_y/c = \gamma(E_y/c - \beta B_z) \quad (109b)$$

$$E'_z/c = \gamma(E_z/c + \beta B_y) \quad (109c)$$

$$B'_x = B_x \quad (109d)$$

$$B'_y = \gamma(B_y + \beta E_z/c) \quad (109e)$$

$$B'_z = \gamma(B_z - \beta E_y/c). \quad (109f)$$

How are these transformation laws related to the Lorentz transformations? We saw earlier that the way to write a Lorentz transformation on a four-vector is to multiply

by a factor of $\Lambda^{\mu'}_{\nu'}$. In order to transform the fields we simply Lorentz transform the field tensor:

$$F_{\mu'\nu'} = \Lambda_{\mu'}^{\rho} \Lambda_{\nu'}^{\sigma} F_{\rho\sigma}. \quad (110)$$

In matrix form, the right-hand side equals

$$\begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ -E_x/c & 0 & -B_z & B_y \\ -E_y/c & B_z & 0 & -B_x \\ -E_z/c & -B_y & B_x & 0 \end{pmatrix} \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (111)$$

which, when multiplied out, equals

$$\begin{pmatrix} 0 & E_x/c & \gamma(E_y/c - \beta B_z) & \gamma(E_z/c + \beta B_y) \\ -E_x/c & 0 & -\gamma(B_z - \beta E_y/c) & \gamma(B_y + \beta E_z/c) \\ -\gamma(E_y/c - \beta B_z) & \gamma(B_z - \beta E_y/c) & 0 & -B_x \\ -\gamma(E_z/c + \beta B_y) & -\gamma(B_y + \beta E_z/c) & B_x & 0 \end{pmatrix}. \quad (112)$$

Comparing this to the list of transformations, we see that we have reproduced them exactly. Evidently, the transformation laws for the components of \mathbf{E} and \mathbf{B} are a generalization of the Lorentz transformations for four-vectors.

Incidentally, this illustrates another strength of index notation: no matter what kind of quantity we consider, we *always* know how to Lorentz transform it! We simply multiply by one factor of $\Lambda^{\mu'}_{\nu'}$ for each index in the object we would like to transform.

4.3.2 Electromagnetic Invariants

In our discussion of four-vectors, we spent a fair amount of time constructing invariant quantities that we could construct out of those vectors. These quantities are important because they have the same value in all inertial frames. We can also construct invariant quantities out of tensors, including the field tensor.

In order to construct a Lorentz invariant out of four-vectors, we made combinations of four-vectors, like $x_{\mu}x^{\mu}$, that contained no free indices. Remember that $x_{\mu}x^{\mu} = \eta^{\mu\nu}x_{\mu}x_{\nu}$. Now, suppose we have some two-index tensor $T_{\mu\nu}$ which we assume, for simplicity, is either symmetric or antisymmetric (so that swapping the indices does nothing or introduces a minus sign, respectively). We can construct an invariant quantity T out of it:

$$T = \eta^{\mu\nu}T_{\mu\nu} = T^{\mu}_{\mu}. \quad (113)$$

In terms of matrices, what we are doing is multiplying T by the metric, and then taking a *trace*[✓]:

$$T = T_{00} - T_{11} - T_{22} - T_{33}. \quad (114)$$

There is another invariant we can construct given $T_{\mu\nu}$:

$$\tilde{T} = \eta^{\mu\rho}\eta^{\nu\sigma}T_{\mu\nu}T_{\rho\sigma} = T_{\mu\nu}T^{\mu\nu}. \quad (115)$$

In components, we have

$$\tilde{T} = T_{00}T^{00} + T_{01}T^{01} + T_{02}T^{02} + \cdots + T_{33}T^{33}. \quad (116)$$

Let's apply this to $F_{\mu\nu}$. Since the diagonal components of $F_{\mu\nu}$ are all zero, $F^\mu{}_\mu = 0$. This is not a very interesting invariant. However, the invariant P that involves two copies of $F_{\mu\nu}$ is interesting:

$$P \equiv \frac{1}{2}F_{\mu\nu}F^{\mu\nu} = \frac{1}{2}(F_{00}F^{00} + F_{01}F^{01} + \cdots + F_{33}F^{33}) = \mathbf{B}^2 - \frac{\mathbf{E}^2}{c^2}. \quad (117)$$

Furthermore, we can construct an invariant Q that involves a copy of the field strength tensor and a copy of the dual tensor:

$$Q \equiv -\frac{1}{4}F_{\mu\nu}\mathcal{F}^{\mu\nu} = -\frac{1}{4}(F_{00}\mathcal{F}^{00} + F_{01}\mathcal{F}^{01} + \cdots + F_{33}\mathcal{F}^{33}) = \frac{1}{c}(\mathbf{E} \cdot \mathbf{B}). \quad (118)$$

Before concluding this section, let's look at one (out of many) interesting results we can conclude just by the invariance of these quantities. Suppose that $P = Q = 0$ in some frame. This means that the magnitude of \mathbf{E} equals the magnitude of \mathbf{B} (up to a factor of c) and that \mathbf{E} and \mathbf{B} are perpendicular. This configuration of fields describes an electromagnetic plane wave. Since P and Q are invariant under Lorentz transformations, if they are 0 in one frame, they will be 0 in all frames. This implies that *an electromagnetic wave in one inertial frame looks like an electromagnetic wave in all inertial frames*.

4.4 Maxwell's Equations

We finally come to the crowning achievement of the tensor formulation of special relativity: Maxwell's equations in Lorentz covariant form. In SI units, they are

$$\partial_\mu F^{\mu\nu} = \mu_0 J^\nu \quad (119a)$$

$$\partial_\mu \mathcal{F}^{\mu\nu} = 0. \quad (119b)$$

[✓] Remember that the trace of a square matrix is simply the sum of the diagonal components of the matrix.

The first one is called the *inhomogeneous Maxwell equation*, and the second one is called the *homogeneous Maxwell equation*. These seemingly simple equations contain all four Maxwell equations. Let's see how.

4.4.1 The Inhomogeneous Maxwell Equation

First, we will calculate the 0th component of the inhomogeneous Maxwell equation. We note that

$$\partial_\mu F^{\mu 0} = \partial_0 F^{00} + \partial_i F^{i0} = 0 + \frac{1}{c} \nabla \cdot \mathbf{E}, \quad (120)$$

and therefore

$$\partial_\mu F^{\mu 0} = \mu_0 J^0 \quad \Rightarrow \quad \frac{1}{c} \nabla \cdot \mathbf{E} = \mu_0 c \rho. \quad (121)$$

Isolating the divergence in this last expression, we find[‡]

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}, \quad (122)$$

Which is *Gauss's Law for electricity*.

Next, let's calculate the spacelike components. Starting with the 1st component, we note that

$$\begin{aligned} \partial_\mu F^{\mu 1} = \partial_0 F^{01} + \partial_i F^{i1} &= \frac{1}{c} \partial_t \left(\frac{-E_x}{c} \right) + \partial_y B_z + \partial_z (-B_y) \\ &= -\frac{1}{c^2} \frac{\partial E_x}{\partial t} + \nabla \times \mathbf{B}|_x. \end{aligned} \quad (123)$$

The 2nd and 3rd components yield similar results. Therefore

$$\partial_\mu F^{\mu i} = \mu_0 J^i \quad \Rightarrow \quad -\frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t} + \nabla \times \mathbf{B} = \mu_0 \mathbf{J}. \quad (124)$$

Solving for the curl of \mathbf{B} , we find

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}, \quad (125)$$

which is the *Ampère-Maxwell Law*.

4.4.2 The Homogeneous Maxwell Equation

We now turn to the Homogeneous Maxwell equation. Analyzing its 0th component produces

$$\partial_\mu \mathcal{F}^{\mu 0} = \partial_0 \mathcal{F}^{00} + \partial_i \mathcal{F}^{i0} = 0 + \nabla \cdot \mathbf{B}, \quad (126)$$

[‡]We use the fact that $\mu_0 \epsilon_0 = \frac{1}{c^2}$.

which implies that

$$\nabla \cdot \mathbf{B} = 0. \quad (127)$$

This is *Gauss's Law for magnetism*.

Analyzing the 1st component of the homogeneous Maxwell equation produces

$$\begin{aligned} \partial_\mu \mathcal{F}^{\mu 1} = \partial_0 \mathcal{F}^{01} + \partial_i \mathcal{F}^{i1} &= \frac{1}{c} \partial_t (-B_x) + \partial_y \left(\frac{-E_z}{c} \right) + \partial_z \left(\frac{E_y}{c} \right) \\ &= -\frac{1}{c} \frac{\partial B_x}{\partial t} + -\frac{1}{c} \nabla \times \mathbf{E}|_x. \end{aligned} \quad (128)$$

Combining this with the analogous results for the 2nd and 3rd components, we find that

$$\partial_\mu \mathcal{F}^{\mu i} = 0 \quad \Rightarrow \quad -\frac{1}{c} \left(\frac{\partial \mathbf{B}}{\partial t} + \nabla \times \mathbf{E} \right) = 0. \quad (129)$$

Isolating the curl of \mathbf{E} , we produce

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad (130)$$

which is *Faraday's Law*.

4.4.3 The Lorentz Force Law

The theory of classical electromagnetism requires us to postulate not only Maxwell's equations, but also the Lorentz force law

$$\mathbf{F} = q(\mathbf{E} + \mathbf{u} \times \mathbf{B}), \quad (131)$$

which tells us how much force a charge feels when subjected to electric and magnetic fields. In Lorentz covariant form, this law is replaced by

$$K^\mu = qU_\nu F^{\mu\nu}. \quad (132)$$

Let's verify this with a calculation.

First, we will calculate the 1st component of this equation:

$$\begin{aligned} K^1 = qU_\nu F^{1\nu} \quad \Rightarrow \quad \gamma F^1 &= q(U_0 F^{10} + U_i F^{1i}) \\ &= q \left(c\gamma \frac{E_x}{c} - \gamma u_y (-B_z) - \gamma u_z B_y \right) \\ &= q\gamma (E_x + \mathbf{u} \times \mathbf{B}|_x). \end{aligned} \quad (133)$$

Combining this with the 2nd and 3rd components, and dividing out the extra factor of γ , we find that

$$\mathbf{F} = q(\mathbf{E} + \mathbf{u} \times \mathbf{B}), \quad (134)$$

as desired.

All this begs the question: what does the 0th component of the covariant form of the Lorentz force law give us? Let's find out:

$$\begin{aligned} K^0 = qU_\nu F^{0\nu} &\Rightarrow \gamma \frac{P}{c} = q(U_0 F^{00} + U_i F^{0i}) \\ &= q\left(0 + \gamma \mathbf{u} \cdot \left(\frac{-\mathbf{E}}{c}\right)\right) \\ &= -q \frac{\gamma}{c} \mathbf{u} \cdot \mathbf{E}. \end{aligned} \quad (135)$$

Therefore, we find that

$$\frac{dE}{dt} = -q \mathbf{u} \cdot \mathbf{E}. \quad (136)$$

This is simply the equation describing the rate of change of a particle's energy due to moving through an electric field. We note that magnetic fields don't come up in this equation at all, which is as it should be because magnetic fields do not change the energy of charged particles, just their trajectories.

5 Conclusion

These notes went over a large amount of material in very few pages. Hopefully, they have given the reader a sense of the beauty and elegance of special relativity as a theory of nature and an understanding of how intimately linked special relativity and classical electromagnetism truly are.

For more information, the author of these notes recommends a number of sources. Both chapter 12 of *Introduction to Electrodynamics* (third edition) by David Griffiths and chapters 11 and 12 of *Classical Electrodynamics* (third edition) by J.D. Jackson offer an in depth look at relativistic kinematics and dynamics, as well as a wealth of information about the relativistic formulation of electromagnetism.