Vectors and Covectors in Special Relativity

Jim Napolitano

March 12, 2010

These notes are meant to accompany the course *Electromagnetic Theory* for the Spring 2010 term at RPI. This material is covered thoroughly in Chapters One and Seven in our textbook *Classical Electrodynamics, 2nd Ed.* by Hans Ohanian, and also in the auxiliary textbook *Principles of Electrodynamics* by Melvin Schwartz. I am attempting here to present a condensed version of this material, along with several worked examples.

Before specializing the discussion to special relativity, we go over two introductory topics. One is the notation we use for vectors written as components, especially the Einstein summation notation. We will use this to come up with "grown up" definitions of scalars, vectors, and tensors. The second is a brief introduction to coordinate-free geometry, which necessitates a discussion of contravariant and covariant vectors. These two topics will be then combined to write down the formalism for special relativity.

Old Math with a New Notation

We think of a vector in ordinary three-dimensional space as an object with direction and magnitude. Typically, when we need to calculate something with a vector, we write it in terms of its coordinates using some coordinate system, say (x, y, z). For example

$$\mathbf{v} = v_x\hat{\mathbf{x}} + v_y\hat{\mathbf{y}} + v_z\hat{\mathbf{z}} \tag{1a}$$

$$v = |\mathbf{v}| = \left[v_x^2 + v_y^2 + v_z^2\right]^{1/2} \tag{1b}$$

It is important to separate the vector object from the vector written in terms of its coordinates. The vector is the same thing, that is in the same direction with the same magnitude, regardless of what coordinate system we use to describe it. Indeed, we will, in a moment, define a vector in terms of how one goes from a description in one coordinate system to a description in another.

Figure 1 shows two sets of axes. Equation 1 describes the vector $\mathbf{v}$ in terms of the unprimed
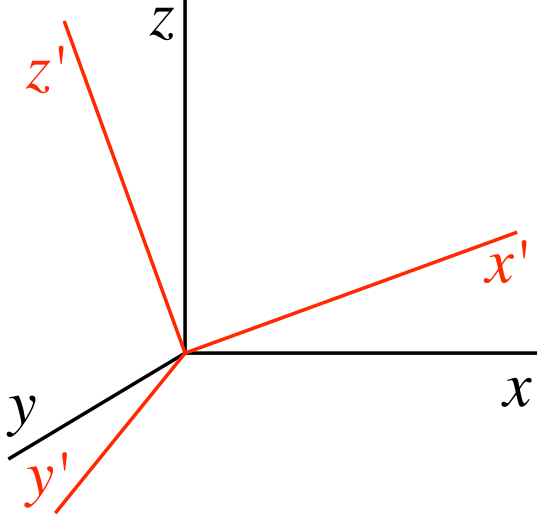
Figure 1: Simple depiction of a default (unprimed) coordinate system, in black. A second (primed) coordinate system is drawn in red. The primed system can be described in terms of some rotation of the unprimed system.

coordinate system. We could just as easily have written the vector in the primed coordinate system. In that case

$$\mathbf{v} = v'_x\hat{\mathbf{x}}' + v'_y\hat{\mathbf{y}}' + v'_z\hat{\mathbf{z}}' \tag{2a}$$

$$v = |\mathbf{v}| = \left[(v'_x)^2 + \left(v'_y\right)^2 + \left(v'_z\right)^2\right]^{1/2} \tag{2b}$$

Since the unit vectors are orthogonal to each other within a particular coordinate system, it is easy enough to write down the coordinates in the primed frame in terms of those in the unprimed frame. We have the linear transformation

$$\hat{\mathbf{x}}' \cdot \mathbf{v} = v'_x = \hat{\mathbf{x}}' \cdot \hat{\mathbf{x}}v_x + \hat{\mathbf{x}}' \cdot \hat{\mathbf{y}}v_y + \hat{\mathbf{x}}' \cdot \hat{\mathbf{z}}v_z \tag{3a}$$

$$\hat{\mathbf{y}}' \cdot \mathbf{v} = v'_y = \hat{\mathbf{y}}' \cdot \hat{\mathbf{x}}v_x + \hat{\mathbf{y}}' \cdot \hat{\mathbf{y}}v_y + \hat{\mathbf{y}}' \cdot \hat{\mathbf{z}}v_z \tag{3b}$$

$$\hat{\mathbf{z}}' \cdot \mathbf{v} = v'_z = \hat{\mathbf{z}}' \cdot \hat{\mathbf{x}}v_x + \hat{\mathbf{z}}' \cdot \hat{\mathbf{y}}v_y + \hat{\mathbf{z}}' \cdot \hat{\mathbf{z}}v_z \tag{3c}$$

which can obviously be written as a matrix if we like.

We need a notation that is both more compact than what we have above, and also one that can easily be generalized to more than three dimensions. Let the (Latin) indices i, j,... represent the numbers 1, 2, and 3, corresponding to the coordinates x , y, and z (or x', y', or z'). Write the components of $\mathbf{v}$ as v^i or v'^i in the two coordinate systems. For example, $v^1 = v_x$, $v'^2 = v'_y$, and so on. Also, define

$$a^{ij} \equiv \hat{\mathbf{x}}'^i \cdot \hat{\mathbf{x}}^j \tag{4}$$

For example, $a^{12} = \hat{\mathbf{x}}' \cdot \hat{\mathbf{y}}$. The a^{ij} define the rotation, and in fact are individually just the cosines of the angle between one axis and another. This allows us to write the transformation

2

(3) as

$$v'^i = \sum_{j=1}^{3} a^{ij} v^j \tag{5}$$

Now here is an important new agreement, called the **Einstein summation convention**. We will agree that whenever an index appears twice, then it is implied that we sum over it. (Sometimes this is called "contraction.") Therefore (5) is written from now on as

$$v'^i = a^{ij} v^j \tag{6}$$

Since a^{ij} and v^j are just ordinary numbers, the order in which we write them is irrelevant. However, in (6) we took the opportunity to write it with the summed index j "adjacent". That is, the two occurrences of j appear next to each other. That means that we can write (6) as a matrix equation with the matrices a and v written in the same order as in (6), viz

$$\begin{bmatrix} v'^1 \\ v'^2 \\ v'^3 \end{bmatrix} = \begin{bmatrix} a^{11} & a^{12} & a^{13} \\ a^{21} & a^{22} & a^{23} \\ a^{31} & a^{32} & a^{33} \end{bmatrix} \begin{bmatrix} v^1 \\ v^2 \\ v^3 \end{bmatrix} \tag{7a}$$

or, if we agree to write matrices as sans serif characters, and leave their dimensionality to context,

$$\mathsf{v}' = \mathsf{a}\mathsf{v} \tag{7b}$$

We emphasize that (6) is the unambiguous way to write the coordinate transformation. Using (7) is handy for carrying through calculations, but relies on the individual factors to be written in the correct order.

Two common vector operations are now easy to write down in our new notation. For two vectors $\mathbf{v}$ and $\mathbf{w}$ we have the "inner" or "dot" product

$$\mathbf{v} \cdot \mathbf{w} = v^1 w^1 + v^2 w^2 + v^3 w^3 = v^i w^i = \mathsf{v}^T \mathsf{w} \tag{8}$$

making use of the "transpose" operation for matrices. The "cross product" is, in terms of components

$$\begin{aligned} (\mathbf{v} \times \mathbf{w})_x &= v_y w_z - v_z w_y \qquad \text{etc} \dots \\ \text{or} \qquad (\mathbf{v} \times \mathbf{w})^i &= \epsilon^{ijk} v^j w^k \end{aligned} \tag{9}$$

(Note that this summation is to be done separately over j and k, so nine terms.) The totally antisymmetric symbol ϵ^{ijk} is discussed on page five of your textbook. Briefly, it equals $+1$ for an even permutation of $ijk = 123$, -1 for an odd permutation, and 0 if any two indices are equal.

You likely remember that you can describe the dot product geometrically. It is equal to the product of the magnitudes of the two vectors, times the cosine of the angle between them.

That means, of course, that it should be independent of which coordinate system we use to write the vector components. We say that the dot product is "invariant" under coordinate rotations. Using our new notation, we can write this mathematically as

$$\mathbf{v} \cdot \mathbf{w} = v^i w^i = (v')^i (w')^i = \left(a^{ij} v^j\right)\left(a^{ik} w^k\right) = a^{ij} a^{ik} v^j w^k \tag{10}$$

In order for this equation to hold, it must be that

$$a^{ij} a^{ik} = \delta^{jk} \tag{11}$$

in which case the (10) becomes

$$(v')^i (w')^i = \delta^{jk} v^j w^k = v^j w^j \tag{12}$$

which just says that the dot product is the same number whether we use the primed or unprimed coordinates. (It doesn't matter that the first summation is over i and the final step is over j. The summation convention turns them into dummy indices.)

Notice carefully the order of the indices in (11). The first of the two indices of a are contracted with each other. To write this as matrix multiplication, recall that the transpose of a matrix just reverses the rows and columns. That is, (11) can be written as

$$\left(a^T\right)^{ji} a^{ik} = \delta^{jk} \qquad \text{or} \qquad \mathsf{a}^T \mathsf{a} = \mathbf{1} \tag{13}$$

since, now, the indices i are adjacent. Matrices with this property are called "orthogonal."

Definitions: Scalars, Vectors, and Tensors

Given some orthogonal transformation a^{ij} we can go on to classify certain objects based on their transformation properties. This is the approach we take in this course, and in fact in any study of physical systems defined on the basis of transformations between different coordinate systems. Our new definition will contain the old definition of a vector, but can be obviously extended to include other systems, in particular those with different transformation properties and different numbers of dimensions.

- A **scalar** is a number K which has the same value in different coordinate systems.

$$K' = K \tag{14}$$

- A **vector** is a set of numbers v^i which transform according to

$$v'^i = a^{ij} v^j \tag{15}$$

4

- A (second rank) **tensor** is a set of numbers T^{ij} which transform according to

$$T'^{ij} = a^{ik}a^{jl}T^{kl} \tag{16}$$

Higher rank tensors (that is, objects with more than two indices) are defined analogously. (See Ohanian, page 13.) Indeed, a vector is a tensor of rank one, and a scalar is a tensor of rank zero.

Further segmentation of these definitions are possible, but we won't discuss them in any length here. For example, depending on whether the determinant of $\mathbf{a}$ is ± 1, we would characterize vectors as polar vectors or axial vectors. Also, it is easy to show that any tensor can be written as the sum of a symmetric part (i.e. $T^{ji} = T^{ij}$) and an antisymmetric part (i.e. $T^{ji} = -T^{ij}$). If and when we need these concepts in our course, we'll discuss them at that time.

Contravariant and Covariant Vectors and Tensors

So that we can get ready for a discussion of special relativity, we need to take up a distinction between different kinds of vectors and tensors based on abstract geometry.

If we use the symbol $\mathbf{A}$ to denote a vector, and $\mathbf{A} \cdot \mathbf{B}$ to denote the inner product between vectors, then we are unnecessarily restricting ourselves to a particular geometry. Indeed, it implies that our geometry is "flat." Two vectors will define a plane surface. You can imagine that there should be more freedom than that. In other words, why can't we have a space with some kind of "curvature", and how would we go about specifying that curvature?

We can remove this restriction by defining two types of vectors. A *contravariant* vector (or just *vector*) $\vec{A}$ will be distinguished from a covariant vector (or just *covector*) $\underset{\rightarrow}{B}$. An inner product will only be defined between a contravariant vector and a covector, namely $\vec{A} \cdot \underset{\rightarrow}{B} = \underset{\rightarrow}{B} \cdot \vec{A}$. If I want to define an angle between two vectors by using the inner product, then I need to somehow determine the geometry. This is done with the metric tensor $\overset{\rightarrow\rightarrow}{g}$ which turns a covector into a vector by the operation $\vec{B} = \overset{\rightarrow\rightarrow}{g} \cdot \underset{\rightarrow}{B}$. The inverse metric tensor $\underset{\rightarrow\rightarrow}{g}$ turns a vector into a covector. Clearly, $\vec{A} \cdot \underset{\rightarrow}{B} = \underset{\rightarrow}{A} \cdot \vec{B}$ implies that $\overset{\rightarrow\rightarrow}{g} \cdot \underset{\rightarrow\rightarrow}{g} = \overset{\rightarrow}{1}$.

Is there a good way to get a pictorial representation of vectors and covectors? Some options are discussed in Amer. Jour. Phys. 65(1997)1037. One of these is depicted in Figure 2. The contravariant vector is a "stick" whose magnitude is the length of the stick, while the covector is like a serving of lasagna. The magnitude of the covector is the density of lasagna noodles.
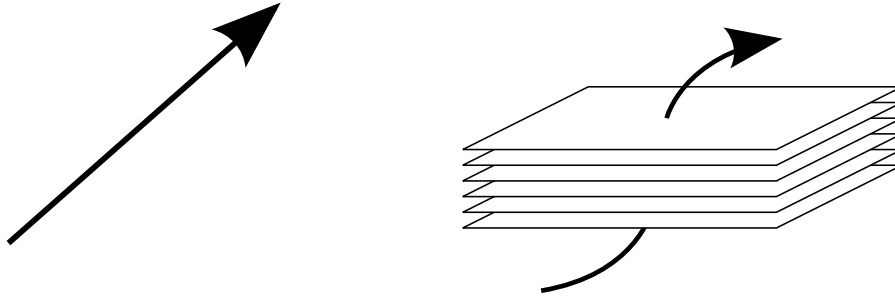
Figure 2: Examples of how you can picture contravariant and covariant vectors. A contravariant vector is a "stick" with a direction to it. Its "worth" (or "magnitude") is proportional to the length of the stick. A covariant vector is like "lasagna." Its worth is proportional to the density of noodles; that is, the closer together are the sheets, the larger is the magnitude of the covector. These and other pictorial examples of visualizing contravariant and covariant vectors are discussed in Am.J.Phys.65(1997)1037.

Figure 3: Pictorial representation of the inner product between a contravariant vector and a covariant vector. The "stick" is imbedded in the "lasagna" and the inner product is equal to the number of noodles pierced by the stick. The inner product shown has the value 5.
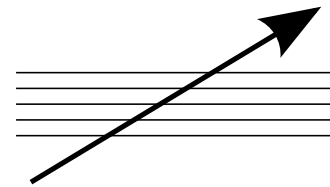
The classic 3D example of a covector is the gradient. You actually think of the gradient for some function in terms of contour lines, along which the function has a constant value. When the contours are close together, the gradient is larger than when they are far apart. (Before the end of this lesson, we will indeed show that a gradient transforms the way a covector should transform, as opposed to a vector, in the case of special relativity.)

This representation lends itself to a nice pictorial representation of the inner product between a vector and covector. See Figure 3. This gives us a handy way to view the inner product without having to resort to the definition of an "angle." In order to define angles we need to define the metric tensor.

Of course, this is all consistent with our usual high school three dimensional geometry. We put all of this contravariant/covariant business behind us just by letting the metric tensor be the identify transformation.

However, the "geometry" of spacetime is a physical situation that is well suited to the distinction between contravariant and covariant vectors. Indeed, the metric tensor has important physical significance in this case.

Application to (Special) Relativity

This formal way of treating vectors is important for a good understanding of Einstein's theory of relativity.

First, some terminology. We work in four-dimensional "spacetime" instead of three dimensional space. In place of a "coordinate system" we speak of a "reference frame." Instead of a "coordinate transformation" we use the Lorentz Transformation to relate different reference frames. Vectors and tensors will be specialized to "four-vectors" and "four-tensors", although we'll frequently just drop the "four-".

We write everything here in terms of coordinates, and will not try to do things purely geometrically. We use upper and lower indices to distinguish components of contravariant and covariant vectors, respectively, namely

$$\textbf{Contravariant vector components}: \quad A^\mu \quad \text{where } \mu = 0, 1, 2, 3$$
$$\textbf{Covariant vector components}: \quad A_\mu \quad \text{where } \mu = 0, 1, 2, 3$$

Inner products will only be between contravariant and covariant vectors or tensor components. That is, we will only be contracting upper indices with lower ones.

The geometry of spacetime is represented by the metric tensor $g_{\mu\nu}$. That is, the invariant spacetime interval is given by

$$ds^2 = dx^\mu dx_\mu = dx^\mu g_{\mu\nu} dx^\nu$$

In principle, $g_{\mu\nu}$ can be a function of spacetime, i.e. $g_{\mu\nu} = g_{\mu\nu}(x^\alpha)$. The equations of Einstein's general theory of relativity are used to determine $g_{\mu\nu}$ in the presence of distributions of matter and energy.

Special relativity is the case of flat spacetime, which is all that we will discuss in this course. The metric tensor for flat spacetime is $g_{\mu\nu} = \eta_{\mu\nu}$ where $\eta_{00} = 1$, $\eta_{11} = \eta_{22} = \eta_{33} = -1$, and all of the other elements are zero. Notice that the (components of the) inverse metric tensor $\eta^{\mu\nu} = \eta_{\mu\nu}$. That is, it is simple to show that

$$\eta^{\mu\alpha}\eta_{\alpha\nu} = \delta^\mu_\nu$$

Don't forget the Einstein summation convention, which now means we sum over repeated indices equal to 0, 1, 2, and 3.

The Lorentz transformation is now written as Ohanian (7.5), namely

$$x'^\mu = a^\mu_{\ \nu} x^\nu \tag{17}$$

$$\text{where} \quad a^\mu_{\ \nu} = a^\mu_{\ \nu}(+V) = \begin{bmatrix} \gamma & -\gamma V/c & 0 & 0 \\ -\gamma V/c & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{18}$$

for boosts in the x^1 direction. Such a transformation indeed leaves

$$ds^2 = dx^\mu \eta_{\mu\nu} dx^\nu = (cdt)^2 - dx^2 - dy^2 - dz^2$$

invariant. This was our original starting point for special relativity.

Note our care in (17) and (18) so that when we mix contravariant and covariant indices, it is clear which is first, and therefore labels the row, and which is second, labeling the column. We know that the inverse Lorentz transformation only amounts to $V \to -V$, and so

$$x^\mu = a^\mu{}_\nu(-V)x'^\nu = a^\mu{}_\nu(-V)a^\nu{}_\alpha(+V)x^\alpha \tag{19}$$

Switching dummy indices α and ν, this means that

$$a^\mu{}_\alpha(-V)a^\alpha{}_\nu(+V) = \delta^\mu_\nu \tag{20}$$

which can be written in terms of 4×4 matrices as

$$\mathsf{a}(-V)\mathsf{a}(+V) = 1 \tag{21}$$

It is worth a few minutes' time to convince yourself that the matrix in (18) satisfies (21).

How do covectors transform? This is easy to answer using the metric tensor. So,

$$x'_\mu = \eta_{\mu\alpha}x'^\alpha = \eta_{\mu\alpha}a^\alpha{}_\beta x^\beta = \eta_{\mu\alpha}a^\alpha{}_\beta \eta^{\beta\nu} x_\nu$$

In other words, the transformation equation for covectors is

$$x'_\mu = a_\mu{}^\nu x_\nu \tag{22}$$
$$\text{where} \quad a_\mu{}^\nu = \eta_{\mu\alpha}a^\alpha{}_\beta \eta^{\beta\nu} \tag{23}$$

It is simple to show that $a_\mu{}^\nu(+V) = a^\mu{}_\nu(-V)$. See Ohanian (7.22). Therefore

$$a_\alpha{}^\mu a^\alpha{}_\nu = \delta^\mu_\nu \tag{24}$$

In other words, Lorentz transformation show exactly the same orthogonality rule, namely "sum over the first index to get the identity matrix", that we encountered with old fashioned three dimensional vectors in (11). We will make much use of (24).

That is pretty much everything we need to proceed with our study of electromagnetism. The rest of this note gives some sample problems, and then their solutions. First, though, let us summarize our fundamental objects on the basis of their fundamental properties under Lorentz transformations between two reference frames moving with relative speed V:

- A **scalar** K is a number such that $K' = K$. That is, it has the same value in all reference frames.

- A **contravariant four-vector** A^μ is a set of four numbers which transform as

$$A'^\mu = a^\mu{}_\nu A^\nu \tag{25}$$

- A **covariant four-vector** A_μ is a set of four numbers which transform as

$$A'_\mu = a_\mu{}^\nu A_\nu \tag{26}$$

Note equations (23) and (24) which relate the matrices $a^\mu{}_\nu$ and $a_\mu{}^\nu$.

- A (second rank) **fully contravariant tensor** $T^{\mu\nu}$ is a set of sixteen numbers which transform as

$$T'^{\mu\nu} = a^\mu{}_\alpha a^\nu{}_\beta T^{\alpha\beta} \tag{27}$$

Higher rank tensors are obvious generalizations.

- A (second rank) **fully covariant tensor** $T_{\mu\nu}$ is a set of sixteen numbers which transform as

$$T'_{\mu\nu} = a_\mu{}^\alpha a_\nu{}^\beta T_{\alpha\beta} \tag{28}$$

There are some obvious generalizations of tensors. Clearly, a vector can be considered a first rank tensor, and a scalar is a zero rank tensor. Higher rank tensors are clearly possible as well, and their definitions straightforward extensions. We can also have mixed contra/covariant tensors. For example, the set of sixteen numbers $T^\mu{}_\nu$ would transform as

$$T'^\mu{}_\nu = a^\mu{}_\alpha a_\nu{}^\beta T^\alpha{}_\beta$$

Finally, remember that if you want to use matrix multiplication to carry out any contractions with tensors, make sure that you write the matrices in the correct order. The correct order is the one which matches adjacent indices.