Massachusetts Institute of Technology
Department of Physics
8.033 Fall 2021

Lecture 9
Some more math: The metric tensor, dual vectors, and tensors more generally

9.1 The scalar product revisited

Similar to Lecture 5, this lecture again largely focuses on mathematical issues. We have introduced you to 4-vectors, and have shown how they can be used to organize a *Lorentz covariant* presentation of some of the laws of physics. In this lecture, we expand the "vocabulary" of mathematical objects that we use to describe quantities in relativistic physics.

We begin by revisiting the the scalar product between two 4-vectors,

$$\vec{a} \cdot \vec{b} = -a^0 b^0 + a^1 b^1 + a^2 b^2 + a^3 b^3 \ . \tag{9.1}$$

It is not difficult to show that $\vec{a} \cdot \vec{b}$ is invariant. Indeed, "scalar product" refers to the fact that a "scalar" in relativistic physics is a quantity that is invariant across Lorentz frames.

As written, there is nothing wrong with Eq. (9.1); indeed, we used it in this form to help understand invariants associated with relativistic energy and momentum. However, from a certain perspective Eq. (9.1) is somewhat "distasteful." It's necessary to write the whole expression out; there's no nice shorthand that let's you write this expression in index notation if we follow this form.

To correct for these shortcomings, we introduce a new mathematical object called the *metric*. The metric is a *tensor*, a mathematical object that we define more completely below. For now, you can regard it is an object with two indices that is represented in a particular Lorentz frame by a matrix. The metric has components $\eta_{\alpha\beta}$ given by

$$\eta_{\alpha\beta} \doteq \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \ . \tag{9.2}$$

As in the previous lecture, we use the symbol "$\doteq$" to stand for "the object on the left-hand side has the components on the right-hand side." Using the metric, you should be able to convince yourself quite easily that Eq. (9.1) is equivalent to

$$\vec{a} \cdot \vec{b} = \sum_{\alpha=0}^{3} \sum_{\beta=0}^{3} \eta_{\alpha\beta} a^\alpha b^\beta = \eta_{\alpha\beta} a^\alpha b^\beta \ . \tag{9.3}$$

The second form, using the Einstein summation convention, is how the invariant scalar product is most commonly written out.

Let's see what the invariance of the scalar product tells us about how the components of the metric transform between reference frames. Suppose that observer $\mathcal{O}$ measures $\vec{a}$ and $\vec{b}$ to have components a^α and b^β, and they use $\eta_{\alpha\beta}$ for metric components. Observer $\mathcal{O}'$ measures these vectors to have components $a^{\mu'}$ and $b^{\nu'}$, and they use $\eta_{\mu'\nu'}$ for metric components. The components of the vectors are related in the usual way by the Lorentz transformation matrix:

$$a^\alpha = \Lambda^\alpha{}_{\mu'} a^{\mu'} \tag{9.4}$$
$$b^\beta = \Lambda^\beta{}_{\nu'} b^{\nu'} \; . \tag{9.5}$$

How do we compute the metric components used by $\mathcal{O}'$? We figure this out by enforcing invariance:

$$\begin{aligned} \vec{a} \cdot \vec{b} &= \eta_{\alpha\beta} a^\alpha b^\beta \\ &= \eta_{\alpha\beta} \left(\Lambda^\alpha{}_{\mu'} a^{\mu'}\right) \left(\Lambda^\beta{}_{\nu'} b^{\nu'}\right) \\ &= \left(\eta_{\alpha\beta} \Lambda^\alpha{}_{\mu'} \Lambda^\beta{}_{\nu'}\right) a^{\mu'} b^{\nu'} \; . \end{aligned} \tag{9.6}$$

This quantity is an invariant provided we transform the components of the metric via the rule

$$\eta_{\mu'\nu'} = \eta_{\alpha\beta} \Lambda^\alpha{}_{\mu'} \Lambda^\beta{}_{\nu'} \; . \tag{9.7}$$

Notice that this is basically just the "line up the indices" rule that we discussed when we introduced index notation. **CAUTION:** if you want to do this analysis using matrix multiplication techniques that you learned in linear algebra, you must be very careful — it is quite easy to go awry. See my comment in the final section of these lecture notes.

I've gone through the calculation of how the metric transforms with some care because I want to make clear the *principle* behind how we transform these components. In a few pages, we are going to apply the ideas discussed here to tensors in general; indeed, the behavior of quantities under transformation is central to our definition of what a tensor is. That said, it must be noted that for the metric the final result is so simple that all the calculation presented above surely will feel like distressing overkill: $\eta_{\alpha\beta}$ is represented by the matrix

$$\begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{9.8}$$

in *all* Lorentz frames. This can be proved by computing Eq. (9.7).

One last detail: you might be wondering what happened, in Eq. (9.3), with the unit vectors which go into the vectors $\vec{a}$ and $\vec{b}$. After all, if $\vec{a} = a^\alpha \vec{e}_\alpha$ and $\vec{b} = b^\beta \vec{e}_\beta$, shouldn't it also be the case that

$$\vec{a} \cdot \vec{b} = \left(a^\alpha b^\beta\right) \left(\vec{e}_\alpha \cdot \vec{e}_\beta\right) \tag{9.9}$$

is exactly equivalent to the form we wrote down in involving $\eta_{\alpha\beta}$?

The answer is certainly *yes*. Comparing Eqs. (9.3) and (9.9) shows us that for these forms to be equivalent, then we must have

$$\vec{e}_\alpha \cdot \vec{e}_\beta = \eta_{\alpha\beta} \; . \tag{9.10}$$

This, at last, allows us to see how the geometric objects $\vec{e}_\alpha$ are, in fact, *unit* vectors: the scalar product of any two unit vectors is zero if $\alpha \neq \beta$; the scalar product is 1 when $\alpha = \beta$ and correspond to one of the spatial directions; and the scalar product is -1 when $\alpha = \beta = t$.

The negative scalar product is what we expect for timelike vectors, so $\vec{e}_t \cdot \vec{e}_t = -1$ should make sense, although it looks starkly different from the "modulus squared" you have seen with unit vectors in previous classes.

As discussed above, $\eta_{\alpha\beta}$ is represented by the matrix (9.8) in *all* reference frames. This means that when we change frames, and then build the unit vectors in the new frame,

$$\vec{e}_{\mu'} = \Lambda^{\alpha}{}_{\mu'} \vec{e}_{\alpha} \;, \tag{9.11}$$

we must have $\vec{e}_{t'} \cdot \vec{e}_{t'} = -1$, $\vec{e}_{x'} \cdot \vec{e}_{x'} = 1$, $\vec{e}_{x'} \cdot \vec{e}_{z'} = 0$, etc. You will test out this expectation on an upcoming problem set.

We wrap up our discussion of the metric with a few comments:

- Note that writing out that matrix over and over again is tedious and tiring. As shorthand, we will often write $\mathrm{diag}(-1, 1, 1, 1)$ rather than the full 4×4 matrix. This is shorthand for "the matrix which has -1, 1, 1, 1 on the diagonal, and has zero everywhere else."

- For reasons that will be clearer in the next section, it is useful to define an *inverse* metric: we define $\eta^{\alpha\beta}$ by the rule that

  $$\eta^{\alpha\beta}\eta_{\beta\gamma} = \delta^{\alpha}{}_{\gamma} \;. \tag{9.12}$$

  Recall that $\delta^{\alpha}{}_{\gamma}$ is known as the Kronecker delta. It equals 1 if $\alpha = \gamma$, and equals 0 otherwise [equivalently, we can say $\delta^{\alpha}{}_{\gamma} \doteq \mathrm{diag}(1, 1, 1, 1)$.] The matrix representation of the components $\eta^{\alpha\beta}$ is exactly the same as the matrix representation of the components $\eta_{\alpha\beta}$ — both are given by $\mathrm{diag}(-1, 1, 1, 1)$.

- The metric is not always going to be as simple as $\mathrm{diag}(-1, 1, 1, 1)$. The metric becomes more complicated when we start using different coordinate systems; and, it becomes *significantly* more complicated when we move from special relativity to general relativity. In these cases, the components of the metric become functions of the coordinates. We will denote the metric by $g_{\alpha\beta}$ when it becomes necessary for us to make it more complicated; we will always use $\eta_{\alpha\beta}$ for the metric that is represented by the matrix $\mathrm{diag}(-1, 1, 1, 1)$. This is the form that we use in special relativity with Cartesian spatial coordinates. (Such coordinates are often called *inertial* coordinates, since an observer at constant Cartesian spatial coordinate moves with constant velocity in all Lorentz frames.)

- Finally, the word "metric" comes from a root that means to *measure*. This is because by using the metric we can write the invariant interval $ds^2 = \eta_{\alpha\beta} dx^{\alpha}\, dx^{\beta}$ — the metric is the mathematical object which introduces a notion of measurable distance between two events, one located at x^{α}, the other at $x^{\alpha} + dx^{\alpha}$. This may seem fairly trivial given what we have discussed so far, but it becomes substantially less trivial when we move into more complicated geometries. In those cases, we will write $ds^2 = g_{\alpha\beta} dx^{\alpha}\, dx^{\beta}$. The behavior of $g_{\alpha\beta}$ is very important for understanding the distance between two coordinate points in these more complicated cases.

9.2 Lowering and raising indices

When we compute $\vec{a} \cdot \vec{b} = \eta_{\alpha\beta} a^\alpha b^\beta$, we say that we are *contracting* the metric with $\vec{a}$ and $\vec{b}$ on the indices α and β. What do we get if we contract the metric with a single vector, on only one index? In other words, what is $\eta_{\alpha\beta} a^\alpha$?

As is the way in mathematics, when we encounter a construction like this, we use it to define something new. In this case, we define a quantity with an index in the "downstairs" position:

$$a_\beta \equiv \eta_{\alpha\beta} a^\alpha \ . \tag{9.13}$$

For reasons that are hopefully obvious, this operation is called *lowering* the index on the vector components a^α. The components in the "downstairs" position are sometimes called *dual* to the components with index "upstairs."

In special relativity using inertial coordinates, lowering the index flips the sign of the zero component: $a_0 = -a^0$, but $a_1 = a^1$, $a_2 = a^2$, $a_3 = a^3$. Lowering the index gives us another way to construct the inner product:

$$\vec{a} \cdot \vec{b} = a_\alpha b^\alpha = a^\alpha b_\alpha \ . \tag{9.14}$$

If the metric lowers an index, then it is hopefully not too surprising that the inverse metric raises it:

$$\eta^{\alpha\beta} a_\alpha = \eta^{\alpha\beta} \left(\eta_{\alpha\mu} a^\mu\right) = \left(\eta^{\alpha\beta} \eta_{\alpha\mu}\right) a^\mu = \delta^\beta{}_\mu a^\mu = a^\beta \ . \tag{9.15}$$

This is why the inverse metric was introduced — it gives us a tool to reverse the lowering operation which the metric performs.

How do the components a_α transform between reference frames? You can probably guess based on the "line up the indices" rule, but to be sure, let's carefully compute how components in the frame of $\mathcal{O}'$ are related to components in the frame of $\mathcal{O}$:

$$\begin{aligned} a_{\alpha'} &= \eta_{\alpha'\beta'} a^{\beta'} \\ &= \left(\Lambda^\mu{}_{\alpha'} \Lambda^\nu{}_{\beta'} \eta_{\mu\nu}\right) \left(\Lambda^{\beta'}{}_\sigma a^\sigma\right) \\ &= \left(\Lambda^\mu{}_{\alpha'} \Lambda^\nu{}_{\beta'} \Lambda^{\beta'}{}_\sigma\right) \eta_{\mu\nu} a^\sigma \\ &= \Lambda^\mu{}_{\alpha'} \delta^\nu{}_\sigma \eta_{\mu\nu} a^\sigma \\ &= \Lambda^\mu{}_{\alpha'} \eta_{\mu\nu} a^\nu \\ &= \Lambda^\mu{}_{\alpha'} a_\mu \ . \end{aligned} \tag{9.16}$$

On the first line, we write the lowering operation with all quantities written using components in the frame of $\mathcal{O}'$. On the second line, we introduce the Lorentz transformation matrices that express those $\mathcal{O}'$-frame quantities in terms of $\mathcal{O}$-frame quantities. On the third line, we rearrange the terms slightly, then on the fourth line we sum over the index β'. This yields the Kronecker delta by combining the second and third Lorentz transformation matrices. On the fifth line, we sum over the index σ, which (thanks to the properties of the Kronecker delta) changes the a^σ to a^ν. On the last line, we lower the index. The end result shows that to transform "downstairs" components, we indeed just "line up the indices."

As mentioned in a previous lecture, "upstairs" components are often called contravariant, and downstairs ones are called covariant. We now see that the metric and inverse metric are the tools we use to flip between the two forms. This holds up in general, including when the metric becomes more complicated than diag$(-1, 1, 1, 1)$.

The 4-vectors we have discussed so far (spacetime displacement, 4-momentum, 4-velocity) are most "naturally" presented with their indices up. This is largely because they descend from the spacetime displacement vector, $\Delta\vec{x} = \Delta x^\alpha \vec{e}_\alpha$, in which the physical quantity we care about is the set of coordinate displacements Δx^α. There are some quantities which are most "naturally" expressed using indices down. The prototypical example of this is the spacetime *gradient*. Suppose that $\phi(\vec{x})$ is a scalar field — that is, it is a field whose value at the event located $\vec{x}$ away from the origin is the same to all inertial observers. Then we define its gradient by

$$A_\alpha = \frac{\partial\phi}{\partial x^\alpha} \equiv \partial_\alpha\phi \; . \tag{9.17}$$

On a problem set, you will show that if $x^{\mu'} = \Lambda^{\mu'}{}_\alpha x^\alpha$, then $A_{\alpha'} = \Lambda^\mu{}_{\alpha'} A_\mu$ — under Lorentz transformations, the gradient behaves like a "downstairs" index quantity.

The metric lets us define a variation on the gradient. Let us define

$$x_\alpha = \eta_{\alpha\mu} x^\mu \; . \tag{9.18}$$

The components of this "downstairs" variant of x^μ are identical, except for the time-like piece, which picks up a minus sign:

$$x_0 = -x^0 = -ct \; ; \qquad x_{1,2,3} = x^{1,2,3} \; . \tag{9.19}$$

We define our variant of the gradient using derivatives with respect to x_α:

$$A^\alpha = \frac{\partial\phi}{\partial x_\alpha} \equiv \partial^\alpha\phi \; . \tag{9.20}$$

It's not hard to show that this quantity transform like an "upstairs" index quantity, hence our association of it with A^α.

One of the places where this is really useful is that we can combine and contract the two notions of gradient to produce a combination of second derivatives that is a Lorentz invariant operator. Let's look at what happens when we act both notions of gradient with the indices contracted onto scalar field ϕ:

$$\partial^\alpha\partial_\alpha\phi = -\frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} + \frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2} \equiv \Box\phi \; . \tag{9.21}$$

You may recognize this combination of derivatives as exactly what we have for quantities that obey a wave equation. Indeed, the combination $\partial^\alpha\partial_\alpha$, often denoted with the "box" symbol $\Box$, is called the *wave operator*. Notice that it has no free indices.

9.3 Tensors

The metric is an example of a family of mathematical objects called *tensors* which are used in many places in physics. They are particularly important in both special and general relativity.

Tensors are geometric objects whose components are represented by quantities with indices on them. The metric tensor is the first example we have seen of a tensor with two indices, but this generalizes — tensors can have an arbitrary number[1] of indices. Their

[1] In my research, I use a tensor with 4 indices more or less daily, and have done work that involves tensors with 5 and 6 indices.

defining characteristic is the transformation law: a quantity is a tensor if it transforms with a transformation matrix "correcting" each of its indices. For example, suppose physics tells us that we care about a quantity with 4 indices, one in the up position and three down: $R^{\mu}{}_{\alpha\beta\gamma}$. This quantity is a tensor if it transforms between reference frames with the rule

$$R^{\mu'}{}_{\alpha'\beta'\gamma'} = R^{\mu}{}_{\alpha\beta\gamma}\Lambda^{\mu'}{}_{\mu}\Lambda^{\alpha}{}_{\alpha'}\Lambda^{\beta}{}_{\beta'}\Lambda^{\gamma}{}_{\gamma'} \ . \tag{9.22}$$

The number of indices used for a tensor's components (and hence the number of transformation matrices used to transform it) tells us the tensor's *rank*. The example (9.22) is a rank-4 tensor. The metric is a rank-2 tensor. Vectors are rank-1 tensors; they transform using one transformation matrix. Scalars — Lorentz invariants — are often considered to be rank-0 tensors; they transform with *no* transformation matrices, since they are the same in all frames. Note that the wave operator we defined in the previous section acts like a scalar (more properly, a "scalar operator," because it defines a combination of derivatives that operate in the same way in all frames).

In 8.033, we will work almost entirely with tensors of rank 0, 1, and 2. (We will briefly discuss higher rank tensors when we move from special relativity to general relativity, but the discussion will be largely qualitative.) Rank-2 tensors are sufficiently important that they are worth some detailed discussion. Many rank-2 tensors can be regarded as a quantity that, in essence, points in two directions at once. For example, in a few lectures we will discuss a quantity called the "stress-energy tensor" which describes the flux of 4-momentum. Components $T^{\alpha\beta}$ of this tensor describe the flux of 4-momentum component p^{α} in the x^{β} direction.

In general, rank-2 tensors in spacetime have 16 components — 4 for each index. However, many rank-2 tensors have symmetry properties that allows us to relate some of the components to each other:

- A tensor $S^{\alpha\beta}$ is *symmetric* if it has the property that $S^{\alpha\beta} = S^{\beta\alpha}$. This reduces the number of independent components from 16 to 10: the four components on the diagonal, plus half of the 12 off-diagonal components. The stress-energy tensor mentioned above has this property; so does the metric, even in the general form $g_{\alpha\beta}$.

- A tensor $A^{\alpha\beta}$ is *antisymmetric* if it has the property that $A^{\alpha\beta} = -A^{\beta\alpha}$. This reduces the number of independent components from 16 to 6. The four components on the diagonal must be zero (this is the only solution to $A^{\alpha\beta} = -A^{\beta\alpha}$ if $\alpha = \beta$), and we have half of the 12 off-diagonal components. We will soon find that an antisymmetric tensor $F^{\alpha\beta}$ allows us to describe electric and magnetic fields in a covariant formulation.

9.4 Aside: Using matrix multiplication to combine tensors and matrices

Once we start working with rank-2 tensors, there is a class of mistakes that 8.033 instructors tend to encounter from students who use their knowledge of linear algebra to work through equations that involve products of tensors. Let me emphasize again that this can be done correctly, but it requires that you be very careful to think through how the multiplication of the different tensors works.

Suppose you need to construct a tensor $A^{\alpha\beta}$ which is given by combining three tensors together. For instance, suppose we have

$$A^{\alpha\beta} = B_{\mu\nu}D^{\alpha\mu}F^{\beta\nu} \ . \tag{9.23}$$

By far the most common mistake we see is that people write this as the following (wrong!) equation:

$$\mathsf{A}_{\text{WRONG}} = \mathsf{B} \cdot \mathsf{D} \cdot \mathsf{F} \,, \tag{9.24}$$

where

$$\mathsf{A}_{\text{WRONG}} = \begin{pmatrix} A^{00} & A^{01} & A^{02} & A^{03} \\ A^{10} & A^{11} & A^{12} & A^{13} \\ A^{20} & A^{21} & A^{22} & A^{23} \\ A^{30} & A^{31} & A^{32} & A^{33} \end{pmatrix}_{\text{WRONG}} \,, \qquad \mathsf{B} = \begin{pmatrix} B_{00} & B_{01} & B_{02} & B_{03} \\ B_{10} & B_{11} & B_{12} & B_{13} \\ B_{20} & B_{21} & B_{22} & B_{23} \\ B_{30} & B_{31} & B_{32} & B_{33} \end{pmatrix} \,, \tag{9.25}$$

with D and F defined similarly.

Why is this wrong? When we represent a rank-2 tensor by a matrix, the first index corresponds to the row, the second index to the column. We need to make sure that when we contract on indices, we are correctly linking up rows and columns of the different objects.

With this in mind, let's carefully examine Eq. (9.23). To produce $A^{\alpha\beta}$, we first contract $B_{\mu\nu}$ on its first index with the second index of $D^{\alpha\mu}$. In matrix form, this means we select column ν of B, we select row α of D, and we combine:

$$B_{\mu\nu} D^{\alpha\mu} \mapsto \mathsf{D} \cdot \mathsf{B} \,. \tag{9.26}$$

We also need to contract the second index of $B_{\mu\nu}$ with the *second* index of $F^{\beta\nu}$. In other words, when we put things in matrix form, we select row μ of B and combine it with row β of F. In the language of matrix multiplication, this means we are multiplying B with the *transpose* of F.

The correct translation of Eq. (9.23) into matrix form is thus

$$\mathsf{A} = \mathsf{D} \cdot \mathsf{B} \cdot \mathsf{F}^T \,, \tag{9.27}$$

where the T superscript denotes matrix transpose. We see that the wrong response is wrong in two ways: it puts the matrices in the wrong order (and since matrix multiplication does not commute, that can have serious consequences), and it uses F rather than its transpose F^T. (In some cases, failing to use the transpose may be harmless because the underlying matrix is symmetric, so the matrix and its transpose are identical. That's a case of getting lucky — you can't count on it working. If the matrix is in fact antisymmetric, you'll wind up with a minus sign that could drive you slightly mad.)

Carefully following the logic described here to combine rank-2 tensors via matrix multiplication will work. However, it must be emphasized that simply working with the index format *always* just works. You don't need to do any of this careful vetting of which index is combining with which, and writing out the matrices accordingly.

It must also be emphasized that this way of mapping index equations into linear algebra becomes more or less impossible to use once we move beyond rank-2 tensors. For instance, as I type up these notes, part of my brain is consumed by a research paper I am writing with a graduate student that is largely concerned with finding solutions to the equation

$$\frac{Dp^\mu}{d\tau} = -\frac{1}{2} R^\mu{}_{\alpha\beta\gamma} u^\alpha S^{\beta\gamma} \,. \tag{9.28}$$

This equation tells us how a body's momentum changes as it moves through spacetime if the body's 4-velocity has components u^α, and the body is itself spinning (the tensor components

$S^{\beta\gamma}$ describe its spin in relativistic language). The operator $D/d\tau$ is a special kind of time derivative, and the tensor components $R^{\mu}{}_{\alpha\beta\gamma}$ describe the action of gravitational tides in general relativity. There is really no way we can put an equation like this into a form that is matrix-like. Instead, we just run through the indices and combine everything by straightforward multiplication and summation of the quantities written out index by index. Using computer algebra tools (we live and breathe with Mathematica), this isn't so bad, as long as everything is set up and defined carefully.