

# Notes on Linear Algebra With and Without Indices

Mathew Calkins

*Center for Cosmology and Particle Physics,  
New York University  
mc8558@nyu.edu.*

December 29, 2025

## Abstract

Some supplementary discussion of vectors, matrices, and indices in Euclidean and Minkowski spaces.

## Contents

<b>1</b>	<b>Euclidean Space</b>	<b>1</b>
1.1	Prelude: Free and Dummy Variables	1
1.2	Vectors and Components	3
1.3	Linear Transformations	4
1.4	Change of Basis	5
<b>2</b>	<b>Minkowski Space</b>	<b>7</b>
2.1	Raised and Lowered Greek Indices	7
2.2	No free indices = Lorentz-invariant	7
2.3	The Invariant Interval And The Inner Product	9
2.4	Lorentz Transformations	11
2.5	Preserving the Metric	12

## 1 Euclidean Space

### 1.1 Prelude: Free and Dummy Variables

Before we get into the interesting mathematics, we clarify a bit of notational vocabulary. Suppose you are doing a calculus problem (solving a PDE by Green's function methods, perhaps) and you end up writing down an equation of the form

$$f(x) = \int_a^b K(x, y)g(y)dy.$$

Take a moment to notice why this equation makes sense. On the LHS we have a single variable  $x$ , while on the RHS we have two variables  $x$  and  $y$ . But these play very different roles. The variable  $x$  is free to be varied or set to some particular value. We could for example evaluate at  $x = 5$

$$f(5) = \int_a^b K(5, y)g(y)dy.$$

or differentiate with respect to  $x$

$$f'(x) = \int_a^b \frac{\partial K(x, y)}{\partial x} g(y)dy.$$

We call  $x$  a *free* variable. But  $y$  is very different. The RHS is not quite a function of  $y$ ; rather,  $y$  instructs us to perform a certain definite integral. We could just as well call it by some other letter, replacing  $y \rightarrow z$  or  $y \rightarrow \Omega$ .

$$\int_a^b K(x, y)g(y)dy = \int_a^b K(x, z)g(z)dz = \int_a^b K(x, \Omega)g(\Omega)d\Omega.$$

We couldn't replace  $y \rightarrow x$ , since this would give a totally different integral in general

$$\int_a^b K(x, y)g(y)dy \neq \int_a^b K(x, x)g(x)dx$$

but otherwise we can replace  $y$  with any other symbol we like. By contrast with the free variable  $x$ , we call  $y$  a *dummy* variable. In this language, we say that the *free variable structure* of our original equality makes sense. Both sides have exactly one free  $x$  and nothing else.

This is not to say that we can never have disagreement between the free variable structure of the two sides of an equation. For example, in solving PDEs we sometimes get equations like

$$f(x) = g(y).$$

But this means something extremely special has happened, and implies that  $f$  and  $g$  must both be constant functions. Outside of that edge case, if you find that you have written  $f(x) = g(y)$  it is a good bet that you have transcribed something wrong earlier in your calculation.

This phenomenon of free and dummy indices also played a role in your courses on linear algebra, though you may not have called them that. For example, given a vector equality

$$\vec{y} = M\vec{x}$$

or in components (assuming for ease of typesetting that everything is in two dimensions)

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

we learned in our youths that  $M$  acts on  $\vec{x}$  by the rule

$$y_i = \sum_{j=1}^2 M_{ij}x_j.$$

Here again we see free and dummy indices, though everything is now discrete. We have a single free index  $i$  on the LHS, while the RHS has a single free index  $i$  plus a dummy index  $j$ . And just as before, we could have called  $j$  whatever we want (except for  $i$ ).

$$\sum_{j=1}^2 M_{ij}x_j = \sum_{k=1}^2 M_{ik}x_k = \sum_{L=1}^2 M_{iL}x_L.$$

We could also run across objects with multiple dummy indices. For example, if we wanted to sum up all of the entries of  $M$  we could compute any of

$$\sum_{i,j=1}^2 M_{ij} = \sum_{a,b=1}^2 M_{ab} = \sum_{X,Y=1}^2 M_{XY}.$$

All of these are just shorthand for summing up every component of  $M$ , so they must all give the same result, namely  $M_{11} + M_{12} + M_{21} + M_{22}$ . By the simplest possible relabeling, we could simply swap  $i$  and  $j$ .

$$\sum_{i,j=1}^2 M_{ij} = \sum_{i,j=1}^2 M_{ji}.$$

This is pure notation. It doesn't require  $M$  to be a symmetric matrix, it's just a relabelling of dummy indices. One of the surprising facts about the index notation used in relativity and differential geometry is that doing this kind of almost trivial relabeling is a key step in many calculations.

## 1.2 Vectors and Components

One of the first things we learn about vectors is that they have *components*. For example, in  $\mathbb{R}^3$  we might variously write

$$\vec{A} = (A_1, A_2, A_3) = \begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix} = (A_x, A_y, A_z) = \begin{bmatrix} A_x \\ A_y \\ A_z \end{bmatrix}.$$

But  $\mathbb{R}^n$  is a slightly unusual vector space. Its elements are **by definition**  $n$ -tuples of real numbers. In any other real vector space, expressing a vector as a list of numbers requires picking a *basis* for that space. The fact that  $\mathbb{R}^n$  doesn't require making this choice in an arbitrary way is reflected in the existence of a *standard basis* for  $\mathbb{R}^n$ , usually written

$$\text{standard basis for } \mathbb{R}^3 = \hat{i}, \hat{j}, \hat{k} \text{ or } \hat{x}, \hat{y}, \hat{z} \text{ or } \hat{e}_1, \hat{e}_2, \hat{e}_3.$$

In terms of the standard basis we can rewrite  $\vec{A} \in \mathbb{R}^3$  in a way that better matches the notation we must use for other vector spaces, namely

$$\begin{aligned} \vec{A} &= \begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix} \\ &= A_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + A_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + A_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\ &= A_1 \hat{e}_1 + A_2 \hat{e}_2 + A_3 \hat{e}_3 \\ &= \sum_{i=1}^3 A_i \hat{e}_i. \end{aligned}$$

This basis has many nice properties that aren't true of general bases. Probably the nicest is that it is an *orthonormal* basis. Each standard basis vector is a unit vector, and any two distinct standard basis vectors are orthogonal as measured by the dot product. We can write both of these properties simultaneously as

$$\hat{e}_i \cdot \hat{e}_j = \delta_{ij} = \begin{cases} 1 & i = j, \\ 0 & i \neq j. \end{cases}$$

Not every basis has this property. For a general real vector space  $V$ , a basis  $\mathcal{B}$  is just an ordered list of vectors

$$\mathcal{B} = (\vec{\epsilon}_1, \dots, \vec{\epsilon}_n) \subset V$$

such that:

- Every vector  $\vec{A} \in V$  can be written as a linear combination

$$\vec{A} = \sum_{i=1}^n A_i \vec{\epsilon}_i.$$

That is,  $\mathcal{B}$  must *span* our vector space.

- For any given  $\vec{A}$ , there is only one set of coefficients  $A_i$  that makes this equality true. Equivalently, the vectors in  $\mathcal{B}$  are *linearly independent*. Equivalently, no one basis vector can be written as a linear combination of the others.

An important elementary theorem of linear algebra is that, for a fixed vector space, every basis is guaranteed to have the same number  $n$  of elements. This number  $n$  is then taken as the abstract definition of the *dimension* of our vector space.

### 1.3 Linear Transformations

Now suppose we are in a general vector space  $V$  (maybe  $\mathbb{R}^n$ , maybe something more exotic) and we have chosen a basis

$$\mathcal{B} = (\vec{\epsilon}_1, \dots, \vec{\epsilon}_n).$$

We have just seen that this lets us describe vectors by lists of numbers, *i.e.*, their components. It also lets us describe linear transformations by numbers by representing them as matrices. To see how this works, suppose we have a linear transformation  $T$  which takes in vectors in  $V$  and returns new vectors in  $V$ . Abstractly, we have

$$T : V \rightarrow V.$$

Let  $\vec{A} \in V$  be any vector and let  $\vec{B}$  be its image under  $T$ .

$$T(\vec{A}) = \vec{B}.$$

We can expand both vectors in terms of our basis

$$T\left(\sum_{i=1}^n A_i \vec{\epsilon}_i\right) = \sum_{i=1}^n B_i \vec{\epsilon}_i$$

and use the fact that  $T$  is linear to pull the summation outside.

$$\begin{aligned} \sum_{i=1}^n B_i \vec{\epsilon}_i &= T\left(\sum_{i=1}^n A_i \vec{\epsilon}_i\right) \\ &= T(A_1 \vec{\epsilon}_1 + \dots + A_n \vec{\epsilon}_n) \\ &= A_1 T(\vec{\epsilon}_1) + \dots + A_n T(\vec{\epsilon}_n) \\ &= \sum_{i=1}^n A_i T(\vec{\epsilon}_i). \end{aligned}$$

Now we introduce some notation in a way which may initially seem arbitrary, but which will shortly be vindicated as it turns out to reproduce formulas we have seen earlier in our mathematical training. We know that  $T(\vec{\epsilon}_i)$  at any fixed  $i$  must be a linear combination of our basis vectors. Denote by  $T_{ji}$  the component of  $T(\vec{\epsilon}_i)$  along the basis vector  $\vec{\epsilon}_j$ .

$$T(\vec{\epsilon}_i) = \sum_{j=1}^n T_{ji} \vec{\epsilon}_j.$$

Take a moment to notice that the index structure here makes sense. On each side we have exactly one free index which we've written as  $i$ , while on the RHS we have an additional dummy index  $j$ . Putting this all together, we have

$$\begin{aligned} \sum_{i=1}^n B_i \vec{\epsilon}_i &= \sum_{i=1}^n A_i T(\vec{\epsilon}_i) \\ &= \sum_{i=1}^n A_i \left( \sum_{j=1}^n T_{ji} \vec{\epsilon}_j \right) \\ &= \sum_{i,j=1}^n A_i T_{ji} \vec{\epsilon}_j, \end{aligned}$$

Now we make one of the most useful moves we can make when manipulating linear-algebraic expressions in index notation: we **relabel our dummy indices** on the RHS. The  $i$  and  $j$  are both being summed over, so it doesn't matter what we call them.

$$\sum_{i,j=1}^n A_i T_{ji} \vec{\epsilon}_j = \sum_{a,b=1}^n A_a T_{ba} \vec{\epsilon}_b = \sum_{M,N=1}^n A_M T_{NM} \vec{\epsilon}_N.$$

We use this freedom to relabel  $i \rightarrow j$ ,  $j \rightarrow i$ . Now our equality becomes

$$\sum_{i=1}^n B_i \vec{\epsilon}_i = \sum_{i,j=1}^n A_j T_{ij} \vec{\epsilon}_i$$

or after slightly rewriting the RHS in a more suggestive way

$$\underbrace{\sum_{i=1}^n B_i \vec{\epsilon}_i}_{\vec{B}} = \underbrace{\sum_{i=1}^n \left( \sum_{j=1}^n T_{ij} A_j \right) \vec{\epsilon}_i}_{T(\vec{A})}$$

Taking a step back, the LHS and RHS are both vectors and we have expanded both of these vectors in the basis  $\mathcal{B}$ . Since these expansions are unique, we can equate the components in this expansion.

$$B_i = \sum_{j=1}^n T_{ij} A_j.$$

But this is something we have seen before. Temporarily setting  $n = 3$  to make typesetting convenient, this is just the relation between column vectors

$$\begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix} = \begin{bmatrix} T_{11}A_1 + T_{12}A_2 + T_{13}A_3 \\ T_{21}A_1 + T_{22}A_2 + T_{23}A_3 \\ T_{31}A_1 + T_{32}A_2 + T_{33}A_3 \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix}$$

which is exactly how we wrote linear transformations in our youth, when we only knew them as matrices. Staring at this for a moment also gives us the familiar expression for matrix multiplication, namely

$$(ST)_{ij} = \sum_{k=1}^n S_{ik} T_{kj}.$$

## 1.4 Change of Basis

Now suppose again we are in a general vector space and we have chosen a basis

$$\text{first basis} = \mathcal{B} = (\vec{\epsilon}_1, \dots, \vec{\epsilon}_n)$$

so that some vector  $\vec{A}$  has basis expansion

$$\vec{A} = \sum_{i=1}^n A_i \vec{\epsilon}_i.$$

Now in  $\mathbb{R}^n$  or elsewhere, we are always free to consider some other basis

$$\text{second basis} = \mathcal{B}' = (\vec{\epsilon}'_1, \dots, \vec{\epsilon}'_n)$$

with respect to which  $\vec{A}$  has some other expansion.

$$\vec{A} = \sum_{i=1}^n A'_i \vec{\epsilon}'_i.$$

Of course  $\vec{A} = \vec{A}$  no matter what basis we choose to expand it in, so these sums must be equal

$$\sum_{i=1}^n A_i \vec{\epsilon}_i = \sum_{i=1}^n A'_i \vec{\epsilon}'_i.$$

This is an important moral point. Changing from one basis to another changes the **components** of vectors, but doesn't change the vectors themselves! A large number of basic results can be proved by starting with tautologies like  $\vec{A} = \vec{A}$  and then expanding the two sides using two different bases. In terms of our earlier vocabulary for geometric spaces, in moving from one basis to another we have performed a **passive transformation**. None of the intrinsic objects in our space have changed, just our descriptions of them using numbers and coordinates.

Now we have two objects whose descriptions change. On one hand, each of our new basis elements must be a linear combination of the old ones. From our earlier experience with linear transformation we might guess that it's more natural to pair up a basis index with the first (not second) index of a matrix, so we write the coefficients as

$$\vec{\epsilon}'_i = \sum_{j=1}^n R_{ji} \vec{\epsilon}_j.$$

On the other hand, we recall that the components of a vector in one basis must also be linearly related to its components to another. We write the coefficients of this expansion as

$$A'_i = \sum_{j=1}^n S_{ij} A_j.$$

How do  $R$  and  $S$  relate? To find out, we started from  $\vec{A} = \vec{A}$  and expand. We first find

$$\begin{aligned} \sum_{i=1}^n A_i \vec{\epsilon}_i &= \sum_{i=1}^n A'_i \vec{\epsilon}'_i \\ &= \sum_{i=1}^n \left( \sum_{j=1}^n S_{ij} A_j \right) \left( \sum_{k=1}^n R_{ik} \vec{\epsilon}_k \right) \\ &= \sum_{i,j,k=1}^n S_{ij} A_j R_{ki} \vec{\epsilon}_k \end{aligned}$$

Notice that, as a matter of index hygiene, in the second expression in parentheses we have freely used  $k$  rather than  $j$  as our dummy index so that we don't have two  $\sum_k$ 's multiplying each other. Also take another second to pause and examine the index structure: both sides have various sets of dummy indices, but neither side has any free indices. Notice also that each dummy index appears under a summation twice. Now we again proceed by a notational triviality, relabeling our dummy indices by replacing  $k \rightarrow i$ ,  $i \rightarrow j$ , and  $j \rightarrow k$ .

$$\begin{aligned} \sum_{i=1}^n A_i \vec{\epsilon}_i &= \sum_{i,j,k=1}^n S_{jk} A_k R_{ij} \vec{\epsilon}_i \\ &= \sum_{i=1}^n \left( \sum_{j,k=1}^n R_{ij} S_{jk} A_k \right) \vec{\epsilon}_i \end{aligned}$$

We play the same trick as before, recognizing that both sides of this equation are vectors expanded in a basis of  $\vec{\epsilon}_i$ 's and equating the coefficients.

$$A_i = \sum_{j,k=1}^n R_{ij} S_{jk} A_k.$$

In terms of matrices, this is exactly the equality (again setting  $n = 3$  for ease of typesetting)

$$\begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix} \begin{bmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix}.$$

So if we write  $S$  for the matrix that describes the transformation of vector components and  $R$  for the matrix that transforms the basis vectors, we see that one must be the inverse of the other. That is,

$$RS = 1 \text{ or equivalently } R = S^{-1}$$

or in terms of indices

$$\sum_{k=1}^n R_{ik} S_{kj} = \delta_{ij}.$$

## 2 Minkowski Space

### 2.1 Raised and Lowered Greek Indices

Now we move from Euclidean space into Minkowski space, the vector space inhabited by *four-vectors*. Our discussion of abstract linear algebra above will largely carry over, but there are two primary notational differences. The less interesting is that we use Greek indices like  $\alpha, \beta, \mu, \nu$  rather than Latin ones like  $i, j, k$ . The more interesting is that the objects that appear around Minkowski space may have these indices superscripted or subscripted, and the difference matters! Just like we can define a vector at varying levels of abstraction, there are a range of answers to “what exactly is the difference between a raised index and a lowered index?” For now, rather than give the completely general answer we will take as a notational axiom that our coordinates on Minkowski space are enumerated with a raised index as

$$(x^0, x^1, x^2, x^3) = (t, x, y, z).$$

The prototypical four-vector is the relativistic velocity, which has components which must therefore have raised indices too.

$$U^\mu := \frac{dx^\mu}{d\tau} \implies (U^0, U^1, U^2, U^3) = \left( \frac{dt}{d\tau}, \frac{dx}{d\tau}, \frac{dy}{d\tau}, \frac{dz}{d\tau} \right).$$

Recall that here  $\tau$  is the proper time along the particle’s worldline, which by construction is the same according to every inertial observer. In general, we write the components of four-vectors using a raised index. Since we are in Minkowski space from this point forward, we will often call four-vectors simply vectors.

On the other hand, we can construct natural objects which carry a lowered index on their components. A scalar field  $\phi$  has a gradient whose  $\mu$ -th component is

$$G_\mu = \frac{\partial \phi}{\partial x^\mu}.$$

We will see later on that the placement of indices in fact tells us how the components of an object change under a Lorentz transformation, which in our above notation is just a change of basis. We call any object whose components carry a lowered index a *covector* or *dual vector* or *one-form*. This is yet another object whose totally general definition we will have to give a bit later. Just like vectors may be assembled into vector fields, we see that computing the gradient of a scalar field at every point in fact yields a *covector field*.

### 2.2 No free indices = Lorentz-invariant

The two physical stories we told to introduce vectors and covectors - the covariant velocity of a particle or the gradient of a scalar field - can be combined to illustrate a further key point about indices. Suppose we have a particle whose trajectory in some coordinate frame (that is, in some choice of basis for Minkowski space) has components

$$x^\mu(\tau) \sim \mu\text{th component of location at } \tau$$

where  $\tau$  is again the proper time along the particle’s worldline. We could imagine measuring the value of the scalar field at the particle’s location at proper time  $\tau$ .

$$\phi(x(\tau)) \sim \text{value of scalar field at particle's location at } \tau.$$

We could then see how this value varies with  $\tau$ . Notice that the result **must be a Lorentz scalar**, since every observer agrees on how proper time passes for the particle and every observer agrees on the value of a scalar

field at some definite event. Directly from the chain rule, we see that this changes at a rate

$$\begin{aligned}\frac{d\phi(x(\tau))}{d\tau} &= \frac{\partial\phi}{\partial t} \frac{dt}{d\tau} + \frac{\partial\phi}{\partial x} \frac{dx}{d\tau} + \frac{\partial\phi}{\partial y} \frac{dy}{d\tau} + \frac{\partial\phi}{\partial z} \frac{dz}{d\tau} \\ &= \frac{\partial\phi}{\partial x^0} U^0 + \frac{\partial\phi}{\partial x^1} U^1 + \frac{\partial\phi}{\partial x^2} U^2 + \frac{\partial\phi}{\partial x^3} U^3 \\ &= \sum_{\mu=0}^3 \frac{\partial\phi}{\partial x^\mu} U^\mu.\end{aligned}$$

There are two morals here. First, notice that our dummy index in the final expression appears exactly twice - once as a superscript in the four-velocity, and once as a subscript in the gradient. This is a feature baked into index notation for Minkowski space. An index will only ever appear as a dummy index of summation in this way. If you find yourself summing over  $\mu$  in an expression that has three  $\mu$ 's

$$\sum_{\mu=0}^3 A^\mu B_\mu C^\mu = \text{something has gone wrong}$$

or two  $\mu$ 's that are both lowered

$$\sum_{\mu=0}^3 A^\mu B^\mu = \text{something has gone wrong}$$

or anything except one raised plus one lowered, **you have made an error** and your results are at best not to be trusted. An index must appear exactly once (in which case it is a free index) or exactly twice with one raised and one lowered (in which case it must be summed over). Because this rule is so fundamental, when there are repeated indices we usually omit the explicit summation symbol, as for example

$$\frac{d\phi(x(\tau))}{d\tau} = \frac{\partial\phi}{\partial x^\mu} U^\mu.$$

This is called the *Einstein summation convention*.<sup>1</sup>

Second, notice that both sides of this equation have no free indices, and that as we argued above the LHS (hence also the RHS) is a Lorentz scalar, with the same value for every inertial observer. This is another feature baked into our index notation: if you pair up and sum over your indices in legal ways to produce a result with no free indices, it is guaranteed to be a Lorentz-invariant object.

Now we clarify an extremely subtle point that came up in recitation: just because something is Lorentz-invariant, it is not necessarily a Lorentz scalar! The quantity  $d\phi/d\tau$  is a Lorentz scalar, as we just noted. But another example of something that doesn't change under a change of reference frame is a **four-vector itself**. Sure, if we have some four-vector  $A$  with components  $A^\mu$  in one frame, then the components  $A'^\mu$  in another frame will generally be different.

$$A'^\mu \neq A^\mu.$$

But  $A$  is not just its components! The components of  $A$  are the coefficients of  $A$  when  $A$  is expanded in the basis vectors that our reference frame gives us. Since  $A^\mu$  has a raised index, to accord with Einstein index notation we must enumerate the basis elements with a lowered index as  $\epsilon_0, \epsilon_1, \epsilon_2, \epsilon_3$ , so that the whole vector is

$$A = A^\mu \epsilon_\mu = A^0 \epsilon_0 + A^1 \epsilon_1 + A^2 \epsilon_2 + A^3 \epsilon_3.$$

But  $A$  is  $A$  is  $A$ , no matter what reference frame we choose and thus what basis we expand it in. So we have furthermore

$$A = A^\mu \epsilon_\mu = A'^\mu \epsilon'_\mu.$$

---

<sup>1</sup>Professor Einstein famously claimed that this notation was his greatest contribution to mathematics.

Under a Lorentz change of reference frames we get new basis vectors  $\epsilon'_\mu$ , and corresponding new components  $A'^\mu$ , but the vector  $A$  itself is unchanged. If you are asked “how does a four-vector change under a Lorentz transformation?” your interrogator is mildly abusing vocabulary, and almost certainly means to ask how the *components* change.

This leads to a related subtlety. It is commonly said that “a four-vector has raised indices.” But we just wrote about  $\epsilon_\mu$  and called these basis four-vectors. So what’s going on? The resolution is just that we need to be more precise with our language. If we are enumerating the  $\mu$ th component of a four-vector, we use a raised index as in  $A^\mu$ . But if we are enumerating the  $\mu$ th four-vector in some basis, we use a lowered index as in  $\epsilon_\mu$ . Even worse, it is almost universal amongst physicists to say things like “consider the four-vector  $A^\mu$ .” As you may have guessed, this is shorthand for “consider the four-vector whose  $\mu$ th component is  $A^\mu$ .”

## 2.3 The Invariant Interval And The Inner Product

Now let’s combine everything we’ve learned so far to re-express the invariant interval. Given some inertial coordinate system  $(t, x, y, z)$  and two events which in this coordinate system have relative spacetime locations  $(\Delta t, \Delta x, \Delta y, \Delta z)$ , the appropriately Lorentz-invariant notion of distance between these events is

$$\Delta s^2 = -\Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2$$

where we have set  $c = 1$ . If we move into some new frame with coordinates  $(t', x', y', z')$  and corresponding new values of  $(\Delta t', \Delta x', \Delta y', \Delta z')$ , the value of the invariant interval must be the same. In other words, it must be a Lorentz scalar. In our new index notation, this means that the expression above must ultimately translate to something with no free indices. To see how this works, we take inspiration from the dot product on Euclidean space  $\mathbb{R}^n$

$$\begin{aligned}\vec{A} \cdot \vec{B} &= [A_1 \ A_2 \ A_3] \begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix} \\ &= [A_1 \ A_2 \ A_3] \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix}\end{aligned}$$

or equivalently

$$\vec{A} \cdot \vec{B} = \sum_{i=1}^3 A_i B_i = \sum_{i,j=1}^3 A_i \delta_{ij} B_j$$

and rewrite the invariant interval as

$$\Delta s^2 = [\Delta t \ \Delta x \ \Delta y \ \Delta z] \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \Delta t \\ \Delta x \\ \Delta y \\ \Delta z \end{bmatrix}.$$

The components  $\Delta x^\mu$  each have a raised index, so if this interval is to be a Lorentz scalar the array in the middle must carry two lowered indices. So we must have in equivalent notation

$$\Delta s^2 = [\Delta x^0 \ \Delta x^1 \ \Delta x^2 \ \Delta x^3] \begin{bmatrix} \eta_{00} & \eta_{01} & \eta_{02} & \eta_{03} \\ \eta_{10} & \eta_{11} & \eta_{12} & \eta_{13} \\ \eta_{20} & \eta_{21} & \eta_{22} & \eta_{23} \\ \eta_{30} & \eta_{31} & \eta_{32} & \eta_{33} \end{bmatrix} \begin{bmatrix} \Delta x^0 \\ \Delta x^1 \\ \Delta x^2 \\ \Delta x^3 \end{bmatrix}$$

where

$$\begin{aligned}\eta_{\mu\nu} &= \begin{cases} -1 & \mu = \nu = 0, \\ 1 & \mu = \nu \neq 0, \\ 0 & \text{otherwise.} \end{cases} \\ &= \text{diag}(-1, +1, +1, +1).\end{aligned}$$

In our compressed Einstein summation notation, we have more simply

$$(\text{interval}) = \Delta x^\mu \eta_{\mu\nu} \Delta x^\nu.$$

Just like in Euclidean space, any object defined in terms of components ought to have a nice intrinsic story. To see the intrinsic story here, we note that our reference frame defining the components of  $\Delta x$  equivalently means the reference frame must be giving us some list of basis vectors  $\epsilon_\mu$  in terms of which we can expand

$$\Delta x = \Delta x^\mu \epsilon_\mu.$$

Now we introduce a new kind of object which we denote by  $\eta$  (sans indices). This will be a map that takes in two four-vectors and returns a scalar, or abstractly

$$\eta : M^4 \times M^4 \rightarrow \mathbb{R}$$

where  $M^4$  is four-dimensional Minkowski space, thought of as a vector space. We also take  $\eta$  to act linearly in both arguments separately, *i.e.*, to be *bilinear*, as an inner product must be. Then if we feed into  $\eta$  the four-vector difference between events  $\Delta x$  for both arguments, we find

$$\begin{aligned}\eta(\Delta x, \Delta x) &= \eta(\Delta x^\mu \epsilon_\mu, \Delta x^\nu \epsilon_\nu) \\ &= \Delta x^\mu \Delta x^\nu \eta(\epsilon_\mu, \epsilon_\nu)\end{aligned}$$

So if we choose our bilinear function  $\eta$  to act on pairs of basis vectors  $\epsilon_\mu, \epsilon_\nu$  according to the rules

$$\eta(\epsilon_\mu, \epsilon_\nu) = \begin{cases} -1 & \mu = \nu = 0, \\ 1 & \mu = \nu \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Then we have

$$\Delta s^2 = \Delta x^\mu \eta_{\mu\nu} \Delta x^\nu = \eta(\Delta x, \Delta x).$$

This is great! We have cooked up an abstract machine  $\eta$  which computes the invariant interval.

But we have seen something very much like this before when we studied the geometry of  $\mathbb{R}^n$ . There we saw that the definition of the length of a vector arose from a more primitive object, namely the dot product, via the rule

$$\|\vec{A}\|^2 = \vec{A} \cdot \vec{A}.$$

where we could abstractly think of the dot product as a bilinear map  $\mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ . This has exactly the same structure as the relation between the invariant interval and our new function  $\eta$ . From this we draw several important morals about the geometry of Minkowski space:

1. Minkowski space is not just a vector space but an *inner product space*, equipped with an inner product  $\eta : M^4 \times M^4 \rightarrow \mathbb{R}$ .
2. The invariant interval should be thought of as the squared distance between two events, as measured by  $\eta$ . This term “squared distance” should be dealt with carefully though; we have seen the invariant interval is negative for timelike-separated events, so in this case it is not literally the square of any real distance.
3. Much like the components of a matrix, the components of an inner product depend on the basis you work in. So inertial reference frames now have yet another characterization: they are the choices of basis on Minkowski space for which  $\eta_{\mu\nu}$  has the form given above. Similarly, Lorentz transformations are the changes of basis which preserve this form.
4. We now know how to compute even more Lorentz invariants. For one, we can take any four-vector  $A$  (not just the displacement between events) and compute its squared length using  $\eta$ .

$$A^2 := \eta(A, A) = \eta_{\mu\nu} A^\mu A^\nu.$$

Again note the abuse of notation; we can perfectly well have  $A^2 < 0$ . Even better, we can compute the inner product between two distinct vectors.

$$\eta(A, B) = \eta_{\mu\nu} A^\mu B^\nu.$$

## 2.4 Lorentz Transformations

Now let's take everything we've learned and apply it to Lorentz transformations. We have seen that these always take the form of linear transformations of our coordinates, for example by a  $xy$  rotation

$$\begin{bmatrix} t' \\ x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta & 0 \\ 0 & \sin \theta & \cos \theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} t \\ x \\ y \\ z \end{bmatrix}$$

or by an  $xt$ -boost

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

Accordingly, four-vectors like the covariant velocity  $U$  transform the same way. For example, differentiating the most recent equation with respect to  $\tau$  along some particle's worldline gives

$$\begin{bmatrix} U'^0 \\ U'^1 \\ U'^2 \\ U'^3 \end{bmatrix} = \begin{bmatrix} \gamma & -\gamma v & 0 & 0 \\ -\gamma v & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} U^0 \\ U^1 \\ U^2 \\ U^3 \end{bmatrix}.$$

To write this as an indexed equality, we see that the LHS has a single free raised index, so the right hand side must as well.

$$U'^\mu = (\text{right hand side?})^\mu.$$

We also recall that matrices taking in vectors and returning new vectors in our old Euclidean notation looked like

$$A_i = \sum_j M_{ij} B_j.$$

So structurally we ought to have something similar. We just have to remember the two differences between Minkowski index notation and the old Euclidean way of doing things: Minkowski notation always use Greek indices, and we only pair up indices when one is raised and the other is lowered. If we make the arbitrary but standard choice of symbol  $\Lambda$  for the matrix implementing our Lorentz transformation, this means the right structure must be

$$U'^\mu = \Lambda^\mu_\nu U^\nu.$$

Be careful about the object  $\Lambda^\mu_\nu$ . Note in particular that, not only is  $\mu$  raised while  $\nu$  is lowered, but that  $\mu$  sits to the left. This makes a difference, just like how in Euclidean notation  $T_{ij} \neq T_{ji}$  unless we happen to be considering a symmetric matrix. For objects with multiple indices in Minkowski notation, both the horizontal and the vertical placement of indices matters!

We now have the language for one of several commonly given definitions for four-vectors. We have seen that different kinds of objects have components which transform in different ways under Lorentz transformations. A *four-vector* could be defined as any observable  $A$  in special relativity with four components which, under a change of reference frame implemented by the Lorentz transformation  $\Lambda$ , transforms as<sup>2</sup>

$$A'^\mu = \Lambda^\mu_\nu A^\nu.$$

But then how do the basis vectors transform? We know that ultimately  $A$  must be some intrinsic object, which only acquires components when expanded in the basis given to our by choice of reference frame.

$$A = A^\mu \epsilon_\mu = A'^\mu \epsilon'_\mu.$$

---

<sup>2</sup>These kind of definitions of different objects are common in the way physicists have culturally agreed to teach differential geometry. This practice leads to the famous adage that a *tensor* is, by definition, something that transforms like a tensor.

We denote by  $\tilde{\Lambda}$  the object under which the basis vectors transform, recalling that we learned above that it is more natural to contract a basis vector index against the first index of a matrix.

$$\epsilon'_\mu = \tilde{\Lambda}^\nu_\mu \epsilon_\nu.$$

Just as before, from the tautology  $A = A$  we can deduce what  $\tilde{\Lambda}$  must be relative to  $\Lambda$ .

$$\begin{aligned} A^\mu \epsilon_\mu &= A'^\mu \epsilon'_\mu \\ &= (\Lambda^\mu_\nu A^\nu) (\tilde{\Lambda}^\rho_\mu \epsilon_\rho) \\ &= \tilde{\Lambda}^\rho_\mu \Lambda^\mu_\nu A^\nu \epsilon_\rho. \end{aligned}$$

Now relabel dummy indices  $\rho \rightarrow \mu, \mu \rightarrow \rho$  on the RHS

$$A^\mu \epsilon_\mu = \tilde{\Lambda}^\mu_\rho \Lambda^\rho_\nu A^\nu \epsilon_\mu$$

and recognize this as an expansion in the basis of  $\epsilon_\mu$ 's, so we can equate the coefficients.

$$A^\mu = \tilde{\Lambda}^\mu_\rho \Lambda^\rho_\nu A^\nu.$$

But this is exactly the statement that, if we take an arbitrary vector  $A$  and multiply by  $\Lambda$  then by  $\tilde{\Lambda}$ , we get  $A$  back. In matrix notation, we have discovered

$$\tilde{\Lambda} \Lambda = 1$$

or in index notation

$$\tilde{\Lambda}^\mu_\rho \Lambda^\rho_\nu = \delta^\mu_\nu.$$

So the basis vectors transform exactly under the inverse matrix.

$$\epsilon'_\mu = (\Lambda^{-1})^\nu_\mu \epsilon_\nu$$

## 2.5 Preserving the Metric

We have so far given several abstract characterizations of Lorentz transformations. They are the transformations built up by composing boosts and rotations, or else those linear transformations of our coordinates that respect the invariant interval. Now that we know the invariant interval comes from the  $\eta$  inner product, we see that another way to think of Lorentz transformations is as linear isometries, *i.e.*, as coordinate transformations

$$x'^\mu = \Lambda^\mu_\nu x^\nu$$

such that the inner product between any two four-vectors measured with  $\eta$  is the same before and after.

$$\eta_{\mu\nu} A^\mu B^\nu = \eta_{\mu\nu} A'^\mu B'^\nu$$

for arbitrary  $A, B$ . We know how four-vectors transform, so if we pick our dummy indices presciently we can write

$$A'^\mu = \Lambda^\mu_\rho A^\rho \text{ and } B'^\nu = \Lambda^\nu_\sigma B^\sigma$$

and plug this into our isometry condition above. This gives

$$\begin{aligned} \eta_{\mu\nu} A^\mu B^\nu &= \eta_{\mu\nu} A'^\mu B'^\nu \\ &= \eta_{\mu\nu} (\Lambda^\mu_\rho A^\rho) (\Lambda^\nu_\sigma B^\sigma) \\ &= \eta_{\mu\nu} \Lambda^\mu_\rho \Lambda^\nu_\sigma A^\rho B^\sigma. \end{aligned}$$

Our next step is to again tastefully relabel our dummy indices on the RHS. We replace  $\rho \rightarrow \mu, \sigma \rightarrow \nu, \mu \rightarrow \rho, \nu \rightarrow \sigma$  to find

$$\eta_{\mu\nu} A^\mu B^\nu = \Lambda^\rho_\mu \Lambda^\sigma_\nu \eta_{\rho\sigma} A^\mu B^\nu.$$

But this is true for arbitrary values of  $A^\mu$  and  $B^\nu$  if and only if

$$\eta_{\mu\nu} = \Lambda^\rho_\mu \Lambda^\sigma_\nu \eta_{\rho\sigma}.$$

So we have found, in index notation, exactly what makes a linear coordinate change in Minkowski space a Lorentz transformation.