1 These notes

Here we introduce tensor notation or relativistic index notation.

We probably previously alluded to this notation multiple times already, e.g., referring to the components of 4-vector A with superscripted (upstairs) indices, like A^μ, and referring to inner products as $A^\mu B_\mu$ or $A_\mu B^\mu$. What does it mean to have subscripted (downstairs) indices? We will define this soon.

The index notation we are introducing — involving superscripts nad subscripts — is also known as **tensor** notation. We will be introducing tensors as well in this chapter.

2 Latin & Greek, Einstein notation

An event or a spacetime point is a 4-vector

$$X^\mu = (ct, x, y, z) \qquad \text{with} \quad \mu = 0, 1, 2, 3;$$
$$\text{so that} \quad X^0 = ct, \; X^1 = x, \; X^2 = y, \; X^3 = z.$$

We have previously complained that the symbol X^μ is often used to represent the complete 4-vector, not just the μ-th component. (There is nothing we can do about it, except for getting used to the indignity of slightly inconsistent notation.)

Similarly for the 4-momentum

$$P^\mu = (E/c, p_x, p_y, p_z) \qquad \text{with} \quad \mu = 0, 1, 2, 3;$$
$$\text{so that} \quad P^0 = E/c, \; P^1 = p_x, \; P^2 = p_y, \; P^3 = p_z.$$

Of course, there are texts that place time as the 4th coordinate, so that μ runs from 1 through 4, instead of running from 0 through 3 like we have. The 4-momentum components in such a convention would be

$$P^1 = p_x, \; P^2 = p_y, \; P^3 = p_z, \; P^4 = \frac{E}{c}.$$

For the Minkowski indices, we usually use lower-case Greek letters, e.g., μ above. Sometimes, in the same calculation you might have to introduce indices that run only over the spatial part, i.e., only from 1 to 3. It is a common convention to use lower-case Latin letters (i, j, k,...), i.e., lower-case letters from the English-language alphabet, to mean indices that run over the three spatial components, excluding the temporal component.

We are familiar with writing Lorentz transformations (from inertial frame Σ to inertial frame $\tilde{\Sigma}$) as the matrix equation

$$\begin{pmatrix} c\tilde{t} \\ \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{pmatrix} = \Lambda \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix}$$

for spacetime coordinates. Here Λ is a 4×4 transformation matrix. In the index notation that we are developing, this is written as

$$\tilde{X}^\mu = \sum_\nu \Lambda^\mu{}_\nu X^\nu .$$

You recognize this as being just the definition of matrix multiplication, with $\Lambda^\mu{}_\nu$ representing the element in the μ-th row and ν-th column of of Λ. The second index is however written downstairs, for reasons that will become clear later on. In the expression on the right, the summation index is ν. A summation index is called a "dummy index" because it can be replaced by any other symbol without changing the meaning of the equation.

All these indices are going to get painful. Let's make our lives easier with the notational trick

Lorentz transformation, in tensor notation

$$\tilde{X}^\mu = \sum_\nu \Lambda^\mu{}_\nu X^\nu = \Lambda^\mu{}_\nu X^\nu . \tag{1}$$

We just omitted the summation symbol! The idea is that, whenever we see an index repeated, we will understand that this index is summed over. This is called the Einstein summation convention. It will save us writing summation symbols with indices under them, but we will have to remember to sum whenever there is an index appearing twice. E.g., the 0-th component of $\tilde{X}$ is

$$\tilde{X}^0 = \Lambda^0{}_\nu X^\nu = \Lambda^0{}_\alpha X^\alpha = \Lambda^0{}_0 X^0 + \Lambda^0{}_1 X^1 + \Lambda^0{}_2 X^2 + \Lambda^0{}_3 X^3 .$$

Since any repeated index is a summation index and hence a dummy index, it doesn't matter whether we call it ν or α!

In our example, the index that is repeated appears once upstairs and once downstairs. This will be true in general: an upstairs index is contracted with a downstairs index.
Summing over the repeated index is often called **contracting** the two indices.

The summation convention is widely used also in situations where no distinction is made between upstairs and downstairs indices. For example, in discussing non-relativistic physics in Euclidean space, you might see the expression $M_{ij}A_j$ to mean $\sum_j M_{ij}A_j$, i.e., a matrix-vector multiplication. You might see this, e.g., in an advanced text on non-relativistic classical mechanics, with i, j running over the three spatial directions, 1, 2, 3.

I read somewhere that Einstein regarded the summation convention to be one of his greatest contributions.

By definition, any 4-vector has to transform the same way under LT's. Thus if A is any 4-vector, it transforms as $\tilde{A}^\mu = \Lambda^\mu{}_\nu A^\nu$. From Eq. (**??**) we note, by differentiating with respect to X^ν on both sides, that

$$\Lambda^\mu{}_\nu = \frac{\partial \tilde{X}^\mu}{\partial X^\nu}$$

so that the transformation equation for a 4-vector with upstairs-index is

$$\tilde{A}^\mu = \Lambda^\mu{}_\nu A^\nu = \frac{\partial \tilde{X}^\mu}{\partial X^\nu} A^\nu \,.$$

3 Indices upstairs and downstairs — contravariant and covariant

It's time to introduce 4-vectors with downstairs indices. The 4-vectors we have used till now, with upstairs indices, are called **contravariant** vectors, and the superscript indices are contravariant indices. Associated with each contravariant 4-vector is a dual 4-vector which we call a **covariant** 4-vector, written with an index downstairs. The subscript indices are covariant indices.

Instead of writing **contravariant vectors** and **covariant vectors**, we might sometimes just call them **vectors** and **covectors**, respectively.

3.1 Raising and lowering

The metric tensor is written either as $g_{\mu\nu}$ or as $g^{\mu\nu}$. Given a contravariant 4-vector A^μ, the corresponding covariant 4-vector is defined as

$$A_\alpha = g_{\alpha\beta} A^\beta$$

where of course the summaton convention is understood. The covariant 4-vector is obtained by contracting the contravariant index with one of the indices of the metric tensor. This is colloquially called "lowering the index."

lowering

You can see that the components of the covariant vector A_μ are closely related to the components of the corresponding contravariant vector A^μ. In our convention, the diagonal elements of $g_{\mu\nu}$ are $(1, -1, -1, -1)$;

therefore

$$A_0 = A^0,\ A_1 = -A^1,\ A_2 = -A^2,\ A_3 = -A^3.$$

The timelike component is unchanged by index lowering, while the spacelike components get their signs rudely flipped. Using the convention that Latin indices represent the spatial components, we could write this as $A_0 = A^0$, $A_i = -A^i$.

For example, the elements of the covariant vector corresponding to a 4-momentum are $P_0 = E/c$, $P_1 = -p_x$, $P_2 = -p_y$, $P_3 = -p_z$.

In the other convention for the metric tensor (positive-trace convention), the temporal component would have its sign flipped while the spatial components would remain unchanged. ($A_0 = -A^0$, $A_i = A^i$)

The covariant vector was obtained from the contravariant vector through multiplication by the matrix $g_{\mu\nu}$. One should be able to obtain the contravariant vector from the corresponding covariant vector through multiplication by the inverse of $g_{\mu\nu}$. Fortunately, the metric tensor is its own inverse. Thus we have

$$A^\rho = g^{\rho\lambda} A_\lambda.$$

We have "raised" an index using the metric tensor.

raising

Note that the inverse of $g_{\mu\nu}$ is written with indices upstairs. This guarantees that an upstairs index is contracted with a downstairs index.

3.2 Inner products and norms

The inner product of 4-vectors A and B can be written as

$$A^0B^0 - A^1B^1 - A^2B^2 - A^3B^3 \ = \ A_0B^0 + A_1B^1 + A_2B^2 + A_3B^3 \ = \ A_\mu B^\mu .$$

The summation convention has been used in the last step. The inner product is thus written in a remarkably simple form, using the notation we've introduced.

It is easy to show (please do) that the following forms are equivalent, all representing the same inner product:

$$A_\mu B^\mu \ = \ A^\mu B_\mu \ = \ g_{\mu\nu} A^\mu B^\nu \ = \ g_{\mu\nu} A^\nu B^\mu \ = \ g^{\mu\nu} A_\mu B_\nu \ = \ g^{\alpha\beta} A_\beta B_\alpha \, .$$

Hopefully, it should be obvious that $A_\mu B^\mu = A_\alpha B^\alpha = A_\beta B^\beta = A_\nu B^\nu = A_\lambda B^\lambda$. A repeated index is summed over and hence is a dummy index — in principle any symbol can be used for it. But please restrict to smaller-case Greek symbols — this is conventional for Minkowski indices running from 0 to 3.

Similarly, you should be able to show (please do!) that the norm of the 4-vector A is

$$A_\mu A^\mu \;=\; A^\mu A_\mu \;=\; g_{\mu\nu} A^\mu A^\nu \;=\; g^{\mu\nu} A_\mu A_\nu \,.$$

3.3 Transformation

We have defined covariant vectors through the index-lowering operation. It is also common to define covariant vectors through their transformation properties.

We have seen that a contravariant vector A^μ transforms under Lorentz transformations as

$$\tilde{A}^\mu = \Lambda^\mu{}_\nu A^\nu = \frac{\partial \tilde{X}^\mu}{\partial X^\nu} A^\nu \,.$$

The corresponding covariant vector transforms as

Transformation of co-variant vectors

$$\tilde{A}_\mu = \left(\Lambda^{-1}\right)^\nu{}_\mu A_\nu = \frac{\partial X^\nu}{\partial \tilde{X}^\mu} A_\nu \qquad (2)$$

Covariant vectors transform oppositely to how contravariant vectors transform.

We can show this transformation rule, taking $A_\sigma = g_{\sigma\beta} A^\beta$ to be the definition of the covariant vector. It's a bit clumsy and is worked out in the box below.

Since $A_\mu = g_{\mu\nu} A^\nu$, in the frame $\tilde{\Sigma}$ we have

$$\tilde{A}_\mu = \tilde{g}_{\mu\nu}\, \tilde{A}^\nu = g_{\mu\nu}\, \tilde{A}^\nu = g_{\mu\nu}\, \Lambda^\nu{}_\sigma\, A^\sigma = g_{\mu\nu}\, \Lambda^\nu{}_\sigma\, g^{\sigma\beta}\, A_\beta$$

Looks like the transformation matrix is $g_{\mu\nu}\, \Lambda^\nu{}_\sigma\, g^{\sigma\beta}$. Remembering the summation convention, this is the product of three matrices, g, Λ and g^{-1}. Symbolically

$$\tilde{A}_\mu = \left(g\Lambda g^{-1}\right)_\mu{}^\beta A_\beta$$

The indices ν and σ are summed over and so don't appear any more, while the μ and β indices are written with their order unchanged, which makes the covariant index of the transformation matrix the first ('row') index, and the contravariant index the second ('column') index. To maintain our convention, let's take the transpose, so that the contravariant index of the transformation matrix becomes the first index:

$$\tilde{A}_\mu = \left(\left(g\Lambda g^{-1}\right)^{\mathsf{T}}\right)^\beta{}_\mu A_\beta$$

Now from the definition of the Lorentz transformation:

$$\Lambda^{\mathsf{T}} g \Lambda = g \quad \implies \quad g\Lambda g^{-1} = \left(\Lambda^{\mathsf{T}}\right)^{-1} = \left(\Lambda^{-1}\right)^{\mathsf{T}}$$

so that $\left(g\Lambda g^{-1}\right)^{\mathsf{T}} = \Lambda^{-1}$.
Thus the transformation matrix is Λ^{-1}, so that $\tilde{A}_\mu = \left(\Lambda^{-1}\right)^\beta{}_\mu A_\beta$.

We have used 'index lowering' to define covariant vectors, and from that proved that covectors transform 'oppositely' as vectors. Often, however, Eq. (??) is taken as the definition of covectors, and the lowering/raising operations are regarded as a notational trick.

To summarize, the transformation of vectors and covectors under a Lorentz transformation Λ are

$$\boxed{\tilde{A}^\mu = \Lambda^\mu{}_\nu A^\nu\,, \qquad \tilde{A}_\mu = \left(\Lambda^{-1}\right)^\beta{}_\mu A_\beta} \tag{3}$$

The transformation equation for covectors is sometimes written as $\tilde{A}_\mu = \Lambda_\mu{}^\beta A_\beta$, where it is understood that $\left(\Lambda^{-1}\right)^\beta{}_\mu = \Lambda_\mu{}^\beta$. This notation relies on taking care of the horizontal placement of indices in addition to their vertical placement. I find this clumsy, especially as one might reasonably decide to interpret $\Lambda_\mu{}^\beta$ as the transpose of $\Lambda^\beta{}_\mu$. In (??), we have used the contravariant (upper) index as the first (row) index in both transformations.

To avoid worrying about horizontal placement of indices altogether, it's convenient to remember instead the following forms of the transformations:

$$\boxed{\tilde{A}^\mu = \frac{\partial \tilde{X}^\mu}{\partial X^\nu} A^\nu\,, \qquad \tilde{A}_\mu = \frac{\partial X^\beta}{\partial \tilde{X}^\mu} A_\beta} \tag{4}$$

This form highlights that a vector (contravariant vector) transforms the same way as spacetime coordinates (spacetime 4-vectors). This is of course how we defined vectors. It also shows that covectors transforms oppositely to (contravariant) spacetime 4-vectors.

Contra and Co — what's with these names?

It's amusing that regular vectors are 'contra' and the artificial new objects we've introduced are 'co', which sounds more positive than 'contra'.
The names contravariant and covariant relate to transformation properties. A coordinate frame can be described by unit vectors, or basis vectors. It turns out that contravariant vectors transform oppositely to the transformation of unit vectors, hence 'contra'. Covariant vectors transform the same way as the unit vectors, hence 'co'. In describing flat Minkowski space, unit vectors are not very necessary, and so will not introduce them.

3.4 The metric tensor and its inverse, notation

We have noted that $g^{\mu\nu}$ is the inverse of $g_{\mu\nu}$. It is denoted by g instead of g^{-1} because the two happen to be the same in special relativity in Cartesian coordinates.

Even when they are not the same (e.g., in general relativity, or if using curvilinear coordinates), using g^{-1} would be unnecessary because the indices appearing upstairs tell us that $g^{\mu\nu}$ is the inverse.

Treating them as matrices, multiplying the two should give the 4×4 identity matrix:

$$g^{\mu\beta} g_{\beta\nu} = \delta^{\mu}_{\nu}$$

where the right side is the Kronecker delta, $= 1$ when $\mu = \nu$ and $= 0$ when $\mu \neq \nu$. In the present context, the Kornecker delta has to be written with one index up and one index down, but the horizontal placement of the indices doesn't matter as it is a symmetric matrix.

$g_{\mu\nu}$ versus $\eta_{\mu\nu}$

It is common to make a notation distinction between the metric tensor of non-flat spacetime (as in general relativity), and the metric tensor in flat spacetime (as in special relativity). In the former case, the metric itself depends on spacetime coordinates, while in the latter, we have the constant metric we are used to in this semester. It is common to use $g_{\mu\nu}$ for the general case and $\eta_{\mu\nu}$ for the special (flat) case; so you will find $\eta_{\mu\nu}$ in many texts. For our purposes, it does not matter.

3.5 Maybe think of covectors as row vectors?

(Optional subsection)

We have introduced of a second type of index and a second type of vector. We have already seen a gain in conciseness due to this notation — the inner product of A and B being written as $A_\mu B^\mu$.

A possibly useful way to think about this is to think about contravariant vectors as column vectors (4×1 matrices) and covariant vectors as row vectors (1×4 matrices). Lower indices are column indices and upper indices are row indices. The contraction (implied summation) of the index in the expression $A_\mu B^\mu$ can be regarded as a multiplication between a row vector and a column vector to yield a 1×1 number:

$$\begin{pmatrix} A_0 & A_1 & A_2 & A_3 \end{pmatrix} \begin{pmatrix} B^0 \\ B^1 \\ B^2 \\ B^3 \end{pmatrix} = A_0 B^0 + A_1 B^1 + A_2 B^2 + A_3 B^3 \,.$$

$$\begin{pmatrix} \tilde{A}^0 \\ \tilde{A}^1 \\ \tilde{A}^2 \\ \tilde{A}^3 \end{pmatrix} = \Lambda \begin{pmatrix} A^0 \\ A^1 \\ A^2 \\ A^3 \end{pmatrix}$$

The Lorentz transformation equation $\tilde{A}^\mu = \Lambda^\mu{}_\nu A^\nu$ can then be regarded as a matrix-vector multiplication — a square matrix multiplies a column vector to give a column vector. The transformation matrix $\Lambda^\mu{}_\nu$ has an upper and a lower index, serving as a row index and a column index respectively. The contraction is between the column index of Λ and the row index of the contravariant vector it is mutliplying — this matches the usual rule of matrix multiplication.

This is satisfying. What about the transformation rule for covectors: $\tilde{A}_\mu = (\Lambda^{-1})^\beta{}_\mu A_\beta$? Now, on the right, the row index of Λ contracts with the column index of the covector. This is the standard multiplication of a row vector from the left with a matrix, made more visible by re-ordering the right side:

$$\tilde{A}_\mu = A_\beta \left(\Lambda^{-1}\right)^\beta{}_\mu = \sum_\beta A_\beta \left(\Lambda^{-1}\right)^\beta{}_\mu .$$

In the row-vector picture, this can be visualized as

$$\begin{pmatrix} \tilde{A}_0 & \tilde{A}_1 & \tilde{A}_2 & \tilde{A}_3 \end{pmatrix} = \begin{pmatrix} A_0 & A_1 & A_2 & A_3 \end{pmatrix} \begin{pmatrix} & & & \\ & \Lambda^{-1} & & \\ & & & \end{pmatrix} .$$

So far this picture has been quite satisfying. While the analogy is useful to some extent, it has its limits, which we next illustrate.

Let's look at the lowering operation — We obtained covariant vectors from contravariant vectors using the metric tensor to lower an index: $A_\beta = g_{\beta\mu} A^\mu$. This equation cannot be regarded as a matrix-vector multiplication. Since both indices of g are down, it would seem to have two column indices, which would be a strange way to describe a matrix. The action of index lowering is to transpose the column vector and then flip the signs of the spatial part:

$$\begin{pmatrix} \tilde{A}_0 & \tilde{A}_1 & \tilde{A}_2 & \tilde{A}_3 \end{pmatrix} = \begin{pmatrix} A_0 \\ A_1 \\ A_2 \\ A_3 \end{pmatrix}^{\mathrm{T}} \times \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

or equivalently, flipping signs of the spatial components first and then taking the transpose. Either way, the lowering operation is difficult to write as a simple matrix multiplication. This illustrates that the row-vector picture is rather limited. The index notation is actually richer and more powerful than the matrix picture, and allows us to express relationships that don't have a nice interpretation in the row-vector/column-vector language.

Elsewhere (outside Minkowski space or relativity), you might also meet this concept of introducing duals to each vector in order to define inner products. For example, consider the Euclidean dot product $\vec{A} \cdot \vec{B}$ between 3-vectors $\vec{A}$ and $\vec{B}$. If you want, you could think of this as a matrix product between the A written as a row vector and B written as a column vector. We could even call the row-vector a covector. There is little advantage to introducing this notation, so it's not usually done.
In quantum mechanics, every state is a vector in Hilbert space, which in Dirac notation we think of as the space of kets. A dual space is introduced: corresponding to every ket $|\phi\rangle$ is a bra $\langle\phi|$. Thinking of kets as column vectors and bras as row vectors, we can think of inner products in quantum mechanics as arising from multiplying a row-vector and a column-vector:

$$\langle\psi|\phi\rangle = \begin{pmatrix} \psi_1^* & \psi_2^* & \dots & \psi_D^* \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_D \end{pmatrix} = \sum_{\alpha=1}^{D} \psi_\alpha^* \phi_\alpha$$

In quantum mechanics, this is a useful and widely accepted way to think about ket-vectors and bra-vectors, as long as the Hilbert-space dimension D is finite. When D is infinite (common in quantum mechanics!), you have to go back to the abstract definition and the column-row picture of kets and bras doesn't work any more.
In discussing relativity and the Minkowski inner product, the column-row picture of contravariant and covariant vectors is not as common, and, beyond very simple manipulations, has limited use. We will soon be doing very funky things with indices. The matrix notation won't be able to keep up.

3.6 The 4-gradient and 4-divergence

In (Euclidean) vector calculus, we use the nabla operator or gradient operator,

$$\vec{\nabla} \equiv \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) \equiv (\partial_x, \partial_y, \partial_z)$$

to define the operations of gradient, divergence, and curl. If f is a scalar field, then the 3-vector $\vec{\nabla} f$ is the gradient of f. If $\vec{V}$ is a vector field, then the scalar field $\vec{\nabla} \cdot \vec{V}$ is the divergence of $\vec{V}$. One can check that, if f is a scalar under rotations, $\vec{\nabla} f$ transforms like a vector under rotations, and that if $\vec{V}$ transforms like a vector under rotations, then $\vec{\nabla} \cdot \vec{V}$ is invariant under rotations.

We can generalize to the Minkowski world. The 4-derivative operator is

$$\partial_\mu = \frac{\partial}{\partial X^\mu}.$$

Notice that we have indexed ∂ downstairs, implying that it is a covariant vector operator. We will justify this shortly.

Given a Lorentz scalar, ϕ, its 4-gradient is the 4-vector

$$\partial_\mu \phi \equiv \left(\frac{\partial \phi}{\partial X^0}, \frac{\partial \phi}{\partial X^1}, \frac{\partial \phi}{\partial X^2}, \frac{\partial \phi}{\partial X^3} \right)$$

The subscript indicates that $\partial_\mu \phi$ is a covector rather than a vector. Why? Consider a Lorentz transformation taking spacetime coordinates X^μ to $\tilde{X}^\mu$. Under this transformation, the 4-gradient is transformed to $\frac{\partial \phi}{\partial \tilde{X}^\mu}$, as ϕ is invariant. To express this quantity in terms of derivatives with respect to the original coordinates X^μ, we use the chain rule for differentiation of implicit functions:

$$\widetilde{\partial_\mu \phi} = \frac{\partial \phi}{\partial \tilde{X}^\mu} = \sum_\rho \frac{\partial X^\rho}{\partial \tilde{X}^\mu} \frac{\partial \phi}{\partial X^\rho} = \frac{\partial X^\rho}{\partial \tilde{X}^\mu} \frac{\partial \phi}{\partial X^\rho} = \left(\Lambda^{-1} \right)^\rho{}_\mu \partial_\rho \phi.$$

Thus, $\partial_\mu \phi$ transforms as a covector, justifying our use of subscripts on the ∂_μ operator when differentiating with contravariant spacetime coordinates X^μ.

If one differentiates with respect to the components of the covariant vector X_μ, one can obtain a contravariant gradient operator:

$$\partial^\mu \equiv \frac{\partial}{\partial X_\mu} = g^{\mu\beta} \partial_\beta.$$

Applying the 4-gradient operator to a 4-vector $M^\mu = (M^0, \vec{M})$ and contracting the index gives a Lorentz scalar, which is the 4-divergence of the vector:

$$\begin{aligned} \partial_\alpha M^\alpha &= \partial_0 M^0 + \partial_1 M^1 + \partial_2 M^2 + \partial_3 M^3 \\ &= \frac{\partial M^0}{\partial x^0} + \frac{\partial M^1}{\partial x^1} + \frac{\partial M^2}{\partial x^2} + \frac{\partial M^3}{\partial x^3} \\ &= \frac{1}{c} \frac{\partial M^0}{\partial t} + \vec{\nabla} \cdot \vec{M}\,. \end{aligned}$$

4-divergences are a convenient way to represent conservation laws through continuity equations. If $\frac{1}{c} M^0$ represents the density of some quantity and the $\vec{M}$ represents the corresponding current density, then the continuity equation for this quantity would be expressed as $\partial_\alpha M^\alpha = 0$. For example, the conservation of electromagnetic charge, of mass, of electromagnetic field energy, etc. could be expressed in this way.

4 Tensors in non-relativistic mechanics (3-tensors)

In 3D Euclidean physics, a scalar is a single-component object which does not require indices (has zero indices); one could say that it has $1 = 3^0$ elements. This is also known as a rank-0 tensor.

A (3-)vector has $3 = 3^1$ elements and has 1 index. A vector is a rank-1 tensor.

Generalizing, a rank-2 tensor has $9 = 3^2$ elements. It needs 2 indices. If T is a rank-2 tensor, we could label it's elements as T_{ij}. It might feel natural to pack the 9 elements into a 3×3 matrix. However, a rank-2 tensor is more than a matrix — it is an object that transforms in a certain way.

Under rotation of the coordinate frame, described by rotation matrix $\mathcal{R}$, a scalar (rank-0 tensor) remains invariant — the rotation matrix $\mathcal{R}$ does not enter into the transformation relation. If f is a scalar, it transforms as $f' = f$.

A vector (rank-1 tensor), transforms with a factor of $\mathcal{R}$, e.g., if $\vec{a} = (a_1, a_2.a_3)$ is a vector, it transforms as

$$\vec{a'} = \mathcal{R}\vec{a}, \qquad \text{or} \quad a'_i = \mathcal{R}_{ij}a_j.$$

We've used the summation convention, even though we are doing Euclidean physics.

Before writing down the transformation for a rank-2 tensor, let's construct an example. If $\vec{a} = (a_1, a_2.a_3)$ and $\vec{b} = (b_1, b_2.b_3)$ are vectors, then the combination of nine numbers

$$T_{ij} = a_i b_j$$

forms a tensor. Not every rank-2 tensor can be decomposed in this way into two vectors, but any two vectors can be combined (via such an *outer product*) to construct a rank-2 tensor.

For the example above we can construct the transformation rule:

$$T'_{ij} = a'_i b'_j = \sum_k \mathcal{R}_{ik} a_k \sum_m \mathcal{R}_{jm} b_m = \sum_{km} \mathcal{R}_{ik} (a_k b_m) \mathcal{R}_{jm} = \sum_{km} \mathcal{R}_{ik} T_{km} (\mathcal{R}^{\mathrm{T}})_{mj}$$
$$\implies \quad T' = \mathcal{R} T \mathcal{R}^{\mathrm{T}}$$

We can write this more concisely using the Einstein summation convention. Of course, then it becomes our responsibility to remember what summations are implied:

$$T'_{ij} = a'_i b'_j = \mathcal{R}_{ik} a_k \sum_m \mathcal{R}_{jm} b_m = \mathcal{R}_{ik} (a_k b_m) \mathcal{R}_{jm} = \mathcal{R}_{ik}$$

We have found the transformation rule $T' = \mathcal{R}T\mathcal{R}^{\mathrm{T}}$ for the special case of rank-2 tensors that are outer products of two vectors. The same transformation rule, using 2 factors of the transformation matrix, holds for any rank-2 tensor, even those which cannot be decomposed into an outer product of two rank-1 vectors.

To summarize the rank-1 and rank-2 cases:

A rank-1 tensor transforms as $\vec{a}' = \mathcal{R}\vec{a}$ or $a'_i = \mathcal{R}_{ij}a_j$;

A rank-2 tensor transforms as $T' = \mathcal{R}T\mathcal{R}^{\mathrm{T}}$ or $T'_{ij} = \mathcal{R}_{ik}\mathcal{R}_{jm}T_{km}$.

A tensor of rank higher than 2 is defined analogously. A rank-k tensor has 3^kelements and its transformation involves k factors of the transformation matrix. For $k > 2$, writing the transformation as matrix multiplication becomes clumsy or impossible, but it is easy to write as a summation of appropriate indices. E.g., if B is a rank-4 tensor, then its elements transforms as

$$B_{ijkl} = \mathcal{R}_{im}\mathcal{R}_{jn}\mathcal{R}_{kp}\mathcal{R}_{lq}B_{mnpq}$$

with summation over repeated indices implied.

4.1 Examples of non-relativistic rank-2 tensors

The best-known example is probably the moment of intertia tensor of an extended object with mass density $\rho(\vec{r})$:

$$I_{ij} = \int d^3r\rho(\vec{r}) \left[r^2\delta_{ij} - r_ir_j\right] .$$

For a rigid body with extended structure, it relates the angular velocity $\vec{\omega}$ to the angular momentum:

$$L_i = \sum_j I_{ij}\omega_j \tag{5}$$

which could also be written as a matrix-vector multiplication.

This may be regarded as the analog of $\vec{p} = m\vec{v}$, for angular motion. When the object is complicated enough, the naive $\vec{L} = I\vec{\omega}$ may not hold, and the more general relation (5) is required. Note that contracting one of the tensor indices with a vector index provides a vector.

Another non-relativistic example is the conductivity tensor. For simple materials, the current density $\vec{J}$ created by an applied electric field $\vec{E}$ is usually just proportional: $\vec{J} = \sigma\vec{E}$. However, there are situations (materials or geometries) when this is no longer true, and an electric field can create currents in both longitudinal and transverse directions. Then we would write

$$J_k = \sum_l \sigma_{kl} E_l.$$

Tensors also appear in the study of elastic properties of matter (the elastic stress tensor or Cauchy tensor) and in the study of fluids (the viscous stress tensor).

5 4-tensors and their transformations

In Minkowski space, a scalar has $1 = 4^0$ element and is invariant under Lorentz transformations. This is a rank-0 tensor. A vector has $4 = 4^1$ elements and its transformation involves 1 factor of the transformation matrix ($\tilde{M}^\alpha = \Lambda^\alpha{}_\beta M^\beta$). This is a rank-1 tensor.

Generalizing, a rank-2 tensor is a 2-index object with $16 = 4^2$ elements. If T has two upstairs (contravariant) indices, then it transforms as

$$\tilde{T}^{\alpha\beta} = \Lambda^\alpha{}_\mu \, \Lambda^\beta{}_\nu \, T^{\mu\nu}$$

which is a straightforward generalization of the Euclidean case. Such an object is obtained when we take an outer product of two 4-vectors (tow rank-1 tensors): e.g., the 16 numbers $A^\mu B^\nu$ will transform as a rank-2 tensor. However, given a rank-2 tensor, it cannot necessarily be decomposed into an outer product of two 4-vectors.

Tensor indices can also be downstairs (covariant), and these indices can be lowered or raised using the metric tensor:

$$g_{\alpha\mu} T^{\alpha\beta} = T_\mu{}^\beta \,, \quad g_{\beta\nu} T^{\alpha\beta} = T^\alpha{}_\nu \,, \quad g_{\alpha\mu} g_{\beta\nu} T^{\alpha\beta} = T_{\mu\nu} \,.$$

If all indices are upstairs (downstairs), a tensor is contravariant (covariant); if some are upstairs and some downstairs, the tensor is of mixed type.

Covariant components of tensors transform like covariant vectors, i.e., with a factor of Λ^{-1} instead of a factor of Λ. Thus the mixed tensor $T^\alpha{}_\nu$ will transform as

$$\tilde{T}^\alpha{}_\beta = \Lambda^\alpha{}_\mu \left(\Lambda^{-1}\right)^\nu{}_\beta \, T^\mu{}_\nu$$

and the totally covariant tensor $T_{\rho\sigma}$ transforms as

$$\tilde{T}_{\alpha\beta} = \left(\Lambda^{-1}\right)^\mu{}_\alpha \left(\Lambda^{-1}\right)^\nu{}_\beta \, T_{\mu\nu}$$

You might find the Λ^{-1}'s clumsy to write, especially to remember the horizontal positioning of the indices. Maybe the following forms are nicer:

$$\tilde{T}^{\alpha\beta} = \frac{\partial \tilde{X}^{\alpha}}{\partial X^{\mu}} \frac{\partial \tilde{X}^{\beta}}{\partial X^{\mu}} T^{\mu\nu} , \qquad \tilde{T}^{\alpha}{}_{\beta} = \frac{\partial \tilde{X}^{\alpha}}{\partial X^{\mu}} \frac{\partial X^{\nu}}{\partial \tilde{X}^{\beta}} T^{\mu}{}_{\nu} ,$$

$$\tilde{T}_{\alpha\beta} = \frac{\partial X^{\mu}}{\partial \tilde{X}^{\alpha}} \frac{\partial X^{\nu}}{\partial \tilde{X}^{\beta}} T_{\mu\nu} .$$

The concept of tensors, with contravariant and covariant indices, is readily generalized to arbitrary rank. For example, the outer product $A^{\alpha\beta}B^{\mu}$ is a rank-3 tensor with all contravariant indices: $A^{\alpha\beta}B^{\mu} = M^{\alpha\beta\mu}$. The outer product $A^{\alpha\beta}B^{\mu}{}_{\nu}$ is a rank-4 tensor with 3 contravariant indices and one covariant index: $A^{\alpha\beta}B^{\mu} = N^{\alpha\beta\mu}{}_{\nu}$. However, if you see the expression $A^{\alpha\beta}B^{\mu}{}_{\beta}$, you will know that the β index is contracted (summed over), so that the result is a rank-2 contravariant tensorr: $A^{\alpha\beta}B^{\mu} = P^{\alpha\mu}$. If you see the expression $A^{\alpha\beta}B^{\beta}{}_{\lambda}$, someone probably made a mistake, because a contravariant index is usually not contracted with another contravariant index.

A rank-r contravariant tensor transforms as

$$\tilde{T}^{\alpha_1\alpha_2\ldots\alpha_r} = \frac{\partial \tilde{X}^{\alpha_1}}{\partial X^{\mu_1}} \frac{\partial \tilde{X}^{\alpha_2}}{\partial X^{\mu_2}} \cdots \frac{\partial \tilde{X}^{\alpha_r}}{\partial X^{\mu_r}} T^{\mu_1\mu_2\ldots\mu_r} ,$$

or as

$$\tilde{T}^{\alpha_1\alpha_2\ldots\alpha_r} = \Lambda^{\alpha_1}{}_{\mu_1} \Lambda^{\alpha_2}{}_{\mu_2} \ldots \Lambda^{\alpha_r}{}_{\mu_r} T^{\mu_1\mu_2\ldots\mu_r} .$$

A rank-r covariant tensor transforms as

$$\tilde{T}_{\beta_1\beta_2\ldots\beta_r} = \frac{\partial X^{\nu_1}}{\partial \tilde{X}^{\beta_1}} \frac{\partial X^{\nu_2}}{\partial \tilde{X}^{\beta_2}} \cdots \frac{\partial X^{\nu_r}}{\partial \tilde{X}^{\beta_r}} T_{\nu_1\nu_2\ldots\nu_r} ,$$

or as

$$\tilde{T}_{\beta_1\beta_2\ldots\beta_r} = \left(\Lambda^{-1}\right)^{\sigma_1}{}_{\beta_1} \left(\Lambda^{-1}\right)^{\sigma_2}{}_{\beta_2} \cdots \left(\Lambda^{-1}\right)^{\sigma_r}{}_{\beta_r} T_{\sigma_1\sigma_2\ldots\sigma_r} .$$

A tensor $T^{\mu_1\mu_2\cdots\mu_r}{}_{\beta_1\beta_2\ldots\beta_s}$ of rank $r+s$, with r contravariant indices and s covariant indices, will transform with r factors of Λ and s factors of Λ^{-1}. Not pretty.

6 Why is tensor notation useful?

Tensor notation leads to concise expressions in electromagnetism and in (relativistic) quantum field theory. For example, in electromagnetism, two of Maxwell's equations become simply $\partial_{\mu}F^{\mu\nu} = J^{\nu}$. However, tensor notation has a deeper benefit than conciseness.

If a physical law is expressed as an equality between Minkowski tensors, it has the same form in all inertial frames.

For example, let us imagine that the equation or physical law

$$P^{\alpha\beta}{}_{\gamma} \;=\; Q^{\alpha\beta}{}_{\gamma} + R^{\alpha\beta}{}_{\gamma} + A^{\alpha\beta} B_{\gamma}$$

holds in frame Σ. Of course, for the equation to make sense, each term needs to have the same number of un-contracted contravariant indices and un-contracted covariant indices, and they have to be the same indices. (In other words, only tensors of the same 'size' can be meaningfully equal or added to each other, and tensor equality implies element-by-element equality.) Do check that this is true in the above made-up example.

Then in frame $\tilde{\Sigma}$, we will observe

$$\tilde{P}^{\alpha\beta}{}_{\gamma} \;=\; \tilde{Q}^{\alpha\beta}{}_{\gamma} + \tilde{R}^{\alpha\beta}{}_{\gamma} + \tilde{A}^{\alpha\beta} \tilde{B}_{\gamma}.$$

The reason is that, when transforming tensors, each uncontracted index brings in one factor of Λ or Λ^{-1}. Thus the same number of factors of Λ and Λ^{-1} appear with each term, and hence get cancelled. The equation in the tilde-d frame ends up having the same form as the equation in the no-tilde frame.

Thus, physical laws can be shown to be frame-invariant by expressing them in tensor form.

Often physicists say that an equation has been expressed in "Lorentz-covariant" form, if it has been written using tensor form as above. You might also hear "explicitly covariant" or "manifestly Lorentz-covariant", meaning that an equation or law is written in terms of Minkowski tensors and hence is automatically known to have the same form in all inertial frames. (This usage of the word covariant is completely different from the same word used to describe downstairs indices. I am deeply sorry, but this double-usage is not my fault.)

7 The Lorentz transformation in index notation

We have previously defined Lorentz transformations as those satisfying the matrix equation

$$\Lambda^{\mathrm{T}} g \Lambda = g \tag{6}$$

where we have used matrix notation: The LT's Λ are represented here as 4×4 matrices, which act on event coordinates (or other 4-vectors) represented as 4×1 column vectors. Applying Λ on the spacetime coordinates of an event, x, gives the coordinates of the same event as observed from another inertial frame: $\tilde{x} = \Lambda x$.

The relation (6) is derived from the condition that the quantity $x^{\mathrm{T}} g x$ is invariant under LT's. The quantity $x^{\mathrm{T}} g x$ is the Minkowski norm of x, or alternately, the interval $c^2 t^2 - |\vec{r}|^2$ from the origin at zero time. Recall that the invariance of this interval follows from the invariance of the speed of light.

We can do the same derivation now in index/tensor notation instead of matrix notation:

If the 4-coordinate x^μ transforms to $\tilde{x}^\alpha$ under a LT, then its norm transforms to

$$\tilde{x}_\mu \tilde{x}^\mu = g_{\mu\nu} \tilde{x}^\mu \tilde{x}^\nu = g_{\mu\nu} \left(\Lambda^\mu{}_\alpha x^\alpha \right) \left(\Lambda^\nu{}_\beta x^\beta \right) = g_{\mu\nu} \, \Lambda^\mu{}_\alpha \, \Lambda^\nu{}_\beta \, x^\alpha \, x^\beta$$

Since the norm has to be invariant, we need this to be equal to the norm in the original (no-tilde) frame, i.e., to

$$x_\sigma x^\sigma = g_{\sigma\lambda} x^\sigma x^\lambda$$

Thus

$$g_{\mu\nu} \, \Lambda^\mu{}_\alpha \, \Lambda^\nu{}_\beta \, x^\alpha \, x^\beta = g_{\sigma\lambda} x^\sigma x^\lambda \, ; \qquad \Longrightarrow \qquad g_{\mu\nu} \, \Lambda^\mu{}_\alpha \, \Lambda^\nu{}_\beta \, x^\alpha \, x^\beta = g_{\alpha\beta} x^\alpha x^\beta .$$

In the last step we have changed dummy indices to match the dummy indices on the 4-vectors on both sides of the equation. This equality has to be true for *any* 4-vector x. Therefore, the coefficients of $x^\alpha x^\beta$ on the two sides have to be equal. Thus we have

$$\boxed{g_{\mu\nu} \, \Lambda^\mu{}_\alpha \, \Lambda^\nu{}_\beta = g_{\alpha\beta}}$$

This is our familiar expression $\Lambda^{\mathrm{T}} g \Lambda = g$ (the definition of Lorentz transformations), written in tensor/index notation instead of matrix notation.