# Notes on Index Notation
## PHYS 471

Index notation is a short-hand method of writing entire systems of equations, based on recognizing consistent patterns in the algebra. As discussed in class, this applies to a wide range of mathematical objects, including:

- Scalars: $\phi$ (no index)
- Vectors: $A^\mu$ (one index upstairs)
- Covariant vectors: $A_\mu$ (one index downstairs)
- Matrix: $\Lambda^\mu{}_\nu$ (two indices, up and/or down)
- Tensor: $R^\mu{}_{\nu\alpha\beta}$ (two or more indices, can be up and/or down)

By now, you should all be familiar with scalars, vectors, and matrices (right?...), as well as their associated operations. Scalars can multiply with other scalars, vectors and matrices. Vectors can combine with other vectors through the dot or cross product[1], and matrices transform vectors into new vectors[2]. In General Relativity, we'll encounter all of these objects, so understanding their structure and their interactions is crucial! (*i.e.* how to manipulate the indices when you do algebraic operations). Once you have the hang of it, you'll find that index notation is the best thing ever to happen to Physics... well, maybe not the best thing. That'd probably be the Principle of Stationary Action. But anyway, it's right up there with the best!

# 1 Einstein Summation Convention

A typical linear transformation from an $n$-component vector $\vec{v}$ to an $n$-compoennt vector $\vec{u}\,'$ with an $n \times n$ matrix $A$ is written

$$\vec{u}' = A\vec{v} \quad \longrightarrow \quad u'_i = \sum_{j=1}^{n} A_{ij} v_j$$

where $u'_i$ are the $n$ components of $\vec{u}'$. So, in his infinite wisdom, Einstein simply took out the $\sum$, and made the convention that **repeated, or paired, indices are implicitly summed over**:

$$u_i = \sum_{j=1}^{n} A_{ij} v_j \quad \longrightarrow \quad u_i = A_{ij} v_j$$

[1] **Q:** What do you get when you cross an apple and a banana? **A:** Apple banana $\sin\theta$.

[2] **Q:** What do you get when you cross an apple with a mountain climber? **A:** Undefined. A mountain climber is a scaler.

In Euclidean space (where lengths of vectors are $u^2 = u_1^2 + y_2^2 + ... + u_n^2$), it doesn't matter if the indices are up or down. But in spacetime (hyperbolic), the meaning of upstairs and downstairs indices are important. The meaning of an index being up or down, and how they relate to one another, will be discussed in a later section, so for now just accept that they are different things.

The rule for spacetime indices is: **paired up and down indices are summed over (from 0 to 3)**:

$$x'^{\nu} = \Lambda^{\mu}{}_{\nu} x^{\nu} = \Lambda^{\mu}{}_{0} x^0 + \Lambda^{\mu}{}_{1} x^1 + \Lambda^{\mu}{}_{2} x^2 + \Lambda^{\mu}{}_{3} x^3$$

In addition to sums across different quantities like the transformation above, we can also have sums *within the same object*. For example, the following

$$F^{\mu}{}_{\mu}$$

is the sum of the diagonal components of the matrix, and is called the *tcontraction* of $F$. This idea extends to higher tensor objects, as well. For example,

$$R^{\alpha}_{\beta\alpha\delta}$$

is a contraction that reduces a rank-(1,3) tensor to a rank-(0,2) tensor. These objects are called the Riemann curvature and Ricci curvature tensor, respectively, which we will become quite familiar with in GR. Again, we'll see how to specifically do this in a few sections.

An **extremely** important thing to remember about index notation is that **the indices are arbitrary symbols**, and do not represent specific numbers. That is, $\mu$ refers to $0, 1, 2, 3$, but so does $\nu$ or $\alpha$ or $\sigma$. When we sum over indices, those indices are called **dummy indicces**, which isn't a very nice name (actually kind of cruel), but just means that the actual index symbol is *irrelevant*. So, if you see a pair of raised and lowered *anythings*, you know you have to sum over all values:

$$x^2 = x^{\mu} x_{\mu}, x^{\nu} x_{\nu}, x^{\text{🐱}} x_{\text{🐱}}, x^{\text{⚽🪟}} x_{\text{⚽🪟}}, ...$$

# 2 Some Examples of Summation Convention

Repeat after me: "paired upstairs/downstairs indices mean we sum over the values of the index!" (did you really repeat it? If not, please do so before reading on). We just imagine the $\Sigma$ symbol is there, but have saved some ink by not writing it. The way in which we can manipulate vectors, *e.g.* coordinate transformations, is a good example with which you're probably familiar.

**Coordinate transformations** between two vectors $x^\mu$ and $x'^\mu$ are performed by the matrix $\frac{\partial x'^\mu}{\partial x^\nu}$ (or $\frac{\partial x^\mu}{\partial x'^\nu}$, depending on the direction of the transformation), which is familiar as a Jacobian-like object. So, if we do a coordinate transformation on a vector $A^\mu \to A'^\mu$, it would be represented by the operation

$$A'^\mu = \frac{\partial x'^\mu}{\partial x^\nu} A^\nu$$

The thing $\mathcal{J}^\mu_\nu = \frac{\partial x'^\mu}{\partial x^\nu}$ is a matrix (one index upstairs and one downstairs), and as mentioned is just the Jacobian. Lorentz transformations are also coordinate transformation, as we can see from the definition of the transformations, $t' = \gamma t - \beta\gamma x, x' = -\beta\gamma t + \gamma x$, *etc...* (try it out).

To be fully technical, whether an object is a scalar, vector, or tensor is determined by its coordinate transformation properties. Specifically:

- **Scalars are invariant under coordinate transformations**, $\phi(x^\mu) = \phi(x'^\mu)$, where $x'^\mu = \frac{\partial x'^\mu}{\partial x^\nu} x^\nu$. Note the scalar is a *function* of the vector $x^\mu$, but does not itself have components.
- **Vectors transform like vectors under coordinate transformations**. This seemingly circular definition simply means that you can always identify a vector if the following is true: $A'^\mu = \frac{\partial x'^\mu}{\partial x^\nu} A^\nu$.
- **Tensors transform like tensors under coordinate transformations**. Again, this means $F'^{\mu\nu} = \frac{\partial x'^\mu}{\partial x^\alpha} \frac{\partial x'^\nu}{\partial x^\beta} F^{\alpha\beta}$

While these definitions seem somewhat odd, we will see certain cases where an object that *looks* like a tensor doesn't transform according to the above rules when we apply a coordinate transformation.

## 2.1 Some Other Terminology

The term "tensor" is commonplace in the fields of particle physics, relativity, and cosmology. This isn't a surprise, because when it comes down to it, almost *everything* is a tensor! The location and number of the index tells you what kind of tensor you're dealing with. We refer to the **rank** of the tensor as the **number of indices**, and the position (upper or lower) is denoted as follows:

$$F^{\mu\nu} \implies \text{Rank } (2,0) \text{ tensor}$$

$$R^\mu{}_{\nu\alpha\beta} \implies \text{Rank } (1,3) \text{ tensor}$$

$$A^\mu \implies \text{Rank } (1,0) \text{ tensor}$$

*etc...*

## 3 The Metric Tensor

Up to now, we've talked about upstairs and downstairs indcies, as if somehow their location was important. In the simplest case, they are associated with the components of contravariant vectors (upstairs) and covariant vectors (downstairs). But what exactly is the difference between a contravariant vector and a covariant vector? How do you go from one to the other? To answer this, we first need to understand some features of **the metric tensor**, $g$, whose components are $g_{\mu\nu}$. This defines the geometry of the spacetime – relativity theory is all about the metric! It is usually a **diagonal matrix** – *i.e.* the only non-zero components are $g_{00}, g_{11}, g_{22}, g_{33}$. So we could loosely write

$$g_{\mu\nu} = \begin{pmatrix} g_{00} & 0 & 0 & 0 \\ 0 & g_{11} & 0 & 0 \\ 0 & 0 & g_{22} & 0 \\ 0 & 0 & 0 & g_{33} \end{pmatrix} = \mathrm{diag}(g_{00}, g_{11}, g_{22}, g_{33})$$

Note this is a bit of sloppy notation, since $g_{\mu\nu}$ technically represents the *components* of the metric tensor $g$. But as you've probably figured out by now, *physicists* are sloppy with notation, and *relativists* are even sloppier with notation and never write the metric as $g$! Get used to it – it is the way of the world.

The "upstairs" metric tensor $g^{\mu\nu}$ is the inverse of $g_{\mu\nu}$, and we can show its components are also inverses:

$$g^{00} = g_{00}^{-1} \quad , g^{11} = g_{11}^{-1} \quad , etc...$$

In flat spacetime, the metric reduces to the Minkowski metric, $g_{\mu\nu} \to \eta_{\mu\nu} = \eta^{\mu\nu} = \mathrm{diag}(+1, -1, -1, -1)$.

A given spacetime will have a metric whose components depend on the coordinate system in which they're defined. Special relativity is simplest when described with extended Cartesian coordinates $(t, x, y, z)$, because there's no preferential coordinate in empty space. But when we start discussing gravity, it will make more sense to speak in extended spherical coordinates, $(t, r, \theta, \phi)$, since we will be discussing how spacetime curves around a mass. Metrics will look different in different coordinate bases, but the fundamental invariants associated with them will not.

Lastly, and very importantly, **the metric is symmetric under exchange of indices**:

$$g_{\mu\nu} = g_{\nu\mu}$$

In fact, there's *one more* important fact about the metric and four-vectors. The **sign convention** is not fixed! Many researchers and textbooks will adopt the convention $(+, -, -, -)$ (*i.e.* time $+$, space $-$), while others will represent things with the convention $(-, +, +, +)$ (*i.e.* time $-$, space $+$). **This does not change the physics of**

**relativity!** It only determines the sign of certain invariants of the theory. For example, using $(+,-,-,-)$ as we've been doing, the invariant square of the four velocity is $u^2 = +1$, but if you adopt $(-,+,+,+)$ it's $u^2 = -1$. If you know the convention going into a problem, though, it's a trivial issue to overcome. Most authors, at the outset, will state their chosen sign convention.

Note that in GR, the metric tensor is **not** the Minkowski metric, which represents spacetime in the absence of matter. Since this is the crux of this course, we'll save discussion for a bit later!

# 4 Raising and Lowering Indices

The placement of the index is important for understanding how the mathematical object in question will play with others. The following section briefly reviews the raising and lowering operations for vectors and tensors, but not scalars (because they don't have indices!).

## 4.1 Vectors

A vector is defined as having one index "up", *i.e.* $x^\mu$. This is technically a **contravariant vector**, but it's probably simpler for now just to call it a vector. The **covariant vector** that corresponds to this is $x_\mu$. What's the difference? The index is down, of course! But what does this mean, and how do we write it?

The act of raising or lowering indices is performed with the **metric tensor**. That is:

$$x^\mu = g^{\mu\nu} x_\nu \quad ; \quad x_\mu = g_{\mu\nu} x^\nu$$

So, if we know the structure of $g^{\mu\nu}$ – and by association $g_{\mu\nu}$ – we can figure out what the raised and lowered (vector and covariant vector) components are.

In the case of flat spacetime, we have $g_{\mu\nu} = \mathrm{diag}(+1,-1,-1,-1)$, so for a four-vector with components $(a,b,c,d)$, the corresponding lowered index vector is

$$x_\mu = g_{\mu\nu} x^\mu = \begin{pmatrix} +1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} a \\ -b \\ -c \\ -d \end{pmatrix}$$

The covariant vector $x_\mu$ is a vector whose spatial entries are the negative of the vector's, but the time component is the same! The product of two four-vectors is formally written

$$g_{\mu\nu} x^\mu x^\nu = x_\nu x^\nu$$

which we can also write as

$$g_{\mu\nu}x^\mu x^\nu = x^\mu x_\mu$$

The order of the up or down index isn't important, only the combination. They are **paired up and down indices**, which means we are summing over the values of the indices ($\mu = 0, 1, 2, 3$).

If we evaluate the product algebraically, we need to take the transpose of the first vector to get a row to ensure the dimensionality of the product is correct:

$$g_{\mu\nu}x^\mu x^\nu = x_\nu x^\nu = (a, -b, -c, -d)\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = a^2 - b^2 - c^2 - d^2$$

which is the length (squared) of the four vector, and just happens to look like a Lorentz invariant quantity!

Why should we introduce this new concept of a covariant vector, when we understand all vector operations perfectly and never mentioned them before? The truth is: you've *always* known about covariant vectors and used them. You just never knew it! In fact, if we evaluate the dot product of two good ol' fashioned Euclidean vectors, we get

$$\vec{A}\cdot\vec{B} = A_x B_x + A_y B_y + A_z B_z$$

We could write this in index notation as

$$\vec{A}\cdot\vec{B} = A^i B_i \quad , \quad i = 1, 2, 3$$

If $A^i$ is the vector, what does the covariant vector $B_i$ look like? To be consistent with the four-dimensional case, we need to define the Euclidean metric $g_{ij} = \text{diag}(+1, +1, +1)$. But this means the components of $B^i$ (vector) will be the same as $B_i$ (covariant vector). So, in good ol' fashioned Euclidean linear algebra, the dot product between two vectors is defined as

$$\vec{A}\cdot\vec{B} = g_{ij}A^i B^j = A_x B_x + A_y B_y + A_z B_z$$

Now you know!

## 4.2 Partial Derivatives

We've just seen that a contravariant vector has all positive components, *e.g.*

$$x^\mu = (x^0, x^1, x^2, x^3)$$

The covariant version has negative spatial components through the lowering operation with the mertic:

$$x_\mu = g_{\mu\nu}x^\nu = (x^0, -x^1, -x^2, -x^3)$$

So it will come as a confusing shock that the partial derivative operator has the opposite sign convention! Why? *WHY????*

The simple reason is that the (partial) derivative operator has the variable which one is differentiating with respect to in the "denominator",

$$\partial_\mu = \frac{\partial}{\partial x^\mu} = \left(\frac{\partial}{\partial x^0}, \frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, \frac{\partial}{\partial x^3}\right)$$

so the index gets flipped from up to down. Because the index is downstairs on this partial, it is a **covariant partial derivative**.

Conversely, the partial with the index upstairs is taken with respect to the *covariant* vector $x_\mu$,

$$\partial^\mu = \frac{\partial}{\partial x_\mu} = \left(\frac{\partial}{\partial x^0}, -\frac{\partial}{\partial x^1}, -\frac{\partial}{\partial x^2}, -\frac{\partial}{\partial x^3}\right)$$

Because the index is upstairs on this partial, it is a **cotravariant partial derivative**.

As usual, the two are related through raising/lowering with the metric:

$$\partial_\mu = g_{\mu\nu}\partial^\nu \quad , \quad \partial^\mu = g^{\mu\nu}\partial_\nu$$

In summary, what makes a vector contravariant or covariant is the placement of the index, and *not* the signs of the components.

## 4.3 The Line Element

The metric is key to defining the **line element**, which is the infinitesimal distance between two points in spacetime:

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu = dx^\mu dx_\mu = dt^2 - dx^2 - dy^2 - dz^2$$

The length of a path in spacetime is

$$s = \int ds = \int \sqrt{g_{\mu\nu}dx^\mu dx^\nu} = \int \sqrt{g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}}d\lambda$$

We can modify this to calculate distances of spatial paths in certain spacetime geometries, like those induced by stars, planets, or black holes. As we will see in this course, the intuitive notion of a distance will change with geometry. This will make more sense when we actually define the metric components and start talking about orbits, so to be continued!...

## 4.4 Tensors

A tensor a mathematical object of higher dimension than a vector, which can have both contravariant (upstairs) and covariant (downstairs) indices. The simplest form of a tensor is a matrix – the metric tensor is one such object, and has two indices to represent rows and columns. But we can image a extended array of numbers that is maybe three dimensional, or four, or five! We can't write them down as nicely as a matrix, but we can write down their essence with – you guessed it – *index notation!* As discussed earlier, the number of indices a tensor has is referred to as its rank.

The electromagnetic field tensor is $F^{\mu\nu}$, which embodies everything you need to know about $\vec{E}$ and $\vec{B}$ fields. We can lower the indices by applying the metric tensor for each index in question:

$$\text{Lower one} \implies F^{\mu}_{\ \alpha} = g_{\nu\alpha}F^{\mu\nu}$$

$$\text{Lower both} \implies F_{\alpha\beta} = g_{\alpha\mu}g_{\beta\nu}F^{\mu\nu}$$

The Riemann tensor tells us everything we need to know about the curvature of spacetime (and hence gravitation), and will likely become your worst nightmare this semester – if not for all time. It is a four-component object that has $4\times4\times4\times4 = 256$ components, though (SPOILER ALERT) most of them turn out to be zero. The fully covariant form is $R_{\mu\nu\alpha\beta}$, and we can raise and lower these at will using the metric:

$$\text{Raise one} \implies R^{\sigma}_{\ \nu\alpha\beta} = g^{\mu\sigma}R_{\mu\nu\alpha\beta}$$

$$\text{Raise two} \implies R^{\sigma\delta}_{\ \ \alpha\beta} = g^{\mu\sigma}g^{\nu\delta}R_{\mu\nu\alpha\beta}$$

Neat?

The tricky thing about working with the Riemann tensor is that the raising and lowering happen with a metric that *isn't flat*. But the rules of play are as described above, so it's smooth sailing once you get the hang of it.

## 4.5 Contraction

We saw earlier that tensors with self-contained upstairs and downstairs index pairs are called contractions, *e.g.*

$$F^{\mu}_{\ \mu} \quad , \quad R^{\alpha}_{\ \beta\alpha\delta} \quad , \ldots$$

As you have probably now surmised, the way to achieve this is through the metric tensor. The first contraction is

$$F^{\mu}_{\ \mu} = g_{\alpha\mu}F^{\mu\alpha} = g_{00}F^{00} + g_{11}F^{11} + g_{22}F^{22} + g_{33}F^{33}$$

We could also write this contraction as

$$F^{\mu}_{\ \mu} = g^{\alpha\mu}F_{\mu\alpha} = g^{00}F_{00} + g^{11}F_{11} + g^{22}F_{22} + g^{33}F_{33}$$

and we will get the same result. Note that $F^{\mu}{}_{\mu}$ is no longer a tensor, but is actually a scalar. Why? Because $g^{00}$ is a number, $F_{00}$ is a number, so their product is a number (and the same holds for the others). We can drop the indices in this case are write

$$F^{\mu}{}_{\mu} = F$$

Similarly, the other contraction listed above is

$$R^{\alpha}{}_{\beta\alpha\delta} = g^{\mu\alpha}R_{\mu\beta\alpha\delta} = g^{00}R_{0\beta0\delta} + g^{11}R_{1\beta1\delta} + g^{22}R_{2\beta2\delta} + g^{33}R_{3\beta3\delta}$$

and so forth. Again, this is no longer a rank-4 tensor, but a rank-2 tensor – *i.e.* it only has two free indices (a matrix!). So we could call this

$$R^{\alpha}{}_{\beta\alpha\delta} = R_{\beta\delta}$$

This will become quite familiar to you later in the course.

Note that if we are dealing with Euclidean space, the metric $g_{\mu\nu} \to g_{ij} = \mathrm{diag}(+1, +1, +1)$, and so the contraction of a tensor would just be

$$F^{i}_{i} = g_{ij}F^{ij} = g_{11}F^{11} + g_{22}F^{22} + g_{33}F^{33} = F^{11} + F^{22} + F^{33}$$

which is simply the sum of its diagonal components. You should recognize as the **trace** of the matrix $F$! The contraction is thus a generalization of the trace to spaces with non-Euclidean metric.

# 5 The Kronecker Delta

Any two objects that are inverses of each other will have a product equal to unity. When dealing with tensors, this places constraints on the components. We represent this through the **Kronecker delta**

$$\delta^{\mu}{}_{\nu} = 1 \;\text{ if } \mu = \nu \;\;, 0 \text{ if } \mu \neq \nu$$

(Note! We might be tempted to interpret the above statement as $\delta^{\mu}{}_{\mu} = 1$, but this is *not correct*. This latter statement is the trace of $\delta^{\mu}{}_{\nu}$, which is $\delta^{\mu}{}_{\mu} = 4$).

In the case of inverse objects (*e.g.* the metric tensor), this is just a way to represent the identity:

$$g_{\mu\alpha}g^{\mu\beta} = \delta_{\alpha}{}^{\beta}$$

The overall effect of the Kronecker delta is basically to “index swap”:

$$A^{\mu}\delta^{\beta}{}_{\mu} = A^{\beta} \;\;, F_{\mu\nu}\delta^{\nu}{}_{\sigma} = F_{\mu\sigma} \;\;, etc...$$

This is useful when you try to simplify an equation, and can help to show you if you have terms that are symmetric or interchangeable.

## 6 Helpful Hints and Tensor Faux Pas

Again, the beauty of index notation is in its power to recognize common properties in equations. But at the same time, it is very tempting to use the notation in a way that is nonsensical and reveals nothing (only a wrong equation!). The following are a few initial helpful tips to remember – you will pick up may more as we work through the semester:

- Look to simplify equations by "matching" pairs indices. Remember that whenever you have paired (dummy) indices, they can be replaced by any symbol. Look for other terms in an equation that might have the same indices, and it will be obvious that they're the same. For example:

$$A^\mu B_\mu + A^\beta B_\beta \quad \Longrightarrow \text{Replace } \beta \to \mu \quad \Longrightarrow \quad \mathrm{A}^\mu \mathrm{B}_\mu + \mathrm{A}^\mu \mathrm{B}_\mu = 2\mathrm{A}^\mu \mathrm{B}_\mu$$

- Use symmetry of objects (if known) to manipulate indices. For example, the following equation involves sums over the metric tensor ($g_{\mu\nu} = g_{\nu\mu}$ symmetric under index exchange) and the electromagnetic field tensor ($F^{\alpha\beta} = -F^{\beta\alpha}$ antisymmetric under index exchange):

$$g_{\mu\alpha} F^{\alpha\beta} + g_{\sigma\mu} F^{\beta\sigma}$$

$$\begin{aligned} g_{\mu\nu} = g_{\nu\mu} \ \text{ symmetric } \quad &\Longrightarrow g_{\mu\alpha} F^{\alpha\beta} + g_{\mu\sigma} F^{\beta\sigma} \\ F^{\mu\nu} = -F^{\nu\mu} \ \text{ antisymmetric } \quad &\Longrightarrow g_{\mu\alpha} F^{\alpha\beta} - g_{\mu\sigma} F^{\sigma\beta} \\ \text{Replace } \sigma \to \alpha \quad &\Longrightarrow g_{\mu\alpha} F^{\alpha\beta} - g_{\mu\alpha} F^{\alpha\beta} = 0 \end{aligned}$$

  What initially seems to be an equation that we can't simplify is actually easy to manipulate and solve.

- Always check that the number, position, and name of free indices are consistent! If one term is a rank-$(m, n)$ tensor, then **all** terms in the equation must be rank-$(m, n)$ tensors with the same free indices!

$$\text{GOOD!} \quad \longrightarrow \quad G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + \Lambda g_{\mu\nu}$$

$$\text{BAD!} \quad \longrightarrow \quad G_{\mu\nu} = R_{\alpha\nu} - \frac{1}{2} g^{\beta\nu} R + \Lambda g_{\mu\nu\rho}$$

- When in doubt: *write it out!* Index notation can get confusing quite easily, so if you find yourself getting flustered, always remember that this is just a short-hand way of representing components of vectors, matrices, and so forth. Expand paired indices as the sum they represent. Explicitly write four-vectors in their coordinate form, and so forth.