The Einstein convention, indices and networks

These notes are intended to help you gain facility using the index notation to do calculations for indexed objects. The advantage of this notation is that it allows you to perform many calculations all at once. Each value of the free indices (see below) represents an equation that you previously would have had to work out on its own.

1 Einstein summation convention

For reasons that I don't fully understand your textbook avoids one convention that is pervasive throughout all of physics and which is extremely useful. This section introduces this convention, the *Einstein summation convention.*

Frequently when we would like to keep track of the components of a vector $\mathbf{v} = (v_1, v_2, v_3)$ we use index notation. In this notation v_i refers to the ith element of the collection $\{v_1, v_2, v_3\} = \{v_i\}_{(i=1,\ldots,3)}$. For example, if we have two vectors $\mathbf{v}$ and $\mathbf{w}$ then we can write their dot product as,

$$\mathbf{v} \cdot \mathbf{w} = \sum_{i=1}^{3} v_i w_i = v_1 w_1 + v_2 w_2 + v_3 w_3.$$

The central notation with the capital greek sigma (Σ) is redundant: firstly, the values that i runs over should be clear either from the context (e.g. 3D vectors) or specified when the vector (or whatever is indexed) is introduced. Secondly, whenever you have a sum over indexed quantities you will use the same label for all quantities summed over, so that in fact, the sigma itself is redundant. Thus Einstein defined

$$v_i w_i \equiv \sum_i v_i w_i,$$

that is, a repeated index (i in this case) *means* take a sum over that index. This notation is used everywhere in physics and is very efficient.

I will use this convention in lecture. On occasion I will also write in the sigma; this is often stylistic, with the intention of emphasizing the sum but it can also be necessary for clarity, for example, when a sum contains three indices that are all the same. There are also instances where you don't want a repeated index to imply a sum and this is usually written out explicitly. An example from your textbook is the definition of the angular momentum of the αth particle in a collection of N particles,

$$\ell_\alpha = \mathbf{r}_\alpha \times \mathbf{p}_\alpha \qquad \text{(no sum)}.$$

Finally we distinguish between free and dummy indices. A free index is one that is not summed over and consequently can take *any* of the values in its allowed range. For example i is a free index in this familiar equation,

$$F_i = m a_i, \qquad (i = 1, 2, 3).$$

This equation is shorthand for the three equations representing the components of Newton's second law. By contrast, a dummy index is one that is summed over and instead it must take *all* of the values in its allowed range. The dot product equation above is an example. Dummy indices have an interesting property: you can rename them at will (hence the name). For example, both $v_i w_i$ and $v_j w_j$ represent exactly the same thing,

$$v_i w_i = v_j w_j = v_1 w_1 + v_2 w_2 + v_3 w_3.$$

This is useful to keep in mind as you will often run out of names for indices and you might want to rename something that you've already written.

By the way, experts sometimes use free indices to point out each other's mistakes; the free indices on the two sides of an equation must agree otherwise the equation doesn't make sense. Here are two examples that you can think through,

$$v_i w_j a^i = F_j$$

and

$$a^i b^j c^k u_i v_j w_k = L_l.$$

Which is "well formed" and which is not?

2 Indices for 3D quantities

There are many more conventions about indices and what they mean in the physics literature. In this section I introduce a few more. As mentioned above the intention of this supplement is primarily to get you up to speed with calculating using indices; for a wonderful exposition of what all of this *means* see Nadir's recent book "An Introduction to Tensors and Group Theory for Physicists."

At this point I'll specialize to consideration of 3D vectors, this will keep things concrete and specify the range of all indices. Many texts and articles distinguish between upper and lower indices. An upper index usually indicates a standard vector and we represent it by a column matrix made up of its components,

$$v^i = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}.$$

While a lower index refers to a "dual" vector and is represented by a row of its components,

$$u_i = (u_1, u_2, u_3).$$

There is a mathematical gadget, called a metric, that converts vectors into dual vectors, $v^i \mapsto v_i$. It turns out that in regular, Euclidean, 3-space this gadget is represented by the matrix

$$\delta_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and so,

$$v_i = \delta_{ij} v^j = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = (v_1, v_2, v_3).$$

This shows that in Euclidean 3D the components of the dual vector v_i corresponding to v^i are the *same* and there is little reason for distinction. (The only substantive distinction that remains is the row vs. column distinction and we simply infer this by context). All of this is just to explain why, in your text, there is no distinction made between upper and lower indices and everything is denoted with a lower index. This convention is perfectly valid but you do want to keep room for a distinction to arise in the future.[1]

The metric introduced above, δ_{ij}, is often called the Kronecker delta and another one of its functions is to contract two vectors into a dot product,

$$v_i \delta_{ij} w_j = v_i w_i = \mathbf{v} \cdot \mathbf{w}.$$

There is a second product of vectors that we have been using extensively, namely the cross product,

$$\mathbf{v} \times \mathbf{w} = (v_y w_z - v_z w_y, v_z w_x - v_x w_z, v_x w_y - v_y w_x).$$

This product can also be captured using the index notation. The key is to appreciate the antisymmetry of this product and to introduce the Levi-Civita epsilon,

$$\epsilon_{ijk} = \begin{cases} 1 & \text{if } (i,j,k) \text{ is } (1,2,3),\ (3,1,2) \text{ or } (2,3,1), \\ -1 & \text{if } (i,j,k) \text{ is } (1,3,2),\ (2,1,3) \text{ or } (3,2,1), \\ 0 & \text{if } i = j \text{ or } j = k \text{ or } i = k. \end{cases}$$

This definition is a little dense but not hard to unpack. One way to think about it is that if you start with whichever index is 1 and read to the right (where if the 1 is in the last slot you consider "to the right" as starting over) do you encounter a 2 or a 3. If you encounter a 2 it's called a cyclic permutation and you are in the top line of our definition, so for example $\epsilon_{312} = 1$. If you encounter a 3 it's an anti-cyclic permutation and the middle line of the definition is the relevant one. If any index is repeated then the Levi-Civita epsilon is zero.

A straightforward but tedious calculation confirms that the Levi-Civita epsilon captures the cross product, if $\mathbf{u} = \mathbf{v} \times \mathbf{w}$ then,

$$u_i = \epsilon_{ijk} v_j w_k.$$

Note the sums over the last two indices and the matching of the free i index on each side. To reiterate, this efficiently captures all three components of the cross product in a single equation.

We need two more formulae to make the index notation as versatile as is necessary in this course. The first one is a relationship between the ϵ and the δ introduced above. The proof of this formula is again a somewhat involved calculation; I leave it to you for when you are in the dentist's waiting room. It is a wonderful formula and it is worth proving it at some point in your life,

$$\epsilon_{ijk} \epsilon_{klm} = \delta_{il} \delta_{jm} - \delta_{im} \delta_{jl}.$$

[1]For example, such a distinction is of value in special relativity

This is used when you have multiple cross products and simplifies them down to dot products frequently. The second formula is not at all new to you it just expresses what you know about derivatives in the index notation. Let us, as usual, denote the position vector by $\mathbf{r} = (x, yz)$ and its index form by r_i. What is the derivative of r_i with respect to position, that is with respect to r_j? Well, that depends on whether $i = j$ or not. If it does then, the derivative is 1 and if it's not then the derivative is 0. We can notate this with the Kronecker δ again. Because δ_{ij} also has the property that it is 1 if $i = j$ and zero otherwise. Then,

$$\frac{\partial r_i}{\partial r_j} = \delta_{ij}.$$

Alright, that does it. We can use these results to calculate all of the vector identities on the front cover of Griffiths' "Introduction to Electrodynamics." I'll do an example so that you get an idea of how it works. Let's derive

$$\nabla \cdot (\mathbf{A} \times \mathbf{B}) = \mathbf{B} \cdot (\nabla \times \mathbf{A}) - \mathbf{A} \cdot (\nabla \times \mathbf{B}).$$

A common and useful index shorthand for nabla (∇) is ∂_i and, of course, $\partial_i = \partial/\partial r_i$. Then we'd like to calculate,

$$\nabla \cdot (\mathbf{A} \times \mathbf{B}) = \partial_i[(\mathbf{A} \times \mathbf{B})]_i,$$

here I am using the square braces to denote the ith component of the vector. So,

$$\partial_i[(\mathbf{A} \times \mathbf{B})]_i = \partial_i(\epsilon_{ijk} A_j B_k),$$

from our formula for the cross product. Now there comes an important point: because we're working with components, they are just regular variables (as opposed to vector variables) and so we can use the product rule for regular variables. We find,

$$\partial_i(\epsilon_{ijk} A_j B_k) = \epsilon_{ijk} B_k \partial_i(A_j) + \epsilon_{ijk} A_j \partial_i(B_k).$$

(Note that ϵ_{ijk} does depend on position so we don't get derivatives of it.) Again we're dealing with components and they multiply as numbers so can commute things around to get,

$$\epsilon_{ijk} B_k \partial_i(A_j) + \epsilon_{ijk} A_j \partial_i(B_k) = B_k \epsilon_{ijk} \partial_i(A_j) + A_j \epsilon_{ijk} \partial_i(B_k).$$

Next we use the properties of the ϵ; a cyclic permutation of its indices doesn't change its value, while an anti-cylclic one incurs a minus sign. Let's perform a cyclic permutation of the ϵ in the first term and an anti-cyclic permutation of the one in the second term to get,

$$B_k \epsilon_{ijk} \partial_i(A_j) + A_j \epsilon_{ijk} \partial_i(B_k) = B_k \epsilon_{kij} \partial_i(A_j) - A_j \epsilon_{jik} \partial_i(B_k).$$

This last equation has recognizable pieces. The first term is a dot product between $\mathbf{B}$ and the curl of $\mathbf{A}$ and the second is the dot product of $\mathbf{A}$ with the curl of $\mathbf{B}$, so indeed,

$$\nabla \cdot (\mathbf{A} \times \mathbf{B}) = \mathbf{B} \cdot (\nabla \times \mathbf{A}) - \mathbf{A} \cdot (\nabla \times \mathbf{B}).$$

Practice doing these calculations on the rest of the formulae on Griffiths' cover.

3 Networks

This is an optional section that I'm still working on. To summarize, some of us have an easier time thinking visually. The index notation can make your eyes cross if you're this type of person. This section gives a visual way to do these calculations.