That One Notation Named After Einstein For Being Lazy

PHY 304 – Advanced Electromagnetism

Ravin Raj

Contents

1 Introduction to Index Notation

While we are used to performing vector manipulations in a given coordinate system, it turns out that abstraction can often make these expressions take on a much simpler form. Specifically, we want to develop a formalism for expressing the components of vectors in a compact fashion so that manipulations of these quantities and the underlying bases can be performed with relative mathematical ease. In his seminal 1916 paper outlining his newfound theory of general relativity [1], Einstein proposed a new convention out of "brevity", though everyone knew him to be one who hated to needlessly carry over expressions from previous lines. It was rapidly adopted by theorists in the field, and despite their best efforts, detractors did not hold out past the mid 1930s[1]. This notation has now been named after him, as the **Einstein summation convention**, sometimes also simply called **index notation**.

Given a region of space which resembles[2] $\mathbb{R}^n$, we can establish a Cartesian coordinate system which then denotes the position of any point in this space by an n-tuple $(a_1, a_2, \cdots, a_n)$. This is equivalent to defining an orthonormal basis $\{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \cdots, \hat{\mathbf{e}}_n\}$ for the space, following which all positions can be described as position vectors formed by linear combinations of this basis. It has been implicitly assumed that all of these vectors are column vectors. Mathematically, we write:

[1] This was around the time when he had successfully fled from Germany and arrived in Princeton to take up a position at the Institute for Advanced Study (where he would continue his work until his passing in 1955). He was known to be a brilliant speaker and author, so it was only natural that he would sway the naysayers eventually.

[2] This qualification is incredibly important, and will be expounded later. As a simple example of how this argument breaks down in special cases, consider the surface of a 3-dimensional sphere. Locally (in small enough patches on the surface), the surface looks like it can be described by a coordinate system of $\mathbb{R}^2$, but it would be egregious to extrapolate this to say that the entire sphere can be described by a single $\mathbb{R}^2$ coordinate system (see *flat earther*). Specifically, it can be shown that there will always be at least one point on the sphere which cannot be uniquely described by $\mathbb{R}^2$. For instance, if we set the origin of the coordinate system at the North Pole, then the South Pole will be a finite line in this coordinate system.

$$\mathbf{r} = a_1\hat{\mathbf{e}}_1 + a_2\hat{\mathbf{e}}_2 + \cdots + a_n\hat{\mathbf{e}}_n \tag{1.1}$$

$$= \sum_{i=1}^{n} a_i\hat{\mathbf{e}}_i \tag{1.2}$$

The coefficients of the linear combination constituting the position vector are thus simply the coordinates of the position in question. Einstein's convention involves dropping the summation symbol, where it is henceforth implied that the occurrence of any index **twice** denotes a sum over that index from 1 to n. For instance, the position vector can now be rewritten as:

$$\mathbf{r} = a_i\hat{\mathbf{e}}_i \tag{1.3}$$

One should note that the choice of label for the index (i in this case) is arbitrary, and can be chosen at will. These are known as **dummy indices**, since they end up being summed over and are thus free to take on any convenient label (e.g. $i, \mu, \$, \heartsuit, \cdots$)[3]. This implies two very important rules:

1. No index can appear on the same side of an equation more than twice.
2. This is an unofficial rule commonly known as the conservation of index – any indices appearing on one side of an equation once must also appear on the other side of the equation once. Such indices are known as **free indices**.

That's really it.

2 The Kronecker Delta and Scalar Products

From this point on, we will work exclusively in 3D so it is always implied that the summations run as $i = 1, 2, 3$. The first object we will meet is the **Kronecker delta** δ_{ij}, which is defined as:

$$\delta_{ij} = \begin{cases} 1\,, & \text{when } i = j \\ 0\,, & \text{when } i \neq j \end{cases} \tag{2.1}$$

While this seems like a rather trivial object, it turns out to be surprisingly useful in much of what follows. Its first primary use appears in the orthonormality relation between basis vectors, where we can succinctly write:

$$\hat{\mathbf{e}}_i \cdot \hat{\mathbf{e}}_j = \delta_{ij} \tag{2.2}$$

This is easy to see, since we imposed that the basis was orthonormal at the start, so the inner products between different basis vectors vanishes, while the inner product of a basis vector with itself gives unity for a normalised (unit) vector. With this, we can use the Kronecker delta to construct the scalar (inner) product between arbitrary vectors as:

[3] In the literature, it is common to use the Latin alphabet for indices which run over the values $\{1, 2, 3\}$, while reserving Greek letters for indices which run over the values $\{0, 1, 2, 3\}$, as is used in relativistic (3+1)-dimensional notation. Any other choice of index will usually prompt strange looks from the audience.

$$\begin{aligned}\mathbf{a}\cdot\mathbf{b} &= (a_i\hat{\mathbf{e}}_i)\cdot(b_j\hat{\mathbf{e}}_j)\\ &= a_ib_j(\hat{\mathbf{e}}_i\cdot\hat{\mathbf{e}}_j)\\ &= \delta_{ij}a_ib_j\\ &= a_ib_i\end{aligned} \tag{2.3}$$

where in the second line, I have used the bilinearity of the scalar product and in the last line, I have invoked the definition of the Kronecker delta from Eq. 2.1 to set the index j to i, since all other values vanish under the implied summation over j. Note that we could have just as well performed the summation over i instead, using the Kronecker delta to set all appearances of i to j. The two results are equivalent, since the final index appearing on the right is a dummy index, and can be renamed at will.

We can similarly also express matrices in index notation, though I will first borrow from the Dirac notation in quantum mechanics to simplify things a little. So far, all the vectors presented (including the unit vectors) have been assumed to be column vectors ($n\times 1$ objects). We can identify these as kets, making the notational switch: $\mathbf{a}\longrightarrow|a\rangle$ for a general vector, and $\hat{\mathbf{e}}_i\longrightarrow|e_i\rangle$ for the unit vectors. Then, it should be clear that the transpose of any of these vectors $|a\rangle$ is simply given by the corresponding bra $\langle a|$. Now, we return to the problem of expressing matrices in index form, by writing:

$$\begin{aligned}\mathbf{M} &= \sum_{i,j}|e_i\rangle\langle e_i|\,\mathbf{M}\,|e_j\rangle\langle e_j|\\ &= \sum_{i,j}|e_i\rangle\,M_{ij}\,\langle e_j|\\ &= M_{ij}\,|e_i\rangle\langle e_j|\end{aligned} \tag{2.4}$$

where in the first line, I have used the completeness relation on either side of the matrix $\mathbf{M}$. In the second line, we define the ij-th matrix element of $\mathbf{M}$ as $M_{ij}=\langle e_i|\mathbf{M}|e_j\rangle$, and in the last line I have invoked the index notation and simply moved the matrix element to the left since it is just a scalar. In this form, we see that every matrix is just a sum over all of its elements M_{ij} multiplied by the corresponding dyad[4] $|e_i\rangle\langle e_j|$.

It is interesting to note that the Kronecker delta has two indices, just like the matrix element M_{ij}, and it is thus tempting to think of it as a matrix. It turns out that this is absolutely the way to think about it, since replacing M_{ij} in Eq. 2.4 with the Kronecker delta returns:

$$\delta_{ij}\,|e_i\rangle\langle e_j| = |e_i\rangle\langle e_i| = \mathbb{1} \tag{2.5}$$

where we use the completeness relation for the last equality. Thus, we see that the Kronecker delta is nothing more than the identity matrix. With these ideas, we can represent matrix multiplication using this notation as:

$$\begin{aligned}&\mathbf{A}\mathbf{x}=\mathbf{b}\\ \iff& A_{ij}x_j=b_i\\ \iff& x_jA_{ij}=b_i\end{aligned}$$

[4]The word dyad is not too commonly seen in physics. Think of it like the matrix equivalent of a unit vector, where it has a one in only one spot and zeroes everywhere else, so it only addresses a single position in a matrix. This is similar to the Cartesian representation of a unit vector, where it just has a one in only one spot and zeroes everywhere else, meaning it only points along one of the coordinate axes.

Interestingly, the fact that the index notation now reduces mathematical objects like vectors and matrices to their scalar components implies that we can now freely permute these quantities around to our hearts' content. This is perfectly valid since these are just scalars and do not have any row or column structure that we need to care about when taking their products. We just have to be careful when re-expressing the final expression back into vector/matrix form, where the summations must be performed in the right order. A good rule of thumb is to keep indices as close together as possible, so it is more easily seen which order the product must be taken in, and where (if applicable) any transpose operations should be applied. In the above expression, I have also implicitly dropped the basis vector $\hat{\mathbf{e}}_i$ on both sides of the equation, which is allowed since we are merely comparing the components of each basis vector.

It should also be noted that the operation of summing over one of the indices in a Kronecker delta is identical to (and sometimes known as) **tensor contraction**[5]. In this operation, we are effectively reducing the rank (number of indices) of the object by two. For instance, the outer product of two vectors $a_i b_j$ forms a matrix, which is a rank-two tensor. In contrast, the inner product of two vectors $a_i b_i$ forms a scalar, which is a rank-zero tensor. These two expressions differ only by the imposition of δ_{ij}. A useful result is the contraction of the Kronecker delta itself, which can also be viewed as the trace of the identity matrix:

$$\delta_{ij}\delta_{ij} = \delta_{ii} = 3 \tag{2.6}$$

3 The Levi-Civita Symbol and Cross Products

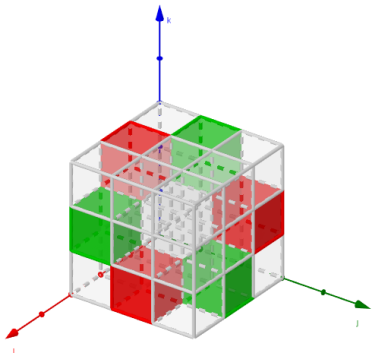

FIGURE 1: Graphical representation of the Levi-Civita symbol, with each of the coordinate axes representing one of the indices. Green spots represent a value of $+1$, while red spots represent a value of -1, with all other spots having zero. Image obtained from [3].

The next most useful symbol we will frequently encounter throughout all of physics is the **Levi-Civita symbol** ε_{ijk}, sometimes also known as the **totally antisymmetric tensor**. In contrast to the Kronecker delta, the Levi-Civita has three indices, so this should not be represented as a 2D matrix but rather a 3D cube with entries in the $3^3 = 27$ possible spots. Its definition is given as follows:

$$\varepsilon_{ijk} = \begin{cases} +1\,, & ijk \text{ is an even permutation of } 123 \\ -1\,, & ijk \text{ is an odd permutation of } 123 \\ 0\,, & \text{otherwise} \end{cases} \tag{3.1}$$

In this definition, the terms even and odd permutation refer to the method of obtaining ijk starting from the triple 123. Even permutations are those which can be obtained through cyclic permutation, such as $123, 231, 312$. Odd permutations are those which can be obtained by cyclic permutation after a single swap, such as $213, 132, 321$. An equivalent definition of the permutation parity is also to count the number of pairwise swaps it takes to obtain the triple ijk starting from 123. It will quickly become clear that the even (odd) permutations are those which take an

[5] If the sudden appearance of the term tensor has shocked you, refer to [2] for one of the best reviews of the subject (in my opinion). Specifically, chapters 2–8 cover Cartesian tensor theory in great detail while remaining accessible to a large audience. A thorough understanding of tensor geometry is not required for the subject matter here though.

even (odd) number of such swaps, hence the names. The symbol also vanishes for all values of ijk which are not permutations of 123, and this specifically occurs when any two (or all three) indices are equal. In this manner we see that any entry of ε_{ijk} along a line connecting two corners of the cube must vanish, which sets 21 of the values to zero, and any entry in the region where $(i+j+k) < 6$ must vanish (by the pigeonhole principle) leaving only six nonzero entries. An intuitive visualisation of the Levi-Civita symbol is provided in Fig. 1, which shows that all the nonzero entries lie in a single plane perpendicular to $\mathbf{n} = (1,1,1)$. This geometry gives the Levi-Civita symbol its most notable characteristic – it changes sign under the permutation of any two indices.

Coming back to its use in index notation, the Levi-Civita symbol provides the natural representation of a cross (vector) product in index notation. Specifically, we can write:

$$\mathbf{a} \times \mathbf{b} = \varepsilon_{ijk} \hat{\mathbf{e}}_i a_j b_k = \varepsilon_{ijk} a_i b_j \hat{\mathbf{e}}_k \tag{3.2}$$

where the unit basis vector can be placed either in the first or last position with no consequence, since these are just cyclic permutations of the indices which do not change the sign of the Levi-Civita symbol. One may also be familiar with the definition of the cross product as a matrix determinant:

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \hat{\mathbf{e}}_1 & \hat{\mathbf{e}}_2 & \hat{\mathbf{e}}_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} \tag{3.3}$$

Comparing Eqs. 3.2 and 3.3 reveals some close similarities, and it turns out that the Levi-Civita is nothing more than the operation which takes a set of vectors, places them into the columns of a matrix, and computes their determinant. From this, we can actually write the explicit form of the Levi-Civita symbol (in 3D) as:

$$\varepsilon_{ijk} \equiv \begin{vmatrix} \delta_{i1} & \delta_{i2} & \delta_{i3} \\ \delta_{j1} & \delta_{j2} & \delta_{j3} \\ \delta_{k1} & \delta_{k2} & \delta_{k3} \end{vmatrix} \tag{3.4}$$

In this form, it becomes much clearer why the Levi-Civita symbol vanishes if any two indices are equal. If this were to occur, at least one row of this matrix would be empty, and so the determinant trivially vanishes. We can now use this form to compute several incredibly useful identities for the Levi-Civita symbol which come up ubiquitously throughout physics.

3.1 Contraction Identities

One may be tempted (or not) to consider the product of two Levi-Civita symbols $\varepsilon_{ijk}\varepsilon_{lrs}$. Following the definition from Eq. 3.4 would give us a horrendous expression, and indeed the resulting expression is a rank-six beast which is almost never useful. Instead, we can consider the simplest case where the product is fully contracted, that is, we wish to evaluate the result of $\varepsilon_{ijk}\varepsilon_{ijk}$. This is rather simple, since it implies that we are simply taking the square of all 27 entries in the symbol and summing over it. We previously already saw that there are only six nonzero entries in the symbol, and they all have values ± 1, so the result is simply:

$$\varepsilon_{ijk}\varepsilon_{ijk} = 6 \tag{3.5}$$

The next case is that of a doubly-contracted product, where two of the indices on the two symbols are set to be equal as $\varepsilon_{ijk}\varepsilon_{ijl}$. First, we note that for a given set of the first two indices, there is (at most) only one nonvanishing element for the last index. Furthermore, the equality of the first two indices in the two occurrences of the symbol implies that any nonzero product must also have the last indices of the two symbols

being equal. We thus expect that the answer is proportional to δ_{kl}. This also means that any contribution to the answer would appear as a square of the Levi-Civita element, which can only be unity since $(\pm 1)^2 = 1$.

Now, since k and l are free indices, they are not being summed over and we only have to worry about summation over i and j. For this, we note that we can fix the index k, which thus also fixes l, and consider how many nonzero combinations for (i, j) remain. This clearly gives two remaining combinations, since neither of these values can be equal to k and they cannot be equal to each other. We thus obtain the result:

$$\varepsilon_{ijk}\varepsilon_{ijl} = 2\delta_{kl} \tag{3.6}$$

Finally, we have the "hardest" case for which the product is only singly contracted as $\varepsilon_{ijk}\varepsilon_{ilm}$, so we expect the result to be a rank-four tensor. To tackle this, we first invoke Eq. 3.4 to write:

$$\varepsilon_{ijk}\varepsilon_{ilm} \equiv \begin{vmatrix} \delta_{i1} & \delta_{i2} & \delta_{i3} \\ \delta_{j1} & \delta_{j2} & \delta_{j3} \\ \delta_{k1} & \delta_{k2} & \delta_{k3} \end{vmatrix} \begin{vmatrix} \delta_{i1} & \delta_{i2} & \delta_{i3} \\ \delta_{l1} & \delta_{l2} & \delta_{l3} \\ \delta_{m1} & \delta_{m2} & \delta_{m3} \end{vmatrix} \tag{3.7}$$

We can simplify this by first noting that the product of two determinants is equal to the determinant of the product. However, before proceeding to write out the entire matrix, we can make some greatly simplifying observations. First, the determinant of a matrix is unaffected by taking its transpose, so we will transpose the second matrix to ensure that the numerical indices match up under matrix multiplication. Then, the entries of the resulting matrix will be of the form:

$$\begin{aligned} (\varepsilon_{ijk}\varepsilon_{ilm})_{ab} &= \delta_{a1}\delta_{b1} + \delta_{a2}\delta_{b2} + \delta_{a3}\delta_{b3} \\ &= \delta_{ac}\delta_{bc} \\ &= \delta_{ab} \end{aligned} \tag{3.8}$$

where $(\varepsilon_{ijk}\varepsilon_{ilm})_{ab}$ represents the element obtained by taking the product of the a-th row of ε_{ijk} with the b-th column of $\varepsilon^{\intercal}_{ilm}$. This is a considerable simplification, and enables us to write:

$$\begin{aligned} \varepsilon_{ijk}\varepsilon_{ilm} &= \left| \begin{pmatrix} \delta_{i1} & \delta_{i2} & \delta_{i3} \\ \delta_{j1} & \delta_{j2} & \delta_{j3} \\ \delta_{k1} & \delta_{k2} & \delta_{k3} \end{pmatrix} \begin{pmatrix} \delta_{i1} & \delta_{l1} & \delta_{m1} \\ \delta_{i2} & \delta_{l2} & \delta_{m2} \\ \delta_{i3} & \delta_{l3} & \delta_{m3} \end{pmatrix} \right| \\ &= \begin{vmatrix} \delta_{ii} & \delta_{il} & \delta_{im} \\ \delta_{ji} & \delta_{jl} & \delta_{jm} \\ \delta_{ki} & \delta_{kl} & \delta_{km} \end{vmatrix} \\ &= \delta_{ii}(\delta_{jl}\delta_{km} - \delta_{jm}\delta_{kl}) - \delta_{il}(\delta_{ji}\delta_{km} - \delta_{jm}\delta_{ki}) + \delta_{im}(\delta_{ji}\delta_{kl} - \delta_{jl}\delta_{ki}) \\ &= 3\delta_{jl}\delta_{km} - 3\delta_{jm}\delta_{kl} - \delta_{jl}\delta_{km} + \delta_{jm}\delta_{kl} + \delta_{jm}\delta_{kl} - \delta_{jl}\delta_{km} \\ &= \delta_{jl}\delta_{km} - \delta_{jm}\delta_{kl} \end{aligned} \tag{3.9}$$

It should also be noted that the two equal indices on the Levi-Civita symbols do not have to be in the same position, since they can always be permuted cyclically to make use of this identity. We only have to ensure that common indices are either in the first or last position[6]. A simple mnemonic to remember this

[6]This is because of the cyclic property of the indices which leave the Levi-Civita symbol unchanged. If the common index is in either the first or last positions, the reduced pair of indices (once the common index is removed) preserves its ordering.

identity is *same same, then different*. Specifically, if we look at the two sets of indices on the Levi-Civitae[7], say (ijk) and (rsi), one of these will have to be cyclically permuted to put both i's in the same position. However, the end result is independent of i since this is a dummy index, and the cyclic permutation does not change the relative order of the remaining two indices, so we can simply ignore i and consider the reduced pair (jk) and (rs). The mnemonic then demands that we pair the indices in the deltas as $(jr)(ks)$ (same same), then as $(js)(kr)$ (different) with the minus sign on the second term. This identity will be repeatedly used in vector calculus, particularly in cases where we have a nested cross product (or curl). In obtaining this result, we have also passed the general solution to the product of two Levi-Civita symbols in the second line, where we can immediately identify the solution as:

$$\varepsilon_{ijk}\varepsilon_{qrs} = \begin{vmatrix} \delta_{iq} & \delta_{ir} & \delta_{is} \\ \delta_{jq} & \delta_{jr} & \delta_{js} \\ \delta_{kq} & \delta_{kr} & \delta_{ks} \end{vmatrix} \tag{3.10}$$

This result comes up far less often, and for good reason too. In general, when performing manipulations on vectors, we expect to obtain a tensorial result of rank zero, one or two. In these cases, it is likely that the intermediate result would contain tensors of higher rank, but the final result will typically be contracted down appropriately. Thus, in any occurrence of a product of two Levi-Civita symbols whereby we have a rank-six tensor, it is usually the case that this can be first contracted at least once to invoke Eq. 3.9, rather than having to use the general case of Eq. 3.10. As an example of how these identities are used, we will now derive the vector triple product rule using index notation.

$$\begin{aligned} \mathbf{A}\cdot(\mathbf{B}\times\mathbf{C}) &= \delta_{ij}A_i(\varepsilon_{jkl}B_kC_l) \\ &= \varepsilon_{ikl}A_iB_kC_l, \quad \text{summing over } j \text{ in the previous delta} \\ &= \varepsilon_{kli}A_iB_kC_l, \quad \text{cyclic permutation of } ikl \text{ in the Levi-Civita symbol} \\ &= B_k\varepsilon_{kli}C_lA_i, \quad \text{moving scalars} \\ &= \delta_{jk}B_j\varepsilon_{kli}C_lA_i, \quad \text{reintroducing a delta through reverse contraction} \\ &= \mathbf{B}\cdot(\mathbf{C}\times\mathbf{A}) \end{aligned} \tag{3.11}$$

While this may have looked like a lengthier-than-necessary derivation, once intuition sets in, derivations like these using the index notation can be done in just a couple of lines. As an example, we will also derive the "BAC-CAB" rule using index notation.

$$\begin{aligned} \mathbf{A}\times(\mathbf{B}\times\mathbf{C}) &= \varepsilon_{ijk}\hat{\mathbf{e}}_iA_j(\varepsilon_{klm}B_lC_m) \\ &= \hat{\mathbf{e}}_iA_jB_lC_m(\delta_{il}\delta_{jm}-\delta_{im}\delta_{jl}), \quad \text{same same, then different} \\ &= \hat{\mathbf{e}}_iA_jB_iC_j - \hat{\mathbf{e}}_iA_jB_jC_i, \quad \text{summing over } l \text{ and } m \\ &= \mathbf{B}(\mathbf{A}\cdot\mathbf{C}) - \mathbf{C}(\mathbf{A}\cdot\mathbf{B}) \end{aligned} \tag{3.12}$$

where I have skipped any contraction steps by automatically summing over one index in the Kronecker delta (i.e. $\delta_{ij}a_j \to a_i$). Clearly, this is far simpler than writing out all terms explicitly and cancelling things out. One is also less prone to algebraic errors in the derivation by using this method, unless of course one's index bookkeeping is subpar, in which case one should reconsider studying vector calculus.

[7]This is probably not the plural of Levi-Civita, especially given Civita is literally someone's name, but this just sounds phonetically correct so let's go with it for now.

4 From Tensor Algebra to Tensor Calculus

Thusfar, we have been considering the algebraic manipulations one can perform on tensors, with components that were assumed to be constants. We can extrapolate our arguments from before to tensors of functions, whose components now have some functional parametrisation. This parametrisation could be spatial (a function of position $\mathbf{r}$), temporal (a function of time t), or both. With functions in each element of a tensor, we can now perform the standard operations of calculus on tensors, namely differentiation and integration. For this, we introduce the Cartesian vector derivative operator, or simply the **gradient operator** ∂_i, defined by:

$$\partial_i \equiv \left\langle \frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \cdots, \frac{\partial}{\partial x_n} \right\rangle = \frac{\partial}{\partial x_i}\hat{\mathbf{e}}_i \overset{d=3}{\equiv} \nabla \tag{4.1}$$

At this point, I must reiterate that the definition above is strictly only valid in Cartesian coordinates. The contents of this section **cannot** be extrapolated to curvilinear coordinate systems (such as cylindrical or spherical coordinates) in general. While the identities may carry over, any results obtained in the Cartesian formalism will not. Now that we have defined the gradient operator in a similar fashion to a vector, it is trivial to define the three basic vector calculus operations:

$$\boldsymbol{\nabla}\phi = \partial_i \phi \tag{4.2}$$

$$\boldsymbol{\nabla} \cdot \mathbf{A} = \partial_i A_i \tag{4.3}$$

$$\boldsymbol{\nabla} \times \mathbf{A} = \partial_i A_j \hat{\mathbf{e}}_k = \hat{\mathbf{e}}_i \partial_j A_k \tag{4.4}$$

where the two forms in the last line can be used interchangeably up to personal preference. With these definitions in place, and the identities from the previous section, we can derive *nearly*[8] every vector calculus identity that is relevant in undergraduate physics. The biggest mistake that students typically make in using the index notation for vector calculus is forgetting that, while the quantities are scalars, the differential operator is still a differential operator. Thus, a great deal of attention should be paid to the application of any derivatives, particularly if the product rule is required. As a demonstration, we will now derive the identity for the divergence of a cross product (relevant in the use of the Poynting vector, Chapter 8).

$$\begin{aligned}
\boldsymbol{\nabla} \cdot (\mathbf{A} \times \mathbf{B}) &= \partial_i(\varepsilon_{ijk} A_j B_k) \\
&= \varepsilon_{ijk}(B_k \partial_i A_j + A_j \partial_i B_k), \quad \text{product rule} \\
&= B_k \varepsilon_{ijk} \partial_i A_j - A_j \varepsilon_{jik} \partial_i B_k, \quad \text{permute once in the 2nd term} \\
&= \mathbf{B} \cdot (\boldsymbol{\nabla} \times \mathbf{A}) - \mathbf{A} \cdot (\boldsymbol{\nabla} \times \mathbf{B}), \quad \text{vector triple products}
\end{aligned} \tag{4.5}$$

Note that while I have used the product rule on the second line, it does not apply to the Levi-Civita symbol since that is merely a mathematical object consisting of constants ± 1. To really drive home the point, I will also include two more derivations which involve a neat trick that comes up repeatedly. First, we will show the usual result that the divergence of any curl vanishes. We proceed as:

[8] I say nearly, because there is specifically one identity that you cannot – the gradient of a scalar product $\boldsymbol{\nabla}(\mathbf{A} \cdot \mathbf{B}) = (\mathbf{A} \cdot \boldsymbol{\nabla})\mathbf{B} + (\mathbf{B} \cdot \boldsymbol{\nabla})\mathbf{A} + \mathbf{A} \times (\boldsymbol{\nabla} \times \mathbf{B}) + \mathbf{B} \times (\boldsymbol{\nabla} \times \mathbf{A})$. The good news is that you wouldn't be able to derive it explicitly either. One simply needs to divine the right hand side and show they are equal.

$$
\begin{aligned}
\boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}\times\mathbf{A}) &= \partial_i\varepsilon_{ijk}\partial_j A_k \\
&= \varepsilon_{ijk}\partial_i\partial_j A_k, \quad \text{move the Levi-Civita out} \\
&= \frac{1}{2}(\varepsilon_{ijk}\partial_i\partial_j A_k + \varepsilon_{jik}\partial_j\partial_i A_k), \quad \text{split into two then rename } i\leftrightarrow j \text{ in the 2nd term} \\
&= \frac{1}{2}(\varepsilon_{ijk}\partial_i\partial_j A_k + \varepsilon_{jik}\partial_i\partial_j A_k), \quad \text{commute derivatives by Schwarz' theorem} \\
&= \frac{1}{2}(\varepsilon_{ijk}\partial_i\partial_j A_k - \varepsilon_{ijk}\partial_i\partial_j A_k), \quad \text{permute the 2nd Levi-Civita} \\
&= 0
\end{aligned}
$$

While it might look annoying, this is an incredibly common and powerful trick. It involves splitting a term up into two identical half-copies of itself, then waving a magic wand on the second term until we get it back to the same form as the first with a minus sign, which shows the original term vanishes. From a symmetry point of view, this cancellation arises because the rank-one tensor $\varepsilon_{ijk}\partial_i\partial_j$ is antisymmetric in the indices (i,j). Thus, any summation over those two indices will always give equal and opposite terms which cancel, regardless of the third index. Similarly, we can use this to show the curl of a gradient vanishes.

$$
\begin{aligned}
\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\phi) &= \varepsilon_{ijk}\hat{\mathbf{e}}_i\partial_j(\partial_k\phi) \\
&= \hat{\mathbf{e}}_i\frac{1}{2}(\varepsilon_{ijk}\partial_j\partial_k\phi + \varepsilon_{ikj}\partial_k\partial_j\phi), \quad \text{split and rename } i\leftrightarrow j \text{ in the 2nd term} \\
&= \hat{\mathbf{e}}_i\frac{1}{2}(\varepsilon_{ijk}\partial_j\partial_k\phi - \varepsilon_{ijk}\partial_j\partial_k\phi), \quad \text{Schwarz' theorem and permute the Levi-Civita} \\
&= \mathbf{0}
\end{aligned}
$$

If you are not sold by this point on index notation, that is perfectly fine. Whether or not someone ends up using it is entirely up to personal preference, and many difficult calculations are often easier to analyse with just one component before abstracting the argument to the general case using index notation.

5 Physics with Index Notation

Now that we have covered the mathematical aspects of this summary, we can finally do some physics. Here, I aim to provide just a few illustrative examples of how index notation can either simplify computation, or provide better intuition on the physics involved.

5.1 Electromagnetism

A very common set of results that we will use in employing index notation for electromagnetism-related calculations is that of derivatives of the position vector, or its norm. For instance, we will frequently have to take the gradient of a scalar potential which scales as r^{-n}, or the divergence of a vector field which goes as $\mathbf{r}/r^n$, for some integer n. While these can be worked out with brute-force methods, they are good exercises for building intuition with the index notation. We first state a near-trivial but important fact about the derivatives of the position vector $\mathbf{r} = r_i\hat{\mathbf{e}}_i$. When considering the general vector derivative of a position vector, we have:

$$
\partial_i r_j = \delta_{ij} \tag{5.1}
$$

This is easy to see since the Cartesian vector derivative is made up of components $\partial/\partial x$, $\partial/\partial y$, $\partial/\partial z$, while the position vector has components x, y, z. Thus, each component of the differential operator acts on only one component of the position vector to give a nonvanishing result. This result is only true when working in an orthonormal basis, and is not general. Another useful result is that of the divergence and curl of the position vector, with the divergence as:

$$\boldsymbol{\nabla} \cdot \mathbf{r} = \partial_i r_i = 3 \tag{5.2}$$

While it looks strange, this is simply the result from Eq. 5.1 with a contraction imposed over the indices (ij) using another Kronecker delta. That is, we are simply taking the trace of the above result. Since the Kronecker delta can be represented by the identity matrix, and the trace of the identity matrix is simply a sum of n 1's, where n is the dimension of the space, it is clear how we arrive at a value of 3. Next, we consider the curl of the position vector:

$$\begin{aligned}
\boldsymbol{\nabla} \times \mathbf{r} &= \hat{\mathbf{e}}_i \varepsilon_{ijk} \partial_j r_k \\
&= \hat{\mathbf{e}}_i \varepsilon_{ijk} \delta_{jk}, \quad \text{using Eq. 5.1} \\
&= \hat{\mathbf{e}}_i \varepsilon_{ijj} = 0, \quad \text{since Levi-Civita vanishes with two equal indices}
\end{aligned} \tag{5.3}$$

With these in place, let's derive one of the two useful results for electromagnetism, namely the gradient of r^{-n}. We proceed as:

$$\begin{aligned}
\boldsymbol{\nabla}\left(\frac{1}{r^n}\right) &= \hat{\mathbf{e}}_i \partial_i (r_j r_j)^{-n/2}, \quad \text{expressing } r = \sqrt{\mathbf{r} \cdot \mathbf{r}} \\
&= \hat{\mathbf{e}}_i \left(-\frac{n}{2}\right)(r_j r_j)^{-(n/2+1)} \partial_i (r_k r_k), \quad \text{chain rule} \\
&= \hat{\mathbf{e}}_i \left(-\frac{n}{2}\right)(r_j r_j)^{-(n/2+1)} \cdot 2 r_k \delta_{ik}, \quad \text{product rule with Eq. 5.1} \\
&= -n (r_j r_j)^{-(n/2+1)} r_i \hat{\mathbf{e}}_i \\
&= -n r^{-(n+2)} \mathbf{r} \qquad (5.4) \\
&= -\frac{n}{r^{n+1}} \hat{\mathbf{r}} \qquad (5.5)
\end{aligned}$$

There are several subtleties here that we should pay attention to. In the first line, we note that the quantity r^n is defined as the n-th power of the norm of the position vector, so the natural way to express it would be to contract the position vector with itself and take the appropriate power of the product. When contracting the position vector, however, we must be careful to assign it a new dummy index since this is not to be confused with the index from the derivative operator. In the second line, we apply the chain rule to differentiate the entire expression in parentheses raised to the power $-n/2$, then apply the derivative operator to the argument of the parentheses. When doing so, we must also ascribe a new index to the quantity contained within the parentheses, since these are contracted to give a scalar. Furthermore, the first rule of index notation states that no index can appear on the same side of the equation more than twice, so keeping the same index j would result in its appearance four times, which violates this rule. In the third line, we apply the product rule when differentiating the contraction of the position vector, and also apply Eq. 5.1 to obtain a Kronecker delta. The remainder of the proof uses elementary operations from before, and the final expression resembles what we expect from intuition, that the derivative should lower the power of r by one. The result in Eq. 5.4 is also an equally useful way to express this, though the power of r may be off-putting to the untrained eye.

Finally, we are in a good position to perform a calculation pertaining to actual physics. As a demonstration, we will attempt to derive the coordinate-free expression of the electric field from a point dipole (Problem 3.36 in [4]). We begin with the coordinate-free expression for the potential (Eq. 3.99 in [4]) of a point dipole:

$$V_{\text{dip}} = \frac{1}{4\pi\epsilon_0}\frac{\hat{\mathbf{r}}\cdot\mathbf{p}}{r^2} \tag{5.6}$$

For the remainder of the derivation, the prefactor of $1/4\pi\epsilon_0$ is irrelevant so we will simply ignore it and insert it at the end to ensure the units make sense. To proceed, we consider the gradient of the potential:

$$\begin{aligned}
\boldsymbol{\nabla}\left(\frac{\hat{\mathbf{r}}\cdot\mathbf{p}}{r^2}\right) &= \boldsymbol{\nabla}\left(\frac{\mathbf{r}}{r^3}\cdot\mathbf{p}\right) \\
&= (\mathbf{p}\cdot\boldsymbol{\nabla})\frac{\mathbf{r}}{r^3} + \underbrace{\left(\frac{\mathbf{r}}{r^3}\cdot\boldsymbol{\nabla}\right)\mathbf{p}}_{\partial_i p_j = 0} + \frac{\mathbf{r}}{r^3}\times\underbrace{(\boldsymbol{\nabla}\times\mathbf{p})}_{\partial_i p_j = 0} + \mathbf{p}\times\underbrace{\left(\boldsymbol{\nabla}\times\frac{\mathbf{r}}{r^3}\right)}_{\varepsilon_{ijk}\partial_j r_k = 0} \\
&= p_i\partial_i\left[\frac{r_j}{(r_k r_k)^{3/2}}\right]\hat{\mathbf{e}}_j
\end{aligned} \tag{5.7}$$

In the second line, I have used the vector identity for the gradient of a dot product (yes, the only one that you cannot derive easily). Of the resulting four terms, three of them vanish. Since the dipole moment $\mathbf{p}$ has no functional dependence, any derivatives of it trivially vanish. For the last term, we note by expanding it in index notation that it is proportional to the curl of $\mathbf{r}$, which we have previously noted in Eq. 5.3 also vanishes. Thus, we are left with a single term to deal with. For this, we focus solely on the derivative and ignore the prefactor and unit vector, where we should note that in the following expression, (i, j) are a free indices so they must be conserved.

$$\begin{aligned}
\partial_i\left[\frac{r_j}{(r_k r_k)^{3/2}}\right] &= \frac{\partial_i r_j}{(r_k r_k)^{3/2}} + r_j\partial_i\frac{1}{r^3}, \quad \text{product rule} \\
&= \frac{1}{r^3}\delta_{ij} - \frac{3r_i r_j}{r^5}, \quad \text{Eq. 5.1, Eq. 5.4} \\
&= \frac{1}{r^3}\left(\delta_{ij} - \frac{3r_i r_j}{r^2}\right)
\end{aligned}$$

where I have dealt with the second term of the product rule using our expression from Eq. 5.4, keeping i as a free index by ignoring the unit vector. We now plug this into Eq. 5.7 and proceed:

$$\begin{aligned}
\boldsymbol{\nabla}\left(\frac{\hat{\mathbf{r}}\cdot\mathbf{p}}{r^2}\right) &= \frac{\hat{\mathbf{e}}_j}{r^3}\left(\delta_{ij}p_i - \frac{3r_i r_j p_i}{r^2}\right) \\
&= \frac{1}{r^3}\left(p_j\hat{\mathbf{e}}_j - \frac{3(\mathbf{p}\cdot\mathbf{r})r_j\hat{\mathbf{e}}_j}{r^2}\right), \quad \text{summing over } i \\
&= \frac{1}{r^3}[\mathbf{p} - 3(\mathbf{p}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}}]
\end{aligned} \tag{5.8}$$

Finally, we append the prefactors to obtain:

$$\mathbf{E}_{\text{dip}} = -\boldsymbol{\nabla}V_{\text{dip}} = \frac{1}{4\pi\epsilon_0 r^3}[3(\mathbf{p}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}} - \mathbf{p}] \tag{5.9}$$

In a similar fashion, one can obtain the coordinate-free expression for the magnetic field of a point magnetic dipole (Problem 5.34 in [4]), starting with its corresponding vector potential. The proof follows very similar logic to the one above, so I won't write it out here.

5.2 Quantum Mechanics

As a demonstration of the versatility of the index notation, let's also do a simple proof from quantum mechanics. Here, we will derive the fundamental angular momentum commutation relation $[L_i, L_j] = i\hbar\varepsilon_{ijk}L_k$, assuming only knowledge of the canonical commutation relation $[r_i, p_j] = i\hbar\delta_{ij}$.

$$
\begin{aligned}
[L_i, L_j] &= [\varepsilon_{ikl}r_k p_l, \varepsilon_{jmn}r_m p_n], \quad \text{using the definition } \mathbf{L} = \mathbf{r} \times \mathbf{p} \\
&= \varepsilon_{ikl}\varepsilon_{jmn}[r_k p_l, r_m p_n], \quad \text{Levi-Civitae are scalars} \\
&= \varepsilon_{ikl}\varepsilon_{jmn}(r_k[p_l, r_m p_n] + [r_k, r_m p_n]p_l), \quad \text{using } [AB, C] = A[B, C] + [A, C]B \\
&= \varepsilon_{ikl}\varepsilon_{jmn}\left(r_k r_m \underbrace{[p_l, p_n]}_{=0} + r_k[p_l, r_m]p_n + r_m[r_k, p_n]p_l + \underbrace{[r_k, r_m]}_{=0} p_l p_n\right) \\
&= i\hbar\varepsilon_{ikl}\varepsilon_{jmn}(r_m p_l \delta_{kn} - r_k p_n \delta_{lm}), \quad \text{using the canonical commutation relation} \\
&= i\hbar(\varepsilon_{lik}\varepsilon_{jmk}r_m p_l - \varepsilon_{ikl}\varepsilon_{njl}r_k p_n), \quad \text{summing over one } \delta \text{ index} \\
&= i\hbar[(\delta_{lj}\delta_{im} - \delta_{lm}\delta_{ij})r_m p_l - (\delta_{in}\delta_{jk} - \delta_{ij}\delta_{kn})r_k p_n], \quad \text{same same, then different} \\
&= i\hbar(r_i p_j - \delta_{ij}r_l p_l - r_j p_i + \delta_{ij}r_k p_k) \\
&= i\hbar(r_i p_j - r_j p_i), \quad \text{relabel } k \to l \text{ since it's a dummy index, then cancel} \\
&= i\hbar\varepsilon_{ijk}L_k, \quad \text{since this is the } k\text{-th component of the cross product}
\end{aligned} \tag{5.10}
$$

Thus, the commutation relation is proven. We can go a step further to have a look at a cool property of quantum mechanical angular momenta, by considering the following cross product:

$$\mathbf{L} \times \mathbf{L} = \varepsilon_{ijk}L_i L_j \hat{\mathbf{e}}_k = (L_i L_j - L_j L_i)\hat{\mathbf{e}}_k = [L_i, L_j]\hat{\mathbf{e}}_k \tag{5.11}$$

Thus, by multiplying the unit vector $\hat{\mathbf{e}}_k$ on both sides of Eq. 5.10 and removing the Levi-Civita symbol[9], we obtain:

$$\mathbf{L} \times \mathbf{L} = i\hbar\mathbf{L} \tag{5.12}$$

This is (in my opinion) one of the strangest things about quantum mechanics, that the cross product of a vector with itself can give a nonzero result, namely itself again. This truly illustrates the concept of why a quantum angular momentum vector is completely unlike a classical angular momentum vector, where the Heisenberg uncertainty relation accords just enough uncertainty in the position and momentum coordinates that a vector can behave like it is not perfectly aligned with itself, when treated in real position space.

[9] This is a horrendous abuse of notation, because we have violated two rules here. The first is that no index can appear on the same side of an equation more than twice. The second is that "removing the Levi-Civita" symbol is not even a mathematically sound operation. What I am really doing here is multiplying a second Levi-Civita onto both sides and totally contracting it with the first, then invoking Eq. 3.5, cancelling the common factor of 6 at the end.

6 A Word of Caution – Curvilinear Coordinates

In this final section, I'd like to elaborate on why the arguments presented here do not generalise well to curvilinear coordinate systems. Up till now, we have been working explicitly in Cartesian coordinates where the unit vectors have always been constant over the entire coordinate system. This is no longer the case in curvilinear coordinates, since the unit vectors in the various directions are now functions of the coordinates themselves. As an example, consider the orthonormal basis vectors in spherical coordinates [4]:

$$\begin{cases} \hat{\mathbf{r}} &= \sin\theta\cos\phi\hat{\mathbf{x}} + \sin\theta\sin\phi\hat{\mathbf{y}} + \cos\theta\hat{\mathbf{z}} \\ \hat{\theta} &= \cos\theta\cos\phi\hat{\mathbf{x}} + \cos\theta\sin\phi\hat{\mathbf{y}} - \sin\theta\hat{\mathbf{z}} \\ \hat{\phi} &= -\sin\phi\hat{\mathbf{x}} + \cos\phi\hat{\mathbf{y}} \end{cases} \tag{6.1}$$

Clearly, these vectors themselves are varying functions of the angles θ and ϕ. If these were to appear in a derivative, we can thus no longer pull them out of the derivative as we did before with the Cartesian basis vectors. Specifically, we observe that $\partial_i(f(x_i)\hat{\mathbf{e}}_i) \neq \hat{\mathbf{e}}_i\partial_i f(x_i)$ for some function f of the coordinates, since the unit vectors themselves have functional dependence. To deal with this, we need a more general view of coordinate systems, which is the ultimate starting point of the field of differential geometry. This will not be dealt with here, since the subject is rather definition-heavy upon introduction and may not be elucidative until the reader is more familiar with handling abstract mathematical objects such as manifolds. A thorough introduction to differential geometry is provided in [5], along with an incredible review of how physics fundamentally arises from the underlying geometric structure of the universe.

References

1. Einstein, A. Die Grundlage der allgemeinen Relativitätstheorie. **Annalen der Physik 354,** 769–822 (1916).
2. Hess, S. **Tensors for Physics** 440. ISBN: 9783319127873 (Springer London, Limited, 2015).
3. Rodríguez-Sánchez, P. **Visualization of a 3D Levi-Civita symbol** 2019. https://pabrod.github.io/levi-civita-en.html.
4. Griffiths, D. J. **Introduction to Electrodynamics** ISBN: 1108420419 (Cambridge University Press, June 2017).
5. Frankel, T. **The Geometry of Physics** 750 pp. ISBN: 1107602602 (Cambridge University Press, Mar. 2011).