Massachusetts Institute of Technology
Department of Physics

Physics 8.07 Fall 2004

Primer on Index Notation

©2003-2004 Edmund Bertschinger. All rights reserved.

1 Introduction

Equations involving vector fields — such as the Maxwell Equations — take a much simpler form if one uses vector notation. For example, $\vec{\nabla} \times \vec{E} = -\partial \vec{B}/\partial t = 0$ (Faraday's Law) is many times longer if written out using components. However, the vector notation is not ideal for all types of calculations with vector fields; for example, $\vec{\nabla} \times (\vec{\nabla} \times \vec{A})$ is not so easily parsed: how does one evaluate it given the components of $\vec{A}$? Moreover, the vector notation hides the subtleties of curvilinear coordinates — there are no coordinates or components at all in the vector expressions. How, then, does one compute $\vec{\nabla} \times (\vec{\nabla} \times \vec{A})$ or even $\vec{\nabla} \times \vec{A}$ if one has components of $\vec{A}$ in spherical coordinates? Index notation is introduced to help answer these questions and to simplify many other calculations with vectors.

In his presentation of relativity theory, Einstein introduced an index-based notation that has become widely used in physics. This notation is almost universally used in general relativity but it is also extremely useful in electromagnetism, where it is used in a simplified manner. These notes summarize the index notation and its use. For a look at the original usage, see Chapter 1 of *The Meaning of Relativity* by Albert Einstein (Princeton University Press, 1979). (Einstein introduces tensors early on; they are similar to vectors but have more indices. Except for a few special tensors that we will introduce as needed, we will not require or use the machinery of tensor calculus in 8.07.)

In these notes, vectors have arrows over the symbols.

2 Basis Vectors, Components, and Indices

The starting point for the index notation is the concept of a basis of vectors. A basis is a set of linearly independent vectors that span the vector space. Consider, for example, some vector $\vec{A}$ in three-dimensional space. The dimensionality of the vector space is

three, meaning that $\vec{A}$ can point in three independent directions (e.g., x, y, and z). Thus three independent vectors are necessary and sufficient to form a basis for $\vec{A}$:

$$\vec{A} = A_x\vec{e}_x + A_y\vec{e}_y + A_z\vec{e}_z \ . \tag{1}$$

It is very important to make a distinction between the **components** of a vector $\{A_x, A_y, A_z\}$ and the **basis vectors** $\{\vec{e}_x, \vec{e}_y, \vec{e}_z\}$. You may be used to thinking of vectors solely in terms of components. However, components only give half the story. Without the other half — the basis vectors — you cannot really understand how vectors are used in physics.

Sometimes names other than $\{\vec{e}_x, \vec{e}_y, \vec{e}_z\}$ are given to the Cartesian basis vectors. Griffiths, for example, uses $\hat{x}$, $\hat{y}$, $\hat{z}$, with the carets reminding us that these are unit vectors. However, in my opinion it is better to use one notation consistently for vectors, so in 8.07 I write all vectors with an arrow over them. In this way, you will never confuse a component with a basis vector.

Another common notation for the Cartesian basis vectors is $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$, where $\vec{e}_1 = \vec{e}_x$, $\vec{e}_2 = \vec{e}_y$, $\vec{e}_3 = \vec{e}_z$. This notation allows us to rewrite equation (1) using an *index* i:

$$\vec{A} = A_1\vec{e}_1 + A_2\vec{e}_2 + A_3\vec{e}_3 = \sum_{i=1}^{3} A_i\vec{e}_i \ . \tag{2}$$

Now, whenever we need to use components and basis vectors, we can either write out the sum with all three terms, or else use the shorter version with the summation symbol. Einstein found it tedious to write long expressions with lots of summation symbols, so he introduced a shorter form of the notation, by applying the following rule and a couple others that will follow later, which together comprise the **Einstein Summation Convention**:

Summation Convention Rule #1

Repeated, doubled indices in quantities multiplied together are implicitly summed.

According to the Einstein Summation Convention, equation (2) may be further abbreviated to

$$\vec{A} = A_i\vec{e}_i \ . \tag{3}$$

Doubled indices in quantities multiplied together are sometimes called **paired indices**.

If the writer's intent is *not* to have the repeated index summed over, then this must be made explicit. For example,

$$\vec{B} = B_i\vec{e}_i \ \ \text{(no sum on } i\text{)} \ . \tag{4}$$

Now, what exactly does this mean? Equation (4) makes sense under two conditions:

- The index i has a particular value, e.g. $\vec{B} = B_2\vec{e}_2$. In this case, the writer must indicate the value ($i = 2$).

- The equation holds true for *any* i.

In the second case, or when the index appears only once, the index i is called a **free index**: it is free to take any value, and the equation must hold for all values. The free index notation is most widely used to denote the equality of two vectors:

$$\vec{A} = \vec{B} \quad \Leftrightarrow \quad A_i = B_i \ . \tag{5}$$

Note that Summation Convention Rule #1 does not apply here (i.e., there is no sum on i) because A_i and B_i are not multiplied together.

Indices that are summed over are called **dummy indices**. Like integration variables in a definite integral, the names of dummy indices are arbitrary. Thus,

$$A_i\vec{e}_i = A_j\vec{e}_j \ . \tag{6}$$

This equation means $\sum_{i=1}^{3} A_i\vec{e}_i = \sum_{j=1}^{3} A_j\vec{e}_j$.

The Einstein Summation Convention is supplemented by the following rules:

Summation Convention Rule #2

Indices that are not summed over (free indices) are allowed to take all possible values unless stated otherwise.

Summation Convention Rule #3

It is illegal to use the same dummy index more than twice in a term unless its meaning is made explicit.

In Rule #3, the word **term** means a product of quantities having indices, like A_iB_j. In that example, i and j are free indices. In A_iB_i they are paired. In $A_iB_iC_i$ they are illegal (by Rule #3), unless it is stated explicitly how the indices are to be handled. So, $\sum_{i=1}^{3} A_iB_iC_i$ is legal.

These rules may seem arcane at first. However, there are only a few types of terms (and corresponding uses of indices) that are used, and you will quickly recognize in them common vector operations such as those discussed next.

3 Vector Operations: Linear Superposition, Dot and Cross Products

The common vector operations are easily represented using index notation. For example,

$$\vec{C} = a\vec{A} + b\vec{B} \quad \Leftrightarrow \quad C_i = aA_i + bB_i \ . \tag{7}$$

Here, a and b are constants (called **scalars** to distinguish them from vectors or vector components) and i is a free index. The dot product of two vectors using components is likewise easy to represent:

$$\vec{A} \cdot \vec{B} = A_iB_i \ . \tag{8}$$

There is, of course, an implied sum on i. Depending on one's preference for numbers or letters, that sum can be over $(1, 2, 3)$ or over (x, y, z). There is no reason that the value of an index has to be restricted to a number; symbols are equally valid. I often use x and 1 interchangeably as index values.

Equation (8) is sometimes taken as the *definition* of the dot product. However, this is unsatisfactory because the right-hand side presupposes some things about the basis vectors. To see this, let us write the dot product as a bilinear (i.e., distributive) operation between *vectors* rather than *components*:

$$\vec{A} \cdot \vec{B} = (A_i \vec{e}_i) \cdot (B_j \vec{e}_j) = A_i B_j (\vec{e}_i \cdot \vec{e}_j) = A_i B_i \ . \tag{9}$$

This equation is an excellent illustration of the index notation, so be sure you understand every step. By Rule #1, there is an implied sum on both i and j where they occur paired. By Rule #3, it is *mandatory* that different indices be used for the expansion of $\vec{A}$ and $\vec{B}$. Secondly, the dot product is *distributive*: $\vec{A} \cdot (b\vec{B} + c\vec{C}\,) = b(\vec{A} \cdot \vec{B}\,) + c(\vec{A} \cdot \vec{C}\,)$. The distributive law is used in the second equal sign in equation (9). In the third equal sign we have used a very important result for the unit-length Cartesian basis vectors $\vec{e}_i$:

Orthonormality Rule #1

The dot product of orthonormal basis vectors is zero unless the two vectors are identical, when it is one.

This rule is so important that it leads us to introduce a new symbol, δ_{ij}, called the **Kronecker delta**:

$$\vec{e}_i \cdot \vec{e}_j = \delta_{ij} \equiv \begin{cases} 1 \ , & i = j \\ 0 \ , & i \neq j \ . \end{cases} \tag{10}$$

Equation (10) implies (as you should check!)

$$\delta_{ij} B_j = B_i \ . \tag{11}$$

Note that here i is a free index and j is a dummy (i.e., paired) index. You should also verify that $\delta_{ij} B_i = B_j$, and show how these relations imply $A_i B_j (\vec{e}_i \cdot \vec{e}_j) = A_i B_i$.

The curl of two vectors is another bilinear operation on vectors, but it produces a vector (in three dimensions) rather than a number (i.e., scalar). (Actually, the curl produces an object called a pseudovector, which differs from a vector in how it behaves under an inversion of coordinates $\vec{x} \to -\vec{x}$, also known as a parity transformation. But the distinction between vectors and pseudovectors is a technicality of no significance at the moment.) Because the curl is a bilinear operation (i.e., linear in both vectors entering the curl), we may write the curl of two vectors as a sum over curls of the basis vectors, by analogy with equation (9):

$$\vec{A} \times \vec{B} = (A_i \vec{e}_i) \times (B_j \vec{e}_j) = A_i B_j (\vec{e}_i \times \vec{e}_j) \ . \tag{12}$$

To evaluate this expression, we need $\vec{e}_i \times \vec{e}_j$. The result is a vector. Any vector can be expanded in the basis vectors. So, we can write

$$\vec{e}_i \times \vec{e}_j = \epsilon_{ijk} \vec{e}_k \tag{13}$$

for some set of numbers ϵ_{ijk}. This equation may look very complicated, but it has much the same meaning as equation (3). Indeed, if one replaces $\vec{A}$ by $\vec{e}_i \times \vec{e}_j$, and uses k as the dummy index for expanding $\vec{A}$, then equation (13) is the *same* as equation (3). The only real difference is that i and j are free indices in equation (13), so there are actually $3 \times 3 = 9$ different vector equations. Thus, instead of 3 components A_k to be specified, there are $3 \times 9 = 27$ components ϵ_{ijk}.

What is ϵ_{ijk}? We can use the dot product to determine it from equation (13):

$$(\vec{e}_i \times \vec{e}_j) \cdot \vec{e}_k = \epsilon_{ijl}\vec{e}_l \cdot \vec{e}_k = \epsilon_{ijl}\delta_{lk} = \epsilon_{ijk} \ . \tag{14}$$

Note that we had to change the dummy index inside equation (13) from k to l when dotting with $\vec{e}_k$, otherwise the notation would have violated our rules (and been ambiguous! — we introduce rules to avoid ambiguity). So, we see that ϵ_{ijk} is the vector triple product of basis vectors. This quantity is called the **Levi-Civita symbol** and has values

$$\epsilon_{ijk} = \begin{cases} 0 \ , & \text{if } i = j \ , \\ +1 \ , & \text{if } (ijk) \text{ is an even permutation of } (123) \ , \\ -1 \ , & \text{if } (ijk) \text{ is an odd permutation of } (123) \ . \end{cases} \tag{15}$$

(Here, xyz may be used in place of 123 with the usual identifications $x \leftrightarrow 1$, $y \leftrightarrow 2$, $z \leftrightarrow 3$.) A permutation of (123) is defined to be a rearrangement of them obtained by exchanging elements of the set. Even permutations have an even number of exchanges; odd permutations have an odd number. For example, (213) is an odd permutation of (123) but (231) is an even permutation. The only other distinct even permutation is (312). The even permutations (231) and (312) are often called **cyclic permutations** of (123) because they are obtained by rolling around in a cycle like links on a bicycle chain. There are three odd permutations of (123): (213), (132), and (321). Thus, the Levi-Civita symbol is zero aside from 6 terms.

Bringing it all together, we can write the cross product of two vectors as

$$\vec{A} \times \vec{B} = \epsilon_{ijk} A_i B_j \vec{e}_k \ . \tag{16}$$

Although there is an implied sum on all three indices, there are only 6 nonzero terms in total.

4 Partial Derivatives

Until now, we didn't have to specify whether the vector operations were being done for single vectors or for *vector fields*. Derivatives imply the concept of a field, i.e. a quantity associated with each point in space. Because the laws of electrodynamics relate vectors at different points in space through derivatives, we will need to consider vector fields and their derivatives.

First we recall the definition of partial derivatives for scalar fields. A scalar field $f(\vec{x})$ is a rule that associates a number (we always assume a real number, unless stated

otherwise) with each point in space (denoted by its position vector $\vec{x}$). I assume that the reader is well versed with the partial derivative, for example $\partial f/\partial x$. (In this derivative, all variables *except* the one in the denominator are held fixed.) Using our index notation, we can regard f as a function of the three coordinates x_i with

$$x_1 \equiv x \ , \quad x_2 \equiv y \ , \quad x_3 \equiv z \ . \tag{17}$$

Then the partial derivative with respect to an arbitrary coordinate is $\partial f/\partial x_i$. I prefer to shorten the notation further, in a way that is unambiguous and simplifies the index tracking:

$$\frac{\partial f}{\partial x_i} \equiv \partial_i f \ . \tag{18}$$

Partial derivatives obey the distributive and Leibnitz (product) rules, for example, $\partial_i(fg) = g\partial_i f + f\partial_i g$. Can we use this to work out the partial derivative of a vector field expanded in components and basis vectors as in equation (3)? Yes! But first we must define the concept of the derivative of a basis vector. In other words, what is $\partial_i \vec{e}_j$?

Before differentiating a basis vector, we emphasize that we are now broadening our concept of vectors to associate a vector with every point in space. This applies also to the basis vectors! Thus, $\vec{e}_i = \vec{e}_i(\vec{x})$. Partial derivatives are now defined by differencing these vectors at adjacent points of space. For example,

$$\partial_1 \vec{e}_i = \lim_{\Delta x_1 \to 0} \left[\frac{\vec{e}_i(x_1 + \Delta x_1, x_2, x_3) - \vec{e}_i(x_1, x_2, x_3)}{\Delta x_1} \right] \ . \tag{19}$$

The usual rules for derivatives apply to functions that evaluate to numbers, but now we want to differentiate vectors. This can be a subtle business, and a rigorous analysis requires introducing rules that convert vectors to numbers before the differentiation takes place. The branch of mathematics called differential geometry deals with these issues. Here we will take a more intuitive approach that is completely valid and easy to understand in Euclidean space but avoids the mathematical formalities.

We assume that space is Euclidean and that vectors can be translated in the usual way from one point to another so as to be subtracted. For example, the vector $\vec{e}_x$ at $(x, y, z) = (0, 0, 0)$ is identical to the vector $\vec{e}_x$ at $(x, y, z) = (-1, 3, 2)$: their difference is zero. If we compare *any* of the Cartesian basis vectors at neighboring points of space, we find that $\vec{e}_i$ at $\vec{x} + \vec{dx}$ equals $\vec{e}_i$ at $\vec{x}$ for *any* $d\vec{x}$. Therefore all differences like those in equation (19) vanish, implying

$$\partial_i \vec{e}_j = 0 \quad \text{for Cartesian orthonormal basis vectors } \{\vec{e}_x, \vec{e}_y, \vec{e}_z\} \ . \tag{20}$$

This is a very important result and is the starting point for all derivative operations on vectors.

Using equation (20) and the Leibnitz rule for partial derivatives, we see at once how to differentiate a vector field:

$$\partial_i \vec{A} = (\partial_i A_j)\vec{e}_j \ . \tag{21}$$

5 Gradient, Divergence, Curl

One can not only differentiate vectors; one can associate a vector operator with partial derivatives themselves, the **gradient**:

$$\vec{\nabla} \equiv \vec{e}_i \partial_i \text{ in Cartesian coordinates .} \tag{22}$$

This vector operator can be used in the same ways as a vector: it can multiply (act upon) a scalar field, it can be dotted into a vector field, or its cross product with a vector field can be taken. The results are called gradient, divergence, and curl. They are obtained simply by applying $\vec{\nabla}$ like a vector, using index notation to represent the vector operations. The gradient of a scalar field $f(\vec{x}\,)$ is

$$\vec{\nabla} f = (\partial_i f)\vec{e}_i \ . \tag{23}$$

The divergence of a vector field $\vec{v}(\vec{x}\,)$ is

$$\vec{\nabla} \cdot \vec{v} = (\vec{e}_i \partial_i) \cdot (v_j \vec{e}_j) = (\partial_i v_j)(\vec{e}_i \cdot \vec{e}_j) = \partial_i v_i \ , \tag{24}$$

where we have used equations (10) and (11). The curl of $\vec{v}(\vec{x}\,)$ is

$$\vec{\nabla} \times \vec{v} = (\vec{e}_i \partial_i) \times (v_j \vec{e}_j) = (\partial_i v_j)(\vec{e}_i \times \vec{e}_j) = \epsilon_{ijk}(\partial_i v_j)\vec{e}_k \ , \tag{25}$$

where we have used equation (13). Note carefully the pairs of summed indices: equations (23)–(25) have no free indices.

An important warning is now needed: **Equations (22)–(25) do not hold in curvilinear coordinates**. Since curvilinear coordinates (e.g. spherical polar coordinates) are very important in electromagnetism, we must generalize the gradient operator — and all vector operations — beyond Cartesian coordinates, and we must do so in a way that is unambiguous in the notation.

6 Curvilinear Coordinates

Cartesian coordinates have the property that they measure distance. This is not true in curvilinear coordinates. For example, the spherical polar coordinates (r, θ, ϕ) do not even all have units of length. Thus, the partial derivatives with respect to polar coordinates mean something very different from the partial derivatives with respect to Cartesian coordinates. This means that equation (22) will not necessarily hold in curvilinear coordinates.

The vector calculus results of the previous section also rely on the basis vector fields. Herein lies the second important difference between curvilinear coordinates and Cartesian coordinates: the basis vectors associated with curvilinear coordinates are not constant. Equation (20) does not hold in curvilinear coordinates, as we will show.

However, the contents of Sections 1-3 holds equally for curvilinear and Cartesian coordinates.

To proceed further, we have to define the basis vectors for curvilinear coordinates. There are a variety of ways to do this. We follow the standard approach and define the basis vectors as follows:

Basis Vector Rule

The basis vectors are a set orthonormal vectors pointing in the directions of increasing coordinate values.

With three coordinates, say (r, θ, ϕ), there are three corresponding basis vectors $(\vec{e}_r, \vec{e}_\theta, \vec{e}_\phi)$. This definition applies, of course, to the Cartesian basis vectors as well. What distinguishes the Cartesian basis vectors from curvilinear basis vectors is that in the latter case the basis vectors depend on position: $\vec{e}_r(\vec{x}\,)$ points in different directions for the points with coordinates $(x, y, z) = (1, 0, 0)$ and $(x, y, z) = (0, 0, 1)$.

Curvilinear basis vectors are so different from Cartesian ones that one ought to use a different index notation for them. For this reason, I recommend writing vector expansions in curvilinear coordinates using indices chosen from the set (a, b, c, d) rather than (i, j, k, l). (Rarely are more than 4 distinct indices needed in an equation.) For example, we can restate the key results of Sections 1–5 using curvilinear indices. For example,

$$\vec{A} = A_a \vec{e}_a \tag{26}$$

and

$$\vec{e}_a \cdot \vec{e}_b = \delta_{ab} \ . \tag{27}$$

The is last result is very important: *In 8.07 we always work with orthonormal basis vectors.* It follows that we can order our basis vectors (as a right-handed triad) so that

$$(\vec{e}_a \times \vec{e}_b) \cdot \vec{e}_c = \epsilon_{abc} \tag{28}$$

where ϵ_{abc} is the same Levi-Civita symbol we introduced before. The only difference is that now, instead of associating $(1, 2, 3)$ with (x, y, z), we associate them with (r, θ, ϕ) or whatever coordinates we have.

The dot product and curl of vectors in curvilinear coordinates follow simply from equations (26)–(28):

$$\vec{A} \cdot \vec{B} = A_a B_a \ , \quad \vec{A} \times \vec{B} = \epsilon_{abc} A_a B_b \vec{e}_c \ . \tag{29}$$

Now, since a vector is independent of the basis used to represent it (just like state vectors in quantum mechanics, if you are taking 8.05), we must be able to express A_a in terms of Cartesian components and vice versa. These rules follow from applying equation (10) or (27) to equation (3) or (26):

$$A_a \equiv \vec{e}_a \cdot \vec{A} = (\vec{e}_a \cdot \vec{e}_i) A_i \ . \tag{30}$$

You should recognize this as equivalent to matrix multiplication: the numbers A_i can be formed into a column "vector", then multiplied by a matrix whose entries are $\vec{e}_a \cdot \vec{e}_i$, yielding a column "vector" with entries A_a. You should verify that summing over the dummy index i in equation (30) is equivalent to multiplication of a matrix and column vector. (I use quotation marks to make a distinction between vectors, which have arrows over the symbols, and column vectors, which do not. This distinction is important because we can also make a column vector whose elements are the basis vectors themselves!)

Note that the appearance of a transformation matrix is not unique to curvilinear coordinates: it occurs whenever we change from one basis to another. For example, if we rotated the Cartesian coordinate system, so that new coordinates $x_{\bar{i}}$ are given in terms of old ones by the rule $x_{\bar{i}} = R_{\bar{i}i}x_i$, then $A_{\bar{i}} = R_{\bar{i}i}x_i$. To maintain orthonormality and preserve lengths, $R_{\bar{i}i}$ must be an orthogonal matrix. The only difference with curvilinear coordinates is that $R_{\bar{i}i}$ is replaced by $R_{ai} \equiv \vec{e}_a \cdot \vec{e}_i$.

7 Vector Calculus in Curvilinear Coordinates

Vector calculus is where we see the real differences between Cartesian and curvilinear coordinates. As mentioned at the beginning of Section 6, equations (20) and (22) no longer hold in general. We must replace them. How?

First, we want $\vec{\nabla} f$ to be a vector field just like $\vec{A}(\vec{x})$. A vector field must be (at each point in space) independent of the basis used to represent it: $\vec{A} = A_i\vec{e}_i = A_a\vec{e}_a$. Thus, we must be able to write

$$\vec{\nabla} f = (\nabla f)_a \vec{e}_a \ , \tag{31}$$

where $(\nabla f)_a$ are *the components of the gradient in curvilinear coordinates*. How do we find them?

There are at lease two ways to work out the gradient. The first is a brute force approach based on the idea of a coordinate transformation. From equations (23), (27) and (31),

$$(\nabla f)_a = \vec{e}_a \cdot \vec{\nabla} f = (\vec{e}_a \cdot \vec{e}_i)\frac{\partial f}{\partial x_i} = (\vec{e}_a \cdot \vec{e}_i)\left(\frac{\partial x_b}{\partial x_i}\right)\frac{\partial f}{\partial x_b} \ . \tag{32}$$

Notice how easy the index notation makes this kind of manipulation even though there are two implied summations in the last expression. Note also that we are regarding the new coordinates x_b (e.g., r, θ, ϕ) as functions of the old coordinates x_i (e.g., x, y, z) and computing the Jacobian matrix $\partial x_b/\partial x_i$. Interestingly, although the Jacobian matrix tells us how to transform partial derivatives, it does not tell us how to transform vector components — we also need the matrix $\vec{e}_a \cdot \vec{e}_i$. (Equation 32 can easily be seen to be equivalent to two matrices and a column vector all being multiplied.)

Given explicit formulae for the coordinate transformation (e.g. $r = \sqrt{x^2 + y^2 + z^2}$) and the dot products of the basis vectors (e.g., $\vec{e}_r \cdot \vec{e}_x = \sin\theta\cos\phi$), it is possible to work

out the components of the gradient using equation (32). However, it is a lot of work. A shorter method is possible if we bring some knowledge of geometry to the task. The differential displacement vector between two nearby points $\vec{x}$ and $\vec{x} + d\vec{x}$ can be written, for orthogonal coordinates,

$$d\vec{x} = \sum_a h_a (dx_a) \vec{e}_a \ , \tag{33}$$

where the h_a's are functions that tell how to convert coordinate differentials to physical lengths. (Note the use of the summation symbol as required by Summation Convention Rule #3.) For example, with spherical polar coordinates, $h_r = 1$, $h_\theta = r$, and $h_\phi = r \sin\theta$: $d\vec{x} = (dr)\vec{e}_r + (r d\theta)\vec{e}_\theta + (r \sin\theta d\phi)\vec{e}_\phi$. This can be derived along lines similar to equation (32), but for coordinates as simple as spherical polar it is easier to appeal to simple geometry as Griffiths does in Section 1.4.1.

Combining equations (31) and (33), we see

$$df \equiv f(\vec{x} + d\vec{x}) - f(\vec{x}) = (\vec{\nabla} f) \cdot d\vec{x} = \sum_a (\nabla f)_a h_a dx_a \ . \tag{34}$$

But from multivariable calculus we have

$$df = \frac{\partial f}{\partial x_a} dx_a \equiv (\partial_a f) dx_a \ . \tag{35}$$

Comparing these two results, we get the components of the gradient:

$$(\nabla f)_a = h_a^{-1} \partial_a f \quad \text{(no sum on } a\text{)} \ . \tag{36}$$

From this, we may write the gradient operator in curvilinear coordinates:

$$\vec{\nabla} = \sum_a \vec{e}_a h_a^{-1} \partial_a \ . \tag{37}$$

As advertised, it differs in general from equation (22).

Expressions for the divergence and curl in curvilinear coordinates follow almost straightforwardly by combining equation (37) with equations (26) and (29). The one complication is that the partial derivative operator, which we have carefully written to the right of $\vec{e}_a$ and h_a^{-1} in equation (37), acts upon the basis vectors as well as components in a vector $\vec{v} = v_a \vec{e}_a$. Now we have

$$\partial_a \vec{e}_b \neq 0 \quad \text{for curvilinear basis vectors} \ . \tag{38}$$

The divergence is

$$\vec{\nabla} \cdot \vec{v} = \sum_a (\vec{e}_a h_a^{-1} \partial_a) \cdot (v_b \vec{e}_b) = \sum_a h_a^{-1} [(\partial_a v_a) + \vec{e}_a \cdot (\partial_a \vec{e}_b) v_b] \ , \tag{39}$$

where there is an implied sum on b in the terms with doubled b. The sum on a must be made explicit because of Summation Convention Rule #3. An expression for the curl follows similarly:

$$\begin{aligned}\vec{\nabla} \times \vec{v} &= \sum_a (\vec{e}_a h_a^{-1} \partial_a) \times (v_b \vec{e}_b) = \sum_a h_a^{-1} [(\partial_a v_b)(\vec{e}_a \times \vec{e}_b) + \vec{e}_a \times (\partial_a \vec{e}_b) v_b] \\ &= \sum_a h_a^{-1} [\epsilon_{abc} (\partial_a v_b) \vec{e}_c + \vec{e}_a \times (\partial_a \vec{e}_b) v_b] \ . \end{aligned} \tag{40}$$

We cannot fully evaluate equations (39) and (40) until expressions for $\partial_a \vec{e}_b$ are known. Problem Set 1 leads you through a calculation of them by writing $\vec{e}_a$ as a linear combination of the Cartesian basis vectors $\vec{e}_i$ and then differentiating.

8 Differences in General Relativity

As mentioned at the beginning of these notes, the index notation gets extensive use in general relativity, which is a classical field theory similar to electromagnetism in many respects. If you study general relativity, please beware the following differences of notation.

In general relativity, the basis is usually not taken to be orthonormal. It is never assumed to be Cartesian (at least not everywhere), because Cartesian coordinates do not exist in curved spaces (and GR is a theory of curved spacetime). So, equations (10) and (27) do not usually hold, and this means that one has to call the dot product of basis vectors something else and keep track of it. Usually this dot product is called the *metric tensor* and denoted g_{ab}: $\vec{e}_a \cdot \vec{e}_b = g_{ab}$. One consequence of $g_{ab} \neq \delta_{ab}$ is that the index notation becomes very cumbersome unless one makes a distinction between vectors and related objects called dual vectors. (Quantum mechanics has the same distinction, where the objects are called Dirac bra and ket vectors.) To distinguish components of vectors from components of dual vectors, in GR one places indices either as subscripts or as superscripts. (By contrast, we use exclusively subscripts in 8.07.) The Einstein Summation Convention rules are slightly modified as a result. In general relativity, paired indices must be one subscript and one superscript. For example, the scalar product of a vector and dual vector is written $A^a B_a$. Expressions like $A_a B_a$ are illegal in general relativity.

The experience you gain with index notation in 8.07 will be very helpful if you take a course in general relativity or string theory.