

Tensors

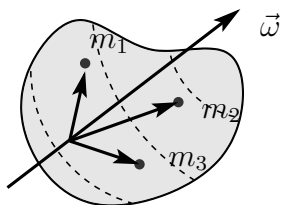
You can't walk across a room without using a tensor (the pressure tensor). You can't align the wheels on your car without using a tensor (the inertia tensor). You definitely can't understand Einstein's theory of gravity without using tensors (many of them).

This subject is often presented in the same language in which it was invented in the 1890's, expressing it in terms of transformations of coordinates and saturating it with such formidable-looking combinations as $\partial x^i / \partial \bar{x}^j$. This is really a sideshow to the subject, one that I will steer around, though a connection to this aspect appears in section 12.8.

Some of this material overlaps that of chapter 7, but I will extend it in a different direction. The first examples will then be familiar.

12.1 Examples

A tensor is a particular type of function. Before presenting the definition, some examples will clarify what I mean. Start with a rotating rigid body, and compute its angular momentum. Pick an origin and assume that the body is made up of N point masses m_i at positions described by the vectors \vec{r}_i ($i = 1, 2, \dots, N$). The angular velocity vector is $\vec{\omega}$. For each mass the angular momentum is $\vec{r}_i \times \vec{p}_i = \vec{r}_i \times (m_i \vec{v}_i)$. The velocity \vec{v}_i is given by $\vec{\omega} \times \vec{r}_i$ and so the angular momentum of the i^{th} particle is $m_i \vec{r}_i \times (\vec{\omega} \times \vec{r}_i)$. The total angular momentum is therefore



$$\vec{L} = \sum_{i=1}^N m_i \vec{r}_i \times (\vec{\omega} \times \vec{r}_i). \quad (12.1)$$

The angular momentum, \vec{L} , will depend on the distribution of mass within the body and upon the angular velocity. Write this as

$$\vec{L} = I(\vec{\omega}),$$

where the function I is called the tensor of inertia.

For a second example, take a system consisting of a mass suspended by six springs. At equilibrium the springs are perpendicular to each other. If now a (small) force \vec{F} is applied to the mass it will undergo a displacement \vec{d} . Clearly, if \vec{F} is along the direction of any of the springs then the displacement \vec{d} will be in the same direction as \vec{F} . Suppose however that \vec{F} is halfway between the k_1 and k_2 springs, and further that the spring k_2 was taken from a railroad locomotive while k_1 is a watch spring. Obviously in this case \vec{d} will be mostly in the x direction (k_1) and is not aligned with \vec{F} . In any case there is a relation between \vec{d} and \vec{F} ,

$$\vec{d} = f(\vec{F}). \quad (12.2)$$

The function f is a tensor.

In both of these examples, the functions involved were *vector valued functions of vector variables*. They have the further property that they are linear functions, *i.e.* if α and β are real numbers,

$$I(\alpha \vec{\omega}_1 + \beta \vec{\omega}_2) = \alpha I(\vec{\omega}_1) + \beta I(\vec{\omega}_2), \quad f(\alpha \vec{F}_1 + \beta \vec{F}_2) = \alpha f(\vec{F}_1) + \beta f(\vec{F}_2)$$

These two properties are the first *definition* of a tensor. (A generalization will come later.) There's a point here that will probably cause some confusion. Notice that in the equation $\vec{L} = I(\vec{\omega})$,

the tensor is the function I . I didn't refer to "the function $I(\vec{\omega})$ " as you commonly see. The reason is that $I(\vec{\omega})$, which equals \vec{L} , is a vector, not a tensor. It is the output of the function I after the independent variable $\vec{\omega}$ has been fed into it. For an analogy, retreat to the case of a real valued function of a real variable. In common language, you would look at the equation $y = f(x)$ and say that $f(x)$ is a function, but it's better to say that f is a function, and that $f(x)$ is the single number obtained by feeding the number x to f in order to obtain the number $f(x)$. In this language, f is regarded as containing a vast amount of information, all the relations between x and y . $f(x)$ however is just a single number. Think of f as the whole graph of the function and $f(x)$ as telling you one point on the graph. This apparently trivial distinction will often make no difference, but there are a number of cases (particularly here) where a misunderstanding of this point will cause confusion.

Definition of "Function"

This is a situation in which a very abstract definition of an idea will allow you to understand some fairly concrete applications far more easily.

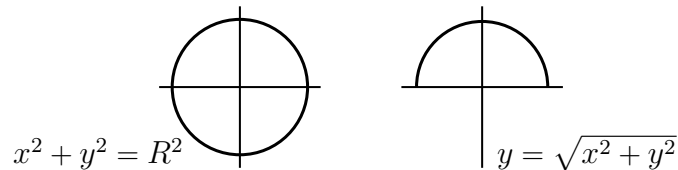
Let X and Y denote sets (possibly the same set) and x and y are elements of these sets ($x \in X$, $y \in Y$). Form a new set F consisting of some collection of ordered pairs of elements, one from X and one from Y . That is, a typical element of the set F is (x_1, y_1) where $x_1 \in X$ and $y_1 \in Y$. Such a set is called a "relation" between X and Y .

If X is the set of real numbers and Y is the set of complex numbers, examples of relations are the sets

$$F_1 = \{(1.0, 7.3 - 2.1i), (-\pi, e + i\sqrt{2}), (3.2 \text{ googol}, 0. + 0.i), (1.0, e - i\pi)\}$$

$$F_2 = \{(x, z) \mid z^2 = 1 - x^2 \text{ and } -2 < x < 1\}$$

There are four elements in the first of these relations and an infinite number in the second. A relation is not necessarily a function, as you need one more restriction. To define a function, you need it to be single-valued. That is the requirement that if $(x, y_1) \in F$ and $(x, y_2) \in F$ then $y_1 = y_2$. The ordinary notation for a function is $y = F(x)$, and in the language of sets we say $(x, y) \in F$. The set F is the function. You can picture it as a graph, containing all the information about the function; it is by definition single-valued. You can check that F_1 above is a function and F_2 is not.



For the real numbers x and y , $x^2 + y^2 = R^2$ defines a *relation* between X and Y , but $y = \sqrt{R^2 - x^2}$ is a *function*. In the former case for each x in the interval $-R < x < R$ you have two y 's, $\pm\sqrt{R^2 - x^2}$. In the latter case there is only one y for each x . The domain of a function is the set of elements x such that there is a y with $(x, y) \in F$. The range is the set of y such that there is an x with $(x, y) \in F$. For example, $-R \leq x \leq R$ is the domain of $y = \sqrt{R^2 - x^2}$ and $0 \leq y \leq R$ is its range.

Equation (12.1) defines a function I . The set X is the set of angular velocity vectors, and the set Y is the set of angular momentum vectors. For each of the former you have exactly one of the latter. The *function* is the set of all the pairs of input and output variables, so you can see why I don't want to call $I(\vec{\omega})$ a function — it's a vector, \vec{L} .

Another physical example of a tensor is the polarizability tensor relating the electric dipole moment density vector \vec{P} of matter to an applied electric field vector \vec{E} :

$$\vec{P} = \alpha(\vec{E})$$

For the vacuum this is zero. More generally, for an isotropic linear medium, this function is nothing more than multiplication by a scalar,

$$\vec{P} = \alpha \vec{E}$$

In a crystal however the two fields \vec{P} and \vec{E} are not in the same direction, though the relation between them is still linear for small fields. This is analogous to the case above with a particle attached to a set of springs. The electric field polarizes the crystal more easily in some directions than in others.

The stress-strain relation in a crystal is a more complex situation that can also be described in terms of tensors. When a stress is applied, the crystal will distort slightly and this relation of strain to stress is, for small stress, a linear one. You will be able to use the notion of a tensor to describe what happens. In order to do this however it will be necessary to expand the notion of “tensor” to include a larger class of functions. This generalization will require some preliminary mathematics.

Functional

Terminology: A *functional* is a real (scalar) valued function of one or more vector variables. In particular, a linear functional is a function of one vector variable satisfying the linearity requirement.

$$f(\alpha \vec{v}_1 + \beta \vec{v}_2) = \alpha f(\vec{v}_1) + \beta f(\vec{v}_2). \quad (12.3)$$

A simple example of such a functional is

$$f(\vec{v}) = \vec{A} \cdot \vec{v}, \quad (12.4)$$

where \vec{A} is a fixed vector. In fact, because of the existence of a scalar product, all linear functionals are of this form, a result that is embodied in the following theorem, the representation theorem for linear functionals in finite dimensions.

Let f be a linear functional: that is, f is a scalar valued function of one vector variable and is linear in that variable, $f(\vec{v})$ is a real number and

$$f(\alpha \vec{v}_1 + \beta \vec{v}_2) = \alpha f(\vec{v}_1) + \beta f(\vec{v}_2) \quad \text{then} \quad (12.5)$$

there is a unique vector, \vec{A} , such that $f(\vec{v}) = \vec{A} \cdot \vec{v}$ for all \vec{v} .

Now obviously the function defined by $\vec{A} \cdot \vec{v}$, where \vec{A} is a fixed vector, is a linear. The burden of this theorem is that all linear functionals are of precisely this form.

There are various approaches to proving this. The simplest is to write the vectors in components and to compute with those. There is also a more purely geometric method that avoids using components. The latter may be more satisfying, but it's harder. I'll pick the easy way.

To show that $f(\vec{v})$ can always be written as $\vec{A} \cdot \vec{v}$, I have to construct the vector \vec{A} . If this is to work it has to work for all vectors \vec{v} , and that means that it has to work for every particular vector such as \hat{x} , \hat{y} , and \hat{z} . It must be that

$$f(\hat{x}) = \vec{A} \cdot \hat{x} \quad \text{and} \quad f(\hat{y}) = \vec{A} \cdot \hat{y} \quad \text{and} \quad f(\hat{z}) = \vec{A} \cdot \hat{z}$$

The right side of these equations are just the three components of the vector \vec{A} , so if this theorem is to be true the only way possible is that its components have the values

$$A_x = f(\hat{x}) \quad \text{and} \quad A_y = f(\hat{y}) \quad \text{and} \quad A_z = f(\hat{z}) \quad (12.6)$$

Now to find if the vectors with these components does the job.

$$f(\vec{v}) = f(v_x \hat{x} + v_y \hat{y} + v_z \hat{z}) = v_x f(\hat{x}) + v_y f(\hat{y}) + v_z f(\hat{z})$$

This is simply the property of linearity, Eq. (12.3). Now use the proposed values of the components from the preceding equation and this is exactly what's needed: $A_x v_x + A_y v_y + A_z v_z = \vec{A} \cdot \vec{v}$.

Multilinear Functionals

Functionals can be generalized to more than one variable. A bilinear functional is a scalar valued function of two vector variables, linear in each

$$\begin{aligned} T(\vec{v}_1, \vec{v}_2) &= \text{a scalar} \\ T(\alpha\vec{v}_1 + \beta\vec{v}_2, \vec{v}_3) &= \alpha T(\vec{v}_1, \vec{v}_3) + \beta T(\vec{v}_2, \vec{v}_3) \\ T(\vec{v}_1, \alpha\vec{v}_2 + \beta\vec{v}_3) &= \alpha T(\vec{v}_1, \vec{v}_2) + \beta T(\vec{v}_1, \vec{v}_3) \end{aligned} \quad (12.7)$$

Similarly for multilinear functionals, with as many arguments as you want.

Now apply the representation theorem for functionals to the subject of tensors. Start with a bilinear functional so that ${}^0_2T(\vec{v}_1, \vec{v}_2)$ is a scalar. This function of two variables can be looked on as a function of one variable by holding the other one temporarily fixed. Say \vec{v}_2 is held fixed, then ${}^0_2T(\vec{v}_1, \vec{v}_2)$ defines a linear functional on the variable \vec{v}_1 . Apply the representation theorem now and the result is

$${}^0_2T(\vec{v}_1, \vec{v}_2) = \vec{v}_1 \cdot \vec{A}$$

The vector \vec{A} however will depend (linearly) on the choice of \vec{v}_2 . It defines a new function that I'll call 1_1T

$$\vec{A} = {}^1_1T(\vec{v}_2) \quad (12.8)$$

This defines a tensor 1_1T , a linear, vector-valued function of a vector. That is, starting from a bilinear functional you can construct a linear vector-valued function. The reverse of this statement is easy to see because if you start with ${}^1_1T(\vec{u})$ you can define a new function of two variables ${}^0_2T(\vec{w}, \vec{u}) = \vec{w} \cdot {}^1_1T(\vec{u})$, and this is a bilinear functional, the same one you started with in fact.

With this close association between the two concepts it is natural to extend the definition of a tensor to include bilinear functionals. To be precise, I used a different name for the vector-valued function of one vector variable (1_1T) and for the scalar-valued function of two vector variables (0_2T). This is overly fussy, and it's common practice to use the same symbol (T) for both, with the hope that the context will make clear which one you actually mean. In fact it is so fussy that I will stop doing it. The *rank* of the tensor in either case is the sum of the number of vectors involved, two ($= 1 + 1 = 0 + 2$) in this case.

The next extension of the definition follows naturally from the previous reformulation. A tensor of n^{th} rank is an n -linear functional, or any one of the several types of functions that can be constructed from it by the preceding argument. The meaning and significance of the last statement should become clear a little later. In order to clarify the meaning of this terminology, some physical examples are in order. The tensor of inertia was mentioned before:

$$\vec{L} = I(\vec{\omega}).$$

The dielectric tensor related \vec{D} and \vec{E} :

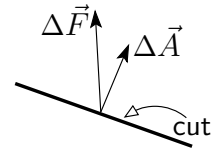
$$\vec{D} = \varepsilon(\vec{E})$$

The conductivity tensor relates current to the electric field:

$$\vec{j} = \sigma(\vec{E})$$

In general this is not just a scalar factor, and for the a.c. case σ is a function of frequency.

The stress tensor in matter is defined as follows: If a body has forces on it (compression or twisting or the like) or even internal defects arising from its formation, one part of the body will exert a force on another part. This can be made precise by the following device: Imagine making a cut in the material, then because of the internal forces, the two parts will tend to move with respect to each other. Apply enough force to prevent this motion. Call it $\Delta\vec{F}$. Typically for small cuts $\Delta\vec{F}$ will be proportional to the area of the cut. The area vector is perpendicular to the cut and of magnitude equal to the area. For small areas you have differential relation $d\vec{F} = S(d\vec{A})$. This function S is called the stress tensor or pressure tensor. If you did problem 8.11 you saw a two dimensional special case of this, though in that case it was isotropic, leading to a scalar for the stress (also called the tension).



There is another second rank tensor called the strain tensor. I described it qualitatively in section 9.2 and I'll simply add here that it is a second rank tensor. When you apply *stress* to a solid body it will develop *strain*. This defines a function with a second rank tensor as input and a second rank tensor as output. It is the elasticity tensor and it has rank four.

So far, all the physically defined tensors except elasticity have been vector-valued functions of vector variables, and I haven't used the n -linear functional idea directly. However there is a very simple example of such a tensor:

$$\text{work} = \vec{F} \cdot \vec{d}$$

This is a scalar valued function of the two vectors \vec{F} and \vec{d} . This is of course true for the scalar product of any two vectors \vec{a} and \vec{b}

$$g(\vec{a}, \vec{b}) = \vec{a} \cdot \vec{b} \quad (12.9)$$

g is a bilinear functional called the metric tensor. There are many other physically defined tensors that you will encounter later. In addition I would like to emphasize that although the examples given here will be in three dimensions, the formalism developed will be applicable to any number of dimensions.

12.2 Components

Up to this point, all that I've done is to make some rather general statements about tensors and I've given no techniques for computing with them. That's the next step. I'll eventually develop the complete apparatus for computation in an arbitrary basis, but for the moment it's a little simpler to start out with the more common orthonormal basis vectors, and even there I'll stay with rectangular coordinates for a while. (Recall that an orthonormal basis is an independent set of orthogonal unit vectors, such as \hat{x} , \hat{y} , \hat{z} .) Some of this material was developed in chapter seven, but I'll duplicate some of it. Start off by examining a second rank tensor, viewed as a vector valued function

$$\vec{u} = T(\vec{v})$$

The vector \vec{v} can be written in terms of the three basis vectors line \hat{x} , \hat{y} , \hat{z} . Or, as I shall denote them \hat{e}_1 , \hat{e}_2 , \hat{e}_3 where

$$|\hat{e}_1| = |\hat{e}_2| = |\hat{e}_3| = 1, \quad \text{and} \quad \hat{e}_1 \cdot \hat{e}_2 = 0 \quad \text{etc.} \quad (12.10)$$

In terms of these independent vectors, \vec{v} has components v_1 , v_2 , v_3 :

$$\vec{v} = v_1 \hat{e}_1 + v_2 \hat{e}_2 + v_3 \hat{e}_3 \quad (12.11)$$

The vector $\vec{u} = T(\vec{v})$ can also be expanded in the same way:

$$\vec{u} = u_1 \hat{e}_1 + u_2 \hat{e}_2 + u_3 \hat{e}_3 \quad (12.12)$$

Look at $T(\vec{v})$ more closely in terms of the components

$$\begin{aligned} T(\vec{v}) &= T(v_1 \hat{e}_1 + v_2 \hat{e}_2 + v_3 \hat{e}_3) \\ &= v_1 T(\hat{e}_1) + v_2 T(\hat{e}_2) + v_3 T(\hat{e}_3) \end{aligned}$$

(by linearity). Each of the three objects $T(\hat{e}_1)$, $T(\hat{e}_2)$, $T(\hat{e}_3)$ is a vector, which means that you can expand each one in terms of the original unit vectors

$$\begin{aligned} T(\hat{e}_1) &= T_{11} \hat{e}_1 + T_{21} \hat{e}_2 + T_{31} \hat{e}_3 \\ T(\hat{e}_2) &= T_{12} \hat{e}_1 + T_{22} \hat{e}_2 + T_{32} \hat{e}_3 \quad \text{or more compactly,} \quad T(\hat{e}_i) = \sum_j T_{ji} \hat{e}_j \\ T(\hat{e}_3) &= T_{13} \hat{e}_1 + T_{23} \hat{e}_2 + T_{33} \hat{e}_3 \end{aligned} \quad (12.13)$$

The numbers T_{ij} ($i, j = 1, 2, 3$) are called the components of the tensor in the given basis. These numbers will depend on the basis chosen, just as do the numbers v_i , the components of the vector \vec{v} . The ordering of the indices has been chosen for later convenience, with the sum on the first index of the T_{ji} . This equation is *the fundamental equation* from which everything else is derived. (It will be modified when non-orthonormal bases are introduced later.)

Now, take these expressions for $T(\hat{e}_i)$ and plug them back into the equation $\vec{u} = T(\vec{v})$:

$$\begin{aligned} u_1 \hat{e}_1 + u_2 \hat{e}_2 + u_3 \hat{e}_3 = T(\vec{v}) &= v_1 [T_{11} \hat{e}_1 + T_{21} \hat{e}_2 + T_{31} \hat{e}_3] \\ &\quad + v_2 [T_{12} \hat{e}_1 + T_{22} \hat{e}_2 + T_{32} \hat{e}_3] \\ &\quad + v_3 [T_{13} \hat{e}_1 + T_{23} \hat{e}_2 + T_{33} \hat{e}_3] \\ &= [T_{11}v_1 + T_{12}v_2 + T_{13}v_3] \hat{e}_1 \\ &\quad + [T_{21}v_1 + T_{22}v_2 + T_{23}v_3] \hat{e}_2 \\ &\quad + [T_{31}v_1 + T_{32}v_2 + T_{33}v_3] \hat{e}_3 \end{aligned}$$

Comparing the coefficients of the unit vectors, you get the relations among the components

$$\begin{aligned} u_1 &= T_{11}v_1 + T_{12}v_2 + T_{13}v_3 \\ u_2 &= T_{21}v_1 + T_{22}v_2 + T_{23}v_3 \\ u_3 &= T_{31}v_1 + T_{32}v_2 + T_{33}v_3 \end{aligned} \quad (12.14)$$

More compactly:

$$u_i = \sum_{j=1}^3 T_{ij}v_j \quad \text{or} \quad \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \quad (12.15)$$

At this point it is convenient to use the summation convention (first* version). This convention says that if a given term contains a repeated index, then a summation over all the possible values of that index is understood. With this convention, the previous equation is

$$u_i = T_{ij}v_j. \quad (12.16)$$

Notice how the previous choice of indices has led to the conventional result, with the first index denoting the row and the second the column of a matrix.

* See section 12.5 for the later modification and generalization.

Now to take an example and tear it apart. Define a tensor by the equations

$$T(\hat{x}) = \hat{x} + \hat{y}, \quad T(\hat{y}) = \hat{y}, \quad (12.17)$$

where \hat{x} and \hat{y} are given orthogonal unit vectors. These two expressions, combined with linearity, suffice to determine the effect of the tensor on all linear combinations of \hat{x} and \hat{y} . (This is a two dimensional problem.)

To compute the components of the tensor pick a set of basis vectors. The obvious ones in this instance are

$$\hat{e}_1 = \hat{x}, \quad \text{and} \quad \hat{e}_2 = \hat{y}$$

By comparison with Eq. (12.13), you can read off the components of T .

$$\begin{aligned} T_{11} &= 1 & T_{21} &= 1 \\ T_{12} &= 0 & T_{22} &= 1. \end{aligned}$$

Write these in the form of a matrix as in Eq. (12.15)

$$(T_{\text{row, column}}) = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

and writing the vector components in the same way, the components of the vectors \hat{x} and \hat{y} are respectively

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

The original equations (12.17), that defined the tensor become the components

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

12.3 Relations between Tensors

Go back to the fundamental representation theorem for linear functionals and see what it looks like in component form. Evaluate $f(\vec{v})$, where $\vec{v} = v_i \hat{e}_i$. (The linear functional has one vector argument and a scalar output.)

$$f(\vec{v}) = f(v_i \hat{e}_i) = v_i f(\hat{e}_i) \quad (12.18)$$

Denote the set of numbers $f(\hat{e}_i)$ ($i = 1, 2, 3$) by $A_i = f(\hat{e}_i)$, in which case,

$$f(\vec{v}) = A_i v_i = A_1 v_1 + A_2 v_2 + A_3 v_3$$

Now it is clear that the vector \vec{A} of the theorem is just

$$\vec{A} = A_1 \hat{e}_1 + A_2 \hat{e}_2 + A_3 \hat{e}_3 \quad (12.19)$$

Again, examine the problem of starting from a bilinear functional and splitting off one of the two arguments in order to obtain a vector valued function of a vector. I want to say

$$T(\vec{u}, \vec{v}) = \vec{u} \cdot T(\vec{v})$$

for all vectors \vec{u} and \vec{v} . You should see that using the same symbol, T , for both functions doesn't cause any trouble. Given the bilinear functional, what is the explicit form for $T(\vec{v})$? The answer is most readily found by a bit of trial and error until you reach the following result:

$$T(\vec{v}) = \hat{e}_i T(\hat{e}_i, \vec{v}) \quad (12.20)$$

(Remember the summation convention.) To verify this relation, multiply by an arbitrary vector, $\vec{u} = u_j \hat{e}_j$:

$$\vec{u} \cdot T(\vec{v}) = (u_j \hat{e}_j) \cdot \hat{e}_i T(\hat{e}_i, \vec{v})$$

which is, by the orthonormality of the \hat{e} 's,

$$u_j \delta_{ji} T(\hat{e}_i, \vec{v}) = u_i T(\hat{e}_i, \vec{v}) = T(\vec{u}, \vec{v})$$

This says that the above expression is in fact the correct one. Notice also the similarity between this construction and the one in equation (12.19) for \vec{A} .

Now take $T(\vec{v})$ from Eq. (12.20) and express \vec{v} in terms of its components

$$\vec{v} = v_j \hat{e}_j, \quad \text{then} \quad T(\vec{v}) = \hat{e}_i T(\hat{e}_i, v_j \hat{e}_j) = \hat{e}_i T(\hat{e}_i, \hat{e}_j) v_j$$

The i component of this expression is

$$T(\hat{e}_i, \hat{e}_j) v_j = T_{ij} v_j$$

a result already obtained in Eq. (12.16).

There's a curiosity involved here; why should the left hand entry in $T(,)$ be singled out to construct

$$\hat{e}_i T(\hat{e}_i, \vec{v})?$$

Why not use the right hand one instead? Answer: No reason at all. It's easy enough to find out what happens when you do this. Examine

$$\hat{e}_i T(\vec{v}, \hat{e}_i) \equiv \tilde{T}(\vec{v}) \quad (12.21)$$

Put $\vec{v} = v_j \hat{e}_j$, and you get

$$\hat{e}_i T(v_j \hat{e}_j, \hat{e}_i) = \hat{e}_i T(\hat{e}_j, \hat{e}_i) v_j$$

The i^{th} component of which is

$$T_{ji} v_j$$

If you write this as a square matrix times a column matrix, the only difference between this result and that of Eq. (12.16) is that the matrix is transposed. This vector valued function \tilde{T} is called the transpose of the tensor T . The nomenclature comes from the fact that in the matrix representation, the matrix of one equals the transpose of the other's matrix.

By an extension of the language, this applies to the other form of the tensor, T :

$$\tilde{T}(\vec{u}, \vec{v}) = T(\vec{v}, \vec{u}) \quad (12.22)$$

Symmetries

Two of the common and important classifications of matrices, symmetric and antisymmetric, have their reflections in tensors. A symmetric tensor is one that equals its transpose and an antisymmetric tensor

is one that is the negative of its transpose. It is easiest to see the significance of this when the tensor is written in the bilinear functional form:

$$T_{ij} = T(\hat{e}_i, \hat{e}_j)$$

This matrix will equal its transpose if and only if

$$T(\vec{u}, \vec{v}) = T(\vec{v}, \vec{u})$$

for all \vec{u} and \vec{v} . Similarly, if for all \vec{u} and \vec{v}

$$T(\vec{u}, \vec{v}) = -T(\vec{v}, \vec{u})$$

then $T = -\tilde{T}$. Notice that it doesn't matter whether I speak of T as a scalar-valued function of two variables or as a vector-valued function of one; the symmetry properties are the same.

From these definitions, it is possible to take an arbitrary tensor and break it up into its symmetric part and its antisymmetric part:

$$\begin{aligned} T &= \frac{1}{2}(T + \tilde{T}) + \frac{1}{2}(T - \tilde{T}) = T_S + T_A & (12.23) \\ T_S(\vec{u}, \vec{v}) &= \frac{1}{2} [T(\vec{u}, \vec{v}) + T(\vec{v}, \vec{u})] \\ T_A(\vec{u}, \vec{v}) &= \frac{1}{2} [T(\vec{u}, \vec{v}) - T(\vec{v}, \vec{u})] \end{aligned}$$

Many of the common tensors such as the tensor of inertia and the dielectric tensor are symmetric. The magnetic field tensor in contrast, is antisymmetric. The basis of this symmetry in the case of the dielectric tensor is in the relation for the energy density in an electric field, $\int \vec{E} \cdot d\vec{D}$.^{*} Apply an electric field in the x direction, then follow it by adding a field in the y direction; undo the field in the x direction and then undo the field in the y direction. The condition that the energy density returns to zero is the condition that the dielectric tensor is symmetric.

All of the above discussions concerning the symmetry properties of tensors were phrased in terms of second rank tensors. The extensions to tensors of higher rank are quite easy. For example in the case of a third rank tensor viewed as a 3-linear functional, it would be called completely symmetric if

$$T(\vec{u}, \vec{v}, \vec{w}) = T(\vec{v}, \vec{u}, \vec{w}) = T(\vec{u}, \vec{w}, \vec{v}) = \text{etc.}$$

for all permutations of $\vec{u}, \vec{v}, \vec{w}$, and for all values of these vectors. Similarly, if any interchange of two arguments changed the value by a sign,

$$T(\vec{u}, \vec{v}, \vec{w}) = -T(\vec{v}, \vec{u}, \vec{w}) = +T(\vec{v}, \vec{w}, \vec{u}) = \text{etc.}$$

then the T is completely antisymmetric. It is possible to have a mixed symmetry, where there is for example symmetry on interchange of the arguments in the first and second place and antisymmetry between the second and third.

^{*} This can be proved by considering the energy in a plane parallel plate capacitor, which is, by definition of potential, $\int V dq$. The Potential difference V is the magnitude of the \vec{E} field times the distance between the the capacitor plates. [$V = Ed$.] (\vec{E} is perpendicular to the plates by $\nabla \times \vec{E} = 0$.) The normal component of \vec{D} related to q by $\nabla \cdot \vec{D} = \rho$. [$A\vec{D} \cdot \hat{n} = q$.] Combining these, and dividing by the volume gives the energy density as $\int \vec{E} \cdot d\vec{D}$.

Alternating Tensor

A curious (and very useful) result about antisymmetric tensors is that in three dimensions there is, up to a factor, exactly one totally antisymmetric third rank tensor; it is called the “alternating tensor.” So, if you take any two such tensors, Λ and Λ' , then one must be a multiple of the other. (The same holds true for the n^{th} rank totally antisymmetric tensor in n dimensions.)

Proof: Consider the function $\Lambda - \alpha\Lambda'$ where α is a scalar. Pick any three independent vectors \vec{v}_{10} , \vec{v}_{20} , \vec{v}_{30} as long as Λ' on this set is non-zero. Let

$$\alpha = \frac{\Lambda(\vec{v}_{10}, \vec{v}_{20}, \vec{v}_{30})}{\Lambda'(\vec{v}_{10}, \vec{v}_{20}, \vec{v}_{30})} \quad (12.24)$$

(If Λ' gives zero for every set of \vec{v} s then it's a trivial tensor, zero.) This choice of α guarantees that $\Lambda - \alpha\Lambda'$ will vanish for at least one set of values of the arguments. Now take a general set of three vectors \vec{v}_1 , \vec{v}_2 , and \vec{v}_3 and ask for the effect of $\Lambda - \alpha\Lambda'$ on them. \vec{v}_1 , \vec{v}_2 , and \vec{v}_3 can be expressed as linear combinations of the original \vec{v}_{10} , \vec{v}_{20} , and \vec{v}_{30} . Do so. Substitute into $\Lambda - \alpha\Lambda'$, use linearity and notice that all possible terms give zero.

The above argument is unchanged in a higher number of dimensions. It is also easy to see that you cannot have a totally antisymmetric tensor of rank $n + 1$ in n dimensions. In this case, one of the $n + 1$ variables would have to be a linear combination of the other n . Use linearity, and note that when any two variables equal each other, antisymmetry forces the result to be zero. These observations imply that the function must vanish identically. See also problem 12.17. If this sounds familiar, look back at section 7.7.

12.4 Birefringence

It's time to stop and present a serious calculation using tensors, one that leads to an interesting and not at all obvious result. This development assumes that you've studied Maxwell's electromagnetic field equations and are comfortable with vector calculus. If not then you will find this section obscure, maybe best left for another day. The particular problem that I will examine is how light passes through a transparent material, especially a crystal that has different electrical properties in different directions. The common and best known example of such a crystal is Iceland spar, a form of calcite (CaCO_3).

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \quad \nabla \times \vec{B} = \mu_0 \vec{j} + \mu_0 \epsilon_0 \frac{\partial \vec{E}}{\partial t} \quad \vec{j} = \frac{\partial \vec{P}}{\partial t} \quad \vec{P} = \alpha(\vec{E}) \quad (12.25)$$

\vec{E} and \vec{B} are the electric and magnetic fields. \vec{P} is the polarization of the medium, the electric dipole moment density. α is the polarizability tensor, relating the amount of distortion of the charges in the matter to the applied electric field. The current density \vec{j} appears because a time-varying electric field will cause a time-varying movement of the charges in the surrounding medium — that's a current.

Take the curl of the first equation and the time derivative of the second.

$$\nabla \times \nabla \times \vec{E} = -\frac{\partial}{\partial t} \nabla \times \vec{B} \quad \nabla \times \frac{\partial \vec{B}}{\partial t} = \mu_0 \frac{\partial \vec{j}}{\partial t} + \mu_0 \epsilon_0 \frac{\partial^2 \vec{E}}{\partial t^2}$$

The two expressions involving \vec{B} are the same, so eliminate \vec{B} .

$$\nabla \times \nabla \times \vec{E} = -\mu_0 \frac{\partial \vec{j}}{\partial t} - \mu_0 \epsilon_0 \frac{\partial^2 \vec{E}}{\partial t^2} = -\mu_0 \frac{\partial^2 \alpha(\vec{E})}{\partial t^2} - \mu_0 \epsilon_0 \frac{\partial^2 \vec{E}}{\partial t^2}$$

I make the assumption that α is time independent, so this is

$$\nabla \times \nabla \times \vec{E} = \nabla(\nabla \cdot \vec{E}) - \nabla^2 \vec{E} = -\mu_0 \alpha \left(\frac{\partial^2 \vec{E}}{\partial t^2} \right) - \mu_0 \epsilon_0 \frac{\partial^2 \vec{E}}{\partial t^2}$$

I am seeking a wave solution for the field, so assume a solution $\vec{E}(\vec{r}, t) = \vec{E}_0 e^{i\vec{k} \cdot \vec{r} - \omega t}$. Each ∇ brings down a factor of $i\vec{k}$ and each time derivative a factor $-i\omega$, so the equation is

$$-\vec{k}(\vec{k} \cdot \vec{E}_0) + k^2 \vec{E}_0 = \mu_0 \omega^2 \alpha(\vec{E}_0) + \mu_0 \epsilon_0 \omega^2 \vec{E}_0 \quad (12.26)$$

This is a linear equation for the vector \vec{E}_0 .

A special case first: A vacuum, so there is no medium and $\alpha \equiv 0$. The equation is

$$(k^2 - \mu_0 \epsilon_0 \omega^2) \vec{E}_0 - \vec{k}(\vec{k} \cdot \vec{E}_0) = 0 \quad (12.27)$$

Pick a basis so that \hat{z} is along \vec{k} , then

$$(k^2 - \mu_0 \epsilon_0 \omega^2) \vec{E}_0 - k^2 \hat{z} \cdot \vec{E}_0 = 0 \quad \text{or in matrix notation,}$$

$$\left[(k^2 - \mu_0 \epsilon_0 \omega^2) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - k^2 \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right] \begin{pmatrix} E_{0x} \\ E_{0y} \\ E_{0z} \end{pmatrix}$$

This is a set of linear homogeneous equations for the components of the electric field. One solution for \vec{E} is identically zero, and this solution is unique *unless* the determinant of the coefficients vanishes. That is

$$(k^2 - \mu_0 \epsilon_0 \omega^2)^2 (-\mu_0 \epsilon_0 \omega^2) = 0$$

This is a cubic equation for ω^2 . One root is zero, the other two are $\omega^2 = k^2 / \mu_0 \epsilon_0$. The eigenvector corresponding to the zero root has components the column matrix $(0 \ 0 \ 1)$, or $E_{0x} = 0$ and $E_{0y} = 0$ with the z -component arbitrary. The field for this solution is

$$\vec{E} = E_{0z} e^{ikz} \hat{z}, \quad \text{then} \quad \nabla \cdot \vec{E} = \rho / \epsilon_0 = ik E_{0z} e^{ikz}$$

This is a static charge density, not a wave, so look to the other solutions. They are

$$E_{0z} = 0, \text{ with } E_{0x}, E_{0y} \text{ arbitrary and } \vec{k} \cdot \vec{E}_0 = \nabla \cdot \vec{E} = 0$$

$$\vec{E} = (E_{0x} \hat{x} + E_{0y} \hat{y}) e^{ikz - i\omega t} \quad \text{and} \quad \omega^2 / k^2 = 1 / \mu_0 \epsilon_0 = c^2$$

This is a plane, transversely polarized, electromagnetic wave moving in the z -direction at velocity $\omega / k = 1 / \sqrt{\mu_0 \epsilon_0}$.

Now return to the case in which $\alpha \neq 0$. If the polarizability is a multiple of the identity, so that the dipole moment density is always along the direction of \vec{E} , all this does is to add a constant to ϵ_0 in Eq. (12.26). $\epsilon_0 \rightarrow \epsilon_0 + \alpha = \epsilon$, and the speed of the wave changes from $c = 1 / \sqrt{\mu_0 \epsilon_0}$ to $v = 1 / \sqrt{\mu_0 \epsilon}$

The more complicated case occurs when α is more than just multiplication by a scalar. If the medium is a crystal in which the charges can be polarized more easily in some directions than in others, α is a tensor. It is a symmetric tensor, though I won't prove that here. The proof involves looking at energy dissipation (rather, lack of it) as the electric field is varied. To compute I will pick a basis, and choose it so that the components of α form a diagonal matrix in this basis.

$$\text{Pick } \vec{e}_1 = \hat{x}, \quad \vec{e}_2 = \hat{y}, \quad \vec{e}_3 = \hat{z} \quad \text{so that} \quad (\alpha) = \begin{pmatrix} \alpha_{11} & 0 & 0 \\ 0 & \alpha_{22} & 0 \\ 0 & 0 & \alpha_{33} \end{pmatrix}$$

The combination that really appears in Eq. (12.26) is $\epsilon_0 + \alpha$. The first term is a scalar, a multiple of the identity, so the matrix for this is

$$(\epsilon) = \epsilon_0 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + (\alpha) = \begin{pmatrix} \epsilon_{11} & 0 & 0 \\ 0 & \epsilon_{22} & 0 \\ 0 & 0 & \epsilon_{33} \end{pmatrix} \quad (12.28)$$

I'm setting this up assuming that the crystal has three different directions with three different polarizabilities. When I get down to the details of the calculation I will take two of them equal — that's the case for calcite. The direction of propagation of the wave is \vec{k} , and Eq. (12.26) is

$$\left[k^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \mu_0 \omega^2 \begin{pmatrix} \epsilon_{11} & 0 & 0 \\ 0 & \epsilon_{22} & 0 \\ 0 & 0 & \epsilon_{33} \end{pmatrix} - \begin{pmatrix} k_1 k_1 & k_1 k_2 & k_1 k_3 \\ k_2 k_1 & k_2 k_2 & k_2 k_3 \\ k_3 k_1 & k_3 k_2 & k_3 k_3 \end{pmatrix} \right] \begin{pmatrix} E_{0x} \\ E_{0y} \\ E_{0z} \end{pmatrix} = 0 \quad (12.29)$$

In order to have a non-zero solution for the electric field, the determinant of this total 3×3 matrix must be zero. This is starting to get too messy, so I'll now make the simplifying assumption that two of the three directions in the crystal have identical properties: $\epsilon_{11} = \epsilon_{22}$. This makes the system cylindrically symmetric about the z -axis and I can then take the direction of the vector \vec{k} to be in the x - z plane for convenience.

$$\vec{k} = k(\vec{e}_1 \sin \alpha + \vec{e}_3 \cos \alpha)$$

The matrix of coefficients in Eq. (12.29) is now

$$\begin{aligned} & k^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \mu_0 \omega^2 \begin{pmatrix} \epsilon_{11} & 0 & 0 \\ 0 & \epsilon_{11} & 0 \\ 0 & 0 & \epsilon_{33} \end{pmatrix} - k^2 \begin{pmatrix} \sin^2 \alpha & 0 & \sin \alpha \cos \alpha \\ 0 & 0 & 0 \\ \sin \alpha \cos \alpha & 0 & \cos^2 \alpha \end{pmatrix} \\ & = \begin{pmatrix} k^2 \cos^2 \alpha - \mu_0 \omega^2 \epsilon_{11} & 0 & -k^2 \sin \alpha \cos \alpha \\ 0 & k^2 - \mu_0 \omega^2 \epsilon_{11} & 0 \\ -k^2 \sin \alpha \cos \alpha & 0 & k^2 \sin^2 \alpha - \mu_0 \omega^2 \epsilon_{33} \end{pmatrix} \end{aligned} \quad (12.30)$$

The determinant of this matrix is

$$\begin{aligned} & (k^2 - \mu_0 \omega^2 \epsilon_{11}) \left[(k^2 \cos^2 \alpha - \mu_0 \omega^2 \epsilon_{11})(k^2 \sin^2 \alpha - \mu_0 \omega^2 \epsilon_{33}) - (-k^2 \sin \alpha \cos \alpha)^2 \right] \\ & = (k^2 - \mu_0 \omega^2 \epsilon_{11}) \left[-\mu_0 \omega^2 k^2 (\epsilon_{11} \sin^2 \alpha + \epsilon_{33} \cos^2 \alpha) + \mu_0^2 \omega^4 \epsilon_{11} \epsilon_{33} \right] = 0 \end{aligned} \quad (12.31)$$

This has a factor $\mu_0 \omega^2 = 0$, and the corresponding eigenvector is \vec{k} itself. In column matrix notation that is $(\sin \alpha \ 0 \ \cos \alpha)$. It is another of those static solutions that aren't very interesting. For the rest, there are factors

$$k^2 - \mu_0 \omega^2 \epsilon_{11} = 0 \quad \text{and} \quad k^2 (\epsilon_{11} \sin^2 \alpha + \epsilon_{33} \cos^2 \alpha) - \mu_0 \omega^2 \epsilon_{11} \epsilon_{33} = 0$$

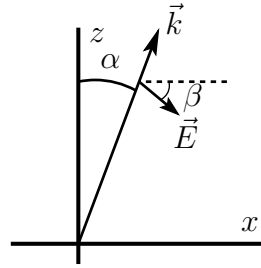
The first of these roots, $\omega^2/k^2 = 1/\mu_0 \epsilon_{11}$, has an eigenvector $\vec{e}_2 = \hat{y}$, or $(0 \ 1 \ 0)$. Then $\vec{E} = E_0 \hat{y} e^{ikz - \omega t}$ and this is a normal sort of transverse wave that you commonly find in ordinary non-crystalline materials. The electric vector is perpendicular to the direction of propagation. The energy flow (the Poynting vector) is along the direction of \vec{k} , perpendicular to the wavefronts. It's called the "ordinary ray," and at a surface it obeys Snell's law. The reason it is ordinary is that the electric vector in this case is along one of the principal axes of the crystal. The polarization is along \vec{E} just as it is in air or glass, so it behaves the same way as in those cases.

The second root has $E_{0y} = 0$, and an eigenvector computed as

$$\begin{aligned}(k^2 \cos^2 \alpha - \mu_0 \omega^2 \epsilon_{11}) E_{0x} - (k^2 \sin \alpha \cos \alpha) E_{0z} &= 0 \\ (k^2 \cos^2 \alpha - k^2 (\epsilon_{11} \sin^2 \alpha + \epsilon_{33} \cos^2 \alpha) / \epsilon_{33}) E_{0x} - (k^2 \sin \alpha \cos \alpha) E_{0z} &= 0 \\ \epsilon_{11} \sin \alpha E_{0x} + \epsilon_{33} \cos \alpha E_{0z} &= 0\end{aligned}$$

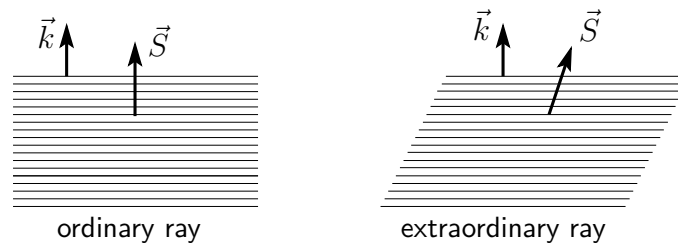
If the two ϵ 's are equal, this says that \vec{E} is perpendicular to \vec{k} . If they aren't equal then $\vec{k} \cdot \vec{E} \neq 0$ and you can write this equation for the direction of \vec{E} as

$$\begin{aligned}E_{0x} &= E_0 \cos \beta, \quad E_{0z} = -\sin \beta, \quad \text{then} \\ \epsilon_{11} \sin \alpha \cos \beta - \epsilon_{33} \cos \alpha \sin \beta &= 0 \\ \text{and so } \tan \beta &= \frac{\epsilon_{11}}{\epsilon_{33}} \tan \alpha\end{aligned}$$

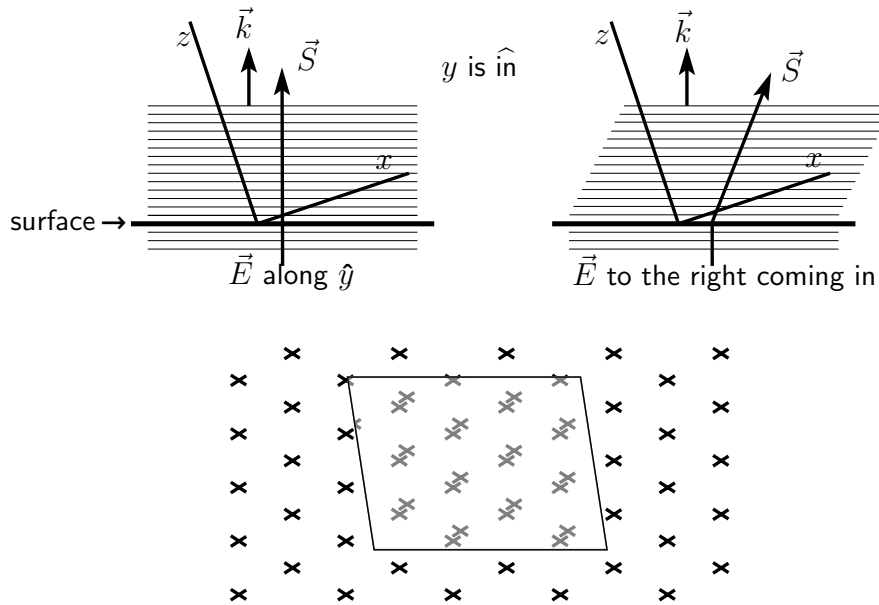


In calcite, the ratio $\epsilon_{11}/\epsilon_{33} = 1.056$, making β a little bigger than α .

The magnetic field is in the $\vec{k} \times \vec{E}$ direction, and the energy flow of the light is along the Poynting vector, in the direction $\vec{E} \times \vec{B}$. In this picture, that puts \vec{B} along the \hat{y} -direction (into the page), and then the energy of the light moves in the direction perpendicular to \vec{E} and \vec{B} . That is *not* along \vec{k} . This means that the propagation of the wave is not along the perpendicular to the wave fronts. Instead the wave skitters off at an angle to the front. The “extraordinary ray.” Snell's law says that the wavefronts bend at a surface according to a simple trigonometric relation, *and they still do*, but the energy flow of the light does *not* follow the direction normal to the wavefronts. The light ray does not obey Snell's law



In calcite as it naturally grows, the face of the crystal is not parallel to any of the x - y or x - z or y - z planes. When a light ray enters the crystal perpendicular to the surface, the wave fronts are in the plane of the surface and what happens then depends on polarization. For the light polarized along one of the principle axes of the crystal (the y -axis in this sketch) the light behaves normally and the energy goes in an unbroken straight line. For the other polarization, the electric field has components along two different axes of the the crystal and the energy flows in an unexpected direction — disobeying Snell's law. The normal to the wave front is in the direction of the vector \vec{k} , and the direction of energy flow (the Poynting vector) is indicated by the vector \vec{S} .



Simulation of a pattern of x's seen through Calcite. See also Wikipedia: birefringence.

12.5 Non-Orthogonal Bases

The next topic is the introduction of more general computational techniques. These will lift the restriction on the type of basis that can be used for computing components of various tensors. Until now, the basis vectors have formed an orthonormal set

$$|\hat{e}_i| = 1, \quad \hat{e}_i \cdot \hat{e}_j = 0 \text{ if } i \neq j$$

Consider instead a more general set of vectors \vec{e}_i . These must be independent. That is, in three dimensions they are not coplanar. Other than this there is no restriction. Since by assumption the vectors \vec{e}_i span the space you can write

$$\vec{v} = v^i \hat{e}_i.$$

with the numbers v^i being as before the components of the vector \vec{v} .

NOTE: Here is a change in notation. Before, every index was a subscript. (It could as easily have been a superscript.) Now, be sure to make a careful distinction between sub- and superscripts. They will have different meanings.

Reciprocal Basis

Immediately, when you do the basic scalar product you find complications. If $\vec{u} = u^j \vec{e}_j$, then

$$\vec{u} \cdot \vec{v} = (u^j \vec{e}_j) \cdot (v^i \vec{e}_i) = u^j v^i \vec{e}_j \cdot \vec{e}_i.$$

But since the \vec{e}_i aren't orthonormal, this is a much more complicated result than the usual scalar product such as

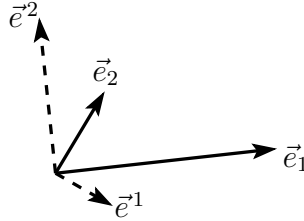
$$u_x v_x + u_y v_y + u_z v_z.$$

You can't assume that $\vec{e}_1 \cdot \vec{e}_2 = 0$ any more. In order to obtain a result that looks as simple as this familiar form, introduce an auxiliary basis: the *reciprocal basis*. (This trick will not really simplify the

answer; it will be the same as ever. It will however be in a neater form and hence easier to manipulate.) The reciprocal basis is defined by the equation

$$\vec{e}_i \cdot \vec{e}^j = \delta_i^j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad (12.32)$$

The \vec{e}^j 's are vectors. The index is written as a superscript to distinguish it from the original basis, \vec{e}_j .



To elaborate on the meaning of this equation, \vec{e}^1 is perpendicular to the plane defined by \vec{e}_2 and \vec{e}_3 and is therefore more or less in the direction of \vec{e}_1 . Its magnitude is adjusted so that the scalar product

$$\vec{e}^1 \cdot \vec{e}_1 = 1.$$

The “direct basis” and “reciprocal basis” are used in solid state physics and especially in describing X-ray diffraction in crystallography. In that instance, the direct basis is the fundamental lattice of the crystal and the reciprocal basis would be defined from it. The reciprocal basis is used to describe the wave number vectors of scattered X-rays.

The basis reciprocal to the reciprocal basis is the direct basis.

Now to use these things: Expand the vector \vec{u} in terms of the direct basis and \vec{v} in terms of the reciprocal basis.

$$\begin{aligned} \vec{u} = u^i \vec{e}_i \quad \text{and} \quad \vec{v} = v_j \vec{e}^j. \quad \text{Then} \quad \vec{u} \cdot \vec{v} &= (u^i \vec{e}_i) \cdot (v_j \vec{e}^j) \\ &= u^i v_j \delta_i^j \\ &= u^i v_i = u^1 v_1 + u^2 v_2 + u^3 v_3. \end{aligned}$$

Notation: The superscript on the components (u^i) will refer to the components in the direct basis (\vec{e}_i); the subscripts (v_j) will come from the reciprocal basis (\vec{e}^j). You could also have expanded \vec{u} in terms of the reciprocal basis and \vec{v} in the direct basis, then

$$\vec{u} \cdot \vec{v} = u_i v^i = u^i v_i \quad (12.33)$$

Summation Convention

At this point, modify the previously established summation convention: Like indices in a given term are to be summed when one is a subscript and one is a superscript. Furthermore the notation is designed so that this is the only kind of sum that should occur. If you find a term such as $u_i v_i$ then this means that you made a mistake.

The scalar product now has a simple form in terms of components (at the cost of introducing an auxiliary basis set). Now for further applications to vectors and tensors.

Terminology: The components of a vector in the direct basis are called the contravariant components of the vector: v^i . The components in the reciprocal basis are called* the covariant components: v_i .

* These terms are of more historical than mathematical interest.

Examine the component form of the basic representation theorem for linear functionals, as in Eqs. (12.18) and (12.19).

$$f(\vec{v}) = \vec{A} \cdot \vec{v} \quad \text{for all } \vec{v}.$$

$$\text{Claim: } \vec{A} = \vec{e}^i f(\vec{e}_i) = \vec{e}_i f(\vec{e}^i) \quad (12.34)$$

The proof of this is as before: write \vec{v} in terms of components and compute the scalar product $\vec{A} \cdot \vec{v}$.

$$\begin{aligned} \vec{v} = v^i \vec{e}_i. \quad \text{Then} \quad \vec{A} \cdot \vec{v} &= (\vec{e}^j f(\vec{e}_j)) \cdot (v^i \vec{e}_i) \\ &= v^i f(\vec{e}_j) \delta_i^j \\ &= v^i f(\vec{e}_i) = f(v^i \vec{e}_i) = f(\vec{v}). \end{aligned}$$

Analogous results hold for the expression of \vec{A} in terms of the direct basis.

You can see how the notation forced you into considering this expression for \vec{A} . The summation convention requires one upper index and one lower index, so there is practically no other form that you could even consider in order to represent \vec{A} .

The same sort of computations will hold for tensors. Start off with one of second rank. Just as there were covariant and contravariant components of a vector, there will be covariant and contravariant components of a tensor. $T(\vec{u}, \vec{v})$ is a scalar. Express \vec{u} and \vec{v} in contravariant component form:

$$\begin{aligned} \vec{u} = u^i \vec{e}_i \quad \text{and} \quad \vec{v} = v^j \vec{e}_j. \quad \text{Then} \quad T(\vec{u}, \vec{v}) &= T(u^i \vec{e}_i, v^j \vec{e}_j) \\ &= u^i v^j T(\vec{e}_i, \vec{e}_j) \\ &= u^i v^j T_{ij} \end{aligned} \quad (12.35)$$

The numbers T_{ij} are called the covariant components of the tensor T .

Similarly, write \vec{u} and \vec{v} in terms of covariant components:

$$\begin{aligned} \vec{u} = u_i \vec{e}^i \quad \text{and} \quad \vec{v} = v_j \vec{e}^j. \quad \text{Then} \quad T(\vec{u}, \vec{v}) &= T(u_i \vec{e}^i, v_j \vec{e}^j) \\ &= u_i v_j T(\vec{e}^i, \vec{e}^j) \\ &= u_i v_j T^{ij} \end{aligned} \quad (12.36)$$

And T^{ij} are the contravariant components of T . It is also possible to have mixed components:

$$\begin{aligned} T(\vec{u}, \vec{v}) &= T(u_i \vec{e}^i, v^j \vec{e}_j) \\ &= u_i v^j T(\vec{e}^i, \vec{e}_j) \\ &= u_i v^j T^i_j \end{aligned}$$

As before, from the bilinear functional, a linear vector valued function can be formed such that

$$\begin{aligned} T(\vec{u}, \vec{v}) = \vec{u} \cdot T(\vec{v}) \quad \text{and} \quad T(\vec{v}) &= \vec{e}^i T(\vec{e}_i, \vec{v}) \\ &= \vec{e}_i T(\vec{e}^i, \vec{v}) \end{aligned}$$

For the proof of the last two lines, simply write \vec{u} in terms of its contravariant or covariant components respectively.

All previous statements concerning the symmetry properties of tensors are unchanged because they were made in a way independent of basis, though it's easy to see that the symmetry properties

of the tensor are reflected in the symmetry of the covariant or the contravariant components (but not usually in the mixed components).

Metric Tensor

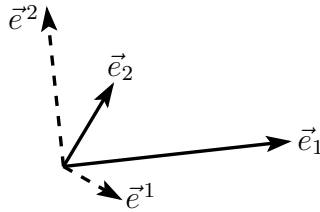
Take as an example the metric tensor:

$$g(\vec{u}, \vec{v}) = \vec{u} \cdot \vec{v}. \quad (12.37)$$

The linear function found by pulling off the \vec{u} from this is the identity operator.

$$g(\vec{v}) = \vec{v}$$

This tensor is symmetric, so this must be reflected in its covariant and contravariant components. Take as a basis the vectors



Let $|\vec{e}_2| = 1$ and $|e_1| = 2$; the angle between them being 45° . A little geometry shows that

$$|\vec{e}^1| = \frac{1}{\sqrt{2}} \quad \text{and} \quad |\vec{e}^2| = \sqrt{2}$$

Assume this problem is two dimensional in order to simplify things.

Compute the covariant components:

$$\begin{aligned} g_{11} &= g(\vec{e}_1, \vec{e}_1) = 4 \\ g_{12} &= g(\vec{e}_1, \vec{e}_2) = \sqrt{2} \\ g_{21} &= g(\vec{e}_2, \vec{e}_1) = \sqrt{2} \\ g_{22} &= g(\vec{e}_2, \vec{e}_2) = 1 \end{aligned} \quad (g_{rc}) = \begin{pmatrix} 4 & \sqrt{2} \\ \sqrt{2} & 1 \end{pmatrix}$$

Similarly

$$\begin{aligned} g^{11} &= g(\vec{e}^1, \vec{e}^1) = 1/2 \\ g^{12} &= g(\vec{e}^1, \vec{e}^2) = -1/\sqrt{2} \\ g^{21} &= g(\vec{e}^2, \vec{e}^1) = -1/\sqrt{2} \\ g^{22} &= g(\vec{e}^2, \vec{e}^2) = 2 \end{aligned} \quad (g^{rc}) = \begin{pmatrix} 1/2 & -1/\sqrt{2} \\ -1/\sqrt{2} & 2 \end{pmatrix}$$

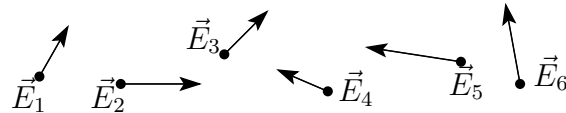
The mixed components are

$$\begin{aligned} g^1_1 &= g(\vec{e}^1, \vec{e}_1) = 1 \\ g^1_2 &= g(\vec{e}^1, \vec{e}_2) = 0 \\ g^2_1 &= g(\vec{e}^2, \vec{e}_1) = 0 \\ g^2_2 &= g(\vec{e}^2, \vec{e}_2) = 1 \end{aligned} \quad (g^r_c) = (\delta^r_c) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (12.38)$$

I used r and c for the indices to remind you that these are the row and column variables. Multiply the first two matrices together and you obtain the third one — the unit matrix. The matrix (g_{ij}) is therefore the inverse of the matrix (g^{ij}) . This last result is *not* general, but is due to the special nature of the tensor g .

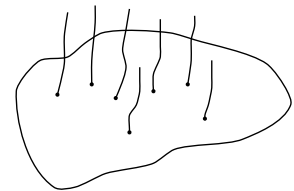
12.6 Manifolds and Fields

Until now, all definitions and computations were done in one vector space. This is the same state of affairs as when you once learned vector algebra; the only things to do then were addition, scalar products, and cross products. Eventually however vector calculus came up and you learned about vector fields and gradients and the like. You have now set up enough apparatus to make the corresponding step here. First I would like to clarify just what is meant by a vector field, because there is sure to be confusion on this point no matter how clearly you think you understand the concept. Take a typical vector field such as the electrostatic field \vec{E} . \vec{E} will be some function of position (presumably satisfying Maxwell's equations) as indicated at the six different points.



Does it make any sense to take the vector \vec{E}_3 and add it to the vector \vec{E}_5 ? These are after all, vectors; can't you always add one vector to another vector? Suppose there is also a magnetic field present, say with vectors \vec{B}_1, \vec{B}_2 etc. , at the same points. Take the magnetic vector at the point #3 and add it to the electric vector there. The reasoning would be exactly the same as the previous case; these are vectors, therefore they can be added. The second case is clearly nonsense, as should be the first. The electric vector is defined as the force per charge at a point. If you take two vectors at two different points, then the forces are on two different objects, so the sum of the forces is not a force on anything — it isn't even defined.

You can't add an electric vector at one point to an electric vector at another point. These two vectors occupy different vector spaces. At a single point in space there are many possible vectors; at this one point, the set of all possible electric vectors forms a vector space because they can be added to each other and multiplied by scalars while remaining at the same point. By the same reasoning the magnetic vectors at a point form a vector space. Also the velocity vectors. You could not add a velocity vector to an electric field vector even at the same point however. These too are in different vector spaces. You can picture all these vector spaces as attached to the points in the manifold and somehow sitting over them.



From the above discussion you can see that even to discuss one type of vector field, a vector space must be attached to each point of space. If you wish to make a drawing of such a system, It is at best difficult. In three dimensional space you could have a three dimensional vector space at each point. A crude way of picturing this is to restrict to two dimensions and draw a line attached to each point, representing the vector space attached to that point. This pictorial representation won't be used in anything to follow however, so you needn't worry about it.

The term "vector field" that I've been throwing around is just a prescription for selecting one vector out of each of the vector spaces. Or, in other words, it is a function that assigns to each point a vector in the vector space at that same point.

There is a minor confusion of terminology here in the use of the word "space." This could be space in the sense of the three dimensional Euclidean space in which we are sitting and doing computations. Each point of the latter will have a vector space associated with it. To reduce confusion (I hope) I shall use the word "manifold" for the space over which all the vector spaces are built. Thus: To each point of the manifold there is associated a vector space. A vector field is a choice of one vector from each of the vector spaces over the manifold. This is a vector field on the manifold. In short: The word "manifold" is substituted here for the phrase "three dimensional Euclidean space."

(A comment on generalizations. While using the word manifold as above, everything said about it will in fact be more general. For example it will still be acceptable in special relativity with four dimensions of space-time. It will also be correct in other contexts where the structure of the manifold is non-Euclidean.)

The point to emphasize here is that most of the work on tensors is already done and that the application to fields of vectors and fields of tensors is in a sense a special case. At each point of the manifold there is a vector space to which all previous results apply.

In the examples of vector fields mentioned above (electric field, magnetic field, velocity field) keep your eye on the velocity. It will play a key role in the considerations to come, even in considerations of other fields.

A word of warning about the distinction between a manifold and the vector spaces at each point of the manifold. You are accustomed to thinking of three dimensional Euclidean space (the manifold) as a vector space itself. That is, the displacement vector between two points is defined, and you can treat these as vectors just like the electric vectors at a point. *Don't!* Treating the manifold as a vector space will cause great confusion. Granted, it happens to be correct in this instance, but in attempting to understand these new concepts about vector fields (and tensor fields later), this additional knowledge will be a hindrance. For our purposes therefore the manifold will not be a vector space. The concept of a displacement vector is therefore *not defined*.

Just as vector fields were defined by picking a single vector from each vector space at various points of the manifold, a scalar field is similarly an assignment of a number (scalar) to each point. In short then, a scalar field is a function that gives a scalar (the dependent variable) for each point of the manifold (the independent variable).

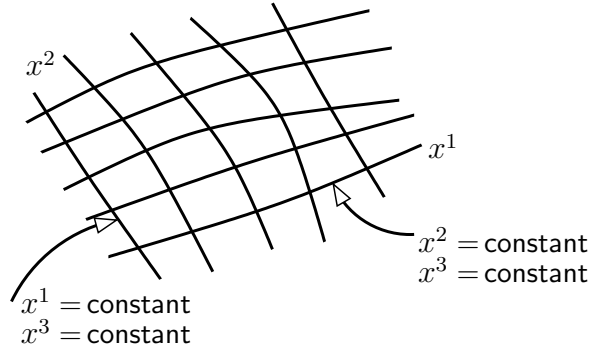
For each vector space, you can discuss the tensors that act on that space and so, by picking one such tensor for each point of the manifold a tensor field is defined.

A physical example of a tensor field (of second rank) is stress in a solid. This will typically vary from point to point. But at each point a second rank tensor is given by the relation between infinitesimal area vectors and internal force vectors at that point. Similarly, the dielectric tensor in an inhomogeneous medium will vary with position and will therefore be expressed as a tensor field. Of course even in a homogeneous medium the dielectric tensor would be a tensor field relating \vec{D} and \vec{E} at the same point. It would however be a constant tensor field. Like a uniform electric field, the tensors at different points could be thought of as “parallel” to each other (whatever that means).

12.7 Coordinate Bases

In order to obtain a handle on this subject and in order to be able to do computations, it is necessary to put a coordinate system on the manifold. From this coordinate system there will come a natural way to define the basis vectors at each point (and so reciprocal basis vectors too). The orthonormal basis vectors that you are accustomed to in cylindrical and spherical coordinates are *not* “coordinate bases.”

There is no need to restrict the discussion to rectangular or even to orthogonal coordinate systems. A coordinate system is a means of identifying different points of the manifold by different sets of numbers. This is done by specifying a set of functions: x^1, x^2, x^3 , which are the coordinates. (There will be more in more dimensions of course.) These functions are real valued functions of points in the manifold. The coordinate axes are defined as in the drawing by



Specify the equations $x^2 = 0$ and $x^3 = 0$ for the x^1 coordinate axis. For example in rectangular coordinates $x^1 = x$, $x^2 = y$, $x^3 = z$, and the x -axis is the line $y = 0$, and $z = 0$. In plan polar coordinates $x^1 = r = \text{a constant}$ is a circle and $x^2 = \phi = \text{a constant}$ is a straight line starting from the origin.

Start with two dimensions and polar coordinates r and ϕ . As a basis in this system we routinely use the unit vectors \hat{r} and $\hat{\phi}$, but is this the only choice? Is it the best choice? Not necessarily. Look at two vectors, the differential $d\vec{r}$ and the gradient ∇f .

$$d\vec{r} = \hat{r} dr + \hat{\phi} r d\phi \quad \text{and} \quad \nabla f = \hat{r} \frac{\partial f}{\partial r} + \hat{\phi} \frac{1}{r} \frac{\partial f}{\partial \phi} \quad (12.39)$$

And remember the chain rule too.

$$\frac{df}{dt} = \frac{\partial f}{\partial r} \frac{dr}{dt} + \frac{\partial f}{\partial \phi} \frac{d\phi}{dt} \quad (12.40)$$

In this basis the scalar product is simple, but you pay for it in that the components of the vectors have extra factors in them — the r and the $1/r$. An alternate approach is to place the complexity in the basis vectors and not in the components.

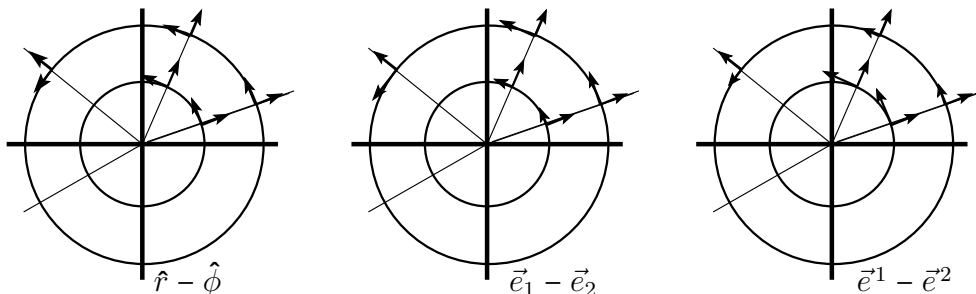
$$\vec{e}_1 = \hat{r}, \quad \vec{e}_2 = r \hat{\phi} \quad \text{and} \quad \vec{e}^1 = \hat{r}, \quad \vec{e}^2 = \hat{\phi}/r \quad (12.41)$$

These are reciprocal bases, as indicated by the notation. Of course the original \hat{r} - $\hat{\phi}$ is self-reciprocal. The preceding equations (12.39) are now

$$d\vec{r} = \vec{e}_1 dr + \vec{e}_2 d\phi \quad \text{and} \quad \nabla f = \vec{e}^1 \frac{\partial f}{\partial r} + \vec{e}^2 \frac{\partial f}{\partial \phi} \quad (12.42)$$

Make another change in notation to make this appear more uniform. Instead of r - ϕ for the coordinates, call them x^1 and x^2 respectively, then

$$d\vec{r} = \vec{e}_1 dx^1 + \vec{e}_2 dx^2 \quad \text{and} \quad \nabla f = \vec{e}^1 \frac{\partial f}{\partial x^1} + \vec{e}^2 \frac{\partial f}{\partial x^2} \quad (12.43)$$



The velocity vector is

$$\frac{d\vec{r}}{dt} = \vec{e}_1 \frac{dr}{dt} + \vec{e}_2 \frac{d\phi}{dt} \quad \text{and Eq. (12.40) is} \quad \frac{df}{dt} = \frac{\partial f}{\partial x^i} \frac{dx^i}{dt} = (\nabla f) \cdot \vec{v}$$

This sort of basis has many technical advantages that aren't at all apparent here. At this point this "coordinate basis" is simply a way to sweep some of the complexity out of sight, but with further developments of the subject the sweeping becomes shoveling. When you go beyond the introduction found in this chapter, you find that using any basis other than a coordinate basis leads to equations that have complicated extra terms that you want nothing to do with.

In spherical coordinates $x^1 = r$, $x^2 = \theta$, $x^3 = \phi$

$$\vec{e}_1 = \hat{r}, \quad \vec{e}_2 = r \hat{\theta}, \quad \vec{e}_3 = r \sin \theta \hat{\phi} \quad \text{and} \quad \vec{e}^1 = \hat{r}, \quad \vec{e}^2 = \hat{\theta}/r, \quad \vec{e}^3 = \hat{\phi}/r \sin \theta$$

The velocity components are now

$$dx^i/dt = \{dr/dt, d\theta/dt, d\phi/dt\}, \quad \text{and} \quad \vec{v} = \vec{e}_i dx^i/dt \quad (12.44)$$

This last equation is central to figuring out the basis vectors in an unfamiliar coordinate system.

The use of x^1 , x^2 , and x^3 (x^i) for the coordinate system makes the notation uniform. Despite the use of superscripts, these coordinates are *not* the components of any vectors, though their time derivatives are.

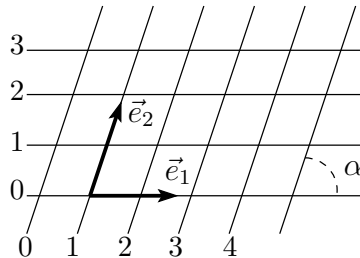
In ordinary rectangular coordinates. If a particle is moving along the x^1 -axis (the x -axis) then $dx^2/dt = 0 = dx^3/dt$. Also, the velocity will be in the x direction and of size dx^1/dt .

$$\vec{e}_1 = \hat{x}$$

as you would expect. Similarly

$$\vec{e}_2 = \hat{y}, \quad \vec{e}_3 = \hat{z}.$$

In a variation of rectangular coordinates in the plane the axes are not orthogonal to each other, but are rectilinear anyway.



Still keep the requirement of Eq. (12.44)

$$\vec{v} = \vec{e}_i \frac{dx^i}{dt} = \vec{e}_1 \frac{dx^1}{dt} + \vec{e}_2 \frac{dx^2}{dt}. \quad (12.45)$$

If the particle moves along the x^1 -axis (or parallel to it) then by the definition of the axes, x^2 is a constant and $dx^2/dt = 0$. Suppose that the coordinates measure centimeters, so that the *perpendicular* distance between the lines is one centimeter. The distance between the points (0, 0) and (1, 0) is then $1 \text{ cm} / \sin \alpha = \csc \alpha \text{ cm}$. If in $\Delta t =$ one second, particle moves from the first to the second of these

points, $\Delta x^1 = \text{one cm}$, so $dx^1/dt = 1 \text{ cm/sec}$. The speed however, is $\csc \alpha \text{ cm/sec}$ because the distance moved is greater by that factor. This means that

$$|\vec{e}_1| = \csc \alpha$$

and this is greater than one; it is not a unit vector. The magnitudes of \vec{e}_2 is the same. The dot product of these two vectors is $\vec{e}_1 \cdot \vec{e}_2 = \cos \alpha / \sin^2 \alpha$.

Reciprocal Coordinate Basis

Construct the reciprocal basis vectors from the direct basis by the equation

$$\vec{e}^i \cdot \vec{e}_j = \delta_j^i$$

In rectangular coordinates the direct and reciprocal bases coincide because the basis is orthonormal. For the tilted basis of Eq. (12.45),

$$\vec{e}_2 \cdot \vec{e}^2 = 1 = |\vec{e}_2| |\vec{e}^2| \cos(90^\circ - \alpha) = (\csc \alpha) |\vec{e}^2| \sin \alpha = |\vec{e}^2|$$

The reciprocal basis vectors in this case are unit vectors.

The direct basis is defined so that the components of velocity are as simple as possible. In contrast, the components of the gradient of a scalar field are equally simple provided that they are expressed in the reciprocal basis. If you try to use the same basis for both you can, but the resulting equations are a mess.

In order to compute the components of $\text{grad } f$ (where f is a scalar field) start with its definition, and an appropriate definition should not depend on the coordinate system. It ought to be some sort of geometric statement that you can translate into any particular coordinate system that you want. One way to define $\text{grad } f$ is that it is that vector pointing in the direction of maximum increase of f and equal in magnitude to $df/d\ell$ where ℓ is the distance measured in that direction. This is the first statement in section 8.5. While correct, this definition does not easily lend itself to computations.

Instead, think of the particle moving through the manifold. As a function of time it sees changing values of f . The time rate of change of f as felt by this particle is given by a scalar product of the particle's velocity and the gradient of f . This is essentially the same as Eq. (8.16), though phrased in a rather different way. Write this statement in terms of coordinates

$$\frac{d}{dt} f(x^1(t), x^2(t), x^3(t)) = \vec{v} \cdot \text{grad } f$$

The left hand side is (by the chain rule)

$$\frac{\partial f}{\partial x^1} \Big|_{x^2, x^3} \frac{dx^1}{dt} + \frac{\partial f}{\partial x^2} \Big|_{x^1, x^3} \frac{dx^2}{dt} + \frac{\partial f}{\partial x^3} \Big|_{x^1, x^2} \frac{dx^3}{dt} = \frac{\partial f}{\partial x^1} \frac{dx^1}{dt} + \frac{\partial f}{\partial x^2} \frac{dx^2}{dt} + \frac{\partial f}{\partial x^3} \frac{dx^3}{dt} \quad (12.46)$$

\vec{v} is expressed in terms of the direct basis by

$$\vec{v} = \vec{e}_i \frac{dx^i}{dt},$$

now express $\text{grad } f$ in the reciprocal basis

$$\text{grad } f = \vec{e}^i (\text{grad } f)_i \quad (12.47)$$

The way that the scalar product looks in terms of these bases, Eq. (12.33) is

$$\vec{v} \cdot \text{grad } f = \vec{e}_i \frac{dx^i}{dt} \cdot \vec{e}^j (\text{grad } f)_j = v^i (\text{grad } f)_i \quad (12.48)$$

Compare the two equations (12.46) and (12.48) and you see

$$\text{grad } f = \vec{e}^i \frac{\partial f}{\partial x^i} \quad (12.49)$$

For a formal proof of this statement consider three cases. When the particle is moving along the x^1 direction (x^2 & x^3 constant) only one term appears on each side of (12.46) and (12.48) and you can divide by $v^1 = dx^1/dt$. Similarly for x^2 and x^3 . As usual with partial derivatives, the symbol $\partial f/\partial x^i$ assumes that the other coordinates x^2 and x^3 are constant.

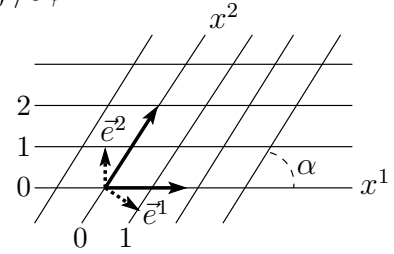
For polar coordinates this equation for the gradient reads, using Eq. (12.41),

$$\text{grad } f = \vec{e}^1 \frac{\partial f}{\partial x^1} + \vec{e}^2 \frac{\partial f}{\partial x^2} = (\hat{r}) \frac{\partial f}{\partial r} + \left(\frac{1}{r} \hat{\phi}\right) \frac{\partial f}{\partial \phi}$$

which is the standard result, Eq. (8.27). Notice again that the basis vectors are not dimensionless. They can't be because $\partial f/\partial r$ doesn't have the same dimensions as $\partial f/\partial \phi$.

Example

I want an example to show that all this formalism actually gives the correct answer in a special case for which you can also compute all the results in the traditional way. Draw parallel lines a distance 1 cm apart and another set of parallel lines also a distance 1 cm apart intersecting at an angle α between them. These will be the constant values of the functions defining the coordinates, and will form a coordinate system labeled x^1 and x^2 . The horizontal lines are the equations $x^2 = 0$, $x^2 = 1$ cm, etc.



Take the case of the non-orthogonal rectilinear coordinates again. The components of $\text{grad } f$ in the \vec{e}^1 direction is $\partial f/\partial x^1$, which is the derivative of f with respect to x^1 holding x_2 constant, and this derivative is *not* in the direction along \vec{e}^1 , but in the direction where $x^2 = \text{a constant}$ and that is along the x^1 -axis, along \vec{e}_1 . As a specific example to show that this makes sense, take a particular f defined by

$$f(x^1, x^2) = x^1$$

$$\text{For this function} \quad \text{grad } f = \vec{e}^1 \frac{\partial f}{\partial x^1} + \vec{e}^2 \frac{\partial f}{\partial x^2} = \vec{e}^1$$

\vec{e}^1 is perpendicular to the x^2 -axis, the line $x_1 = \text{constant}$, (as it should be). Its magnitude is the magnitude of \vec{e}^1 , which is one.

To verify that this magnitude is correct, calculate it directly from the definition. The magnitude of the gradient is the magnitude of $df/d\ell$ where ℓ is measured in the direction of the gradient, that is, in the direction \vec{e}^1 .

$$\frac{df}{d\ell} = \frac{\partial f}{\partial x^1} \frac{dx^1}{d\ell} = 1 \cdot 1 = 1$$

Why 1 for $dx^1/d\ell$? The coordinate lines in the picture are $x^1 = 0, 1, 2, \dots$. When you move on the straight line perpendicular to $x^2 = \text{constant}$ (\vec{e}^1), and go from $x^1 = 1$ to $x^2 = 2$, then both Δx^1 and Δs are one.

Metric Tensor

The simplest tensor field beyond the gradient vector above would be the metric tensor, which I've been implicitly using all along whenever a scalar product arose. It is defined at each point by

$$g(\vec{a}, \vec{b}) = \vec{a} \cdot \vec{b} \quad (12.50)$$

Compute the components of g in plane polar coordinates. The contravariant components of g are from Eq. (12.41)

$$g^{ij} = \vec{e}^i \cdot \vec{e}^j = \begin{pmatrix} 1 & 0 \\ 0 & 1/r^2 \end{pmatrix}$$

Covariant:

$$g_{ij} = \vec{e}_i \cdot \vec{e}_j = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}$$

Mixed:

$$g^i_j = \vec{e}^i \cdot \vec{e}_j = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

12.8 Basis Change

If you have two different sets of basis vectors you can compute the transformation on the components in going from one basis to the other, and in dealing with fields, a different set of basis vectors necessarily arises from a different set of coordinates on the manifold. It is convenient to compute the transformation matrices directly in terms of the different coordinate functions. Call the two sets of coordinates x^i and y^i . Each of them defines a set of basis vectors such that a given velocity is expressed as

$$\vec{v} = \vec{e}_i \frac{dx^i}{dt} = \vec{e}'_j \frac{dy^j}{dt} \quad (12.51)$$

What you need is an expression for \vec{e}'_j in terms of \vec{e}_i at each point. To do this, take a particular path for the particle — along the y^1 -direction (y^2 & y^3 constant). The right hand side is then

$$\vec{e}'_1 \frac{dy^1}{dt}$$

Divide by dy^1/dt to obtain

$$\vec{e}'_1 = \vec{e}_i \frac{dx^i}{dt} \bigg/ \frac{dy^1}{dt}$$

But this quotient is just

$$\vec{e}'_1 = \vec{e}_i \left. \frac{\partial x^i}{\partial y^1} \right|_{y^2, y^3}$$

And in general

$$\vec{e}'_j = \vec{e}_i \frac{\partial x^i}{\partial y^j} \quad (12.52)$$

Do a similar calculation for the reciprocal vectors

$$\text{grad } f = \vec{e}^i \frac{\partial f}{\partial x^i} = \vec{e}'^j \frac{\partial f}{\partial y^j}$$

Take the case for which $f = y^k$, then

$$\frac{\partial f}{\partial y^j} = \frac{\partial y^k}{\partial y^j} = \delta_j^k$$

which gives

$$\vec{e}'^k = \vec{e}^i \frac{\partial y^k}{\partial x^i} \quad (12.53)$$

The transformation matrices for the direct and the reciprocal basis are inverses of each other, In the present context, this becomes

$$\begin{aligned} e'^k \cdot \vec{e}'_j &= \delta_j^k = \vec{e}'^\ell \frac{\partial y^k}{\partial x^\ell} \cdot \vec{e}_i \frac{\partial x^i}{\partial y^j} \\ &= \delta_i^\ell \frac{\partial y^k}{\partial x^\ell} \frac{\partial x^i}{\partial y^j} \\ &= \frac{\partial y^k}{\partial x^i} \frac{\partial x^i}{\partial y^j} = \frac{\partial y^k}{\partial y^j} \end{aligned}$$

The matrices $\partial x^i / \partial y^j$ and its inverse matrix, $\partial y^k / \partial x^i$ are called Jacobian matrices. When you do multiple integrals and have to change coordinates, the determinant of one or the other of these matrices will appear as a factor in the integral.

As an example, compute the change from rectangular to polar coordinates

$$\begin{aligned} x^1 &= x & y^1 &= r & x^2 &= y & y^2 &= \phi \\ x &= r \cos \phi & r &= \sqrt{x^2 + y^2} & y &= r \sin \phi & \phi &= \tan^{-1} y/x \end{aligned}$$

$$\begin{aligned} \vec{e}'_j &= \vec{e}_i \frac{\partial x^i}{\partial y^j} \\ \vec{e}'_1 &= \vec{e}_1 \frac{\partial x^1}{\partial y^1} + \vec{e}_2 \frac{\partial x^2}{\partial y^1} = \hat{x} \frac{\partial x}{\partial r} + \hat{y} \frac{\partial y}{\partial r} \\ &= \hat{x} \cos \phi + \hat{y} \sin \phi = \hat{r} \\ \vec{e}'_2 &= \vec{e}_1 \frac{\partial x^1}{\partial y^2} + \vec{e}_2 \frac{\partial x^2}{\partial y^2} = \hat{x} \frac{\partial x}{\partial \phi} + \hat{y} \frac{\partial y}{\partial \phi} \\ &= \hat{x}(-r \sin \phi) + \hat{y}(r \cos \phi) = r \hat{\phi} \end{aligned}$$

Knowing the change in the basis vectors, the change of the components of any tensor by computing it in the new basis and using the linearity of the tensor itself. The new components will be linear combinations of those from the old basis.

A realistic example using non-orthonormal bases appears in special relativity. Here the manifold is four dimensional instead of three and the coordinate changes of interest represent Lorentz transformations. Points in space-time ("events") can be described by rectangular coordinates (ct, x, y, z) , which are concisely denoted by x^i

$$i = (0, 1, 2, 3) \quad \text{where} \quad x^0 = ct, \quad x^1 = x, \quad x^2 = y, \quad x^3 = z$$

The introduction of the factor c into x^0 is merely a question of scaling. It also makes the units the same on all axes.

The basis vectors associated with this coordinate system point along the directions of the axes.

This manifold is *not* Euclidean however so that these vectors are not unit vectors in the usual sense. We have

$$\vec{e}_0 \cdot \vec{e}_0 = -1 \quad \vec{e}_1 \cdot \vec{e}_1 = 1 \quad \vec{e}_2 \cdot \vec{e}_2 = 1 \quad \vec{e}_3 \cdot \vec{e}_3 = 1$$

and they are orthogonal pairwise. The reciprocal basis vectors are defined in the usual way,

$$\vec{e}^i \cdot \vec{e}_j = \delta_j^i$$

so that

$$\vec{e}^0 = -\vec{e}_0 \quad \vec{e}^1 = \vec{e}_1 \quad \vec{e}^2 = \vec{e}_2 \quad \vec{e}^3 = \vec{e}_3$$

The contravariant (also covariant) components of the metric tensor are

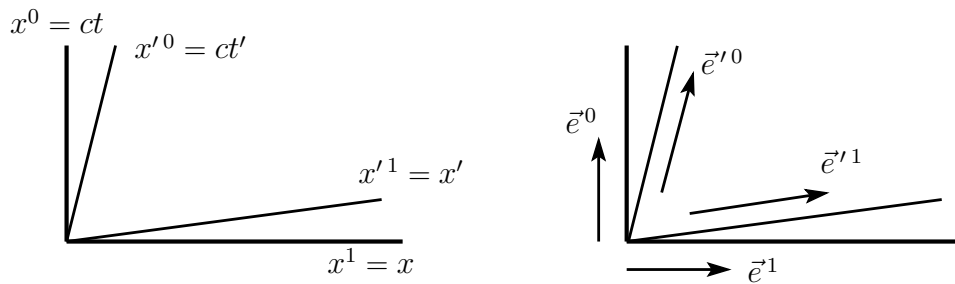
$$g^{ij} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = g_{ij} \quad (12.54)$$

An observer moving in the $+x$ direction with speed v will have his own coordinate system with which to describe the events of space-time. The coordinate transformation is

$$\begin{aligned} x'^0 &= ct' = \frac{x^0 - \frac{v}{c}x^1}{\sqrt{1 - v^2/c^2}} = \frac{ct - \frac{v}{c}x}{\sqrt{1 - v^2/c^2}} \\ x'^1 &= x' = \frac{x^1 - \frac{v}{c}x^0}{\sqrt{1 - v^2/c^2}} = \frac{x - vt}{\sqrt{1 - v^2/c^2}} \\ x'^2 &= y' = x^2 = y \quad x'^3 = z' = x^3 = z \end{aligned} \quad (12.55)$$

You can check that these equations represent the transformation to an observer moving in the $+x$ direction by asking where the moving observer's origin is as a function of time: It is at $x'^1 = 0$ or $x - vt = 0$, giving $x = vt$ as the locus of the moving observer's origin.

The graph of the coordinates is as usual defined by the equations (say for the x'^0 -axis) that x'^1 , x'^2 , x'^3 are constants such as zero. Similarly for the other axes.



Find the basis vectors in the transformed system by using equation (12.52).

$$\vec{e}'_j = \vec{e}_i \frac{\partial x^i}{\partial y^j}$$

In the present case the y^j are x'^j and we need the inverse of the equations (12.55). They are found by changing v to $-v$ and interchanging primed and unprimed variables.

$$x^0 = \frac{x'^0 + \frac{v}{c}x'^1}{\sqrt{1 - v^2/c^2}} \quad x^1 = \frac{x'^1 + \frac{v}{c}x'^0}{\sqrt{1 - v^2/c^2}}$$

$$\begin{aligned}\vec{e}'_0 &= \vec{e}_i \frac{\partial x^i}{\partial x'^0} = \vec{e}_0 \frac{1}{\sqrt{1-v^2/c^2}} + \vec{e}_1 \frac{v/c}{\sqrt{1-v^2/c^2}} \\ \vec{e}'_1 &= \vec{e}_i \frac{\partial x^i}{\partial x'^1} = \vec{e}_0 \frac{v/c}{\sqrt{1-v^2/c^2}} + \vec{e}_1 \frac{1}{\sqrt{1-v^2/c^2}}\end{aligned}\quad (12.56)$$

It is easy to verify that these new vectors point along the primed axes as they should. They also have the property that they are normalized to plus or minus one respectively as are the original untransformed vectors. (How about the reciprocal vectors?)

As an example applying all this apparatus, do the transformation of the components of a second rank tensor, the electromagnetic field tensor. This tensor is the function that acts on the current density (four dimensional vector) and gives the force density (also a four-vector). Its covariant components are

$$(F_{ij}) = F(\vec{e}_i, \vec{e}_j) = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & B_z & -B_y \\ E_y & -B_z & 0 & B_x \\ E_z & B_y & -B_x & 0 \end{pmatrix} \quad (12.57)$$

where the E 's and B 's are the conventional electric and magnetic field components. Compute a sample component of this tensor in the primed coordinate system.

$$\begin{aligned}F'_{20} &= F(\vec{e}'_2, \vec{e}'_0) = F\left(\vec{e}_2, \vec{e}_0 \frac{1}{\sqrt{1-v^2/c^2}} + \vec{e}_1 \frac{v/c}{\sqrt{1-v^2/c^2}}\right) \\ &= \frac{1}{\sqrt{1-v^2/c^2}} F_{20} + \frac{v/c}{\sqrt{1-v^2/c^2}} F_{21}\end{aligned}$$

or in terms of the E and B notation,

$$E'_y = \frac{1}{\sqrt{1-v^2/c^2}} \left[E_y - \frac{v}{c} B_z \right]$$

Since v is a velocity in the $+x$ direction this in turn is

$$E'_y = \frac{1}{\sqrt{1-v^2/c^2}} \left[E_y + \frac{1}{c} (\vec{v} \times \vec{B})_y \right] \quad (12.58)$$

Except possibly for the factor in front of the brackets. this is a familiar, physically correct equation of elementary electromagnetic theory. A charge is always at rest *in its own reference system*. In its own system, the only force it feels is the electric force because its velocity with respect to itself is zero. The electric field that it experiences is \vec{E}' , not the \vec{E} of the outside world. This calculation tells you that this force $q\vec{E}'$ is the same thing that I would expect if I knew the Lorentz force law, $\vec{F} = q[\vec{E} + \vec{v} \times \vec{B}]$. The factor of $\sqrt{1-v^2/c^2}$ appears because force itself has some transformation laws that are not as simple as you would expect.

Exercises

- 1** On the three dimensional vector space of real quadratic polynomials in x , define the linear functional $F(f) = \int_0^1 dx f(x)$. Suppose that 1 , x , and x^2 are an orthonormal basis, then what vector \vec{A} represents this functional F so that $\vec{A} \cdot \vec{v} = F(f)$, where the vector \vec{v} means a quadratic polynomial, as $f(x) = a + bx + cx^2$. Ans: $\vec{A} = 1 + \frac{1}{2}x + \frac{1}{3}x^2$
- 2** In the preceding example, take the scalar product to be $\langle f, g \rangle = \int_{-1}^1 dx f(x)g(x)$ and find the vector that represents the functional in this case. Ans: $\vec{A} = \frac{1}{2} + \frac{3}{2}x$
- 3** In the first example, what if the polynomials can have arbitrary order? Ans: $\vec{A} = -\frac{1}{x} \ln(1-x)$ (Not quite right because this answer is not a polynomial, so it is not a vector in the original space. There really is no correct answer to this question as stated.)

Problems

12.1 Does the function T defined by $T(v) = v + c$ with c a constant satisfy the definition of linearity?

12.2 Let the set X be the positive integers. Let the set Y be all real numbers. Consider the following sets and determine if they are relations between X and Y and if they are functions.

$$\begin{aligned} &\{(0, 0), (1, 2.0), (3, -\pi), (0, 1.0), (-1, e)\} \\ &\{(0, 0), (1, 2.0), (3, -\pi), (0, 1.0), (2, e)\} \\ &\{(0, 0), (1, 2.0), (3, -\pi), (4, 1.0), (2, e)\} \\ &\{(0, 0), (5, 5.5), (5., \pi) (3, -2.0) (7, 8)\} \end{aligned}$$

12.3 Starting from the definition of the tensor of inertia in Eq. (12.1) and using the defining equation for components of a tensor, compute the components of I .

12.4 Find the components of the tensor relating \vec{d} and \vec{F} in the example of Eq. (12.2)

12.5 The product of tensors is defined to be just the composition of functions for the second rank tensor viewed as a vector variable. If S and T are such tensors, then $(ST)(v) = S(T(v))$ (by definition) Compute the components of ST in terms of the components of S and of T . Express the result both in terms of index notation and matrices. Ans: This is matrix multiplication.

12.6 (a) The two tensors 1_1T and ${}^1_1\tilde{T}$ are derived from the same bilinear functional 0_2T , in the Eqs. (12.20)–(12.22). Prove that for arbitrary \vec{u} and \vec{v} ,

$$\vec{u} \cdot {}^1_1T(\vec{v}) = {}^1_1\tilde{T}(\vec{u}) \cdot \vec{v}$$

(If it's less confusing to remove all the sub- and superscripts, do so.)

(b) If you did this by writing everything in terms of components, do it again without components and just using the nature of these as functions. (If you did in without components, do it again using components.)

12.7 What is the significance of a tensor satisfying the relation $\tilde{T}[T(\vec{v})] = T[\tilde{T}(\vec{v})] = \vec{v}$ for all \vec{v} ? Look at what effect it has on the scalar product of two vectors if you let T act on both. That is, what is the scalar product of $T\vec{u}$ and $T\vec{v}$?

12.8 Carry out the construction indicated in section 12.3 to show the dielectric tensor is symmetric.

12.9 Fill in the details of the proof that the alternating tensor is unique up to a factor.

12.10 Compute the components of such an alternating tensor in two dimensions.

Ans: The matrix is anti-symmetric.

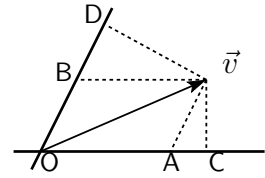
12.11 Take an alternating tensor in three dimensions, pull off one of the arguments in order to obtain a vector valued function of two vector variables. See what the result looks like, and in particular, write it out in component form.

Ans: You should find a cross product here.

12.12 Take a basis $\vec{e}_1 = 2\hat{x}$, $\vec{e}_2 = \hat{x} + 2\hat{y}$. Compute the reciprocal basis. Find the components of the vectors $\vec{A} = \hat{x} - \hat{y}$ and $\vec{B} = \hat{y}$ in each basis and compute $\vec{A} \cdot \vec{B}$ three different ways: in the \vec{e}_1 - \vec{e}_2

basis, in its reciprocal basis, and mixed (\vec{A} in one basis and \vec{B} in the reciprocal basis). A fourth way uses the \hat{x} - \hat{y} basis.

12.13 Show that if the direct basis vectors have unit length along the directions of \vec{OA} and \vec{OB} then the components of \vec{v} in the direct basis are the lengths OA and OB . What are the components in the reciprocal basis?



12.14 What happens to the components of the alternating tensor when a change of basis is made? Show that the only thing that can happen is that all the components are multiplied by the same number (which is the determinant of the transformation). Compute this explicitly in two dimensions.

12.15 A given tensor (viewed as a function of two vector variables) has the property that it equals zero whenever the two arguments are the same, $T(\vec{v}, \vec{v}) = 0$. Show that it is antisymmetric. This is also true if it is a function of more than two variables and the above equation holds on some pair of arguments. Consider $\vec{v} = \alpha\vec{u} + \beta\vec{w}$

12.16 If the components of the alternating tensor are (in three dimensions) e_{ijk} where $e_{123} = 1$, Compute e^{ijk} . Compute

$$e^{ijk}e_{lmk}, \quad e^{ijk}e_{ljk}, \quad e^{ijk}e_{ijk}$$

12.17 In three dimensions three non-collinear vectors from a point define a volume, that of the parallelepiped included between them. This defines a number as a function of three vectors. Show that if the volume is allowed to be negative when one of the vectors reverses that this defines a trilinear functional and that it is completely antisymmetric, an alternating tensor. (Note problem 12.15.) If the units of volume are chosen to correspond to the units of length of the basis vectors so that three one inch long perpendicular vectors enclose one cubic inch as opposed to 16.387 cm^3 then the functional is called "the" alternating tensor. Find all its components.

12.18 Find the direct coordinate basis in spherical coordinates, also the reciprocal basis.
Ans: One set is $\hat{r}, \hat{\theta}/r, \hat{\phi}/r \sin \theta$ (now which ones are these?)

12.19 Draw examples of the direct and reciprocal bases (to scale) for the example in Eq. (12.45). Do this for a wide range of angles between the axes.

12.20 Show that the area in two dimensions enclosed by the infinitesimal parallelogram between x^1 and $x^1 + dx^1$ and x^2 and $x^2 + dx^2$ is $\sqrt{g}dx^1 dx^2$ where g is the determinant of (g_{ij}) .

12.21 Compute the transformation laws for the other components of the electromagnetic field tensor.

12.22 The divergence of a tensor field is defined as above for a vector field. T is a tensor viewed as a vector valued function of a vector variable

$$\text{div } T = \lim_{V \rightarrow 0} \frac{1}{V} \oint T(d\vec{A})$$

It is a vector. Note: As defined here it requires the addition of vectors at two different points of the manifold, so it must be assumed that you can move a vector over parallel to itself and add it to a vector at another point. Had I not made the point in the text about not doing this, perhaps you wouldn't have noticed the assumption involved, but now do it anyway. Compute $\text{div } T$ in rectangular coordinates

using the $\hat{x}, \hat{y}, \hat{z}$ basis.

Ans: $\partial T^{ij}/\partial x^j$ in coordinate basis.

12.23 Compute $\text{div } T$ in cylindrical coordinates using both the coordinate basis and the usual unit vector $(\hat{r}, \hat{\phi}, \hat{z})$ basis. This is where you start to see why the coordinate basis has advantages.

12.24 Show that $g^{ij}g_{j\ell} = \delta_\ell^i$.

12.25 If you know what it means for vectors at two different points to be parallel, give a definition for what it means for two tensors to be parallel.

12.26 Fill in the missing steps in the derivation following Eq. (12.24), and show that the alternating tensor is unique up to a factor.

12.27 The divergence of a vector field computed in a coordinate basis is $\text{div } \vec{F} = \partial F^i/\partial x^i$. The divergence computed in standard cylindrical coordinates is Eq. (9.15). Show that these results agree.

12.28 From Eq. (12.31) verify the stated properties of the $\omega = 0$ solution.