Vectors

January 13, 2013

The simplest tensors are scalars, which are the measurable quantities of a theory, left invariant by symmetry transformations. By far the most common non-scalars are the vectors, also called rank-1 tensors. Vectors hold a distinguished position among tensors – indeed, tensors must be defined in terms of vectors. The reason for their importance is that, while tensors are those objects that transform linearly and homogeneously under a given set of transformations, we require vectors in order to define the action of the symmetry in the first place. Thus, vectors cannot be defined in terms of their transformations.

In the next subsection, we provide an axiomatic, algebraic definition of vectors. Then we show how to associate two distinct vector spaces with points of a manifold. Somewhat paradoxically, one of these vector spaces is called the space of vectors while the other is called the space of 1 -forms. Fortunately, the existence of a metric on the manifold allows us to relate these two spaces in a 1-1, onto way. Moreover, the metric allows us to define an inner product on each of the two vectors spaces. Therefore, we discuss the properties of metrics in some detail.

After the geometric description of vectors and forms, we turn to transformations of vectors. Using the action of a group on a vector space to define a linear representation of the group, we are finally able to define outer products of vectors and give a general definition of tensors in terms of their transformation properties.

1 Vectors as algebraic objects

Alternatively, we can define vectors algebraically. Briefly, a vector space is defined as a set of objects, $V = \{\mathbf{v}\}$, together with a field $\mathcal{F}$ of numbers (general R or C) which form a commutative group under addition and permit scalar multiplication. The scalar multiplication must satisfy distributive laws.

More concretely, being a group under addition guarantees the following:

1. V is closed under addition. If $\mathbf{u}, \mathbf{v}$ are any two elements of V, then $\mathbf{u} + \mathbf{v}$ is also an element of V.

2. There exists an additive identity, which we call the zero vector, $\mathbf{0}$.

3. For each element $\mathbf{v}$ of V there is an additive inverse to $\mathbf{v}$. We call this element $(-\mathbf{v})$.

4. Vector addition is associative, $\mathbf{w} + (\mathbf{u} + \mathbf{v}) = (\mathbf{w} + \mathbf{u}) + \mathbf{v}$

In addition, addition is commutative, $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.

The scalar multiplication satisfies:

1. Closure: $a\mathbf{v}$ is in V whenever $\mathbf{v}$ is in V and a is in $\mathcal{F}$.

2. Scalar identity: $1\mathbf{v} = \mathbf{v}$

3. Scalar and vector zero: $0\mathbf{v} = \mathbf{0}$ for all $\mathbf{v}$ in V and $a\mathbf{0} = \mathbf{0}$ for all a in $\mathcal{F}$.

4. Distributive 1: $(a + b)\,\mathbf{v} = a\mathbf{v} + b\mathbf{v}$

5. Distributive 2: $a\,(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$

6. Associativity: $(ab)\,\mathbf{v} = a\,(b\mathbf{v})$

All of the familiar properties of vectors follow from these. An important example is the existence of a basis for any finite dimensional vectors space. We prove this in several steps as follows.

First, define linear dependence. A set of vectors $\{\mathbf{v}_i \mid i = 1, \ldots, n\}$ is *linearly dependent* if there exist numbers $\{a_i \mid i = 1, \ldots, n\}$, not all of which are zero, such that the sum $a_i\mathbf{v}_i$ vanishes,

$$a_i\mathbf{v}_i = 0$$

As set of vectors is *linearly independent* if it is not dependent. Now suppose there exists a maximal linearly independent set of vectors. By this we mean that there exists some finite number n, such that we can find one or more linearly independent sets containing n vectors, but there do not exist any linearly independent sets containing $n+1$ vectors. Then we say that n is the *dimension* of the vector space.

In an n-dimensional vector space, and collection of n independent vectors is called a *basis*. Suppose we have a basis,

$$B = \{\mathbf{v}_i \mid i = 1, \ldots, n\}$$

Then, since every set with $n+1$ elements is linearly dependent, the set

$$\{\mathbf{u}\} \cup B = \{\mathbf{u}, \mathbf{v}_i \mid i = 1, \ldots, n\}$$

is dependent, where $\mathbf{u}$ is any nonzero vector in V. Therefore, there exist numbers a_i, b, not all zero, such that

$$b\mathbf{u} + a_i\mathbf{v}_i = 0$$

Now suppose $b = 0$. Then we have a linear combination of the $\mathbf{v}_i$ that vanishes, $a_i\mathbf{v}_i = 0$, contrary to our assumption that they form a basis. Therefore, b is nonzero, and we can divide by it. Adding the inverse to the sum $a_i\mathbf{v}_i$ we can write

$$\mathbf{u} = -\frac{1}{b}a_i\mathbf{v}_i$$

This shows that every vector in a finite dimensional vector space V can be written as a linear combination of the vectors in any basis. The numbers $u_i = -\frac{a_i}{b}$ are called the components of the vector $\mathbf{u}$ in the basis B.

Prove that two vectors are equal if and only if their components are equal.

Notice that we have chosen to write the labels on the basis vectors as subscripts, while we write the components of a vector as superscripts. This choice is arbitrary, but leads to considerable convenience later. Therefore, we will carefully maintain these positions in what follows.

Often vector spaces are given an inner product. An inner product on a vector space is a symmetric bilinear mapping from pairs of vectors to the relevant field, $\mathcal{F}$,

$$g : V \times V \to \mathcal{F}$$

Here the Cartesian product $V \times V$ means the set of all ordered pairs of vectors, $(\mathbf{u}, \mathbf{v})$, and bilinear means that g is linear in each of its two arguments. Symmetric means that $g(\mathbf{u}, \mathbf{v}) = g(\mathbf{v}, \mathbf{u})$.

There are a number of important consequences of inner products.

Suppose we have an inner product which gives a nonnegative real number whenever the two vectors it acts on are identical:

$$g(\mathbf{v}, \mathbf{v}) = s^2 \geq 0$$

where the equal sign holds if and only if $\mathbf{v}$ is the zero vector. Then g is a *norm* or *metric* on V – it provides a notion of length for each vector. If the inner product satisfies the triangle inequality,

$$g(\mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v}) \leq g(\mathbf{u}, \mathbf{u}) + g(\mathbf{v}, \mathbf{v})$$

then we can also define angles between vectors, via

$$\cos\theta = \frac{g(\mathbf{u}, \mathbf{v})}{\sqrt{g(\mathbf{u}, \mathbf{u})\, g(\mathbf{v}, \mathbf{v})}}$$

If the number s is real, but not necessarily positive, then g is called a *pseudo-norm* or a *pseudo-metric*. We will need to use a pseudo-metric when we study relativity.

If $\{\mathbf{v}_i\}$ is a basis, then we can write the inner product of any two vectors as

$$\begin{aligned} g(\mathbf{u}, \mathbf{v}) &= g\left(a^i \mathbf{v}_i, b^j \mathbf{v}_j\right) \\ &= a^i b^j g\left(\mathbf{v}_i, \mathbf{v}_j\right) \end{aligned}$$

so if we know how g acts on the basis vectors, we know how it acts on any pair of vectors. We can summarize this knowledge by defining the matrix

$$g_{ij} \equiv g\left(\mathbf{v}_i, \mathbf{v}_j\right)$$

Now, we can write the inner product of any two vectors as

$$g(\mathbf{u}, \mathbf{v}) = a^i g_{ij} b^j = g_{ij} a^i b^j$$

It's OK to think of this as sandwiching the metric, g_{ij}, between a row vector a^i on the left and a column vector b^j on the right. However, index notation is more powerful than the notions of row and column vectors, and in the long run it is more convenient to just note which sums are required. A great deal of computation can be accomplished without actually carrying out sums. We will discuss inner products in more detail in later Sections.

2 Vectors in space

In order to work with vectors in physics, it is most useful to think of them as geometric objects. This approach allows us to associate one or more vector spaces with each point of a manifold, which in turn will allow us to discuss motion and the time evolution of physical properties.

Since there are spaces – for example, the spacetimes of general relativity or on the surface of a sphere – where we want to talk about vectors but can't draw them as arrows because the space is curved, we need a more general, abstract definition. To define vectors, we need three things:

1. A manifold, $\mathcal{M}$, that is, a topological space which in a small enough region looks like a small piece of R^n. Manifolds include the Euclidean spaces, R^n, but also things like the 2-dimensional surface of a sphere or a doughnut or a saddle.

2. Functions on $\mathcal{M}$. A function on a manifold assigns a number to each point of the space.

3. Curves on the space. A curve is a mapping from the reals into the space. Such maps don't have to be continuous or differentiable in general, but we are interested in the case where they are. If M is our space and we have a curve $C : R \rightarrow M$, then for any real number λ, $C(\lambda)$ is a point of M. Changing λ moves us smoothly along the curve in M. If we have coordinates x^i for M we can specify the curve by giving $x^i(\lambda)$. For example, $(\theta(\lambda), \varphi(\lambda))$ describes a curve on the surface of a sphere.

Given a space together with functions and curves, there are *two* ways to associated a space of vectors with each point. We will call these two spaces *forms* and *vectors*, respectively, even though both are vector spaces in the algebraic sense. The existence of two distinct vector spaces associated with a manifold leads us to introduce some new notation. From now on, vectors from the space of vectors will have components written with a raised index, v^i, while the components of forms will be written with the index lowered, ω_i. The convention is natural if we begin by writing the indices on coordinates in the raised position and think of a derivative with respect to the coordinates as having the index in the lowered position. The benefits of this convention will quickly become evident. As an additional aid to keeping track of which vector space we mean, whenever practical we will name forms with Greek letters and vectors with Latin letters.

The definitions are as follows:

A *form* is defined for each function as a linear mapping from curves into the reals. The vector space of forms is denoted V_*.

A *vector* is defined for each curve as a linear mapping from functions into the reals. The vector space of vectors is denoted V^*.

Here's how it works. For a form, start with a function and think of the form, ω_f, as the differential of the function, $\omega_f = df$. Thus, for each function we have a form. The form is defined as a linear mapping on curves, $\omega_f : f \to R$. We can think of the linear mapping as integration along the curve C, so

$$\omega_f(C) = \int_C df = f(C(1)) - f(C(0))$$

In coordinates, we know that df is just

$$df = \frac{\partial f}{\partial x^i} dx^i$$

If we restrict the differentials dx^i to lie along the curve C, we have

$$df = \frac{\partial f}{\partial x^i} \frac{dx^i}{d\lambda} d\lambda$$

We can think of the coordinate differentials dx^i as a basis, and the partial derivatives $\frac{\partial f}{\partial x^i}$ as the components of the vector ω_f.

This argument shows that integrals of the differentials of functions are forms, but the converse is also true – any linear mapping from curves to the reals may be written as the integral of the differential of a function. The proof is as follows. Let ϕ be a linear map from differentiable curves to the reals, and let the curve C be parameterized by $s \in [0,1]$. Break C into N pieces, C_i, parameterized by $s \in \left[\frac{i-1}{N}, \frac{i}{N}\right]$, for $i = 1, 2, ...N$. By linearity, $\phi(C)$ is given by

$$\phi(C) = \sum_{i=1}^{N} \phi(C_i)$$

By the differentiability (hence continuity) of ϕ, we know that $\phi(C_i)$ maps C_i to a bounded interval in R, say, (a_i, b_i), of length $|b_i - a_i|$. As we increase N, each of the numbers $|b_i - a_i|$ tends monotonically to zero so that the value of $\phi(C_i)$ approaches arbitrarily close to the value a_i. We may therefore express $\phi(C)$ by

$$\begin{aligned} \phi(C) &= \lim_{N\to\infty} \sum_{i=1}^{N} \phi(C_i) \\ &= \lim_{N\to\infty} \sum_{i=1}^{N} \frac{a_i}{|b_i - a_i|} ds \end{aligned}$$

where ds replaces $\lim_{N\to\infty} \frac{1}{N}$. Notice that as N becomes large, a_i becomes small, since the average value of $\phi(C_i)$ is $\frac{\phi(C)}{N}$. The normalized expression

$$\frac{a_i}{|b_i - a_i|}$$

therefore remains of order $\phi(C)$ and we may define a function $f(s)$ as the piecewise continuous function

$$f(s_i) = \frac{a_i}{|b_i - a_i|}, \quad s_i \in \left[\frac{i-1}{N}, \frac{i}{N}\right]$$

Then $f(s)$ becomes smooth in the limit as $N \to \infty$, and $\phi(C)$ is given by

$$\phi(C) = \int f(s)\, ds$$

The fundamental theorem of calculus now shows that if we let $F = \int_C f(s)\, ds$, then

$$\phi(C) = \int_C dF$$

so that the linear map on curves is the integral of a differential.

For vectors, we start with the curve and think of the corresponding vector as the tangent to the curve. But this "tangent vector" isn't an intrinsic object within the space – straight arrows don't fit into curved spaces. So for each curve we define a vector as a linear map from functions to the reals – the directional derivative of f along the curve C. The directional derivative can be defined just using $C(\lambda)$:

$$v(f) = \lim_{\delta\lambda \to 0} \frac{f(C(\lambda + \delta\lambda) - f(C(\lambda)))}{\delta\lambda}$$

It is straightforward to show that the set of directional derivatives forms a vector space. In coordinates, we're used to thinking of the directional derivative as just

$$\mathbf{v} \cdot \nabla f$$

and this is just right if we replace $\mathbf{v}$ by the tangent vector, $\frac{dx^i}{d\lambda}$:

$$v(f) = \frac{dx^i}{d\lambda} \frac{\partial f}{\partial x^i}$$

We can abstract v as the differential operator

$$v = \frac{dx^i}{d\lambda} \frac{\partial}{\partial x^i}$$

and think of $\frac{dx^i}{d\lambda}$ as the components of v and $\frac{\partial}{\partial x^i}$ as a set of basis vectors.

For both forms and vectors, the linear character of integration and differentiation guarantee the algebraic properties of vector spaces, while the usual chain rule applied to the basis vectors,

$$\begin{aligned} dx^i &= \frac{\partial x^i}{\partial y^k} dy^k \\ \frac{\partial}{\partial x^i} &= \frac{\partial y^k}{\partial x^i} \frac{\partial}{\partial y^k} \end{aligned}$$

together with the coordinate invariance of the formal symbols ω and v, shows that the components of ω and v transform to a new coordinate system $y^i\left(x^k\right)$ according to

$$\begin{aligned} \tilde{v}^k &= \frac{dy^k}{d\lambda} = \frac{\partial y^k}{\partial x^m} \frac{dx^m}{d\lambda} = \frac{\partial y^k}{\partial x^m} v^m \\ \tilde{\omega}_k &= \frac{\partial}{\partial y^k} = \frac{\partial x^k}{\partial y^i} \frac{\partial}{\partial x^k} = \frac{\partial x^k}{\partial y^i} \omega_k \end{aligned}$$

Since the Jacobian matrix, $J^k{}_m$, and its inverse are given by

$$\begin{aligned} J^k{}_m &= \frac{\partial y^k}{\partial x^m} \\ \bar{J}^k{}_m &= \frac{\partial x^k}{\partial y^m} \end{aligned}$$

we can write the transformation laws for vectors and forms as

$$\tilde{v}^k = J^k{}_m v^m \quad (1)$$

$$\tilde{\omega}_k = \bar{J}^m{}_k \omega_m \quad (2)$$

In general, any object which transforms according to eq.(1) is called *contravariant*, while any object which transforms according to eq.(2) is called *covariant*. There are two uses of the word covariant – here the term refers to transformation with the inverse Jacobian matrix, but the term is also used generically to refer to tensors. Thus, any object which transforms linearly under a group action may be said to transform covariantly under the group. The context usually makes clear which meaning is intended.

The geometric approach taken here shows that, corresponding to the two types of transformation there are two types of geometric object. Both of these are familiar from vector calculus – the vectors that are used in introductory physics and denoted by arrows, $\overrightarrow{v}$ are vectors, while differentials of functions, df, are forms. We shall show below that we can pass from one type to the other whenever our space has one other bit of structure: a metric.

Prove that

$$\begin{aligned} J^k{}_m &= \frac{\partial y^k}{\partial x^m} \\ \bar{J}^k{}_m &= \frac{\partial x^k}{\partial y^m} \end{aligned}$$

are actually inverse to one another.

Prove that the set of directional derivatives satisfies the algebraic definition of a vector space.

Prove that the set of forms satisfies the algebraic definition of a vector space.

There is a natural duality between the coordinate basis for vectors and the coordinate basis for forms. We define the bracket between the respective basis vectors by

$$\left\langle \frac{\partial}{\partial x^j}, dx^i \right\rangle = \delta^i_j$$

This induces a linear map from $V^* \times V_*$ into the reals,

$$\langle , \rangle : V^* \times V_* \to R$$

given by

$$\begin{aligned} \langle v, \omega \rangle &= \left\langle v^j \frac{\partial}{\partial x^j}, \omega_i dx^i \right\rangle \\ &= v^j \omega_i \left\langle \frac{\partial}{\partial x^j}, dx^i \right\rangle \\ &= v^j \omega_i \delta^i_j \\ &= v^i \omega_i \end{aligned}$$

If we pick a particular vector v, then $\langle v, \cdot \rangle$ is a linear mapping from forms to the reals. Since there is exactly one linear map for each vector v, it is possible to define the space of vectors V^* as the set of linear mappings on forms. The situation is symmetric – we might also choose to define forms as the linear maps on vectors. However, both vectors and forms have intrinsic geometric definitions, independently of one another.

Notice that all of the sums in this section involve one raised and one lowered index. This must always be the case, because this is the only type of "inner product" that is invariant. For example, notice that if we transform between $(\tilde{v}, \tilde{\omega})$ and (v, w) the bracket is invariant:

$$\begin{aligned} \langle \tilde{v}, \tilde{\omega} \rangle &= \tilde{v}^i \tilde{\omega}_i \\ &= \left(J^i{}_m v^m \right) \left(\bar{J}^n{}_i \omega_n \right) \\ &= \bar{J}^n{}_i J^i{}_m v^m \omega_n \\ &= \delta^n_m v^m \omega_n \\ &= v^m \omega_m \\ &= \langle v, \omega \rangle \end{aligned}$$