# Physics 303/573

Notes on Linear Algebra 1

September 7, 2016

## 1 Introduction

Physics is a description of the world around us; thousands of years of experience has shown that the appropriate language for detailed quantitative descriptions is mathematics; this first became clear with Newton's astonishing synthesis of terrestrial and planetary dynamics–he found that the same mathematical formulae describe the motion of objects on earth such as falling apples and the motion of the planets. Think about it–with a few simple symbols, you can compute the results of countless observations. It really led us to realize that we can understand the world around us, that things don't just happen at random but according to simple useful rules. Why did the chicken cross the road? Because it wanted to. Why does the moon orbit the earth? Because of Newton's laws.

Today's lecture is devoted to the mathematical background that you will need for the course, in particular to vectors and Linear Algebra. Linear Algebra is a wonderful subject: on the one hand, it is simple–it requires nothing more than arithmetic–but on the other hand, it teaches how to think abstractly and how to generalize.

## 2 Vectors

A vector is an arrow–it has a length (magnitude) and a direction. An example is a displacement vector: how do you get there from here. The key property of vectors is that they can be added: namely, you can combine two vectors to get a third vector. It's easy to see how for displacement vectors–go from here (O) to some place (A), and then continue on to another place (B). If we call the displacement vector from (O) to (A) $\vec{v}_A$ and the displacement vector from (A) to (B) $\vec{v}_B$, then the displacement vector from (O) to (B) is just $\vec{v}_A + \vec{v}_B$. It is clear that there is a notion of a zero vector $\vec{0}$: It's just the instructions not to go anywhere. Similarly, it's clear what minus a vector is: the vector you get by going backwards along the displacement vector. Furthermore, addition is commutative: it doesn't matter in which order we add vectors $\vec{v} + \vec{w} = \vec{w} + \vec{v}$ (see Fig. 1). It is also associative $(\vec{v} + \vec{w}) + \vec{u} = \vec{v} + (\vec{w} + \vec{u})$: (see Fig. 2).

If we can add vectors, we can add a vector to itself–and it makes sense to talk about multiplying a vector by a number (also called a scalar); $n\vec{v}$ is just the vector $\vec{v}$ added to itself $n$

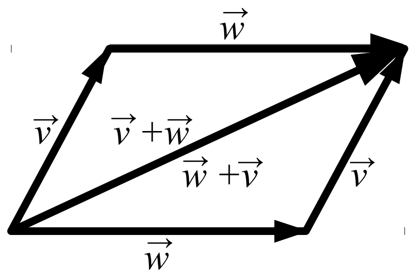


Fig. 1: Vector addition is commutative.

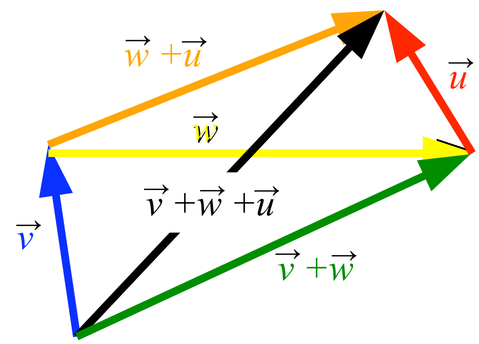


Fig. 2: Vector addition is associative.

times. Multiplying a vector by a number means just multiplying the length by the number and not changing the direction. What vector do we get if we multiply any vector by 0?

This is the basic linear structure–adding vectors and multiplying them by scalars. We call $a\vec{v} + b\vec{w}$, where $a$ and $b$ are scalars and $\vec{v}$ and $\vec{w}$ are vectors, a *linear combination* of the vectors $\vec{v}, \vec{w}$.

Just as we can think of numbers as points on the real number line, or elements of the *space of numbers*, we can think of vectors as elements of a space of vectors, or *vector space*. This is a language that is often used in mathematics.

## 2.1 Linear Operators

A linear operator is any operation on vectors that respects the linear structure:

$$L(a\vec{v} + b\vec{w}) = aL(\vec{v}) + bL(\vec{w}) \ . \tag{2.1}$$

This means that if we take a linear combination $a\vec{v} + b\vec{w}$ of two vectors $\vec{v}, \vec{w}$ and then apply the linear operator $L$ to the result, we get the same vector as if we first apply $L$ to $\vec{v}, \vec{w}$ and then take the linear combination $aL(\vec{v}) + bL(\vec{w})$.

Examples of linear operators are easy–multiplying by a scalar, for example, or rotating a vector by a fixed angle, or doing a combination of both. Other examples are more subtle: If $a$ and $b$ are fixed time-independent numbers, and the vectors vary as functions of time, then the time derivative is a linear operator. Linear operators need not take vectors into vectors–just into anything that we know how to add and multiply by scalars, for example, ordinary numbers.

## 2.2 Basis of a vector space

We can ask if several vectors are *linearly independent*–can we find some linear combination so that it gives the zero vector $\vec{0}$; if yes, they are linearly dependent–at least one can be written as a linear combination of the others; if not, they are linearly independent. We call the largest number of possible linearly independent vectors excluding $\vec{0}$ the dimension $d$ of the vector space, and say the vector space is $d$-dimensional. In general, $d$ can be infinite. For finite $d$ we can choose a *basis*[1], namely a particular set $\{\hat{e}_i\}$, $i = 1...d$ of $d$ linearly independent non-zero vectors in terms of which we express an arbitrary vector $\vec{v}$:

$$\vec{v} = \sum_{i=1}^{d} v^i \hat{e}_i =: v^i \hat{e}_i \ ; \tag{2.2}$$

we often leave off the explicit summation symbol (this notation was introduced by Einstein and is called the Einstein summation convention). The coefficients $v^i$ are called the *components* of the vector in the given basis.

The basis is very useful, as it lets us work with the components, which are ordinary numbers, rather than the more abstract notion of vectors; however, it hides the fact that a vector can be described in any basis with (in general) different components. We can prove a simple theorem:

Theorem: The expression for a vector in terms of a basis is unique. Proof: Imagine that there were two different expressions $\vec{v} = \sum_i v^i \hat{e}_i = \sum_i w^i \hat{e}_i$ for the same vector. Then take the difference–this is zero, since we are taking the difference of a vector and itself:

$$\vec{0} = \vec{v} - \vec{v} = \sum_i (v^i - w^i)\hat{e}_i \tag{2.3}$$

Unless all the $v^i = w^i$ we find that the basis was not linearly independent, which would be a contradiction. As a corollary, we see that all the components of the zero vector $\vec{0}$ are zero.

If we write

$$\vec{v} = v^i \hat{e}_i \ , \quad \vec{w} = w^i \hat{e}_i$$

then because of the linear structure, we have:

$$a\vec{v} + b\vec{w} = \sum_{i=1}^{d} (av^i + bw^i)\hat{e}_i \ , \tag{2.4}$$

where I wrote the summation out in the last line just as a reminder (I'll do that in a few equations below as well, but mostly I will use the Einstein summation convention). So to add two vectors, in any basis, you just add the components. Similarly, to multiply a vector by a scalar, just multiply all the components by the scalar.

The basis is particularly useful because of the linear structure–if we know how a linear operator acts on the basis, we know how it acts on any vector. Consider a linear operator that takes vectors into vectors. We can act on each of the basis vectors to get a new vector, and then write it in terms of its components and the basis:

$$L(\hat{e}_i) = \sum_{j=1}^{d} L_i{}^j \hat{e}_j \ , \tag{2.5}$$

[1] With proper care, this can even be done for certain infinite dimensional vector spaces, such as the space of periodic functions on the unit interval–think of the Fourier series decomposition of such functions.

that is, acting with $L$ on a basis vector $\hat{e}_i$ , we get a new vector which we express in terms of the basis. Then on an arbitrary vector we have:

$$L(\vec{v}) = L(v^i\hat{e}_i) = v^i L(\hat{e}_i) = v^i \sum_{j=1}^{d} L_i{}^j \hat{e}_j = \sum_{i,j=1}^{d} v^i L_i{}^j \hat{e}_j \ , \tag{2.6}$$

which many of you may recognize as matrix multiplication of the row vector $v^i$ with the matrix $L_i{}^j$ ($L$ sits to the right, so it is the transpose of what you are more used to).

## 2.3 Length and Scalar product

An important notion is the *length* or magnitude of a vector, which we write as $|\vec{v}|$; it has certain properties: the length of a vector times a number is the absolute value of the number times the length of the vector: $|a\vec{v}| = |a||\vec{v}|$, the only vector with length zero is the zero vector $\vec{0}$, and the length obeys the triangle inequality:

$$|\vec{v}| + |\vec{w}| \geq |\vec{v} + \vec{w}|,$$

with equality only if the two vectors point in the same direction. Taking the length of a vector is *not* a linear operation–the triangle inequality measures the violation of the linear relation (2.1).

A vector with length 1 is called a unit vector; if we take any nonzero vector and divide it by its length, we get a unit vector:

$$\hat{v} = \left(\frac{1}{|\vec{v}|}\right) \vec{v} \ . \tag{2.7}$$

We will work with vectors that come equipped with a scalar product, which takes two vectors and spits out a number. It satisfies a number of important properties; it is commutative, and for three vectors $\vec{v}, \vec{w}, \vec{u}$

$$(a\vec{v} + b\vec{w}) \cdot \vec{u} = a(\vec{v} \cdot \vec{u}) + b(\vec{w} \cdot \vec{u}) \ . \tag{2.8}$$

It obeys the Schwarz inequality:

$$(\vec{v} \cdot \vec{v})(\vec{w} \cdot \vec{w}) \geq (\vec{v} \cdot \vec{w})^2 \ .$$

We will always choose

$$|\vec{v}| = \sqrt{\vec{v} \cdot \vec{v}} \ ; \tag{2.9}$$

then the Schwarz inequality implies the triangle inequality. The extent to which the Schwarz inequality is not an equality is measured by the angle $\theta$ between the vectors:

$$\cos\theta = \frac{\vec{v} \cdot \vec{w}}{|\vec{v}||\vec{w}|} \tag{2.10}$$

Two (nonzero) vectors are perpendicular or orthogonal to each other if their scalar product vanishes: $\vec{v} \cdot \vec{w} = 0$ implies $\vec{v}$ is perpendicular to $\vec{w}$.

## 2.4 Orthonormal basis

It is very useful to use an orthonormal basis, that is to choose all our basis vectors to have length 1 and to be mutually perpendicular:

$$\hat{e}_i \cdot \hat{e}_j = \delta_{ij} \tag{2.11}$$

where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$. Then it is easy to show that the component of a vector along a basis vector is just its scalar product with the basis vector:

$$\vec{v} = v^i \hat{e}_i \;\Leftrightarrow\; v^i = \vec{v} \cdot \hat{e}_i \tag{2.12}$$

Proof: Just calculate it!

Notice that in equation (2.12), I am mixing subscripts and superscripts–we can do this because we are using an orthonormal basis (2.11); from now on, it won't make a difference whether I write a subscript or a superscript on the components of a vector, *etc.* I comment on this further below.

Then we can write the scalar product of two vectors in terms of the components in a basis

$$\vec{v} \cdot \vec{w} = v^i w^i \ . \tag{2.13}$$

In particular, we can compute the length (magnitude) of a vector in an orthonormal basis and find

$$|\vec{v}| = \sqrt{v^i v^i}$$

In three dimensions, it is common to give special names to unit vectors along the coordinate axes:

$$\hat{e}_x = \hat{x} = \hat{\imath} \quad , \quad \hat{e}_y = \hat{y} = \hat{\jmath} \quad , \quad \hat{e}_z = \hat{z} = \hat{k} \quad . \tag{2.14}$$

An easy calculation shows that the matrix corresponding to a linear transformation is given by

$$L_{ij} = L(\hat{e}_i) \cdot \hat{e}_j \tag{2.15}$$

Proof:

$$L(\hat{e}_i) \cdot \hat{e}_j = \sum_k L_i{}^k \hat{e}_k \cdot \hat{e}_j = \sum_k L_i{}^k \delta_{kj} = L_{ij}$$

## 2.5 Rotations

Orthogonal transformations are linear transformations that preserve the lengths of vectors and the angles between them; in particular, this means that they take an orthonormal basis into another orthonormal basis. Orthogonal transformations consist of rotations and reflections. If $\hat{e}'_i = R(\hat{e}_i) = R_i{}^j \hat{e}_j$, then

$$\delta_{ij} = \hat{e}'_i \cdot \hat{e}'_j = \sum_{m,n} R_i{}^m \hat{e}_m \cdot R_j{}^n \hat{e}_n = \sum_{m,n} R_i{}^m R_j{}^n \delta_{mn} = \sum_m R_i{}^m R_{jm} = \sum_m R_i{}^m R^T_{mj} = (RR^T)_{ij}$$

where $R^T$ is the transpose of the matrix $R$, and $RR^T$ is the matrix product of $R$ and $R^T$. (For clarity, I've again put in the explicit summations–in the Einstein summation convention, I wouldn't bother.)

Using (2.15), we can find a simple expression for $R$:

$$R_{ij} = \hat{e}'_i \cdot \hat{e}_j \tag{2.16}$$

We can find the transformation of any vector using the linear structure (see 2.6):

$$R(\vec{v}) = v^i R(\hat{e}_i) = v^i R_i{}^j \hat{e}_j$$

Here we think of the linear operator $R$ as *actively* transforming the vector $\vec{v}$ to a new vector $R(\vec{v})$.

However, we can also think of changing just the basis and not the vector; this is called the *passive* view, and we will use it most of the time. It is not quite as natural to visualize, but it is actually easier to work with. Then we have

$$\vec{v} = v^i \hat{e}_i = \vec{v}\,' := v'^i \hat{e}'_i = v'^i R(\hat{e}_i) = v'^i R_i{}^j \hat{e}_j$$

which implies

$$v^i = v'^j R_j{}^i = (R^T)^i{}_j v'^j \quad \Leftrightarrow \quad v'^i = R^i{}_j v^j \tag{2.17}$$

where we use the orthogonality of $R$: $R^T R = 1$ in the last step of (2.17).

Notice that the $R^T R = 1$ implies that the determinant of $R$ squares to 1; rotations are orthogonal transformations with determinant 1, whereas orthogonal transformations with determinant $-1$ involve a reflection.

If we restrict our attention to orthonormal bases and orthogonal transformations that preserve them and hence $\delta_{ij}$, we do not need to worry whether we write indices upstairs or downstairs; more generally, we have to be careful about this. Here we consider on orthonormal bases and orthogonal transformations.

## 2.6 Cross product

Up until now, everything that we have done makes sense in any (finite) number of dimensions (most of it works in infinite dimensions too, but you have to be more careful). Now we will consider some linear structure that is special to three dimensions, though in a broader sense, it can also be generalized to higher dimensions.

The cross product or vector product is a product that takes two vectors and gives back a vector (actually, there is a slight subtlety that I'll discuss briefly later); it can be defined by the following properties:

$$\begin{aligned} \text{Antisymmetry:} \quad & \vec{v} \times \vec{w} = -\vec{w} \times \vec{v} \\ \text{Linearity:} \quad & (a\vec{v} + b\vec{w}) \times \vec{u} = a(\vec{v} \times \vec{u}) + b(\vec{w} \times \vec{u}) \\ & \hat{x} \times \hat{y} = \hat{z} \;\; , \quad \hat{y} \times \hat{z} = \hat{x} \;\; , \quad \hat{z} \times \hat{x} = \hat{y} \end{aligned} \tag{2.18}$$

Using these properties, we can find the cross product of any two vectors:

$$\begin{aligned}\vec{v} \times \vec{w} &= (v_x\hat{x} + v_y\hat{y} + v_z\hat{z}) \times (w_x\hat{x} + w_y\hat{y} + w_z\hat{z}) \\ &= v_x(w_y\hat{z} - w_z\hat{y}) + v_y(-w_x\hat{z} + w_z\hat{x}) + v_z(w_x\hat{y} - w_y\hat{x}) \\ &= (v_yw_z - v_zw_y)\hat{x} + (v_zw_x - v_xw_z)\hat{y} + (v_xw_y - v_yw_x)\hat{z} \qquad (2.19)\end{aligned}$$

This definition is not obviously independent of the basis that we use–in fact, it is for rotations, but changes sign for reflections.

Another way to define the cross product makes this clear. We begin by defining the alternating tensor:

$$\epsilon^{ijk} = \begin{cases} +1 & \text{if } ijk = xyz, yzx, zxy \text{ (even permutation of } xyz) \\ -1 & \text{if } ijk = yxz, zyx, xzy \text{ (odd permutation of } xyz) \\ 0 & \text{otherwise, that is, if any two of } ijk \text{ are equal} \end{cases} \qquad (2.20)$$

A crucial property of the alternating tensor is that any object with the same symmetry properties, that is which is totally antisymmetric in three indicies, must be proportional to the alternating tensor.

The determinant of any $3 \times 3$ matrix $M_{ij}$ can be written in terms of the alternating tensor as:

$$\det M = \frac{1}{6}\,\epsilon^{ijk}\epsilon^{pqr}M_{ip}M_{jq}M_{kr} \qquad (2.21)$$

Actually, there is a stronger identity that holds:

$$(M^{-1})^{rk}\det M = \frac{1}{2}\,\epsilon^{ijk}\epsilon^{pqr}M_{ip}M_{jq} \qquad (2.22)$$

where $(M^{-1})$ is the inverse matrix of $M$:

$$(M^{-1})M = 1 \quad \Leftrightarrow \quad (M^{-1})^{ij}M_{jk} = \delta^i{}_k$$

We now define the cross product on basis vectors by:

$$\hat{e}_i \times \hat{e}_j = \epsilon_{ijk}\hat{e}_k \qquad (2.23)$$

You can easily check that this agrees with (2.18) for the usual Cartesian basis vectors. Then the cross product of two general vectors is

$$\vec{v} \times \vec{w} = v^i\hat{e}_i \times w^j\hat{e}_j = v^iw^j\epsilon_{ijk}\hat{e}_k \qquad (2.24)$$

Choosing $i = x, j = y, k = z$, we see that (2.23) implies

$$\hat{e}_k = \frac{1}{2}\,\epsilon^{ijk}\,\hat{e}_i \times \hat{e}_j \qquad (2.25)$$

We can also take the scalar product with another unit vector and find that (2.23) implies

$$(\hat{e}_i \times \hat{e}_j) \cdot \hat{e}_k = \epsilon_{ijk} \qquad (2.26)$$

Let's see how the cross product transforms under a orthogonal transformation. Since any object with the same symmetries as the alternating tensor is proportional to it, we have $\epsilon'_{ijk} = c\epsilon_{ijk}$ for some constant $c$. The right hand side of (2.25) gives:

$$\frac{1}{2}\,\epsilon'^{ijk}\,\hat{e}'_i \times \hat{e}'_j = \frac{1}{2}\,c\,\epsilon^{ijk}\,R_i{}^m R_j{}^n \hat{e}_m \times \hat{e}_n = \frac{1}{2}\,c\,\epsilon^{ijk} R_i{}^m R_j{}^n \epsilon_{mnp}\hat{e}_p$$

Using the symmetry properties of the alternating tensor and (2.22) we can rewrite this as

$$c\,(R^{-1})^p{}_k(\det R)\hat{e}_p$$

But for an orthogonal transformation, $R^{-1} = R^T$ and $\det R = \pm 1$, so we find

$$\hat{e}'_k = c\,(\det R) R_k{}^p \hat{e}_p \quad \Rightarrow \quad c = \det R \tag{2.27}$$

that is, the cross product is defined in a way independent of the basis if the alternating tensor transforms as:

$$\epsilon'_{ijk} = R_i{}^p R_j{}^q R_k{}^r\,\epsilon_{pqr} = (\det R)\,\epsilon_{ijk} \tag{2.28}$$

## 2.7 Triple product

The triple product of three vectors $\vec{v}, \vec{w}, \vec{u}$ is simply $\vec{v} \cdot (\vec{w} \times \vec{u})$. However, it is very simple using the alternating tensor; from (2.24) we find

$$\vec{v} \cdot (\vec{w} \times \vec{u}) = \epsilon_{ijk} w^i u^j v^k = \epsilon_{ijk} v^i w^j u^k \quad , \quad etc. \tag{2.29}$$

This makes all the symmetries immediately obvious.

## 2.8 Some simple identities and their uses

The alternating tensor obeys certain identities; for example,

$$\epsilon_{ijk}\epsilon^{ijm} = 2\delta_k{}^m \quad , \quad \epsilon_{ijk}\epsilon^{imn} = \delta_j{}^m \delta_k{}^n - \delta_j{}^n \delta_k{}^m \tag{2.30}$$

which can easily be checked by trying out a few cases and using rotational invariance to deduce the rest–you can do this as an extra-credit homework problem. There is also a similar expression involving six terms on the left hand side for $\epsilon_{ijk}\epsilon^{mnp}$–see if you can work it out.