Tomi Johnson[1]

Keble College - Michaelmas 2014

CP3&4: Mathematical methods I&II

Tutorial 3 - Vectors and matrices I

Prepare full solutions to the 'problems' with a self assessment of your progress on a cover page.
Leave these at Keble lodge by 5pm on Monday of 2nd week.
Look at the 'class problems' in preparation for the tutorial session.
Suggested reading: RHB 7 and the lecturer's problem sets.

Goals

- Start to think about vectors more abstractly than merely as arrows pointing towards points in coordinate space.
- Identify and close gaps in your understanding of basic operations involving three dimensional vectors, and their relation to points in coordinate space.
- Learn how to describe lines, planes and surfaces using vectors and vector operations.
- Become proficient in using vectors and vector operations, in advance of their use in formulating geometric and kinematic problems.

Problems

Let us start by examining mathematically what we mean by a real vector $|v\rangle$, such as could be used to describe a velocity, force, spatial translation, or magnetic field. Intuitively, a vector $|v\rangle$, e.g. a specific velocity, is one of a set $\mathcal{V}$ of many vectors, e.g. the set of all possible velocities, which we call a vector space. Again, matching our intuitions, this vector space can be supposed to satisfy a few reasonable requirements: adding two vectors gives another vector in the vector space, multiplying a vector by a real constant gives another vector in the vector space, there is a zero vector $|0\rangle$ such that $|v\rangle + |0\rangle = |v\rangle$, and there are also the usual properties of associativity, commutativity and distributivity. That's all we need to begin talking about vectors abstractly.

From its featuring in the definition of a vector space, the notion of adding some vectors to obtain another is an important one. It leads us to the idea of a complete basis $\mathcal{B}$, a set of basis vectors $|e_i\rangle$ from which all other vectors in the vector space $\mathcal{V}$ can be built via linear superposition $|v\rangle = \sum_i v_i |e_i\rangle$ (a more fancy way of saying it: the basis vectors span the vector space). This idea builds on every day intuition e.g. to get to the Sheldonian from Keble, go South down Parks Rd for 500m then West along Broad St for 50m. To avoid redundancy, we can insist that no basis vector can be built from the others via linear superposition, i.e. they are linearly independent. This is equivalent to there being no solution to $|0\rangle = \sum_i v_i |e_i\rangle$ other than the so-called trivial solution $v_i = 0$ and it implies that a decomposition $|v\rangle = \sum_i v_i |e_i\rangle$ is unique. A vector space $\mathcal{V}$ will permit a maximum number N of linearly independent vectors and this number will equal the minimum number of vectors required to form a basis i.e. the exact number of vectors required to form a linearly independent complete basis (all of this can be proved without introducing any more assumptions). N is called the dimension of the vector space.

[1] These problems were compiled by Prof. D. Jaksch based on problem sets by Prof. N. Harnew and past Oxford Prelims exam questions.

Given some basis $\mathcal{B}$ it is then possible to represent a vector $|v\rangle$ by its components v_i, written as a column vector

$$\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix}.$$

To save space, column vectors are usually typeset as $\mathbf{v} = (v_1, \ldots, v_N)^T$ where ^{T} indicates the transpose. Even the transpose sign ^{T} is often left out when it is obvious the object is a column vector. Due to the difficulty of writing in bold font by hand, column vectors $\mathbf{v}$ are also written $\underline{v}$ or $\vec{v}$. It doesn't matter which you use, just be consistent. If two vectors are added, i.e. $|w\rangle = z_u |u\rangle + z_v |v\rangle$ with $\mathbf{u} = (u_1, \ldots, u_N)^T$ and $\mathbf{v} = (v_1, \ldots, v_N)^T$ in basis $\mathcal{B}$, then the column vector representing the vector sum in the same basis $\mathcal{B}$ is that with the components of the original column vectors added $\mathbf{w} = (z_u u_1 + z_v v_1, \ldots, z_u u_N + z_v v_N)^T$.

The difference between what I have called a vector $|v\rangle$ and a column vector $\mathbf{v}$ is quite subtle. The vector $|v\rangle$ is the more fundamental object, while the column vector $\mathbf{v}$ (and its components v_i) is a representation of that object for a particular choice of basis $\mathcal{B}$. The vector $|v\rangle$ doesn't depend on which basis $\mathcal{B}$ you have chosen, but column vector $\mathbf{v}$ does. Change basis $\mathcal{B}$ and your column vector representation $\mathbf{v}$ of un-changed vector $|v\rangle$ must change. When writing a column vector representation $\mathbf{v}$ you should always state (if not obvious) the basis that you are using, otherwise the vector $|v\rangle$ being represented is not clear. RHB makes the same distinction as I have done, denoting what I have called $|v\rangle$ by $\mathbf{v}$ and what I have called $\mathbf{v}$ by sans-serif symbol v. The lecturer often deals with the fundamental vectors (what I call $|v\rangle$) being itself a column vector and thus denoted $\mathbf{v}$. I represent what I call $|v\rangle = \sum_i v_i |e_i\rangle$ and the lecturer represents what he calls $\mathbf{v} = \sum_i v_i \hat{\mathbf{e}}_i$ by the same column vector $(v_1, \ldots, v_N)^T$. However, the lecturer has to call this by a name other than $\mathbf{v}$, since the basis $\{\hat{\mathbf{e}}_i\}$ being used might be different to the usual, i.e. $\hat{\mathbf{e}}_1 = (1, 0, \ldots, 0)^T$, $\hat{\mathbf{e}}_2 = (0, 1, \ldots, 0)^T$, etc and thus it is possible that $(v_1, \ldots, v_N)^T$ is not the same as $\mathbf{v}$. To try and avoid this confusion is why I call denote the fundamental object of a vector by a different symbol, even if it could be written as a column vector. All these approaches are valid, if you are careful and clear, and in practice you will probably use the lecturer's notation in your first year. However, it is instructive to use the (Dirac) notation I have introduced here for as much as the problem set as possible to grasp the concepts and prepare yourself for second year where this notation will be used a lot.

So far we have defined vectors and this is enough to use them. However, we are used to assigning some values of length to a single vector or angles to pairs of vectors. It is clear that to do this we need an extra ingredient in addition to the vector space, some quantities or functions of vectors. Next we will introduce the minimal number of quantities and their minimal properties in order to be able to talk about lengths and angles. In normal coordinate space we go in the other direction, already knowing what a length and angle is but seeking a mathematical expression. However, by starting with the mathematics and making minimal assumptions we will recover the mathematical expressions for length and angles in coordinate space but also show when the same intuitive concepts of length and angle can be applied in more abstract settings.

In this vein, let's introduce a scalar function of two vectors $f(|v\rangle, |u\rangle) = \langle u|\, v\rangle$, called the inner product, and see what this enables us to do. The only single-vector quantity available is $\langle v|\, v\rangle$ and since we want this to give us a length we require it to be non-negative $\langle v|\, v\rangle \geq 0$ with equality only for $|v\rangle = |0\rangle$. This enables us to think of $||\, |v\rangle\, || = \sqrt{\langle v|\, v\rangle}$ as the length (or norm) of a vector, with $|0\rangle$ having zero length but every other vector a positive length. Let's see if we can also use this function to define an angle between two vectors. In familiar real vector spaces, the angle between vectors $|v\rangle$ and $|u\rangle$ is the same as that between $|u\rangle$ and $|v\rangle$, so the inner product might be expected to be symmetric $\langle u|\, v\rangle = \langle v|\, u\rangle$. Also, if we want length to have the usual properties, e.g. that two copies of a vector added have twice the length of the original vector $||2\,|v\rangle\,|| = 2||\,|v\rangle\,||$, then we require that $f(|v\rangle, |u\rangle)$ is linear in its arguments. Amazingly, after only this, we are already in a position to define an angle between two vectors, a fact that is revealed by answering the following question.

1. $*$ Prove the Cauchy-Schwarz inequality $|\langle u|\, v\rangle| \leq ||\,|v\rangle\,||\,||\,|u\rangle\,||$, and determine when the equality holds.

Solution: Here are three methods of increasing length. Method 1: Let $|a\rangle$ and $|b\rangle$ be vectors of unit length, $|a\rangle = |u\rangle / ||\,|u\rangle\,||$, $|b\rangle = |v\rangle / ||\,|v\rangle\,||$. Since it is a length squared $||\,|a\rangle \pm |b\rangle\,||^2 \geq 0$. Expanding this using the linearity and symmetry of the inner product, we obtain $|\langle a|\, b\rangle| \leq 1$, which rearranges to the Cauchy-Schwarz inequality. The length is zero and thus equality holds only when $|a\rangle = \pm|b\rangle$, i.e. when $|u\rangle$ and $|v\rangle$ are parallel, $|v\rangle \propto |u\rangle$. Method 2: For the case $|u\rangle \neq |0\rangle$, consider the quadratic function $p(z) = ||z\,|u\rangle + |u\rangle\,||^2 = z^2||\,|u\rangle\,||^2 + 2z\,\langle u|\, v\rangle + ||\,|v\rangle\,||^2$, where we have used the linearity and symmetry of the inner product. The lowest value taken by the quadratic is $p(-\langle u|\, v\rangle / ||\,|u\rangle\,||^2)$ (complete the square to show this), which must be larger than zero as the function is a length squared. Setting $p(-\langle u|\, v\rangle / ||\,|u\rangle\,||^2) \geq 0$ rearranges to the Cauchy-Schwarz inequality. Equality holds when $p(-\langle u|\, v\rangle / ||\,|u\rangle\,||^2) = 0$, which is when $|v\rangle = \langle u|\, v\rangle\,|u\rangle / ||\,|u\rangle\,||^2$, which occurs when $|u\rangle$ and $|v\rangle$ are parallel, $|v\rangle \propto |u\rangle$. It is also simple to show directly using linearity that the Cauchy-Schwarz inequality holds for the yet unexamined case $|u\rangle = |0\rangle$. Method 3: Divide up $|v\rangle = |v_\parallel\rangle + |v_\perp\rangle$, where $|v_\parallel\rangle = \langle v|\, u\rangle\,|u\rangle / ||\,|u\rangle\,||^2$ and $|v_\perp\rangle = |v\rangle - |v_\parallel\rangle$, assuming $||\,|u\rangle\,||^2 \neq 0$. Later we'll think of these as the projection of $|v\rangle$ onto the direction of $|u\rangle$ and onto the space orthogonal to $|u\rangle$, respectively. The linearity and symmetry of the scalar product means that $||\,|v\rangle\,||^2 = ||\,|v_\parallel\rangle\,||^2 + ||\,|v_\perp\rangle\,||^2 + 2\,\langle v_\parallel|\, v_\perp\rangle$. The same two properties also enable us to show $\langle v_\parallel|\, v_\perp\rangle = 0$ (later we'll come to think of this as their being orthogonal) and so $||\,|v\rangle\,||^2 = ||\,|v_\parallel\rangle\,||^2 + ||\,|v_\perp\rangle\,||^2 \geq ||\,|v_\parallel\rangle\,||^2$, where we have used the non-negativity of length and equality holds only when $|v_\perp\rangle = 0$. This rearranges to the Cauchy-Schwarz inequality. $|v_\perp\rangle = 0$ and thus the equality holds when $|v\rangle = |0\rangle$ or when $|v\rangle \propto |u\rangle$. It is also simple to show directly using linearity thatthe Cauchy-Schwarz inequality holds for the yet unexamined case $|u\rangle = |0\rangle$.

We see that, following only from the basic assumptions we have made about it, the inner product $\langle u|\, v\rangle$ is constrained to be between $-||\,|v\rangle\,||\,||\,|u\rangle\,||$, obtained when $|u\rangle = c\,|v\rangle$ with c non-positive, and $||\,|v\rangle\,||\,||\,|u\rangle\,||$, obtained when $|u\rangle = c\,|v\rangle$ with c non-negative. It follows therefore that we can think of a quantity $\langle u|\, v\rangle / ||\,|v\rangle\,||\,||\,|u\rangle\,|| = \cos(\theta_{vu})$ as being the fractional projection of $|u\rangle$ onto $|v\rangle$ and defining an angle θ_{vu} between any two non-zero vectors that is entirely in keeping with our intuition. For example, the angle between a vector and itself is 0 and the angle between a vector and its negative is π.

In fact, all of the usual properties of vectors follow from the few minimal assumptions we have made above. Here is another example of a notion we intuitively expect to be true.

2. $*$ Prove the Pythagorean theorem, i.e. $||\,|v\rangle + |u\rangle\,||^2 = ||\,|v\rangle\,||^2 + ||\,|u\rangle\,||^2$ iff $|v\rangle$ and $|u\rangle$ are orthogonal. [Note that 'iff' stands for 'if and only if', so whenever used you must prove both the 'if' and 'only if' parts. It is a common mistake to only show one direction.]

Solution: By the linearity of the inner product, $||\,|v\rangle + |u\rangle\,||^2 = ||\,|v\rangle\,||^2 + ||\,|u\rangle\,||^2 + 2\,\langle u|\, v\rangle$. The last term is zero iff the vectors are orthogonal, leaving Pythagoras' theorem.

3. $*$ Prove the triangle inequality $||\,|v\rangle + |u\rangle\,|| \leq ||\,|v\rangle\,|| + ||\,|u\rangle\,||$ and determine when the equality holds.

Solution: By the linearity of the inner product, $||\,|v\rangle + |u\rangle\,||^2 = ||\,|v\rangle\,||^2 + ||\,|u\rangle\,||^2 + 2\,\langle u|\, v\rangle$. Combining this with the Cauchy-Schwarz inequality gives $||\,|v\rangle + |u\rangle\,||^2 \leq ||\,|v\rangle\,||^2 + ||\,|u\rangle\,||^2 + 2||\,|v\rangle\,||\,||\,|u\rangle\,|| = (||\,|v\rangle\,|| + ||\,|u\rangle\,||)^2$. The triangle inequality follows. Equality occurs when $\langle u|\, v\rangle = ||\,|v\rangle\,||\,||\,|u\rangle\,||$, which is when $\theta_{vu} = 0$.

The notion of angle and therefore orthogonality brought about by introducing an inner product also allows us to impose an incredibly useful structure to vectors, as we shall see. Accordingly, we will almost always automatically restrict ourselves to bases $\mathcal{B}$ comprising orthonormal basis vectors $\langle e_j | e_i \rangle = \delta_{ij}$, often without making such an assumption explicit. As before, any vector is decomposed as $|v\rangle = \sum_i v_i |e_i\rangle$. Evaluating $\langle e_i | v \rangle$, we find it equal to the component v_i, and thus $v_i |e_i\rangle$ equal to $\langle e_i | v \rangle |e_i\rangle$, the projection of $|v\rangle$ onto the vector $|e_i\rangle$. The decomposition $|v\rangle = \sum_i v_i |e_i\rangle$ then reveals the vector is equal to the sum of its projections onto orthogonal vectors $|e_i\rangle$, with each projection having value v_i. The column vector $\mathbf{v} = (v_1, \ldots, v_N)^T$ representing $|v\rangle$ in this basis is made up of components corresponding to the various projections. This interpretation is not possible without orthonormality.

Decomposing two vectors $|v\rangle = \sum_i v_i |e_i\rangle$, $|u\rangle = \sum_i u_i |e_i\rangle$ in the same orthonormal basis, it follows that the inner product takes on the simple form $\langle u | v \rangle = \sum_i u_i v_i$ (try and show this). It is remarkable that it is only for this case of an orthonormal basis that the inner product can be solely expressed in terms of the components in a basis (in other cases we would need to also know the inner products of the basis vectors). Thus only for orthonormal bases are we able to deduce all the properties (including lengths and angles) of vectors $|v\rangle$ solely from their column vector representations $\mathbf{v}$. This fact suggests the introduction of an operation, called $\mathbf{u} \cdot \mathbf{v}$, the scalar (or dot) product, between column vectors $\mathbf{u}$ and $\mathbf{v}$ that, for an orthonormal basis, equals the inner product $\mathbf{u} \cdot \mathbf{v} = \langle u | v \rangle$ of the two vectors being represented. This operation is defined by $\mathbf{u} \cdot \mathbf{v} = (u_1, \ldots, u_N)^T \cdot (v_1, \ldots, v_N)^T = \mathbf{u}^T \mathbf{v} = \sum_i u_i v_i$ (though other conventions exist). Accordingly, we can introduce the length (magnitude) of a column vector as $v = |\mathbf{v}| = \sqrt{\mathbf{v}^T \mathbf{v}} = || \, |v\rangle \, ||$ and obtain the angle θ_{uv} from $\cos(\theta_{uv}) = \mathbf{u}^T \mathbf{v} / uv$, which equal the length and angle of the real vectors so long as the basis being used is orthonormal. Thus the choice of an orthonormal basis allows us to deal entirely with column vector representations and use the scalar product in place of the inner product.

We have established that, provided we use a fixed orthonormal basis, we may deal solely with column vectors $\mathbf{v}$ representing vectors $|v\rangle$, without referring directly to $|v\rangle$, and replace the inner product with the scalar product. We now spend the rest of this problem set using such a representation and you could go back an do equivalents of the previous three questions also within this representation (see class problems). You are probably very used to thinking of a vector, e.g. velocity, force, spatial translation, or magnetic field, by its representation in an orthonormal basis. You may even feel the previous three and a bit pages it took to arrive at the decision to use such a representation were a bit of a waste of a time. However, the notion of a vector, as used in physics, is a very general one. During your course you'll come to describe functions as vectors, and you'll deal with vectors with complex components to represent quantum mechanical states. Normal intuition will fail you, but the mathematical intuition gained through the pedantic exploration of the last few pages will hold up. More immediately, when considering changes in basis, you will probably appreciate the ability to retreat from the column vector representation $\mathbf{v}$ in a particular basis to the more fundamental vectors $|v\rangle$.

Becoming even more specific, for the remainder of the problem set we will consider three-dimensional vector spaces, mainly coordinate spaces. These are standard in mechanics and electromagnetism, for obvious reasons. It is always possible to write the choice of orthonormal basis as $\mathcal{B} = \{|e_1\rangle, |e_2\rangle, |e_3\rangle\}$ so the basis vectors are represented by column vectors $\hat{\mathbf{e}}_1 = (1, 0, 0)^T$, $\hat{\mathbf{e}}_2 = (0, 1, 0)^T$, $\hat{\mathbf{e}}_3 = (0, 0, 1)^T$, where the $\hat{}$ indicates that the vector has unit length (is a unit vector).

For this particular case of 3D there is an additional function of a pair of column vectors that is of particular usefulness. The vector (cross) product is defined in the column vector representation as

$$\mathbf{u} \times \mathbf{v} = \mathbf{u} \wedge \mathbf{v} = \begin{pmatrix} u_2 v_3 - u_3 v_2 \\ u_3 v_1 - u_1 v_3 \\ u_1 v_2 - u_2 v_1 \end{pmatrix} = \begin{vmatrix} \hat{\mathbf{e}}_1 & \hat{\mathbf{e}}_2 & \hat{\mathbf{e}}_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix},$$

and can be written as a matrix determinant (a concept to be explored more in two tutorials' time). The vector product is distributive $(\mathbf{u} + \mathbf{v}) \times \mathbf{w} = \mathbf{u} \times \mathbf{w} + \mathbf{v} \times \mathbf{w}$*, but is not associative* $(\mathbf{u} \times \mathbf{v}) \times \mathbf{w} \neq \mathbf{u} \times (\mathbf{v} \times \mathbf{w})$ *and is anti-commutative* $\mathbf{u} \times \mathbf{v} = -\mathbf{v} \times \mathbf{u}$*. The vector product can also be written as* $\mathbf{u} \times \mathbf{v} = uv|\sin(\theta_{uv})|\hat{\mathbf{n}}$ *where* $\hat{\mathbf{n}}$ *is a vector of unity length that is orthogonal to both* $\mathbf{u}$ *and* $\mathbf{v}$*. Finally, also specific to 3D, the triple scalar product is given by* $\{\mathbf{uvw}\} = \mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$ *and has the useful property that* $\{\mathbf{uvw}\} = -\{\mathbf{vuw}\} = \{\mathbf{wuv}\}$*. It can also be written as a determinant*

$$\{\mathbf{uvw}\} = u_1(v_2 w_3 - v_3 w_2) + u_2(v_3 w_1 - v_1 w_3) + u_3(v_1 w_2 - v_2 w_1) = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}.$$

The magnitude of $\{\mathbf{uvw}\}$ *is the volume of the parallelepiped formed by vectors with representations* $\mathbf{u}$*,* $\mathbf{v}$ *and* $\mathbf{w}$*. The problem with introducing the vector and triple scalar products above is that their meaning is not entirely independent of the choice of basis* $\mathcal{B}$*. In particular, the sign of the result of the product depends on something called the handedness of* $\mathcal{B}$*. To use these products to represent physical quantities unambiguously, one needs to fix the handedness of the basis being used. We explore this in the next question.*

4. Think of describing a 3D vector space with reference to an x, y and z axis. We write the unit vectors along the positive directions of the x, y and z axis as $|i\rangle$, $|j\rangle$ and $|k\rangle$, and represent them by $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$, respectively. One possible choice is to set $\mathcal{B} = \{|i\rangle, |j\rangle, |k\rangle\}$, respectively. Another choice is to set $\mathcal{B} = \{|j\rangle, |i\rangle, |k\rangle\}$. Show that these choices lead to different physical meaning of the vector represented by $\mathbf{u} \times \mathbf{v}$, where $\mathbf{u} = u_x\hat{\mathbf{i}} + u_y\hat{\mathbf{j}} + u_z\hat{\mathbf{k}}$ and $\mathbf{v} = v_x\hat{\mathbf{i}} + v_y\hat{\mathbf{j}} + v_z\hat{\mathbf{k}}$. Discuss why this leads one to define two types of basis, right and left-handed, and give definitions for these types. Which orderings of $|i\rangle$, $|j\rangle$ and $|k\rangle$ are conventionally said to give a right or left-handed set?

Solution: For the two choices of basis orderings, $\mathbf{u} \times \mathbf{v}$, written in terms of $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$, differs by a minus sign and thus represents a different vector. Also, $\{\hat{\mathbf{i}}\hat{\mathbf{j}}\hat{\mathbf{k}}\} = 1$ for the former case and $\{\hat{\mathbf{i}}\hat{\mathbf{j}}\hat{\mathbf{k}}\} = -1$ for the latter case. All ordered basis choices related by rotation have the same sign, those by reflection and rotation have opposite signs. Thus there are two types of orthonormal basis, which we refer to as left and right-handed. Convention chooses $\{\hat{\mathbf{i}}\hat{\mathbf{j}}\hat{\mathbf{k}}\} = 1$ to signify right-handedness and thus $\mathcal{B} = \{|i\rangle, |j\rangle, |k\rangle\}$ is a right-handed basis. Thus $\{|k\rangle, |i\rangle, |j\rangle\}$ and $\{|j\rangle, |k\rangle, |i\rangle\}$ are also right-handed, with the other three orderings left-handed.

It boils down to this: if you represent physics in a 3D vector space and you wish to use the vector or triple scalar product then you need to fix the handedness of your basis $\mathcal{B}$ *(as well as making it orthonormal). It matters. The usual way to resolve the handedness ambiguity is to fix/always have a right-handed basis* $\mathcal{B}$*. Many physical relationships involving the vector or triple scalar product, e.g.* $m\ddot{\mathbf{r}} = q\dot{\mathbf{r}} \times \mathbf{B}$ *for the motion of a particle in a magnetic field, assume a right-handed basis is being used, often without mentioning it, and would be incorrect if a left-handed basis was used. Basically, never use a left-handed coordinate system and you can forget about this complication!*

Let us now specialise to thinking about the particular case of three-dimensional vectors represented by column vectors (bold roman letters) in a fixed right-handed basis $\mathcal{B}$. We will think of each column vector corresponding to a point (capital roman letter) in 3D coordinate space or, equivalently, arrows (drawn) from the origin O to the point. As is common, we will often use slightly imprecise language and treat the vector, column vector, point and arrow as equivalent for the ease of discussion.

Let's go through two of the common objects we deal with in coordinate space, lines and planes, and explore how they are represented and analysed using the language of vectors. Firstly, consider a line. All points $\mathbf{r}$ on a line in 3D space fulfill $\mathbf{r} = \mathbf{a} + \lambda\mathbf{b}$ where $\mathbf{a}$ is a point on the line, $\mathbf{b}$ points in the direction of the line, and λ is a real parameter. Implicitly this equation can be rewritten as $(r_x - a_x)/b_x = (r_y - a_y)/b_y = (r_y - a_z)/b_z$. Get to grips with lines in the next three questions.

5. $*$ Points A and B correspond to column vectors $\mathbf{a}$ and $\mathbf{b}$. Find the column vector $\mathbf{g}$ corresponding to the midpoint G of the straight line connecting the two vectors.

Solution: $\mathbf{g} = (\mathbf{a} + \mathbf{b})/2$.

6. The vertices of triangle ABC correspond to column vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$. Find the column vector $\mathbf{g}$ corresponding to the centroid G of the triangle. The centroid is the intersection of the three lines from each vertex to the midpoint of the opposite side.

Solution: Using the definition of the centroid as existing on three lines we must have that $\mathbf{g} = \mathbf{a} + (\mathbf{g}_{bc} - \mathbf{a})\lambda_{bc} = \mathbf{b} + (\mathbf{g}_{ca} - \mathbf{b})\lambda_{ca} = \mathbf{c} + (\mathbf{g}_{ab} - \mathbf{c})\lambda_{ab}$ with λ_{bc} the fractional distance of the centroid from vertex $\mathbf{a}$ to midpoint $\mathbf{g}_{bc} = (\mathbf{b} + \mathbf{c})/2$ and similarly for the other quantities. A 3D vector equation is equivalent to three scalar equations and we can solve for the three unknowns to get $\lambda_{bc} = \lambda_{ca} = \lambda_{ab} = 2/3$ and thus $\mathbf{g} = (\mathbf{a} + \mathbf{b} + \mathbf{c})/3$. The above proves the existence of a centroid, i.e. the existence of an intersection, as well as where it must be. If we were to assume its existence then we could use symmetry to immediately state that the solution must have $\lambda_{bc} = \lambda_{ca} = \lambda_{ab} = \lambda$, meaning we would only have to solve one equation to obtain the value of the single parameter $\lambda = 2/3$. This is also the only value of λ for which the expression for $\mathbf{g}$ would be symmetric in $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, hence could also be arrived at through symmetry arguments.

7. $*$ Show that the points $(1, 0, 1)$, $(1, 1, 0)$ and $(1, -3, 4)$ lie on a line. Give the equation of the line in the form $\mathbf{r} = \mathbf{a} + \lambda\mathbf{b}$.

Solution: The line through the first two points is given by $\mathbf{a} = (1, 1, 0)$, $\mathbf{b} = (0, 1, -1)$. It goes through the third point for $\lambda = -4$.

8. $*$ Two small objects travel with equal speed. The first starts from the point $(2, 3, 3)$ and travels in the direction of $(-1, -1, 0)$, while the second starts at the same time from $(3, 2, 1)$ and travels in the direction of $(-2, 0, 2)$. Determine whether or not they collide.

Solution: Their paths intersect but they do not collide.

9. Derive an expression for the shortest distance ℓ between the two non-parallel lines $\mathbf{r}_i = \mathbf{a}_i + \lambda_i\mathbf{b}_i$, for $i = 1, 2$. What about the case that the lines are parallel i.e. $\mathbf{b}_i = \mathbf{b}$? Find the shortest distance between the lines

$$\frac{x-2}{2} = y - 3 = \frac{z+1}{2} \quad \text{and} \quad x + 2 = \frac{y+1}{2} = z - 1\,.$$

Solution: For non-parallel lines $\ell = |\{(\mathbf{a}_1 - \mathbf{a}_2)\mathbf{b}_1\mathbf{b}_2\}|/|\mathbf{b}_1 \times \mathbf{b}_2|$. For parallel lines $\ell = |(\mathbf{a}_1 - \mathbf{a}_2) \times \mathbf{b}|/|\mathbf{b}|$. For the given lines $\ell = \sqrt{18}$.

The points $\mathbf{r}$ on a plane in 3D are given by $\mathbf{r} = \mathbf{a} + \lambda\mathbf{b} + \mu\mathbf{c}$ with $\mathbf{a}$ a point on the plane, $\mathbf{b}$ and $\mathbf{c}$ vectors parallel to the plane, and λ and μ real parameters. The implicit form of this equation is $\mathbf{r} \cdot \hat{\mathbf{n}} = d$ with $\hat{\mathbf{n}}$ a unit vector normal to the plane and d the shortest distance from the plane to the origin O.

10. Find the shortest distance d_{P} from a plane to a point P corresponding to column vector $\mathbf{p}$.

Solution: $d_{\mathrm{P}} = |\mathbf{p} \cdot \hat{\mathbf{n}} - d|$.

11. $*$ Find the equation of the line passing through $\mathbf{a} = (1, 2, 3)$ perpendicular to the plane $x - 2y + z = 1$.

Solution: $\mathbf{r} = (1, 2, 3) + \lambda(1, -2, 1)$.

The question of whether two vectors are linearly dependent is simple. Are they identical up to a multiplicative factor? Further, in three dimensions, it is impossible to have four linearly independent vectors. Hence, for 3D coordinate space, it is usual to discuss linear independence for sets of three vectors and it boils down to a question of coplanarity, are all the vectors parallel to the same plane? In fact, many things boil down to a question of coplanarity. Let's explore this in the next two questions.

12. Show that for 3D vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ that the following are equivalent:

- The vectors are linearly dependent.
- The vectors are coplanar.
- The vectors cannot form a basis.

Solution: If $\mathbf{b} = \mu\mathbf{c}$, with some constant μ, the vectors are linearly dependent and all three vectors are parallel to the plane defined by plane vectors $\mathbf{a}$ and $\mathbf{b}$ (or $\mathbf{a}$ and $\mathbf{c}$), i.e. coplanar. If instead $\mathbf{b} \neq \mu\mathbf{c}$, then every vector parallel to the plane defined by plane vectors $\mathbf{b}$ and $\mathbf{c}$ can be written as a linear combination of the form $\mu\mathbf{b} + \lambda\mathbf{c}$. Since the three vectors are co-planar iff $\mathbf{a}$ is parallel to the plane, we can thus write that the vectors are coplanar iff $\mathbf{a} = \mu\mathbf{b} + \lambda\mathbf{c}$ i.e. they are linearly dependent.

While you are not currently asked to show the following, it is interesting to note that, using the definition of the triple scalar product, we can add to this list:

- *The parallelepiped formed by the vectors has zero volume.*
- $\{\mathbf{abc}\} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = 0.$

Note that, other than the part involving 3D-only quantity $\{\mathbf{abc}\}$, everything generalises to $N \neq 3$ dimensions. A plane must be replaced by a $N - 1$ dimensional object and a parallelepiped by the N dimensional equivalent and the determinant by its N-dimensional equivalent. You can use these equivalences to prove things quickly, as seen below.

13. $*$ Prove that the three vectors $\mathbf{a} = \hat{\mathbf{i}} + \hat{\mathbf{j}} + 2\hat{\mathbf{k}}$, $\mathbf{b} = 2\hat{\mathbf{i}} + 3\hat{\mathbf{j}} - \hat{\mathbf{k}}$, $\mathbf{c} = \hat{\mathbf{i}} + 2\hat{\mathbf{j}} - 3\hat{\mathbf{k}}$ are coplanar.

Solution: The triple scalar product is $\{\mathbf{abc}\} = 0$ and hence they all lie in a plane.

Finally, away from the physical discussion of lines and planes etc. sometime it is worth just getting proficient at solving abstract vector equations. This and some of the class problems are good examples.

14. If $\mathbf{a} \times \mathbf{b} = \mathbf{a} - \mathbf{b}$, prove that $\mathbf{a} = \mathbf{b}$.

Class problems

15. Prove the equivalents of the Cauchy-Schwarz inequality, triangle inequality and parallelogram equality for column vectors representations $\mathbf{v}$ in an orthonormal reference basis $\mathcal{B}$, noting what is simpler in this case. Specifically, the equivalent expressions are given by

$$|\mathbf{u} \cdot \mathbf{v}| \leq |\mathbf{u}||\mathbf{v}|,$$
$$|\mathbf{v} + \mathbf{u}|^2 = |\mathbf{v}|^2 + |\mathbf{u}|^2,$$
$$|\mathbf{v} + \mathbf{u}| \leq |\mathbf{v}| + |\mathbf{u}|.$$

16. Prove the parallelogram equality using Dirac notation $||\,|v\rangle + |u\rangle\,||^2 + ||\,|v\rangle - |u\rangle\,||^2 = 2(||\,|v\rangle\,||^2 + ||\,|u\rangle\,||^2)$ or implicitly using representations in an orthonormal basis $|\mathbf{v} + \mathbf{u}|^2 + |\mathbf{v} - \mathbf{u}|^2 = 2(|\mathbf{v}|^2 + |\mathbf{u}|^2)$.

17. Show that for the the case of an orthonormal basis, i.e. $\langle v_j|\, v_i\rangle = \delta_{ij}$, a vector decomposition $|v\rangle = \sum_i v_i\,|v_i\rangle$ satisfies Bessel's inequality $||\,|v\rangle\,||^2 = |\mathbf{v}|^2 \geq \sum_i |v_i|^2$.

18. Find a vector along the line of intersection of the planes $x + 3y - z = 5$ and $2x - 2y + 4z = 3$.

Solution: Any multiple of $\mathbf{b} = (10, -6, -8)^T$.

19. A line intersects a plane at an angle $\alpha = \pi/6$. The line is defined by $\mathbf{r} = \mu\hat{\mathbf{n}}$ and the plane by $\mathbf{r} \cdot \hat{\mathbf{m}} = 0$, with $\hat{\mathbf{n}}$ and $\hat{\mathbf{m}}$ unit vectors. Calculate the shortest distance ℓ from the plane to the point on the line with $\mu = 2$.

Solution: $\ell = 1$ if the angle is taken between line and plane and $\ell = \sqrt{3}$ if the angle is taken between the line and the normal to the plane.

20. Calculate the shortest distance between the planes $x + 2y + 3z = 1$ and $x + 2y + 3z = 5$.

21. Identify the following surfaces:

(a) $|\mathbf{r}| = k$

(b) $\mathbf{r} \cdot \hat{\mathbf{u}} = \ell$

(c) $\mathbf{r} \cdot \hat{\mathbf{u}} = m|\mathbf{r}|$ for $-1 \leq m \leq +1$.

(d) $|\mathbf{r} - (\mathbf{r} \cdot \hat{\mathbf{u}})\hat{\mathbf{u}}| = n$.

Here k, ℓ, m and n are fixed scalars and $\hat{\mathbf{u}}$ is a fixed unit vector.

22. Which of the following set of vectors are i) linearly independent and ii) orthogonal (or both) and explain why:

(a) $\mathbf{a} = (0, 1, 0)^T$, $\mathbf{b} = (1, 0, 0)^T$, and $\mathbf{c} = (0, 0, 1)^T$,

(b) $\mathbf{a} = (0, 1, 1)^T$, $\mathbf{b} = (1, 1, 1)^T$, and $\mathbf{c} = (0, 0, 1)^T$,

(c) $\mathbf{a} = (1, 1, 1)^T$, $\mathbf{b} = (1, -1, 1)^T$, and $\mathbf{c} = (1, 1, -2)^T$,

(d) $\mathbf{a} = (1, 0, 1)^T$, $\mathbf{b} = (2, 3, 1)^T$, and $\mathbf{c} = (1, 6, -1)^T$.

23. Various methods exist to construct an orthonormal set of vectors from a non-orthogonal set of vectors (provided they are not linearly independent). Use the Gram-Schmidt method to construct three orthonormal vectors for the example in the previous question that features three linearly independent but non-orthogonal vectors.

24. Using index notation, show which of the statements below about general vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ are true and which are false (with $c = |\mathbf{c}|$). Use this question to familiarise yourself with when index notation is needed and when known properties of scalar and vectors products can be used instead.

(a) $\mathbf{c}\cdot(\mathbf{a}\times\mathbf{b}) = (\mathbf{b}\times\mathbf{a})\cdot\mathbf{c}$,

(b) $\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = (\mathbf{a}\times\mathbf{b})\times\mathbf{c}$,

(c) $\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = (\mathbf{a}\cdot\mathbf{c})\mathbf{b} - (\mathbf{a}\cdot\mathbf{b})\mathbf{c}$,

(d) $\mathbf{d} = \lambda\mathbf{a} + \mu\mathbf{b}$ implies $(\mathbf{a}\times\mathbf{b})\cdot\mathbf{d} = 0$,

(e) $\mathbf{a}\times\mathbf{c} = \mathbf{b}\times\mathbf{c}$ implies $\mathbf{c}\cdot\mathbf{a} - \mathbf{c}\cdot\mathbf{b} = c|\mathbf{a}-\mathbf{b}|$,

(f) $(\mathbf{a}\times\mathbf{b})\times(\mathbf{c}\times\mathbf{b}) = \mathbf{b}[\mathbf{b}\cdot(\mathbf{c}\times\mathbf{a})]$.

(g) $(\mathbf{a}\times\mathbf{b})\dot{(}\mathbf{c}\times\mathbf{d}) = (\mathbf{a}\cdot\mathbf{c})(\mathbf{b}-\mathbf{d}) - (\mathbf{b}\cdot\mathbf{c})(\mathbf{a}-\mathbf{d})$.

25. For a given ordered set of vectors $\{\mathbf{a},\mathbf{b},\mathbf{c}\}$ cyclic permutations are $\{\mathbf{a},\mathbf{b},\mathbf{c}\}$, $\{\mathbf{b},\mathbf{c},\mathbf{a}\}$ and $\{\mathbf{c},\mathbf{a},\mathbf{b}\}$ while anti-cyclic permutations are $\{\mathbf{a},\mathbf{c},\mathbf{b}\}$, $\{\mathbf{b},\mathbf{a},\mathbf{c}\}$, and $\{\mathbf{c},\mathbf{b},\mathbf{a}\}$. How many permutations are needed to get from the initial set of vectors to a cyclic permutation and how many to get to an anti-cyclic permutation? How can we make use of these definitions in the above examples?

In the near future you'll be studying solid state physics, where atoms are located at a regular translationally symmetric lattice points $\mathbf{R} = l\mathbf{a}+m\mathbf{b}+n\mathbf{c}$, for integer l, m and n. This structure is revealed by scattering waves off the the lattice and observing peaks when waves are scattered by momenta $\mathbf{K}$ satisfying $\exp(\mathrm{i}\mathbf{K}\cdot\mathbf{R}) = 1$. It is possible to write $\mathbf{K} = h\mathbf{a}' + k\mathbf{b}' + l\mathbf{c}'$, for integer h, k and l, where $\mathbf{a}'$, $\mathbf{b}'$ and $\mathbf{c}'$ are called reciprocal lattice vectors. It's worth spending one question getting to grips with the definition of these.

26. The vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ are not coplanar. Verify that the expressions

$$\mathbf{a}' = \frac{2\pi\mathbf{b}\times\mathbf{c}}{\{\mathbf{abc}\}}, \qquad \mathbf{b}' = \frac{2\pi\mathbf{c}\times\mathbf{a}}{\{\mathbf{abc}\}}, \qquad \mathbf{c}' = \frac{2\pi\mathbf{a}\times\mathbf{b}}{\{\mathbf{abc}\}},$$

define a set of vectors $\mathbf{a}'$, $\mathbf{b}'$ and $\mathbf{c}'$ with the following properties:

(a) $\mathbf{a}'\cdot\mathbf{a} = \mathbf{b}'\cdot\mathbf{b} = \mathbf{c}'\cdot\mathbf{c} = 2\pi$,

(b) $\mathbf{a}'\cdot\mathbf{b} = \mathbf{a}'\cdot\mathbf{c} = \mathbf{b}'\cdot\mathbf{a} = \mathbf{b}'\cdot\mathbf{c} = \mathbf{c}'\cdot\mathbf{a} = \mathbf{c}'\cdot\mathbf{b} = 0$,

(c) $\{\mathbf{a}'\mathbf{b}'\mathbf{c}'\} = (2\pi)^3/\{\mathbf{abc}\}$,

(d) $\mathbf{a} = (2\pi\mathbf{b}'\times\mathbf{c}')/\{\mathbf{a}'\mathbf{b}'\mathbf{c}'\}$.

27. Figure 1 shows a Mathematica script attempting to calculate of the length of a column vector with complex components. The built-in Mathematica function `Norm` gives a different result for its length than naively applying definition above for real column vectors. Explore the difference between the two ways of calculating the length of a column vector and discuss which is a better approach.

```
In[1]:= a = {i, 2, 4 + 2 i};
```

We wish to work out the length of this vector

```
In[2]:= Print["|a| = ", Sqrt[a.a] // Simplify]
```

$|a| = \sqrt{15 + 16\,i}$

When using the built in *Mathematica* function the result is different

```
In[3]:= Print["|a| = ", Norm[a] // Simplify]
```

$|a| = 5$

Figure 1: Calculating the length of a complex vector in Mathematica.

There's nothing in our definition of a vector space $\mathcal{V}$ that stops us from allowing them to be built using complex rather than real numbers: letting $z|v\rangle$ be a vector in the space if $|v\rangle$ is, where z may be complex. In this case we'll be dealing with components v_i in $|v\rangle = \sum_i v_i |v_i\rangle$ that could be complex. This last question highlights that if we allow complex numbers then something needs to change in our definition of an inner product $f(|v\rangle, |u\rangle) = \langle u|\, v\rangle$. The first thing we wanted from an inner product was a concept of length. Nothing changes here: The only single-vector quantity available is $\langle v|\, v\rangle$ and since we want this to give us a length we require it to be non-negative $\langle v|\, v\rangle \geq 0$ with equality only for $|v\rangle = |0\rangle$. This enables us to think of $|| \, |v\rangle \, || = \sqrt{\langle v|\, v\rangle}$ as the length (or norm) of a vector, with $|0\rangle$ having zero length but every other vector a positive length. However, it now becomes impossible to insist that $f(|v\rangle, |u\rangle)$ is linear in both its first and second arguments as then $f(z|v\rangle, z|v\rangle) = z^2 f(|v\rangle, |v\rangle)$, i.e. $||z|v\rangle|| = z||\,|v\rangle||$, would mean that both lengths could not be real for arbitrary complex z. The simplest thing to do to fix this is let the inner product be linear in its first argument but conjugate linear in its second $f(|v\rangle, z_u|u\rangle + z_w|w\rangle) = z_u^* f(|v\rangle, |u\rangle) + z_w^* f(|v\rangle, |w\rangle)$. Alternatively one can say that $f(|v\rangle, |u\rangle)$ is linear in its first argument and $f(|u\rangle, |v\rangle) = (f(|v\rangle, |u\rangle))^*$, meaning it is not symmetric. For real numbers this makes no difference, so we tend to use the complex vector space notation anyway. If we do have complex numbers it means that $f(z|v\rangle, z|v\rangle) = |z|^2 f(|v\rangle, |u\rangle)$, i.e. $||z|v\rangle|| = |z|||v\rangle||$, thus solving our problem. The only other thing it changes is that decomposing two vectors $|v\rangle = \sum_i v_i |v_i\rangle$, $|u\rangle = \sum_i u_i |v_i\rangle$ in the same orthonormal basis, it follows that the inner product is $\langle u|\, v\rangle = \sum_i u_i^* v_i$. We can let this define anew what we mean by the scalar product $\mathbf{u} \cdot \mathbf{v} = (u_1, \ldots, u_N)^T \cdot (v_1, \ldots, v_N)^T = \mathbf{u}^\dagger \mathbf{v} = \sum_i u_i^* v_i$, where we have defined the symbol $\dagger$ to meant the conjugate transpose (transpose the elements and then conjugate them). Again, other conventions exist.

In the next vectors and matrices problem set we'll work entirely with complex vector spaces. Why bother generalising in this way? It turns out that using complex numbers make things more concise and some mathematics, e.g. of quantum mechanics, is most easily expressed by making use of them (it is possible to represent quantum mechanics using only real numbers, since all measurable results will be real, but it's a total nightmare that is never attempted other than by those who just wanted to check it could be done).