Physics 605: Abstract Linear Algebra

Due: never

Abstract linear algebra pervades advanced physics, from quantum mechanics to electricity and magnetism. It is also one of the first places that physics students start to look at what constitutes a proof and how mathematics "really" works.

1 What is a vector?

It used to be that people started thinking about an object using examples. This is how we tend to do things in physics too. Mathematicians, however, have found a better way. They start with an abstract definition of the properties they want the objects they are defining to have, then prove that examples exist. This is an advantage because they are then able to find examples of the objects that are not obvious.

Consider a vector. A mathematician would define an object called a **vector space**, V, and two operations: addition and multiplication by a scalar. That scalar could be either a real number, a complex number of some other number-like entity (like the integers or integers modulo 5). Unless otherwise specified, however, assume its a complex number. We'll follow Dirac, in that elements of V will be denoted $|v\rangle$ or $|w\rangle$.

The elements of a vector space have the following properties:

1. Addition commutes: $|v\rangle + |w\rangle = |w\rangle + |v\rangle$.

2. Addition is associative: $|v\rangle + (|w\rangle + |u\rangle) = (|v\rangle + |w\rangle) + |u\rangle$.

    You may think this is silly but not every operation is associative: $\vec{x} \times (\vec{y} \times \vec{z}) \neq (\vec{x} \times \vec{y}) \times \vec{z}$ where $\vec{x}$, $\vec{y}$ and $\vec{z}$ are three-dimensional vectors and $\times$ is the cross product. It is better to be safe than sorry.

3. There is an element, $|0\rangle$ such that $|0\rangle + |v\rangle = |v\rangle + |0\rangle = |v\rangle$.

    We do not have to specify that $|0\rangle$ be unique; the other axioms will allow us to prove it.

4. For every element $|v\rangle$ there is an inverse, $|-v\rangle$, such that $|v\rangle + |-v\rangle = |0\rangle$.

   Again, we will be able to prove that inverses are unique.

5. Scalar multiplication is associative: for two scalars r and s, $r(s\,|v\rangle) = (rs)\,|v\rangle$.

6. Scalars distribute: $(r+s)\,|v\rangle = r\,|v\rangle + s\,|v\rangle$.

7. Vectors also distribute: $r(|v\rangle + |w\rangle) = r\,|v\rangle + r\,|w\rangle$.

8. $1\,|v\rangle = |v\rangle$.

1.1 Examples of vector spaces

Now, of course, we've defined what we want a vector space to *be*. The question a mathematician would ask is: are there any examples? Luckily, it turns out there are a lot of examples, many of which we are completely familiar with.

1. The set of column vectors,

$$|v\rangle = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_d \end{pmatrix} \tag{1}$$

   forms a vector space. This is the notion of vectors most physicists are comfortable with.

2. The set of row vectors,

$$|v\rangle = \begin{pmatrix} v_1 & v_2 & \cdots & v_d \end{pmatrix} \tag{2}$$

   forms a vector space. Notice that there is a map taking column vectors to row vectors which is given by

$$|v\rangle = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_d \end{pmatrix} \to |v\rangle = \begin{pmatrix} v_1 & v_2 & \cdots & v_d \end{pmatrix}. \tag{3}$$

Because of this very natural map, many people intuitively think of column vectors as the same as row vectors. Nevertheless, this equivalence isn't always true; it is useful to distinguish them. We will see later that there is an interesting correspondence between $|v\rangle$ and $\langle v|$.

3. The set of smooth, bounded functions $f(x)$ such that $f(0) = f(L) = 0$ is also a vector space. In particular, note that addition of functions is closed. This is one of the first surprises, at least if you haven't seen this before.

4. The set of complex numbers forms a vector space. Note that there is also another operation – multiplication – which means the set of complex numbers have more structure than just a vector space. More generally, they form a **field**. But they are still a vector space.

5. The set of $n-$tuples of real numbers $(x_1, \cdots, x_n)$ forms a vector space with addition $(x_1, \cdots) + (y_1, \cdots) = (x_1 + y_1, \cdots)$ and scalar multiplication $s(x_1, \cdots) = (sx_1, \cdots)$.

6. The set of polynomials of degree n in one variable forms a vector space.

1.2 Things you can prove

There are many things you know are true that can be proven from this initial set of assumptions. For example, you can prove that there is only one zero in a vector space (the definition only requires at least one zero to exist). Here follow a few simple theorems and proofs, so you can see what we usually mean by a proof:

Theorem: Zeros are unique. In other words, there is no other element $|0'\rangle$ such that $|0'\rangle + |v\rangle = |v\rangle$ (other than $|0\rangle$ which is there by assumption).

> Assume that there is another zero, called $|0'\rangle$. Then $|0\rangle + |v\rangle = |v\rangle = |0'\rangle + |v\rangle$ must also be true. Add the element $|-v\rangle$ (remember that, by assumption, $|v\rangle + |-v\rangle = |0\rangle$) to both sides of this equation. You obtain $|0\rangle = |0'\rangle + |0\rangle = |0'\rangle$ which shows that $|0\rangle = |0\rangle'$. Suppose $|v\rangle + |-v\rangle = |0'\rangle$ – you would have obtained the same result: $|0\rangle = |0'\rangle$ then as well.

Theorem: Additive inverses are unique.

Assume that $|v'\rangle$ and $|-v\rangle$ are both inverses of $|v\rangle$. This implies $|v'\rangle + |v\rangle = |0\rangle$ (from the previous proposition, there is only one zero). Now add $|-v\rangle$ to both sides to obtain $|v'\rangle = |-v\rangle$.

Theorem: $0\,|v\rangle = |0\rangle$.

First, $0\,|v\rangle + 0\,|v\rangle = (0+0)\,|v\rangle = 0\,|v\rangle$. If $0\,|v\rangle \neq |0\rangle$, then it has a unique inverse, which we can add to both sides of the equation. Consequently, we find that $0\,|v\rangle = |0\rangle$, which is a contradiction of our initial assumption. We then see that $0\,|v\rangle = |0\rangle$.

Theorem: $-1\,|v\rangle = |-v\rangle$.

We compute that $|v\rangle + (-1\,|v\rangle) = (1-1)\,|v\rangle = 0\,|v\rangle = |0\rangle$.

A few important things to notice about these proofs. First, they use a lot of words; math isn't just manipulating equations and doing algebra. To a large extent, math is about defining some objects and working out all the logical consequences. Second, things that are "obvious" still need to be proved; many obvious things turn out to be false. Third, assuming something *is* true, then finding a contradiction with the assumptions, is an important way to show that things cannot exist.

2 Putting some structure on vector spaces

2.1 Linear independence

Two vectors are said to be linearly independent if $r\,|v\rangle + s\,|w\rangle = 0$ implies that $r = s = 0$. A basis for a vector space is a set of linearly independent vectors, $\{|1\rangle, |2\rangle, \cdots\}$ such that *any* vector is a linear combination of the basis vectors. Linear independence has nothing to do with being orthogonal – in fact, we haven't even defined what the word "orthogonal" means, at this point.

A vector space is **finite dimensional** if there is a finite basis of linearly independent vectors, $\{|1\rangle, |2\rangle, \cdots, |n\rangle\}$. Interestingly, you can prove that if one basis has n vectors in it, then all bases have exactly n vectors. Suppose, in addition to a basis with n elements, we have $n+1$ vectors sitting around, $\{|1'\rangle, \cdots, |n+1'\rangle\}$ arranged so that the first n really are linearly independent.

Then we can write

$$|i\rangle = \sum_{j=1}^{n} c_{ij} |j'\rangle . \tag{4}$$

That is, each element of our basis is also writable as a linear combination of our first n vectors. Now consider the $|n+1'\rangle$. We know that

$$|n+1'\rangle = \sum_{i=1}^{n} d_i |i\rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} d_i c_{ij} |j'\rangle . \tag{5}$$

In other words, $|n+1'\rangle$ has just been written as a linear combination of the first n primed vectors. Hence, any set of $n+1$ vectors cannot be linearly-independent. A similar argument tells us that a set of vectors less than n can't be a basis. This allows us to define the notion of dimension. We call dim V the dimensional of vector space V, where n is the maximal number of linearly independent vectors in a basis.

There is nothing in the definition of a vector space that requires that it be finite dimensional. In fact, we've already seen an example of an infinite-dimensional vector space, the space of functions $f(x)$ such that $f(0) = 0$ and $f(L) = 0$. It is easy enough to show that the set of functions $\sin(n\pi x/L)$ are in this vector space for all integers n and are linearly independent. However, unlike a finite dimensional vector space, trying to write any function $f(x)$ satisfying these constraints in terms of the linearly independent vectors $\sin(n\pi x/L)$ requires us to think about infinite sums. Where there are infinite sums, we must also worry about how series converge. These are considerations to save for a later chapter.

2.2 Inner products

An inner product is a map from two vectors to a scalar. The scalar technically needs to be an element of a field (say, the real numbers or the complex numbers). As a general rule, we will assume it is always the complex numbers unless stated otherwise. The inner product then has the following properties:

1. $\langle v|w\rangle = \langle w|v\rangle^*$, where * is the complex conjugate.

2. $\langle v+w|u\rangle = \langle v|w\rangle + \langle v|u\rangle$.

3. $\langle v|rw\rangle = r\langle v|w\rangle$ for a scalar r.

4. $\langle v|v\rangle \geq 0$ with $\langle v|v\rangle = 0$ if and only if $|v\rangle = |0\rangle$.

We denote $\|v\| \equiv \sqrt{\langle v|v\rangle}$ and call it the "norm" of $|v\rangle$. In fact, we will see on the homework that we can define a norm independently from an inner product. For class, we will just jump right to inner products.

Examples

1. Consider the vector space of real $n-$tuples, $(x_1, \cdots, x_n)$. We can define an inner product $(x_1, \cdots, x_n) \cdot (y_1, \cdots, y_n) = \sum_{i=1}^{n} x_i y_i$. This is called the "dot product," and is what most undergraduate physicists think of when they think of inner products.

2. We already learned that the space of smooth, bounded functions $f(x)$ such that $f(0) = f(L) = 0$ forms a vector space. In addition, we can define an inner product by

$$\langle f|g\rangle = \int_0^L dx\ f^*(x)g(x). \tag{6}$$

To prove it is an inner product, you only have to demonstrate that it satisfies the four axioms for inner products. This allows us to define a norm by $\|f\|^2 = \int_0^L dx\ |f(x)|^2$, which is called the L_2 norm. In quantum mechanics, the wave functions satisfying appropriate boundary conditions form a vector space, the inner product gives quantum amplitudes and the L_2 norm is related to the normalization of the wave function. There is an entire family of normals called L_p norms given by

$$lfl_p^p = \int_0^L dx\ |f(x)|^p. \tag{7}$$

Theorem: Two orthogonal, nonzero vectors are linearly independent.

Let $|1\rangle$ and $|2\rangle$ be orthogonal vectors. Then let $r\,|1\rangle + s\,|2\rangle = |0\rangle$. Let's compute $\langle 1|0\rangle = r\,\langle 1|1\rangle$. Since $|1\rangle$ is nonzero, $\langle 1|1\rangle \neq 0$ so $r = 0$. Similarly, $\langle 2|0\rangle = s\,\langle 2|2\rangle = 0$ implies $s = 0$.

2.3 Orthonormal Bases

Definition: An **orthonormal basis** is a basis such that $\langle n|m\rangle = \delta_{nm}$, where δ_{nm} is the Kronecker delta.

1. Let's lead off with a simple example first. Consider the vector space of $n-$tuples. Then $\{(1,0,\cdots 0),(0,1,0,\cdots,0),\cdots(0,\cdots,0,1)\}$ is an orthonormal basis under the dot product.

2. The traditional 3D unit vectors, $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ are an orthonormal basis.

3. Now for something more abstract, consider the vector space of bounded, periodic functions $f(x)$ periodic in the interval between 0 and L. The functions $|n\rangle = \sqrt{1/L}e^{i2\pi nx/L}$ form an orthonormal basis using the inner product of Eq. 6 for integers n. To see this, we note that

$$\langle n|m\rangle = \frac{1}{L}\int_0^L dx\ e^{-i2\pi nx/L}e^{i2\pi mx/L} = \delta_{nm}. \tag{8}$$

Consequently, they are linearly independent. Moreover – and this is not obvious – any periodic function can be written as a linear combination of $|n\rangle$. In other words,

$$|f\rangle = \sum_{n=-\infty}^{\infty} c_n\,|n\rangle = \frac{1}{\sqrt{L}}\sum_{n=-\infty}^{\infty} c_n e^{i2\pi nx/L}. \tag{9}$$

This is just the formula for the Fourier series of a periodic function. Fourier's theorem tells us that any periodic function can be written as a sum of exponentials, and therefore it ensures that the resulting set of functions is complete. Therefore the $|n\rangle$ constitute a basis. The idea of orthogonal functions will play an important role in Fourier analysis and solving both ordinary and partial differential equations.

Any finite-dimensional basis can be made into an orthogonal (and hence, orthonormal) basis through the Gram-Schmidt procedure. Suppose you have a basis, $\{|e_1\rangle,\cdots,|e_n\rangle\}$. We want to define an orthogonal basis, $\{|1\rangle,\cdots,|n\rangle\}$. First, define $|1\rangle = |e_1\rangle$. Then define $|2\rangle$ by subtracting any component of $|e_2\rangle$ along $|1\rangle$, remembering we have to normalize $|1\rangle$ in the process. The result is

$$|2\rangle = |e_2\rangle - \frac{|1\rangle\langle 1|e_2\rangle}{\langle 1|1\rangle}. \tag{10}$$

Similarly, we can obtain the i^{th} step in this process from the $(i-1)^{th}$. Indeed, the general formula is

$$|i\rangle = |e_i\rangle - \sum_{j=1}^{i-1}|j\rangle\frac{\langle j|e_i\rangle}{\langle j|j\rangle}. \tag{11}$$

Another example

Consider the set of polynomials in one variable, x, up to order n, given by $\sum_{i=0}^{n} c_i x^i$. This forms a vector space, as discussed before. The set of monomials $\{x^0, x^1, \cdots, x^n\}$ forms a basis. We can define an inner product on this vector space by $\langle p|q\rangle = \int_{-1}^{1} dx p^*(x)q(x)$. To find an orthogonal set, we can use the Gram-Schmidt procedure. In that case, we find

$$\begin{aligned} |1\rangle &= 1 \\ |2\rangle &= x - 1 \cdot \frac{\int_{-1}^{1} dx\ 1 \cdot x}{\int_{-1}^{1} dx\ 1^2} = x \\ |3\rangle &= x^2 - \frac{1}{3} \\ &\vdots \end{aligned} \tag{12}$$

We can actually follow this procedure for n as large as we want. The set of orthonormal polynomials we end up with is called the **Legendre polynomials**.

3 The dual space

Once you have a notion of inner product, there is automatically a notion of the "dual space." This ends up being a little abstract so that it can cover situations that are more unusual than you may be used to. However, on finite-dimensional vector spaces, it all turns out to be far more concrete.

Definition: A bounded linear functional is a linear map, $\mathcal{L}$ from a vector space V to the scalars. **Linear** means that $\mathcal{L}(s|v\rangle + r|w\rangle) = s\mathcal{L}|v\rangle + r\mathcal{L}|w\rangle$. **Bounded** means that $|\mathcal{L}|v\rangle| \leq M\|v\|$ for some $M > 0$.

Definition: The dual space of V, denoted V^*, is defined as the vector space of bounded *linear* functionals.

Any vector $|v\rangle$ can be turned into an element of V^* using the inner product: $\mathcal{L}_v(|w\rangle) \equiv \langle v|w\rangle$. As is standard, we're going to just write $\langle v|$ for $\mathcal{L}_v$. This identification between elements of the vector space and corresponding elements of the dual space has a particular notation, †. We generally write

$$|v\rangle^\dagger = \langle v|\,. \tag{13}$$

Amazingly, if dim V is finite then dim $V^* =$ dim V. That is, any object in V has a unique, corresponding object in V^*.

> We don't need to get into the proof too much, but it is useful to sketch it out. Basically, let $\{|1\rangle, \cdots |n\rangle\}$ be a basis for V. Define a map $\mathcal{L}_i(|v\rangle) = \langle 1|v\rangle$. This is an element of the dual space. You can prove that these maps form a basis for V^* and, therefore, dim V must be n as well.

As a side effect, anything in the dual space $\langle v|$ can be turned back into a vector, $|v\rangle$. This justifies, a little, the use of Dirac notation.

For an infinite dimensional vector space, the dual space can be larger than the vector space. We need an additional assumption in order to make sense of the dual space. It turns out, however, that with the right additional assumptions there is an analogue to the identification between V and V^*.

4 Linear operators

Much of linear algebra is about maps from one vector space to another (or back to itself). A linear operator, $\mathcal{L} : V \to W$, from a vector space V to a vector space W has the following property: $\mathcal{L}\left(r\,|u\rangle + s\,|v\rangle\right) = r\mathcal{L}\,|u\rangle + s\mathcal{L}\,|v\rangle$. Because of this property, a linear operator can be entirely defined by what it does to an orthonormal basis. Let $\{|1\rangle, \cdots\}$ be an orthonormal basis for V, and $|v\rangle = \sum_n c_n\,|n\rangle$. Then

$$\mathcal{L}\,|v\rangle = \sum_n c_n \mathcal{L}\,|n\rangle\,. \tag{14}$$

For finite-dimensional vector spaces, things are even more straightforward. Suppose that dim $V = N$ and dim $W = M$ and let $|m'\rangle$ forms an orthonormal basis for W. Then we can write any destination vector as $|w\rangle = \sum_{m=1}^{M} d_m\,|m'\rangle$. Then we can formally compute the coefficients d_m by

$$d_m = \sum_{n=1}^{N} \langle m'|\mathcal{L}|n\rangle \langle n|v\rangle \tag{15}$$

$$= \begin{pmatrix} \langle 1'|\mathcal{L}|1\rangle & \cdots & \langle 1'|\mathcal{L}|N\rangle \\ \vdots & \ddots & \vdots \\ \langle M'|\mathcal{L}|1\rangle & \cdots & \langle M'|\mathcal{L}|N\rangle \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_N \end{pmatrix}. \tag{16}$$

Indeed, *any linear map between finite-dimensional vector spaces can be regarded as a matrix*. Obviously, we can also turn any matrix into a linear map using Eq. (15).

Finally, putting Eqs. (14) and (15), we see that we have

$$|v\rangle = \sum_{n=1}^{N} |n\rangle \langle n|v\rangle \,. \tag{17}$$

Since we also know that $|v\rangle = 1\,|v\rangle$, where 1 is the identity map (the linear map that does nothing), we see that we can make the identification

$$1 = \sum_{n=1}^{N} |n\rangle \langle n| \,. \tag{18}$$

In fact, we can now write any matrix as a linear operator using

$$\mathcal{L} = \sum_{n=1}^{N} \sum_{m=1}^{M} L_{mn} \, |m'\rangle \langle n| \,, \tag{19}$$

where L_{mn} are the components of the matrix.

In addition, suppose we have two linear maps, $\mathbf{L}_1 : V_1 \to V_2$ and $\mathbf{L}_2 : V_2 \to V_3$ from a vector space V_1 to V_2 to V_3 having dimensions N_1, N_2 and N_3. We can use Eq. (15) twice on the vector $\mathbf{L}_1\mathbf{L}_2 \,|v\rangle$ to prove

$$\langle e_I^{(3)}|\mathbf{L}_2\mathbf{L}_1|e_K^{(1)}\rangle = \sum_{J=1}^{N_2} \langle e_I^{(3)}|\mathbf{L}_2|e_J^{(2)}\rangle \langle e_J^{(2)}|\mathbf{L}_1|e_K^{(1)}\rangle \,. \tag{20}$$

We can also define something called the *adjoint* of an operator. We say that $\mathbf{L}^\dagger : W \to V$ is the adjoint of $\mathbf{L} : V \to W$ if $\langle \mathbf{L}^\dagger v|w\rangle = \langle v|\mathbf{L}w\rangle$ for any $|w\rangle$ and $|v\rangle$, where $|\mathbf{L}v\rangle \equiv \mathbf{L}\,|v\rangle$. You will show on your homework

$$\langle e_j|\mathbf{L}^\dagger e_i\rangle = \left(\langle e_i|\mathbf{L}|e_j\rangle\right)^{*T} , \tag{21}$$

where ^{T} is the transpose of a matrix and * the complex conjugate. We usually write ^{*T} as †.

The adjoint of an operator is related to the dual space induced by the inner product; that is why we use the same symbol, †, for both. This correspondence allows us to make beautiful statements like this:

$$\langle w|\mathbf{L}|v\rangle^\dagger = \left(|v\rangle\right)^\dagger \mathbf{L}^\dagger \left(\langle w|\right)^\dagger = \langle v|\mathbf{L}^\dagger|w\rangle \,. \tag{22}$$

Definition: A linear operator is **self-adjoint** if $\mathbf{L}^\dagger = \mathbf{L}$. This implies that the matrix associated with $\mathbf{L}$ is equal to its complex conjugate-transpose.

Self-adjoint linear operators play a special role in quantum mechanics as observables.

5 Eigenvalues and eigenvectors

5.1 Basic definitions and theorems

There is a class of linear algebra problems that arise over and over, both in mathematics and in physics. The problem is to solve

$$\mathbf{L}\,|v\rangle = \lambda\,|v\rangle\,, \tag{23}$$

where λ is a scalar, called an **eigenvalue** of $\mathbf{L}$ and $|v\rangle$ is an element of some vector space, called an **eigenvector**. A special class of problems are to find eigenvectors whose eigenvalue are $\lambda = 0$: $\mathbf{L}\,|v\rangle = 0$ because many linear differential equations are cast in this form.

In this section, we'll collect some basic results that most undergraduates first learn about in quantum mechanics.

The kernal, or **null space**, of $\mathbf{L}$ is the subspace of vectors such that $\mathbf{L}\,|v\rangle = 0$.

Theorem: If $|v\rangle$ belongs to a finite-dimensional vector space, $\mathbf{L}\,|v\rangle = 0$ has a solution such that $|v\rangle \neq 0$ if and only if $\det\mathbf{L} = 0$, where det is the determinant of the matrix corresponding to the operator $\mathbf{L}$.

Theorem: Let $\mathbf{1}$ is the identity operator, $\mathbf{1}\,|v\rangle = |v\rangle$. The eigenvalues of $\mathbf{L}$ satisfy the polynomial equation $p(\lambda) = \det(\mathbf{L} - \lambda\mathbf{1}) = 0$.

Theorem: Let $p(\lambda) = \det(\mathbf{L} - \lambda\mathbf{1})$. Then $p(\mathbf{L}) = 0$. The polynomial $p(\mathbf{L})$ is the characteristic polynomial.

5.2 Self-adjoint operators

Theorem: The eigenvalues of a self-adjoint operator are real. This is true even for infinite-dimensional vector spaces.

> Let $\mathbf{L}\,|v\rangle = \lambda\,|v\rangle$. Then compute $\langle v|\mathbf{L}v\rangle = \lambda\,\langle v|v\rangle$. Taking the complex conjugate of both sides gives $\lambda\,\langle v|v\rangle = \langle \mathbf{L}v|v\rangle = \langle v|\mathbf{L}^\dagger v\rangle = \lambda^*\,\langle v|v\rangle$. Since $\langle v|v\rangle \neq 0$, $\lambda = \lambda^*$.

Theorem: Two eigenvectors of a self-adjoint operator with different eigenvalues are orthogonal. This is true for infinite-dimensional vector spaces.

> We will prove this for self-adjoint operators. Let $\mathbf{L}|v\rangle = \lambda|v\rangle$ and $\mathbf{L}|w\rangle = \omega|w\rangle$ where $\lambda \neq \omega$. Then $\langle w|\mathbf{L}|v\rangle = \lambda\langle w|v\rangle$ but $\langle w|\mathbf{L}v\rangle = \langle \mathbf{L}w|v\rangle = \omega\langle w|v\rangle$. Therefore, $(\lambda - \omega)\langle w|v\rangle = 0$, and since $\lambda \neq \omega$, $\langle w|v\rangle = 0$.

Definition: A unitary transformation, $\mathbf{U}$, is an operator such that $\mathbf{U}^{-1} = \mathbf{U}^\dagger$.
Definition: A **diagonal** operator $\mathbf{D}$ has the property that $\mathbf{D}|i\rangle = \lambda_i|i\rangle$, where $|i\rangle$ is a member of an orthogonal basis. That is, being diagonal is a property of a basis and an operator together.
Theorem: For any self-adjoint operator $\mathbf{L}$, there is a unitary transformation $\mathbf{U}$ such that

$$\mathbf{L} = \mathbf{U}^\dagger \mathbf{D}\mathbf{U} \tag{24}$$

where $\mathbf{D}$ is a diagonal operator.

This is what it means to diagonalize an operator. The matrix associated with a diagonal operator can be written as

$$\langle i|\mathbf{D}|j\rangle = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \lambda_d \end{pmatrix} \tag{25}$$

Theorem: Let $\mathbf{D}_1$ and $\mathbf{D}_2$ be diagonalizable in the same finite, orthogonal basis, $\{|1\rangle, \cdots, |d\rangle\}$. Then $[\mathbf{D}_1, \mathbf{D}_2] = \mathbf{D}_1\mathbf{D}_2 - \mathbf{D}_2\mathbf{D}_1 = 0$ (the commutator vanishes). Consequently, if $[\mathbf{L}_1, \mathbf{L}_2] \neq 0$, $\mathbf{L}_1$ and $\mathbf{L}_2$ cannot be simultaneously diagonalized.

> First, we need to prove that two diagonal operators commute. Let $\mathbf{D}_1|i\rangle = \lambda_i|i\rangle$ and $\mathbf{D}_2|i\rangle = \omega_i|i\rangle$. We have $(\mathbf{D}_1\mathbf{D}_2 - \mathbf{D}_2\mathbf{D}_1)|i\rangle = (\lambda_1\omega_1 - \omega_1\lambda_1)|i\rangle$. Therefore, $[\mathbf{D}_1, \mathbf{D}_2] = 0$ since it is zero for any basis vector.
>
> Second, we note that $[\mathbf{D}_1, \mathbf{D}_2] = \mathbf{U}^\dagger[\mathbf{L}_1, \mathbf{L}_2]\mathbf{U}$ so $[\mathbf{D}_1, \mathbf{D}_2] = 0$ if and only if $[\mathbf{L}_1, \mathbf{L}_2] = 0$. Consequently, if $[\mathbf{L}_1, \mathbf{L}_2] \neq 0$, both operators could not be diagonalizable in the same basis.

5.3 What can we do with this?

One important thing we can do with a self-adjoint operator is find functions that are orthogonal. Consider the space of functions in one variable, θ, that are periodic in 2π, $f(\theta)$. Define the inner product

$$\langle f|g\rangle = \int_0^{2\pi} \frac{d\theta}{2\pi} f^*(\theta)g(\theta). \tag{26}$$

The operator

$$\mathbf{L} = \frac{1}{i}\frac{d}{d\theta} \tag{27}$$

is self-adjoint. (Prove it!).

Let's find its eigenvalues and eigenvectors. We have

$$\begin{aligned} \frac{1}{i}\frac{d}{d\theta}|f\rangle &= \lambda|f\rangle \\ \rightarrow \frac{df}{d\theta} &= i\lambda f. \end{aligned} \tag{28}$$

The solutions to this equation are $f(\theta) = e^{i\lambda\theta}$. However, notice that this function $f(\theta)$ is not in the vector space unless λ is an integer. Hence, we have eigenvectors $f_n = e^{in\theta}$ for integers n and eigenvalues $\lambda_n = n$.

Since these have different eigenvalues, they are all guaranteed to be orthogonal. Importantly, we don't have to do the integral to know this! We have just proved

$$\int_0^{2\pi} \frac{d\theta}{2\pi} e^{i(m-n)\theta} \propto \delta_{mn}. \tag{29}$$

Here is an interesting question. Is the set of functions, $\{e^{in\theta}\}$ a basis for the vector space of periodic functions? To know this, we need to know that all functions can be written as a sum of these exponentials. This is what Fourier's theorem tells us. But to understand this, we need to think a little harder about infinite-dimensional vector spaces – also known as Hilbert spaces to most physicists. This will be a topic of a future set of lectures.