1. Black-body Radiation and the Quantization of Light

One of the first indications that nature is quantized came from looking at the thermal emission from so-called "black bodies". If you make a box and hold the walls at fixed temperature, then the intensity of light as a function of frequency inside the box will be a universal function of temperature, independent of what the walls are made of. This is very different from our usual life experience, where different materials have different colors because how they interact with light depends on the frequency of the light. An ideal black body would absorb all light falling on it, and the light it emits would match the spectrum of our cavity.

A basic challenge of late 19th century physics was to understand this black-body radiation. At low frequencies, everything made sense, with brightness increasing with frequency, and higher temperatures emitting more light. However, a classical model would predict that the brightness would just keep increasing with frequency, resulting in an infinite amount of power getting radiated. That obviously can't happen in real life, and the observed spectrum rolls over and starts to drop at high frequencies. Classical physics could not explain this, and the fact that the energy diverges at high frequencies was known as the "ultraviolet catastrophe", so physicists of the time were well aware there was a serious problem.

Before we see how quantization solves the UV catastrophe, let's look at the classical prediction. The key fact that we'll need is that if we observe the energy in a state at some temperature T, the *relative* probability of observing either energy E_1 or E_2 is $\exp(-(E_1 - E_2)/kT)$ where k is Boltzmann's constant, and T is the temperature in Kelvin. If you look at this, you'll see that for energy differences much less than kT, the two values are roughly equally likely, while for large energy differences, the high-energy state is exponentially disfavored. For those of you, like me, who can't jump very well, it's *possible* for nature to give you enough of a boost to dunk a basketball, it's just *exceedingly* unlikely. Say you need to jump a metre, g is 10, and you weigh 100 kg. Then your potential energy at dunking height is 1*100*10=1 kJ. How likely are you to get this "for free"? Well, assuming you're playing at something like room temperature, T is around 300K, and $k = 1.38e-23$, so the typical thermal energy is around $300 * 1.38e-23 \sim 4e-23$ Joules. The odds that the universe conspires to boost you is then $\exp(-1e3/4e-23) = \exp(-2.5e25)$, or converting to base 10, $10 * *(-2.5e25 * \log_{10}(e)) = 10^{-1.1e25}$. Just to be clear, your odds of dunking are not one in 10^{25}, they are one in $10^{10^{25}}$.

Now let's ask the question "What *is* our typical heigh boost?" The relative probability of observing energy $E = mgz$ (assuming $E = 0$ is the ground and we're at height z) is $P_{rel}(E) = \exp(-E/kT)$. These are relative probabilities, so to get an actual probability, we need to normalize by our total integrated probability, $\int_0^\infty \exp(-E/kT)dE$. We can then calculate the expectation of the energy:

$$E_{typ} = \frac{\int_0^\infty E \exp(-E/kT)dE}{\int_0^\infty \exp(-E/kT)dE}$$

These integrals aren't hard to do, and the answer is that the average energy we'll see is just $< E >= kT$. Note that the mass of whatever is getting bounced around doesn't matter, it's a constant energy of kT. Our typical height is then kT/mg, which for a person is tiny (4e-26 metres). It's *not* tiny for say a nitrogen molecule, where the typical height will be around 8 km. If

you've ever wondered why the air pressure is lower at the top of a mountain, but not zero, this is why. Note also that there's nothing special about height - any way we have of putting energy into a system is going to give us around kT of energy, so if we have n ways of putting energy in, then we'll have nkT Joules on average.

Now that we've worked out the typical energy, we can apply that to a box at fixed temperature, and ask how many ways we can put energy into the light. First, consider a string of length L, tied down at both ends. Since the ends can't move, I can't make the string vibrate with arbitrary wavelengths - only modes that don't move at the ends of the string will work. The allowed wavelengths are then $2L/\lambda$, or if we use frequency instead of wavelength, we have $2L\nu/c$, where c is the speed at which vibrations move in the string. Random vibrations of the string can be broken into terms that have wavelengths of $\lambda/2$, λ, $3\lambda/2$..., and classical physics tells us we can expect *each* of these vibrations to have energy $\sim kT$.

Let's look at a 3D box now. Once again, we need to fit integer numbers of wavelengths into our box. Unlike the 1D string, though, we can pick different numbers of wavelengths to put in the x, y, and z directions, so for some frequency ν, the *total* number of ways I can pick a wave is something like $(2L\nu/c)^3$. If I divide by L^3 to get the energy per unit volume, we now have something like $8\nu^3/c^3$. If you want to ask about the number of possible modes with frequency between ν and $\nu + \delta\nu$, then we can differentiate w.r.t ν, to get $24\nu^2/c^3$. That's the number of distinct modes per unit volume, so if each of those modes has energy kT, then the energy per unit volume per unit frequency should be something like $24kT\nu^2/c^3$. We've been pretty sloppy about counting our modes, so we wouldn't expect the 24 to be exactly right, but it is actually pretty close - the true value is 8π[1]. Our final prediction for the energy density of black-body radiation is then:

$$\rho E_{BB,classical} = 8\pi kT\nu^2/c^3 \tag{1}$$

This works great at low frequencies, where it's known as the Rayleigh-Jeans law. However, you can see that the energy density just keeps increasing with frequency, so this picture can't be correct across all frequencies. Lots of people tried lots of things to try to fix this problem, but it was Planck who figured it out.

At high frequencies, the black body curve is experimentally seen to follow the Wien approximation law, an exponential decay of $\nu^2 \exp(-h\nu/kT)$ for some constant h. People tried several different ways to match the (well motivated) Rayleigh-Jeans law with the (seemingly arbitrary) Wien tail. Planck's big insight was that these could agree if each mode could only have certain allowed energies, instead of the continuum we expect from classical physics. You might start by guessing that the allowed energies for each mode are the same, but if you work through the math (a great exercise for the reader!), you'll find that the UV catastrophe doesn't go away. Planck tried the next obvious thing, which is that the allowed energies in each state are proportional to the frequency, so $E = nh\nu$ for integer n and constant h. If we re-do the energy per state caculation,

[1]We were lucky to get this close. Instead of $8L^3$ we should have used the volume of a sphere: $4\pi/3L^3$, but there are also two polarizations of light. Instead of $3 * 8$, we should have had $2 * 3 * 4\pi/3 = 8\pi$.

we now have

$$\frac{\int_0^\infty E\exp(-E/kT)}{\int_0^\infty \exp(-E/kT)} \rightarrow \frac{\sum_{n=0}^\infty nh\nu\exp(-nh\nu/kT)}{\sum_{n=0}^\infty \exp(-nh\nu/kT)}$$

We've just replaced E in the integral by $nh\nu$, and now we sum over integer values of n. Let's start with the sum on the bottom. We can re-write it as a geometric series $\sum(\exp(-h\nu/kT))^n$. Since the sum of x^n is $\frac{1}{1-x}$[2], the denominator sums to $\frac{1}{1-\exp(-h\nu/kT)}$. We can sum the numerator as well, using the fact[3] that $\sum_{n=0}^\infty nx^n = x/(1-x)^2$. Plugging everything in, we get that the energy per mode is now:

$$\frac{h\nu\exp(-h\nu/kT)/(1-\exp(-h\nu/kT))^2}{1/(1-\exp(-h\nu/kT))} = \frac{h\nu}{\exp(h\nu/kT)-1}$$

For $h\nu << kT$, this expression reduces to the classical limit of kT. However, for $h\nu >> kT$, the energy-per-mode gets exponentially suppressed. Qualitatively, when kT is much less than $h\nu$, then you have a hard time exciting a mode and so a given mode usually has zero energy, instead of kT. We can replace our classical energy-per-mode kT in our classical black-body expression, Equation 1, to get our new, quantum expression:

$$\rho E_{BB} = \frac{8\pi h\nu^3/c^3}{\exp(-h\nu/kT)-1} \tag{2}$$

This is the energy density inside our black-body cavity, but a more common expression is the brightness - how much power per unit solid angle per unit frequency does a black body radiate? To get that, we'll need to multiply by c (since the energy density is travelling at the speed of light) and divide by 4π since there are 4π steradians in a sphere, leaving us with:

$$I_\nu(\nu,T) = \frac{2h\nu^3}{c^2(\exp(h\nu/kT)-1)} \tag{3}$$

Planck's Law (Equation 3) is one of the fundamental results of physics. It established that light was quantized into packets of energy - what we now know are photons. When Planck worked this out, he measured not one but two fundamental constants of nature for the first time - both h and k. Statistical mechanics also that $R = N_A k$ where R is the ideal gas constant in moles and N_A is Avogadro's number. Avogadro's number in turn gives you the mass of atoms. I find it amazing that some of (if not the) first measurements of atomic weights were derived by looking at the properties of light!

Entertainingly, Planck himself did not immediately realize the fundamental importance of his work. In his words, it was "a purely formal assumption ... actually I did not think much about it ...". If you ever feel like you're struggling to understand quantum mechanics, I hope you can take comfort in the fact that Planck himself also didn't fully realize what he was doing.

---

[2]As a reminder, $\sum x^n = 1 + x + x^2 + ... = 1 + x\sum x^n$. Since the sum has to be the same on both sides, you can solve for it to get $1/(1-x)$

[3]We have that $\sum_{0/1}^\infty nx^n = \sum_0^\infty (n+1)x^{n+1} = nx\sum(x^n) + x\sum(x^n) = x\sum nx^n + x/(1-x)$. We've used the fact that the first term in the series is zero when $n = 0$ so we can get away with starting the sum at either $n = 0$ or $n = 1$. Solving for the sum, we have $\sum_0^\infty nx^n = x/(1-x)^2$.

2. COBE-FIRAS Measurement of the Cosmic Microwave Background

It's actually really hard to make a good black body. If you want to see the spectrum of emitted radiation, you need to look inside the black body, which means you don't have a sealed cavity any more. Materials have all sorts of atomic transitions, which will alter the spectrum of emitted light for many setups. The best black body spectrum in nature comes not from the Earth, but from space.

Immediately after the big bang, the universe was hot and dense, and it has been slowly cooling and expanding ever since. When the universe was young, well before any stars or galaxies had formed, it was uniform, hot, and dense - ideal conditions to get nearly perfect black-body radiation. As the universe expands, black body radiation remains black body, just with increasingly lower temperatures. We can look at microwave frequencies and see the light left over from the big bang, which now has a temperature of only 2.73 degrees above absolute zero. This radiation is called the cosmic microwave background (CMB).

The best measurement[4] of the absolute temperature of the CMB was carried out by the FIRAS experiment on board the COBE satellite. The FIRAS measurement is shown in Figure 1, along with the theoretical prediction for a black body with temperature 2.725K. The agreement is exquisite - the error bars are too small to see on the plot. When these results were first presented at a meeting of the American Astronomical Society, the audience broke out into spontaneous applause. This work was recognized with a Nobel Prize in 2006.

3. Discussion

Planck's results on black-body radiation gave us the first indication that the microscopic world was fundamentally different from the world we're used to. Unlike in classical physics, where we can add arbitrary amounts of energy to a system, light comes in discrete packets of energy that depend on their frequency. We can only add/subtract multiples of that amount. This discrete nature, which Planck orignially thought of as a mathematical convenience, became crystal clear through studies of the photoelectric effect.

It would take another quarter-century of experimental and theoretical work before a clear picture of quantum physics came into view, but it all started with Planck's studies of black-body radiation.

---

[4]Canada led many of the first measurements of the CMB temperature. The first measurements of molecules at space were carried out at the Dominion Radio Astronomy Observatory (DRAO) in Penticton, BC. While analyzing various population levels, Andrew McKellar realized they only made sense if there was a background temperature of approsimately 2.3K. He didn't realize this was the CMB, but in fact his answer was closer than the original direct measurement by Penzias and Wilson, who found 3.5K. Herb Gush at UBC led a group using rockets to try to measure the spectrum of the CMB. They got the right answer and released their results around the same time as COBE, but the satellite data were orders of magnitude more precise, and so the rocket data languished in obscurity.

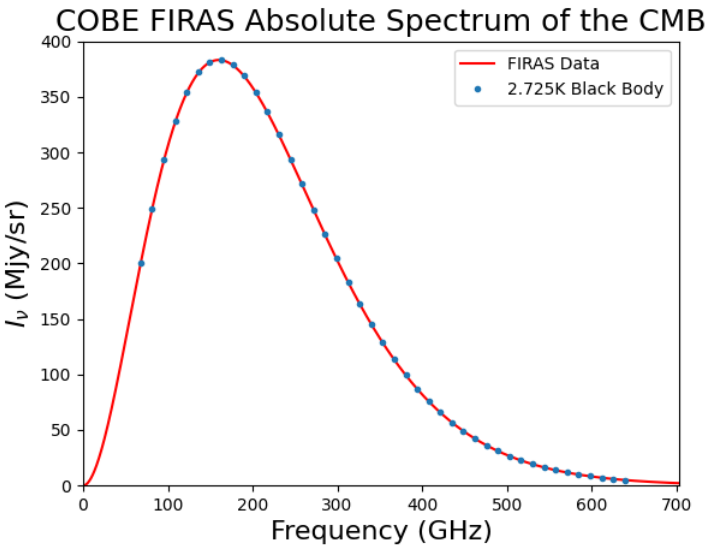

Fig. 1.— Results from the COBE-FIRAS measurement of the CMB absolute spectrum. $x-$axis is frequency in GHz, $y-$axis is specific intensity, in mega-Janksies per steradian. Blue points are the experimental data. The red curve is the theoretical prediction (including amplitude) for a black body with temperature 2.725K. The agreement is spectacular - the error bars are plotted, but they are much smaller than the points in the plot.

1. Expectations

: $< f(x) >= \int f(x) PDF(x) dx$ where PDF is the probability density of x. If x can only take on discrete values, $< f(x) >= \sum f(x) P(x)$. These are correct when the PDF/probabilities are properly normalized, i.e. $\int PDF(x)dx = 1$ if continuous, and $\sum P(x) = 1$ if discrete. If we only have relative probabilities $\tilde{P}$, then we would have to normalize: $P(x) = \tilde{P}/\sum \tilde{P}$. For black body, statistical mechanics told us the relative probabilities of states, $\tilde{P} = e^{-nh\nu/kT}$. To get a proper expectation, we had to normalize to get true probabilities, hence

$$\langle E \rangle = \frac{\sum nh\nu \exp -nh\nu/kT}{\sum \exp -nh\nu/kT}$$

Very important - if I take a particular state and measure its energy in one black body, I'll get some multiple of $h\nu$. If I measure it again, *I will get the same value* unless something has changed. When we talk about expectations, especially in quantum, we mean that if we prepared thousands of copies of a black body (by which we mean they are statistically identical, but are different realizations), pick out a single state, and measure the energy of that state in *each* black body, then the average of all those measurements gives us the expectation value.

2. Magnetic Moments/Angular Momentum

In SI units, the magnetic moment μ of a current loop is the current times the area IA. If we have a particle with charge q moving in a circle of radius r with velocity v then the area is just πr^2. The current is slightly more complicated - the particle goes around the loop $v/2\pi r$ times per second. Since the current is the charge per second, if we sit and watch, we'll see that the charge per second is the charge per particle times particles per second. For our setup, we have $I = qv/2\pi r$, so the magnetic moment is

$$\mu = IA = \frac{qv}{2\pi r}\pi r^2 = \frac{qvr}{2}$$

The angular momentum is $L = mr \times v$, and since for circular motion, v is always perpendicular to r, then the magnitude of L is mrv. If we take the ratio of the magnetic moment to the angular momentum we have:

$$\frac{\mu}{L} = \frac{qvr/2}{mrv} = \frac{q}{2m}$$

That is, the only thing that matters is the ratio of the charge to the mass. If I had a macroscopic spinning ball of electrons and I change the distribution of the elctrons or mess with how they're moving, I'll get different values for L and for μ, but the *ratio* would always be the same. In a classical world, if I had a ball with fixed total charge and mass, the only way to change μ/L is to move some of the charges/mass around relative to each other. For instance, if I moved all the charge (but not the mass) to the center of my ball, then μ would be zero but I'd still have angular momentum.

Of course, in the quantum world, this classical picture is just wrong, and any classical picture we tried to make would just fall over, especially for an electron. However, decent theories of particle

physics do need to be able to predict the ratio. Protons/neutrons and electrons have very different ratios, which tells us that there's something fundamentally different about protons/neutrons. In fact, you might even wonder why a neutron has a magnetic moment at all, given it has no net charge!

3. Polarizers vs. Stern-Gerlach

We think of light as a wave phenomenon and electrons/atoms as particles, but we'll see that light can act like particles (which we call photons), and electrons/atoms act like waves. While this situation is strange, happily for us, nature treats electrons and photons (mostly) the same. That means we can reproduce a lot of the Stern-Gerlach results using polarized filters and light. An important difference is that electrons are spin-1/2 particles, while photons are spin-1. Essentially, the "opposite" of an electron with spin pointing up is an electron with spin pointing down, but the electric field in light oscillates up and down, the "opposite" of up-and-down (vertically polarized) light is left-and-right (horizontally polarized) light. The equivalent of $|+z\rangle$ and $|-z\rangle$ for electrons is $|+Q\rangle$ and $|-Q\rangle$ for light, where $|+Q\rangle$ has the electron vertically polarized, and $|-Q\rangle$ is horizontally polarized. The equivalent of $|+x\rangle$ and $|-x\rangle$ is $|+U\rangle$ and $|-U\rangle$ where the $E-$field is rotated 45° relative to either vertical or horizontal.

With this in mind, we can reproduce many of the Stern-Gerlach experiments with light. We just have to remember to divide all the angles by 2. A polarizing sheet passes a single polarization of light, and so is equivalent to a Stern-Gerlach machine splits the silver atoms into $|+z\rangle$ and $|-z\rangle$, but then blocks the $|-z\rangle$ atoms, so we're left with pure $|+z\rangle$ states. If we hold two polarizers at 90° to each other, then that's equivalent to a Stern-Gerlach machine that passes $|+z\rangle$ followed by one that passes $|-z\rangle$, and we would not expect anything to make it through. Indeed, that is what we see. If we put a $|+x\rangle$ after a $|+z\rangle$, we'd expect half the particles coming out of $|+z\rangle$ to make it through $|+x\rangle$. The light equivalent is to rotate a second polarizer by 45°. Try that, and you'll see things get a bit darker, but not too much.

Of course, we could put a $|-z\rangle$ SG after our $|+x\rangle$ one, and half of what makes it through $|+x\rangle$ should come out $|-z\rangle$. A $|+z\rangle\,|+x\rangle\,|-z\rangle$ SG sequence lets 25% of the particles that came through $|+z\rangle$ through, while if we pull the $|+x\rangle$ machine, nothing comes through. To see this in action, I suggest you hold a pair of polarizers together rotated at 90°, then hold a third polarizer at 45° and stick before, after, and between the two polarizers.

This behavior is easy to explain classically - the field *strength* projected onto an axis rotated by 45° is reduced by a sqrt(2). Since power goes like the field squared, the power is reduced by a factor of 2. Add another polarizer and the field is down by another sqrt(2). Where things get wacky is if we turn down our light source so only one photon at a time makes it through. I could set up a detector that puts out a sound proportional to the amount of energy coming in. What we'd hear is a bunch of clicks, where each click is the same loudness (since all photons of a given frequency have the same energy).

If I put a 45° polarizer after a vertical polarizer, I'd get half as many clicks, but they would remain the same loudness. When a photon comes out of the first polarizer, it happily says "yup,

I'm vertically polarized". If we put a bunch of vertical polarizers together, one after another, once a photon makes it through the first one and has decided it's vertically polarized, it will stay vertically polarized and make it through all the polarizers. However, if it hits a 45° polarizer, the photon will flip a coin to decide if it's +45° or −45°. If it's +45° then the photon makes it through unscathed - no energy is lost, and the vertical field strength before the polarizer is *the same* as the 45° field strength after the polarizer. Of course, if the photon's coin flip comes up −45°, then *all* the energy is lost. Rather remarkably, after a photon makes it through the 45° polarizer, it has lost all memory that it used to be a $|+Q\rangle$ photon. Now the photon has happily decided it is $|+U\rangle$ *and no experiment will ever show it used to be* $|+Q\rangle$. Any memory it could have had has been completely lost. Now when the photon hits the $|-Q\rangle$ polarizer, all the photon knows is right now it's $|+U\rangle$. The photon flips another coin, and if that comes up heads, the photon decides it's now horizontally polarized/$|-Q\rangle$ and makes it through, again with no energy loss. There's absolutely no way to explain this classically - instead of losing sqrt(2) in field strength at each polarizer, the photon either makes it through at full strength, or gets absorbed.

We could (but won't) also do the analog of SG experiment 4 with light. There are polarized beam splitters/combiners. To do experiment 4, we would start with say a vertically polarized beam, send it through a splitter at 45° so $|+U\rangle$ goes one direction, and $|-U\rangle$ goes another. We could then use mirrors to send those beams down different paths, and then recombine them with a beam combiner, giving us once again a single beam. Since we haven't messed with the beam, what comes out remains vertically polarized (*i.e.* in the $|+Q\rangle$ state), so if we feed it through a horizontal polarizer, the whole beam is blocked. However, if we block *one* of the paths, then what comes out of the combiner will be either $|+U\rangle$ or $|-U\rangle$, depending on which path we blocked. Now some light *will* make it through the final horizontal polarizer. This remains true even when we go back to the single-photon limit. The probability the photon makes it through is *not* equal to the probability it goes through the $|+U\rangle$ path plus the probability it goes through the $|-U\rangle$ path. Somehow the photon is going through *both* paths, and what gets added at the end is not probability, but a complex amplitude. What *is* correct is that we can calculate the amplitudes of going through both paths, and we *can* add those amplitudes together. The final probability is the square of the absolute value of that amplitude. When a photon can go through both paths, the $|+Q\rangle$ amplitudes constructively interfere, while the $|-Q\rangle$ paths destructively interfere.

Of course, our photons (and our silver atoms) are very shy. If we watch them to find out *which* path they followed, the interference goes away, as shown in Figure 1. Photons and atoms behave identically in this way - we can send either through an experiment where and individual photon/particle can go down multiple paths, have it interfere with itself, and get an interference pattern out the end. How we do that splitting is different between photons and particles, but the *way* in which the paths interfere is identical and is governed by the same math. If we try to measure which path the particle/photon followed, then the interference goes away. We'll see this more quantitatively later on, but it is absolutely impossible to make *any* measurement that would tell us which path it actually followed without destroying the interference.

Fig. 1.— A scientifically accurate quantum particle. When sent through two paths, the particle interferes with itself. However, when we measure which of the two paths the particle took, the above represents the new interference pattern.

In these notes, we'll concentrate on drawing analogies to electromagnetic (EM) fields to hopefully develop a little intuition about why some of the things in quantum mechanics (QM) work the way they do. If we're careful, the analogies work very well - after all, if we take quantum electrodynamics in the suitable classical limit, we have to end up with Maxwell's equations. There's no way to stare at Maxwell's equations hard enough and say "gee, I guess the world must really be quantum, and QM must behave this way." Instead, we'll say "here's how this particular thing works in QM, and here's how it works in EM. Now look - the math is exactly the same!"

1. Waves vs. States

Before we start, picture a violinist playing a note. How would you describe the motion of the string? Well, you've got two choices. You could specify the position and velocity at every point along the string, then use Newton's laws to work out what will happen as the string vibrates. Alternatively, since a string can only vibrate with (half-)integer periods, you could instead specify the amplitude and phase of each harmonic. The first way requires specifying two continuous functions, while the second way just requires specifying a handful of numbers. It doesn't matter which way you pick because both descriptions carry the full information about the motion of the string. QM is very similar. We could specify the probability amplitude of a particle's amplitude to be in a particular location, then evolve that function using the Schrodinger equation - this is the Schrodinger wave mechanics path. Equivalently, we could specify the amplitude for the particle to be in a range of states (equivalent to the different harmonics of the violin string), and watch how those states evolve - this is the Heisenberg matrix mechanics path. Although they were developed separately, Schrodinger eventually showed that the two paths are indeed mathematically equivalent. In this class we start with matrix mechanics, because the math is a lot easier. As we draw analogies with electromagnetic waves, you may have more intuition for the wave mechanics description, but keep in mind that what is true for one path must also be true for the other.

2. Quantum States through an Electromagnetic Lens

With that out of the way, let's go through some of the properties of EM fields, and see how the relate to QM. For starters, the energy density of an EM field is proportional to $E^2 + B^2$, the *square* of the electric/magnetic fields. Since photons are quantized in units of energy, that means the probability of finding a photon in a location has to be proportional to the field strengh squared. It is most decidedly *not* proportional to the field strength. In EM we generally try to work out the fields - it is the fields after all that add linearly together, and have phases. After that, we can square the fields to see where the energy is, which has to be where the photons are. The analog of the fields in QM is the wave function, or the *amplitude* to be in a state. We will put all our work into figuring out the amplitudes to be in different states. It is the amplitudes that add when we combine states *without measuring along the way.* Then once we've worked out all the amplitudes, we square the final amplitude to figure out where we're likely to find a particle/what value(s) of an observable we're likely to measure.

Now consider a EM plane wave propagating in the $+z$ direction in a vacuum. The most general way of writing the associated electric field for a wave at a single frequency $\nu = 2\pi\omega$ uniform in space is:

$$E(\vec{x},t) = Re\left[(c_x\hat{x} + c_y\hat{y})e^{i\omega(t-z/c)}\right] \tag{1}$$

where both c_x and c_y can be complex. If I set $c_x = 1$ and $c_y = 0$ then we get

$$E(\vec{x},t) = \cos\left(\omega(t-z/c)\right)\hat{x} \tag{2}$$

If you recall from the last set of lecture notes, this corresponds to a pure $+Q$ EM wave. Also note that if I made c_x complex, nothing would really change. I would pick up a phase shift in the wave, which is equivalent to shifting our zero point in time, or our reference point in space where $z = 0$. These aren't really fundamental, so if I sit and watch the wave go by, all I would say is "I have a $+Q$ EM wave." Similarly, I could set $c_x = 0$ and $c_y = 1$. Now I would have a wave with E purely in the y direction - I would call this a $-Q$ EM wave.

Life gets more interesting when I let both c_x and c_y be non-zero. If I have $c_x = c_y = \frac{1}{\sqrt{2}}$, then my E-field is

$$E(\vec{x},t) = \frac{1}{\sqrt{2}}\left(\hat{x} + \hat{y}\right)\cos\left(\omega(t-z/c)\right) \tag{3}$$

We would call this a $+U$ wave, since the $E-$field is oscillating at 45° to the $x-$axis. Once again, an overall phase change doesn't affect the big picture physically - I would still have a $+U$ wave. If I flip the sign of c_y, I get the same thing but now instead of the $E-$field going at 45° to the *right* of our $+x$-axis, we have if oriented at 45° to the *left*. Once again, an overall phase doesn't change anything. I'd also like to point out that we get the same answer if we flip the sign of c_x, since that is equal to flipping the sign of c_y, then introducing an overall phase of -1. There's nothing special about flipping the sign of the $+Q/x-$component or the sign of the $-Q/y-$component. As long as we flip only one, we can go from $+U$ to $-U$. If we flip *both*, we have just put in an overall phase, and we stay in a $+U$ state.

Finally, let's have[1] $c_x = 1/\sqrt{2}$ and $c_y = i/\sqrt{2}$. Now our wave changes quite a lot. We get

$$E(\vec{x},t) = \cos\left(\omega(t-x/c)\right)\hat{x} + \sin\left(\omega(t-x/c)\right)/\sqrt{2}\hat{y} \tag{4}$$

Instead of the electric field oscillating in a plane, now we have the electric field sweeping clockwise around in a circle as the wave propagates. If I flip the sign of c_y (or c_x - once again it's equivalent), now I still have a field going around a circle, but this time it's in a counter-clockwise direction. These circular modes are known as $+V$ and $-V$. Not surprisingly, once again an overall phase does not change the big picture - we'll still have the $E-$field going in circles.

Let's look and see where we're at. I can specify the coefficients c_x and c_y, and that in turn specifies the behavior of our EM wave. An *overall* phase change doesn't affect the qualitative behavior of our wave, but a *relative* phase change between c_x and c_y can change the picture

[1] We don't want to change the average energy in the field with these rotations. For the linearly polarized states, the E-field oscillates between -1 and 1, so the average *squared* field is 1/2. For this circular mode, the field strength doesn't change, so to keep the average squared field to be 1/2, we need to have a $1/\sqrt{2}$ in the field strength.

dramatically. To make the analogy to our spin-1/2 states even clearer, we'll relabel c_x and c_y as c_{+Q} and c_{-Q}. I get complete information about my wave by specifying the *relative* amplitudes and phases of c_{+Q} and c_{-Q}. Once I have these two numbers, I have everything, so I don't need to specify anything else. Similarly, I could have instead told you $c_{+U} = (c_{+Q} + c_{-Q})/\sqrt{2}$ and $c_{-U} = (c_{+Q} - c_{-Q})/\sqrt{2}$. In both cases, all I need to do is tell you the relative amplitudes and phases - if I tell you that in the $\pm U$ basis, you can work out what it is in the $\pm Q$ basis, since $c_{+Q} = (c_{+U} + C_{-U})/\sqrt{2}$ and $c_{-Q} = (c_{+U} - c_{-U})/\sqrt{2}$. I could even have told you the coefficients in the $\pm V$ basis where $c_{+V} = (c_{+Q} + ic_{-Q})/\sqrt{2}$ and $c_{-V} = (c_{+Q} - ic_{-Q})/\sqrt{2}$. As long as I tell you the relative amplitudes and phases of the two components in any of these basis sets, you have complete information about what the wave is doing. Sometimes a particular basis is easier to work in than others - if so, by all means pick that one! However, when you ask yourself any physical question, you'll get the same answer no matter which basis set you use.

I hope this analogy helps develop the intuition for spin-1/2 systems. We can let $+Q$ correspond to $|+Z\rangle$ and $-Q$ correspond to $|-Z\rangle$. I find it more intuitive to think of the amplitudes along the $+x$ and $+y$ directions as independent. For spin-1/2 systems, it's the $|\pm z\rangle$ (or equivalently $|\pm x\rangle$ or $|\pm y\rangle$) amplitudes that are independent. I can pick out a pure state in any direction just by specifying the relative amplitudes and phases of the $|\pm Z\rangle$ states. Again, there's nothing special about the $z-$direction. I could just as well have specified the $|\pm Y\rangle$ states instead, or any other basis I care to choose.

Finally, now that we know we can write down the amplitudes in a particular basis, we can see how to calculate probabilities. The *strength* of an $E-$field in a particular direction corresponds to finding the amplitude of a state along a basis vector. The *energy* goes like the field squared, so if I grab a linear polarizer, orient it some direction, and ask "what fraction of the energy makes it through?", the answer is going to be set by the field in that direction *squared*. For simplicity, let's say we have a linearly polarized EM wave once again propagating in the $+z$ direction, but at some random orientation in the $xy-$plane. We find the $x-$component by taking $\vec{E} \cdot \hat{x}$, and the $y-$component by taking $\vec{E} \cdot \hat{y}$. We can then express the field in by adding the unit vectors back in:

$$\vec{E} = \left(\vec{E} \cdot \hat{x}\right) \hat{x} + \left(\vec{E} \cdot \hat{y}\right) \hat{y} \tag{5}$$

Similarly, if I have a silver atom in some spin state $|S\rangle$, then the amplitude c_{+Z} to be in $|+Z\rangle$ is $\langle +Z|S\rangle$. If I send that atom through a Stern-Gerlach machine, the probability I measure its spin to be $|+Z\rangle$ is $c^*_{+Z}c_{+Z}$. Similarly, the amplitude to be in $|-Z\rangle$ is $c_{-Z} = \langle -Z|S\rangle$, and we can write our state in the same way we wrote the field in Equation 5:

$$|S\rangle = |+Z\rangle \langle +Z|S\rangle + |-Z\rangle \langle -Z|S\rangle \tag{6}$$

Yet again, there is nothing special about the $|\pm Z\rangle$ basis, so we could just as well have used $|\pm X\rangle$ or $|\pm Y\rangle$ instead of $|\pm Z\rangle$. We'd be writing down the state in a different basis, so the components would change, but the state itself doesn't care what basis you use.

3. Spin Basis States

I'm not a fan of how the book derives the $|\pm Y\rangle$ states in the $|\pm Z\rangle$ basis. If this isn't making sense to you, go back and look through the book Section 1.5. We'll accept that the $|+X\rangle$ state in the $Z-$basis is $\frac{1}{\sqrt{2}}|+Z\rangle + \frac{1}{\sqrt{2}}|-Z\rangle$. If we have a $|+Y\rangle$ state, we know the probability of measuring its angular moment along the $z-$axis will give us $+\hbar/2$ half the time, and $-\hbar/2$ half the time. Similarly, if we measure along the $x-$axis, we'll once again find a 50-50 split. We know the amplitude of a $|+Y\rangle$ state in the $|+Z\rangle$ direction has to be $\frac{1}{\sqrt{2}}$ times some phase, since we have a 50% change of measuring $+\hbar/2$ when we go measure the angular momentum in the $z-$direction. Similarly, we know the amplitude in the $|-Z\rangle$ must also be $\frac{1}{\sqrt{2}}$, again up to a phase factor. Now, any *overall* phase can't matter, so we can freely pick the phase that makes the $|+Z\rangle$ term be $\frac{1}{\sqrt{2}}$. That leaves us with

$$|+Y\rangle = \frac{1}{\sqrt{2}}|+Z\rangle + \frac{\alpha}{\sqrt{2}}|-Z\rangle \tag{7}$$

where α is a complex number with magnitude 1, but phase to be determined. We also know the amplitude squared to be in the $|+X\rangle$ state must be $\frac{1}{2}$, so we can write:

$$\langle +X| + Y\rangle^* \langle +X| + Y\rangle = \frac{1}{2} \tag{8}$$

Since we're working in the $|\pm Z\rangle$ basis, we can sub in our $z-$basis representations of $\langle +X|$ and $|+Y\rangle$ to get the amplitude:

$$\langle +X| + Y\rangle = \left(\langle +Z|/\sqrt{2} + \langle -Z|/\sqrt{2}\right)\left(|+Z\rangle/\sqrt{2} + \alpha|-Z\rangle/\sqrt{2}\right) \tag{9}$$

If we multiply these terms out, recalling that $\langle +Z| + Z\rangle = 1$ and $\langle +Z| - Z\rangle = 0$, then we have

$$\langle +X| + Y\rangle = \frac{1+\alpha}{2} \tag{10}$$

We know that thing squared had better be equal to $\frac{1}{2}$, and that the magnitude of α is 1. Let's see where that leaves us:

$$\langle +X| + Y\rangle^* \langle +X| + Y\rangle = \left(\frac{1+\alpha^*}{2}\right)\left(\frac{1+\alpha}{2}\right) = \frac{1+\alpha^*\alpha+\alpha+\alpha^*}{4} \tag{11}$$

Since the magnitude of α is one, we're left with

$$\frac{2+\alpha+\alpha^*}{4} = \frac{1}{2} \tag{12}$$

The only way that is true is if $\alpha + \alpha^* = 0$. For that to happen, the real part of α must be zero, and so α must be either $+i$ or $-i$ (since those are the only pure imaginary numbers with amplitude 1). In fact these two solutions correspond to the $|+Y\rangle$ and $|-Y\rangle$ states. The fact that we use right-handed coordinates (*i.e.* we have $\hat{x}\times\hat{y} = \hat{z}$ and not $-\hat{z}$) means that the $|+Y\rangle$ state corresponds to $\alpha = +i$. We can now write down the $|\pm Y\rangle$ states in the $|\pm Z\rangle$ basis:

$$|\pm Y\rangle = \frac{1}{\sqrt{2}}|+Z\rangle \pm \frac{i}{\sqrt{2}}|-Z\rangle \tag{13}$$

Incidentally, even this simple example shows why we absolutely have to use complex numbers in quantum mechanics. If we say that a state measured to have angular momentum along one axis has a 50% change of getting either $+\hbar/2$ or $-\hbar/2$ along either of the other two axes, and we accept that specifying a pair of amplitudes in the positive and negative directions is a complete description, then the only way that can work is if (at least) one of the sets of basis vectors has complex amplitudes. We just can't explain even very simple behaviors without complex amplitudes.

Notes on Linear Algebra for Quantum Mechanics

By Jonathan Sievers, McGill University

Quantum mechanics obeys the rules of linear algebra. It didn't necessarily have to be this way, and when Heisenberg first developed what is now known as matrix mechanics, he didn't realize the operators he had written down were matrices. In this note, we will try to lay out some of the basics of linear algebra relevant to quantum mechanics.

1. Matrix Multiplication

The distinctive nature of linear algebra comes from how we multiply two matrices. If we have

$$\mathrm{C} = \mathrm{AB} \tag{1}$$

and matrices A and B have sizes n, k and k, m, respectively, then

$$C_{ij} = \sum_k A_{ik} B_{kj} \tag{2}$$

For matrix multiplication, the second dimension of A has to be the same as the first dimension of B (these are often referred to as the "interior dimensions"). Except in special cases, matrix multiplication is not commutative $\mathrm{AB} \neq \mathrm{BA}$. It is, however, associative: $\mathrm{ABC} = (\mathrm{AB})C = \mathrm{A(BC)}$. Addition works like you'd expect. If $\mathrm{C} = \mathrm{A} + \mathrm{B}$ then $C_{ij} = A_{ij} + B_{ij}$, and A and B must have the same sizes. Multiplication by a scalar also works as you'd expect. For scalar α, $\mathrm{C} = \alpha\mathrm{A}$ has $C_{ij} = \alpha A_{ij}$. Finally, the preceeding properties also guarantee that matrix math is distributive: $\mathrm{A(B + C)} = \mathrm{AB} + \mathrm{AC}$.

Matrix-vector and vector-vector operations are just special cases of matrix-multiplication where one (or more) of the dimensions is one. If B is a vector (which we'll write b to make it clear it's a vector), then m is 1, and we can drop the j from Equation 2 since we know it is always 1[1]. In that case we have:

$$c_i = \sum_k A_{ik} bk \tag{3}$$

Similarly, we can have A be a vector, in which case $n = 1$, we'll refer to it as a, and we have:

$$c_j = \sum_k a_k \mathrm{B}_k j \tag{4}$$

Standardly, we refer to $n \times 1$ vectors (*i.e.* a) as row vectors, and $1 \times m$ vectors (*i.e.* b) as column vectors. By convention, if you just say "vector", you usually mean a column vector.

An even more special case of matrix multiplication is when both matrices are vectors. Since the interior dimensions need to line up, we can either have a row vector times a column vector

[1] Or 0, depending on if you were raised in a C-like or FORTRAN-like environment

which produces a scalar (or, depending on your situation, possibly a 1×1 matrix). This is called the *inner* or *dot* product. We can also have a column vector times a row vector, in which case our output is a matrix. This is called the *outer* product. In fact, if you go back and look at the definition of matrix multiplication in Equation 2, you can see that you can just sum the outer product of every column of A with the corresponding row of B to get C. Equivalently, we can also say that every entry in C is the dot product of the corresponding row of A and column of B. When you think about matrix multiplication as a set of outer products, notice that we could re-order the columns of A, and as long as we re-order the columns of B in the same way, the product AB is unchanged.

It is useful to remember that two vectors must have the same dimensions if you want the inner product, but there is no such requirement for outer products. Any old column and row vector will happily make an outer product. It will be square if they're the same dimensions, but it doesn't have to be.

2. Transposes

Every matrix A has a unique transpose A^T with:

$$A^T_{ji} = A_{ij} \tag{5}$$

If you take the tranpose of the transpose of a matrix, you end up with the original matrix again.

A key identity, which again just drops out of the definition of matrix multiplication, is that

$$(\mathrm{AB})^T = \mathrm{B}^T\mathrm{A}^T \tag{6}$$

$$(\mathrm{ABC})^T = \mathrm{C}^T\mathrm{B}^T\mathrm{A}^T \tag{7}$$

and so on. If we have column vectors a and b, then the inner product is:

$$a \cdot b = a^T b = b^T a \tag{8}$$

Since the vectors have to have the same length to take an inner product, it doesn't matter which order you put them in. The outer product of a with b is ab^T, and now it very much does matter which order you use:

$$\left(ab^T\right)^T = \left(b^T\right)^T a^T = ba^T \tag{9}$$

In words, the outer product of a^T and b is the transpose of the outer product of b^T and a.

The distance d between two points in Cartesian geometry is $d^2 = dx^2 + dy^2 + dz^2$. In higher dimensions, you just keep adding the square of the distances. If you had a vector containing the separation between two points, then the inner product gives you exactly that squared distance. If $d = [dx, dy, dz, ...]$ then:

$$d^T d = \sum_k d_k d_k = dx^2 + dy^2 + dz^2 + ... \tag{10}$$

When we talk about the "length" of a vector, this is usually what we mean[2].

3. The Identity Matrix and Matrix Inverses

There is an identity matrix I, which has ones on the diagonal and zeros everywhere else:

$$I_{ij} = \delta_{ij} \tag{11}$$

where δ_{ij} is the Kronecker δ function: one if $i == j$ and zero otherwise. For suitably sizes identity matrices, we have:

$$\mathrm{A} = \mathrm{AI} \tag{12}$$

$$\mathrm{A} = \mathrm{IA} \tag{13}$$

which you can show by, once again, going back to the definition of matrix multiplication. While we often talk about "the" identity matrix, there are an arbitrarily large number of them of different sizes. The size will be determined by how it is being used - if A is $n \times m$, then in IA, I must be $n \times n$ while in AI, I must be $m \times m$.

"Most" square matrices have an inverse A^{-1}, defined to be

$$\mathrm{AA}^{-1} = \mathrm{I} \tag{14}$$

In principle, we could have separate left and right inverses:

$$\mathrm{AA}_r^{-1} = \mathrm{I} \tag{15}$$

$$\mathrm{A}_l^{-1}\mathrm{A} = \mathrm{I} \tag{16}$$

However, we can take advantage of the fact that matrix multiplication is associative to show that the left and right inverses are the same. We can show this by looking at the product $\mathrm{A}_l^{-1}\mathrm{AA}_r^{-1}$:

$$\mathrm{A}_l^{-1}\mathrm{AA}_r^{-1} = \left(\mathrm{A}_l^{-1}\mathrm{A}\right)\mathrm{A}_r^{-1} = \mathrm{IA}_r^{-1} = \mathrm{A}_r^{-1} \tag{17}$$

$$\mathrm{A}_l^{-1}\mathrm{AA}_r^{-1} = \mathrm{A}_l^{-1}\left(\mathrm{AA}_r^{-1}\right) = \mathrm{A}_l^{-1}\mathrm{I} = \mathrm{A}_l^{-1} \tag{18}$$

So the left and right inverses have to be the same, and we can just talk about "the" inverse. Two matrices are said to commute if AB = BA, and we have now shown that a matrix always commutes with its inverse (if it exists): $\mathrm{AA}^{-1} = \mathrm{A}^{-1}\mathrm{A} = \mathrm{I}$ [3]. One corollary of this result is that the dot product of the i^{th} row of A against the j^{th} column of A^{-1} is zero unless $i = j$, but also the dot product of the j^{th} *row* of A against the i^{th} *column* of A^{-1} is also 0 unless $i = j$. This follows naturally from interpreting matrix multiplication as the dot product of each row of the first matrix against each row of the second matrix.

---

[2]More generally, one can talk about different norms of a vector. This is the Ł^2-norm, which is by far the most commonly used one in physics.

[3]Our proof only relied on the existence of an identity operator and associativity, so in general an object will commute with its inverse (if it exists) in any group, even if the operator is not commutative. Matrix multiplication is just a special case of this more general result.

4. Eigenvalues and Eigenvectors

A particularly useful way of looking at matrices is through their *eigenvalues* and *eigenvectors*. Normally, they are introduced by solving the following for vector v and scalar λ for some matrix A:

$$\mathrm{A}v = \lambda v \tag{19}$$

Instead, we'll just assume that we can write A as follows:

$$\mathrm{A} = \mathrm{V}\Lambda\mathrm{V}^{-1} \tag{20}$$

for some matrix V and diagonal matrix Λ. It isn't obvious we could do this, but it turns out that in quantum mechanics, we actually can do this for every matrix where we'd want to[4]. Of course, it's not particularly surprising we can factor a matrix this way - if A is $n \times n$, then there are n^2 entries in A, n^2 entries in V, and n entries in Λ, so we have $n^2 + n$ numbers we can pick to match the n numbers in A. That's more flexibility than we really want, so we'll decree every column of V has length 1, *i.e.* $v_i^T v_i = 1$ if v_i is the i^{th} column of V. We get to pick the first $n-1$ entries this way, but the last entry has to make the length 1, so we only get to pick $n-1$ numbers. Since we have $n-1$ numbers per column, n columns, and n numbers in Λ, we have a total of n^2 numbers we get to pick, to match the n^2 numbers in A.

We won't go into how to calculate the eigenvalues/eigenvectors of a matrix - that is, in general, a very hard problem, and people devote their careers to improving eigenvalue codes. Unless you, too, wish to devote your career, we suggest you use someone else's routines (*e.g.* numpy.linalg.eig/numpy.linalg.eigh). We will, however, show what you can do once you have the eigenvalues/eigenvectors. Again, let v_i be the i^{th} eigenvector. Then:

$$\mathrm{V}^{-1}v_i = [000...010...00] \tag{21}$$

where the non-zero entry is in the i^{th} position. A diagonal matrix times a vector just rescales each element of the vector by the corresponding diagonal element, so

$$\Lambda\mathrm{V}^{-1}v_i = [000...0\lambda_i 0...00] \tag{22}$$

Finally, multiplying a matrix by a vector with one non-zero element just rescales the corresponding column of the matrix by the value of that element. This gives:

$$\mathrm{A}v_i = \mathrm{V}\Lambda\mathrm{V}^{-1}v_i = \mathrm{V}[000...0\lambda_i 0...0] = \lambda_i v_i \tag{23}$$

This is one of the most fundamental results in linear algebra. If we multiply a matrix by an eigenvector, we get the eigenvector scaled by the corresponding eigenvalue back. Normally, if we multiply a matrix by a vector, we get a new vector pointing in a different direction, *unless that vector is an eigenvector*. If the vector is an eigenvector, it continues to point in the same direction, hence the name[5].

---

[4]Every symmetric/Hermitian matrix has can be factored this way. Some non-symmetric matrices can't. The official name for these matrices is *defective* (seriously).

[5]Eigen is German for "own".

One immediate consequence of Equation 23 is that we can trivially carry out repeated multiplications of a matrix times an eigenvector. We have

$$\mathrm{A}^2 v_i = \mathrm{A}(\mathrm{A}v_i) = \mathrm{A}\lambda_i v_i = \lambda_i^2 v_i \tag{24}$$

$$\mathrm{A}^n v_i = \lambda_i^n v_i \tag{25}$$

If we have a vector that is the sum of two eigenvalues $x = \alpha_i v_i + \alpha_j v_j$ then we have:

$$\mathrm{A}x = \alpha_i \mathrm{A}v_i + \alpha_j \mathrm{A}v_j = \alpha_i \lambda_i v_i + \alpha_j \lambda_j v_j \tag{26}$$

In words, we stretch each eigenvector out by its eigenvalue and sum them together. In general, this will point in a different direction from the original vector, *unless* the eigenvalues are the same. Similarly,

$$\mathrm{A}^n x = \alpha_i \lambda_i^n v_i + \alpha_j \lambda_j^n v_j \tag{27}$$

Now what happens if we have an arbitrary vector x? We'd like to be able to express it as a sum of eigenvectors, *i.e.* $x = \sum_i \alpha_i v_i$, but that's just $x = \mathrm{V}b$ for some vector b. The entries of b tell us how much of each eigenvector is in x. Fortunately, we know exactly how to work out b:

$$\mathrm{V}b = x \tag{28}$$

$$\mathrm{V}^{-1}\mathrm{V}b = \mathrm{V}^{-1}x \tag{29}$$

$$\mathrm{I}b = \mathrm{V}^{-1}x \tag{30}$$

$$b = \mathrm{V}^{-1}x \tag{31}$$

We can express x in terms a sum of eigenvectors just by taking $\mathrm{V}^{-1}x$. Then we can multiply by an arbitrary power of A:

$$\mathrm{A}^n x = \mathrm{V}\Lambda^n \mathrm{V}^{-1}x \tag{32}$$

This expression will be key in quantum mechanics when we work out time evolution. We can, for instance, define $e^{\mathrm{A}}x$ as $(\mathrm{A} + \mathrm{A}^2/2 + \mathrm{A}^3/3! + ...)x$. If x is an eigenvector v_i, that gives us

$$\lambda_i v_i + \lambda_i^2/2 v_i + \lambda_i^3/3! v_i + ... \tag{33}$$

$$= \left(\lambda_i + \lambda_i^2/2 + \lambda_i^3/3 + ...\right) v_i \tag{34}$$

$$= e^{\lambda_i} v_i \tag{35}$$

If we had many eigenvectors in our vector, we can take advantage of this to write down:

$$e^{\mathrm{A}}x = \mathrm{V}e^{\Lambda}\mathrm{V}^{-1}x \tag{36}$$

Now that we have some experience with eigenvalues/eigenvectors, we can use this familiarity to see which matrices might not have inverses. Assuming the eigendecomposition exists, then if we have a zero eigenvalue,

$$\mathrm{A}v = 0v = 0 \tag{37}$$

so we'll get zero when we multiply our matrix by the corresponding eigenvector. Since a matrix times a vector of zeros is always zero, there's no way to write down a matrix inverse that will give

us our starting vector back. Normally, $\mathrm{A}^{-1}\mathrm{A}x = x$, but for an eigenvector with zero eigenvalue we have

$$\mathrm{A}^{-1}\mathrm{A}v_i = \mathrm{A}^{-1}\left(\mathrm{A}v_i\right) = \mathrm{A}^{-1}0 = 0 \tag{38}$$

Any matrix that has at least one zero eigenvalue then cannot have an inverse. As far as linear algebra is concerned, there's nothing particularly special about a zero eigenvalues, since if we take the matrix $\mathrm{A} + \alpha\mathrm{I}$ then:

$$\left(\mathrm{A} + \alpha\mathrm{I}\right)v_i = \lambda_i v_i + \alpha v_i = \left(\lambda_i + \alpha\right)v_i \tag{39}$$

In words, if we add α times the identity matrix, we just shift all the eigenvalues by α and the eigenvectors remain unchanged. If we set $\alpha = -\lambda_i$, then the matrix $\mathrm{A} - \lambda_i\mathrm{I}$ will now have a 0 for the i^{th} eigenvalue, and will no longer have an inverse.

We'll note here that the eigenvalues of a diagonal matrix are just the diagonal entries and the eigenvector matrix is just the identity since we can always write:

$$\Lambda = \mathrm{I}\Lambda\mathrm{I}^{-1} \tag{40}$$

for arbitrary diagonal matrix Λ, and Equation 40 is just Equation 23 with $\mathrm{V} = \mathrm{I}$.

4.1. Symmetric Matrices

An important sub-class of matrices are *symmetric* matrices, *i.e.* $A_{ij} = A_{ji}$ or $\mathrm{A} = \mathrm{A}^T$. Symmetric matrices arise over and over in physics, so we'll look at some of their properties here. First, note that

$$\left(\mathrm{A}v_i\right)^T = v_i^T\mathrm{A}^T = v_i^T\mathrm{A} = \lambda_i v_i^T \tag{41}$$

In words, if we have an eigenvector of A, then if we put that eigenvector on the left side of A, it remains an eigenvector. Now let's take two eigenvectors and form $v_j^T\mathrm{A}v_i$. That gives:

$$v_j^T\mathrm{A}v_i = \left(v_j^T\mathrm{A}\right)v_i = \lambda_j v_j^T v_i \tag{42}$$

$$v_j^T\mathrm{A}v_i = v_j^T\left(\mathrm{A}v_i\right) = \lambda_i v_j^T v_i \tag{43}$$

Since the two expressions have to be equal, we have

$$lambda_i v_j^T v_i = \lambda_j v_j^T v_i \tag{44}$$

$$\left(\lambda_i - \lambda_j\right)v_j^T v_i = 0 \tag{45}$$

If the two eigenvectors have different eigenvalues, then the only way Equation 45 can be true is if $v_j^T v_i = 0$. The eigenvectors must be orthogonal. If the eigenvalues are all distinct, then this has to be true for every pair of eigenvectors, so we have

$$v_j^T v_i = \delta_{ij} \tag{46}$$

That means the matrix of eigenvectors is very special, because

$$\mathrm{V}^T\mathrm{V} = \mathrm{I} \tag{47}$$

or, multiplying on the right by V^{-1},

$$\mathrm{V}^T = \mathrm{V}^{-1} \tag{48}$$

If our matrix is symmetric, then the inverse of the eigenvector matrix is just the transpose of that matrix. We call such matrices *orthogonal*. It also turns out that if we have repeated eigenvalues, we can always mix and match the eigenvectors to once again make V orthogonal. How we decide to do that is not unique, but we always can, and computational routines always will return an orthogonal matrix for V, even if there are repeated eigenvalues. It is also true that the eigenvalues of a symmetric matrix are real, but we'll hold off on that proof for now.

One important fact about orthogonal matrices that we'll point out here is that they don't change the *length* of a vector[6]. If $x' = \mathrm{V}^T x$, We have:

$$x'^T x = (\mathrm{V}x)^T \mathrm{V}x = x^T \mathrm{V}^T \mathrm{V}x = x^t x \tag{49}$$

In fact, the inner product of *any* pair of vectors isn't changed if you multiply both of them by the same orthogonal matrix. Since the dot product of two vectors is the product of the length of the vectors times the cosine of the angle between them, that means the angle between any pair of vectors isn't changed if we multiply by an orthogonal matrix. If I take an arbitrary set of vectors and multiply them all by an orthogonal matrix, then their *relative* geometry hasn't changed at all. You can think of a rotation into a new reference frame as a multiplication by an orthogonal matrix, and in fact the terms rotation matrix and orthogonal matrix are often used interchangeably.

To hammer this point home, we'll use an example that I hope in intuitive to Montrealers. Let's say you're hanging out at the Roddick gates and get a case of the late-night munchies, so you decide to head to Schwarz's. How do you get there? Well, you go along Sherbrooke for 1.1km until you hit Saint Laurent, then you turn left and walk another 750m, at which point you can get your greasy fix. What direction did you head on Sherbrooke? Well, if you ask a local, they'd say you went East (obviously), and then north on Saint-Laurent. Someone from out of town would be very confused by this, though, since if you look at a map, we actually went north-northeast on Sherbrooke, and west-northwest on Saint Laurent. We won't agree on what we call the direction we headed, but we'll always agree on how far we went and what directions we turned. If you ever get swamped by all the basis changes we'll do in quantum mechanics, just come back to this - when we change bases, we're just multiplying by an orthogonal matrix, which is nothing more complicated than switching between island north and true north. Some calculations might be *easier* in one basis or another, but the physics (the directions to Schwarz's) won't change.

The transpose/inverse of an orthogonal matrix is also orthogonal since $\mathrm{V}\mathrm{V}^T = \mathrm{V}^T\mathrm{V} = \mathrm{I}$. The product of two orthogonal matrices (say V and Q) must also be orthogonal: $\mathrm{V}\mathrm{Q}^T\mathrm{V}\mathrm{Q} = \mathrm{Q}^T\mathrm{V}^T\mathrm{V}\mathrm{Q} = \mathrm{Q}^T\mathrm{I}\mathrm{Q} = \mathrm{Q}^T\mathrm{Q} = \mathrm{I}$. Intuitively I hope this becomes obvious, because rotating from one basis to another and then rotating from that basis to a third is equivalent to rotating directly from the first basis to the third.

---

[6] A fun side note is that Parseval's theorem/Plancherel theorem, which state that Fourier transforms preserve the power in a function, are just special cases of this since it turns out you can write a Fourier transform operator as an orthogonal matrix.

5. Determinants and Singular Value Decomposition

Let's think about matrices in geometric terms. Let's say we write down a bunch of points that define an $n-$dimensional sphere. In two dimensions, we do this by evaluating $[\cos(\theta), \sin(\theta)]$ for a bunch narrowly spaced values of θ between 0 and 2π. We can multiply each of those vectors by our matrix to end up with a new set of vectors that will trace out a new shape. If our matrix is symmetric, the new shape will be an ellipse with principal axes along the eigenvectors, and semi-minor/semi-major axis lengths given by the eigenvalues. The volume of the new ellipse, relative to the $n-$sphere we started with by the product of the eigenvalues, since each axis gets changed by its eigenvalue. We call the product of eigenvalues the *determinant*, and you can think about it as the volume of a matrix. If the determinant of a matrix is zero, then at least one of the eigenvalues is zero. The determinant of a diagonal matrix is just the product of the diagonal elements.

When you think about determinants in this geometric way, it becomes obvious (although not rigorously proved) that the determinant of the product of two matrices is the product of their determinants. Similarly, the order in which you multiply doesn't matter for determinants:

$$|\mathrm{AB}| = |\mathrm{BA}| = |\mathrm{B}||\mathrm{A}| \tag{50}$$

where $||$ denotes the determinant of a matrix. If we have an inverse, the determinant of the inverse must be one over the determinant of the original matrix since:

$$1 = |\mathrm{A}\mathrm{A}^{-1}| = |\mathrm{A}||\mathrm{A}^{-1}| \tag{51}$$

A consequence of this is that the determinant of even a non-symmetric matrix is still just the product of its eigenvalues since

$$|\mathrm{A}| = |\mathrm{V}\Lambda\mathrm{V}^{-1}| = |\mathrm{V}||\Lambda||\mathrm{V}^{-1}| = ||\mathrm{V}||\Lambda|/|\mathrm{V}| = |\Lambda| \tag{52}$$

Since an orthogonal matrix doesn't change the length of a vector, the magnitude of its determinant must be 1. More formally, since taking the transpose of a matrix can't change its determinant because the eigenvalues don't change

$$|\mathrm{V}\mathrm{V}^{-1}| = |\mathrm{V}\mathrm{V}^{T}| = |\mathrm{V}||\mathrm{V}| = |\mathrm{V}|^2 = |\mathrm{I}| = 1 \tag{53}$$

If V is real, then the only allowed values are ± 1.

An alternative factorization of a matrix is the *singular value decomposition* or SVD. For a general matrix, we can write down

$$\mathrm{A} = \mathrm{U}\Sigma\mathrm{V}^{T} \tag{54}$$

where $\mathrm{U}^{T}\mathrm{U} = \mathrm{I}$ and $\mathrm{V}^{T}\mathrm{V} = \mathrm{I}$, and Σ is diagonal. If A is symmetric, then U=V and the SVD reduces to the eigendecomposition. If A is square but not symmetric, then U and V are square and orthogonal, but not equal to each other. If A is rectangular, then one of U or V is square, the other is rectangular, and Σ matches the smaller dimension of A. If A is square, then up to a possible sign flip the determinant of A is also the product of the elements of Σ, known as the singular values. If A is rectangular, then the determinant doesn't exist, but we can still get an idea of how it is stretching out vectors by looking at the singular values.

I've brought up SVD here because it gives us another way of looking at non-symmetric square matrices that I find useful. I can always insert the identity matrix in an expression so:

$$\mathrm{A} = \mathrm{USV}^T = \left(\mathrm{UV}^T\right)\left(\mathrm{VSV}^T\right) \tag{55}$$

The second term in Equation 55 is a symmetric matrix and the first term is an orthogonal matrix (since the product of two orthogonal matrices is always an orthogonal matrix). Since I mostly deal with symmetric matrices, I find this description useful (or at least comforting). Multiplying by a non-symmetric matrix is just multiplying by a symmetric matrix, with a bonus rotation added on top. Since we can always take the SVD, we can always think of a matrix this way even if it's defective and doesn't have an eigendecomposition.

Before moving on, I would be remiss in my duties if I didn't point out some extremely important practical notes on using SVD. Some languages (looking at you, python) define the SVD as USV, without the tranpose on V. If you're going to use the SVD, *always check*. If you are taking the SVD of a rectangular matrix, then the default behavior is often to return square versions of U and V (once again side-eying python, but in this case the original sin lies I believe with MATLAB). There will be a flag to say "don't be stupid" and return the rectangular versions. For numpy, you add an extra 0 as an input argument, *i.e.*

```
u,s,v=numpy.linalg.svd(A,0)
```

I have never once not wanted the 0, so I suggest you add it by default. It can be the difference between nearly instant results and crashing your computer.

6. Commuting Matrices

In general, matrices do not commute. That is, we expect that:

$$\mathrm{AB} \neq \mathrm{BA} \tag{56}$$

You can see this just by computing the $[0,0]$ element of AB and BA:

$$(\mathrm{AB})[0,0] = \sum A_{0,k}B_{k,0} \tag{57}$$

$$(\mathrm{BA})[0,0] = \sum B_{0,k}A_{k,0} \tag{58}$$

There's no reason to think the first row of A times the first column of B is equal to the first column of A times the first row of B, except under special circumstances.

If AB does equal BA, then the pair of matrices is said to *commute*. One important case is that of diagonal matrices: all pairs of diagonal matrices commute. That's pretty easy to see since

$$(\mathrm{AB})_{ii} = (\mathrm{BA})_{ii} = A_{ii}B_{ii} \tag{59}$$

and of course scalars commute. The identity matrix commutes with all matrices, since

$$\mathrm{AI} = \mathrm{IA} = \mathrm{A} \tag{60}$$

It is *not* true in general that a diagonal matrix commutes with a matrix, since a diagonal B rescales the columns of A in AB, but rescales the rows in BA.

Good life advice when you're trying to show something in linear algebra is to write down the eigenvalues and eigenvectors. In that case we have:

$$\mathrm{AB} = \mathrm{V_A}\Lambda_\mathrm{A}\mathrm{V_A^{-1}}\mathrm{V_B}\Lambda_\mathrm{B}\mathrm{V_B^{-1}} \tag{61}$$

If the eigenvectors of the two matrices are the same $\mathrm{V} \equiv \mathrm{V_A} = \mathrm{V_B}$, then Equation 61 turns into

$$\mathrm{AB} = \mathrm{V}\Lambda_\mathrm{A}\mathrm{V^{-1}}\mathrm{V}\Lambda_\mathrm{B}\mathrm{V^{-1}} = \mathrm{V}\Lambda_\mathrm{A}\Lambda_\mathrm{B}\mathrm{V^{-1}} = \mathrm{V}\Lambda_\mathrm{B}\Lambda_\mathrm{A}\mathrm{V^{-1}} = \mathrm{V}\Lambda_\mathrm{B}\mathrm{V^{-1}}\mathrm{V}\Lambda_\mathrm{A}\mathrm{V^{-1}} = \mathrm{BA} \tag{62}$$

If two matrices share the same *eigenvectors*, they commute. They do not need to share the same eigenvalues. In the case where all the eigenvalues are distinct, this is both necessary and sufficient. When there are repeated eigenvalues, we have freedom to rotate the eigenvectors that span the repeated eigenvalues. As an example, the eigenvalues of the identity matrix are all one, so any vector is an eigenvector of I. That explains why I commutes with everything - when every vector is an eigenvector, then you're free to use the eigenvectors from any matrix. Since I could have picked the eigenvectors to be the same, then I commutes with every matrix. More generally, if there exists a matrix P that diagonalizes both A and B (*i.e.* $\mathrm{A} = \mathrm{PD_AP^{-1}}$ for diagonal D and similarly for B) then A and B commute. Note that it is not true that if A and B commute and B and C commute that A and C commute. Put I in for B, then that has to commute with any A and C, but of course, arbitrary A and C would not commute. If the eigenvalues are all distinct, though then A commuting with B means A and B have the same eigenvectors, and B commuting with C means they have the same eigenvectors, so A and C share eigenvectors and hence commute.

We can qualitatively think of the commutator of A and B, defined to be $\mathrm{AB} - \mathrm{BA}$, to be a measure of how misaligned the principal axes of A and B are. The closer the eigenvectors, the smaller the commutator. The *eigenvalues* matter a lot less for commuting.

7. Complex Matrices

There's no way around complex matrices in quantum mechanics. You can't get interference without complex arithmetic, so now we'll go back over some of what we've discussed for real matrices through a complex lens. When moving to complex matrices, what should the inner product look like? For real values, the squared length of a vector x is just the dot product, or $x^T x$. For a complex-valued vector, we want the inner product to return $\sum |x_i|^2 = \sum (x_i)^* x_i$. To get this, we need the inner product of x with itself to be not be the transpose $x^T x$ but the *conjugate-transpsose* $x^{*T} x$. This conjugate-transpose operation is so common, especially in quantum, that we use a dagger as shorthand, *i.e.* $x^{*T} x = x^\dagger x$. We need to be a little more careful with the inner product of two distinct vectors because $x^\dagger y = (y^\dagger x)^*$, so swapping the order changes the output to its complex conjugate.

Now that we know what we want the inner product to do, we can work out what a rotation matrix looks like. We'll need to use the fact that $(\mathrm{AB})^\dagger = \mathrm{B^\dagger A^\dagger}$ (just like transpose). If we let $x' = \mathrm{V}x$, then we want

$$x^\dagger x = x'^\dagger x' = (\mathrm{V}x)^\dagger (\mathrm{V}x) = x^\dagger \mathrm{V^\dagger V} x \tag{63}$$

For this to work, we need $\mathrm{V}^\dagger\mathrm{V} = \mathrm{I}$. If we have a complex matrix, it is orthogonal if its inverse is its conjugate-transpose, not just its transpose. We call such matrices *unitary*. All of their eigenvalues must have a magnitude of one, since if we put in an eigenvector for x, then we would pick up a factor of $\lambda^*\lambda$. We can't say anything about *phases*, though.

Quantum mechanics requires that we get a real value when taking the expectation of an operator. Mathematically that will correspond to $x^\dagger \mathrm{A} x = c$ where c is strictly real for *any* complex x. Let's look and see what that implies about our matrix. We can take the complex transpose to get:

$$\left(x^\dagger \mathrm{A} x\right)^\dagger = x^\dagger \mathrm{A}^\dagger x = c \tag{64}$$

Since this has to be true for *any* x, that means that $\mathrm{A} = \mathrm{A}^\dagger$[7] . We call a matrix that is equal to its own conjugate transpose *Hermitian*, and every operator in quantum mechanics will have to be Hermitian. Not surprisingly, the fact that a matrix is Hermitian implies it has real eigenvalues since

$$v^\dagger \mathrm{A} v = v^\dagger \lambda v = \lambda \tag{65}$$

Since $x^\dagger \mathrm{A} x$ is real for any x, it also has to be true for an eigenvectors, hence λ must be real.

Now that we know we are going to work with Hermitian matrices with real eigenvalues, let's see what we can say about the eigenvectors. We have

$$\mathrm{A}^\dagger = \mathrm{A} = \mathrm{V}\Lambda\mathrm{V}^{-1} \tag{66}$$

$$\mathrm{A}^\dagger = \left(\mathrm{V}\Lambda\mathrm{V}^{-1}\right)^\dagger = \mathrm{V}^{\dagger -1}\Lambda\mathrm{V}^\dagger \tag{67}$$

For both equations to be true, it had better be the case that the matrices on the right agree, so we have

$$\mathrm{V}^\dagger = \mathrm{V}^{-1} \tag{68}$$

This also implies that $\mathrm{V}^{\dagger -1} = \mathrm{V}$, which you can see by just taking the inverse of Equation 68. Since the inverse of V is its conjugate-transpose, then the matrix of eigenvectors is unitary and the inner product of any distinct pair of eigenvectors is zero. The *magnitude* of the eigenvalues must be one, but they are once again allowed to have arbitrary *phase*.

We've now shown that if $x^\dagger \mathrm{A} x$ is real for any vector x, that A must be Hermitian, that it has real eigenvalues, and that its eigenvectors are orthogonal. Once we switch to complex values, a symmetric matrix corresponds to a Hermitian matrix, not a complex symmetric one. With this background, we can now convert bras/kets *etc.* into linear algebra notation.

---

[7] In case you aren't convinced, consider a vector that has a single one and is zero everywhere else. $x^\dagger \mathrm{A} x$ just picks out the corresponding diagonal entry of A, so the diagonal entries must all be real. Now consider a vector with two ones, say in the i^{th} and j^{th} entries. Then $x^\dagger \mathrm{A} x$ gives us $A_{ii} + A_{ij} + A_{ji} + A_{jj}$. Since the diagonal entries are real, then $A_{ij} + A_{ji}$ must also be real, which is only true if the complex parts of A_{ij} and A_{ji} are equal and opposite. A similar exercise shows that the real part of A must be symmetric, so $A_{ij} = A^*_{ji}$ for every i, j.

8. Dirac Notation in Linear Algebra

Let's say we have a particle in a state Ψ in quantum mechanics. If we want to write down that state, we have to pick a basis to write it down in. The values in the state are then the amplitudes to find the particle in a given eigenstate in our basis, if we were to carry out a measurement. In the case of Stern-Gerlach, if we were to write down a state in the $|+z\rangle, |-z\rangle$ basis, then the $|+z\rangle$ component would be the amplitude to find our particle in the $|+z\rangle$ state, etc. By convention, we treat kets as column vectors. The length of every state has to be one, since when we make a measurement, we have to get *something*. When we have a particle in a given state, we know that both the amplitude and the probability of finding it in that state must be 1. In Dirac notation, that means that

$$\langle\Psi|\Psi\rangle = 1 \tag{69}$$

. In linear algebra, if we want to say that the length of a vector Ψ is 1, we would mean that $\Psi^\dagger\Psi = 1$. Since we decided that a ket is a column vector, the only way this works is if the corresponding bra in Equation 69 is the conjugate-transpose of our vector.

We know the expectation of a function is

$$\langle f(x)\rangle = \sum_i f(x_i)P(x_i) \tag{70}$$

where x can only take on discrete values x_i, and $P(x_i)$ is the probability that $x = x_i$[8]. If we have an observable (say angular momentum along an axis, energy...) that takes on discrete values, then we can write an arbitrary wave function Φ as a superposition of the states Ψ_i that correspond to the possible values of the observable. For simplicity let's assume energy, and we'll label the energy of Ψ_i to be E_i. That gives:

$$\Phi = \sum_i c_i |\Psi_i\rangle \tag{71}$$

and expected energy E_ϕ:

$$\langle E_\Phi\rangle = \sum_i E_i c_i^* c_i \tag{72}$$

since the probability to be in state Ψ_i is $c_i^* c_i$. We can write this as:

$$\langle E_\Phi\rangle = \begin{bmatrix} c_1^* & c_2^* & c_3^* & ... \end{bmatrix} \begin{bmatrix} E_1 & 0 & 0... \\ 0 & E_2 & 0... \\ 0 & 0 & E_3... \\ . & & \\ . & & \\ . & & \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ . \\ . \\ . \end{bmatrix} \tag{73}$$

If you carry out these multiplications, you'll see that Equation 73 is identically equal to our original expresion Equation refeqn:expected. So, if I've written down my state in the energy basis, I can find the expected energy with the matrix product

$$\Phi^\dagger \hat{E} \Phi \tag{74}$$

[8]This turns into an integral in the continuous limit, but then we're dealing with finite-sized matrices here so we'll treat the sum as discrete.

where Φ is the vector of $\begin{bmatrix} c_1^* & c_2^* & c_3^* & ... \end{bmatrix}$ and $\hat{E}$ is the diagonal matrix of the energy values. If I had written down Φ in a *different* basis, where the energy states were described as a set of vectors V_i, then I could rotate Φ into the energy basis by multiplying by the bras of those vectors, which is equivalent to multiplying by the matrix $V^\dagger$. That means my expected energy is now

$$\langle E \rangle = (V^\dagger \Phi))))^\dagger \hat{E} V^\dagger \phi \tag{75}$$

$$= \Phi^\dagger V \hat{E} V^\dagger \Phi \tag{76}$$

Equation **??** tells us that if we have an observable with real values, we can find the expected value of that operator by sandwiching a Hermitian matrix with the wave function. The eigenvectors are the representation of the pure states of our observable and the eigenvalues or the corresponding values.

Supplementary Notes for lectures 6 (and possibly more) for McGill Phys 357.

1. Exponentials as Limits

We're going to show that large powers of $1+\delta$ turn into exponentials, but perhaps via a somewhat roundabout path. First, let's go back to the derivative of an exponential with some base a, a^x. Going back to the fundamental definition of a derivative, we have:

$$\frac{d(a^x)}{dx} = \lim_{dx\to 0} \frac{a^{x+dx} - a^x}{dx} = a^x \lim_{dx\to 0} \frac{a^{dx} - 1}{dx} \tag{1}$$

By factoring out the a^x, we're left with a number that is a function of a in the rightmost term. If that limit equals one for some value of a, which we'll call e, then we have a function that is equal to its own derivative. For that to be true, we need:

$$\lim_{dx\to 0} \frac{e^{dx} - 1}{dx} = 1 \tag{2}$$

$$\lim_{dx\to 0} e^{dx} = 1 + dx \tag{3}$$

$$e = \lim_{dx\to 0} (1+dx)^{1/dx} \tag{4}$$

The more usual way of writing this is taking $n \equiv \frac{1}{dx}$, in which case we have:

$$e = \lim_{n\to\infty} (1 + \frac{1}{n})^n \tag{5}$$

We can of course raise this number to some power x:

$$e^x = \left(\lim_{n\to\infty} (1+\frac{1}{n})^n\right)^x = \lim_{n\to\infty} \left(1+\frac{1}{n}\right)^{nx} \tag{6}$$

We can clean this up by making the substitution[1] $nx \to n$, leaving us with

$$e^x = \lim_{n\to\infty} \left(1+\frac{x}{n}\right)^n \tag{7}$$

From the book, we have that a rotation $\hat{R}(d\phi)$ by a small angle $d\phi$ about the $z-$axis is

$$\hat{R}(d\phi\mathbf{k}) = 1 - \frac{i}{\hbar}\hat{J}_z d\phi \tag{8}$$

(Townsend eqn. 2.29). When we write a macroscopic rotation by angle ϕ as a sequence of N rotations by angle $d\phi = \phi/N$, then we can write the infinitessimal rotation

$$\hat{R}(\phi/N\mathbf{k}) = 1 - \frac{i}{\hbar}\phi/N \tag{9}$$

[1]Yes, I should really say $n' = nx$, write out in terms of n' and then drop the primes to leave the expression we get, but that's a lot of steps to accomplish the same thing.

Since we can get our macroscopic rotation by taking a sequence of N small rotations as N goes to infinity, we have

$$\hat{R}(\phi\mathbf{k}) = \lim_{N\to\infty}\left(1 - \frac{i}{\hbar}\hat{J}_z\phi/N\right)^N = e^{-\frac{i}{\hbar}\hat{J}_z\phi} \tag{10}$$

where we invoke Equation 7 to carry out the final step. I find this a much more natural way of seeing why the limit turns into an exponential. Yes, Townsend's approach of expanding the power using the binomial theorem, taking the terms as the number of steps goes to infinity and matching them with the Taylor series for e^x will *work*, but it doesn't really explain *why*. Forms like Equation7 show up pretty frequently in physics, so it's worth having that expression in the back of your mind for the day you need it.

2. Simple Rotation Derivation

The book goes through a relatively abstract derivation of rotation operators, but it may also be useful to see a more direct derivation in a particular coordinate system. We'll start with the observation that a 90-degree rotation about the $z-$axis rotates $|+x\rangle$ into $|+y\rangle$, and $|+y\rangle$ into $|-x\rangle$. If we're working in the $|\pm z\rangle$ basis, we can write the states as:

$$|+x\rangle = \frac{1}{\sqrt{2}}[1,1]^T \tag{11}$$

$$|+y\rangle = \frac{1}{\sqrt{2}}[1,i]^T \tag{12}$$

$$|-x\rangle = \frac{1}{\sqrt{2}}[1,-1]^T \tag{13}$$

$$|-y\rangle = \frac{1}{\sqrt{2}}[1,-i]^T \tag{14}$$

We know the same rotation matrix has to take $|+x\rangle$ to $|+y\rangle$, so we can write

$$\begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{i}{\sqrt{2}}\end{bmatrix} = R\begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix} \tag{15}$$

Similarly, since we also have to take $|+y\rangle$ to $|-x\rangle$, we have:

$$\begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{-1}{\sqrt{2}}\end{bmatrix} = R\begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{i}{\sqrt{2}}\end{bmatrix} \tag{16}$$

The laws of linear algebra let us merge these equations to get:

$$\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ i & -1\end{bmatrix} = R\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & i\end{bmatrix} \tag{17}$$

We multiply on the right by the inverse of the 2x2 matrix to get:

$$R(\pi/2) = \begin{bmatrix}1 & 0\\ 0 & i\end{bmatrix} \tag{18}$$

We can write this in a slightly more useful way by expressing terms as complex exponentials:

$$R(\pi/2) = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\pi/2} \end{bmatrix} \tag{19}$$

We've derived this by ensuring that $|+x\rangle$ goes into $|+y\rangle$, and $|+y\rangle$ goes into $|-x\rangle$, but it's a good exercise to check that this same matrix also takes $|-x\rangle$ into $|-y\rangle$, and $|-y\rangle$ into $|+x\rangle$. If that doesn't work, we've made a mistake somewhere.

We're always free to rotate states by an overall phase. If we want to make the $|+z\rangle$ and $|-z\rangle$ states more symmetric, we can add in an extra phase of $-\pi/4$, which gives us:

$$R(\pi/2) = \begin{bmatrix} e^{-i\pi/4} & 0 \\ 0 & e^{i\pi/4} \end{bmatrix} \tag{20}$$

We can generalize this to

$$R(\theta) = \begin{bmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{bmatrix} \tag{21}$$

To see this, picture splitting our $\pi/2$ rotation into N smaller rotations. Each of those rotations must be the same, so the product of N of the tiny rotations must equal Equation 20. Since the rotation matrices are diagonal, the product of N matrices is just a diagonal matrix with entries raised to the N^{th} power. If we let $N\theta = \pi/2$ then the only way to get Equation 20 for an arbitrary value of N is if the rotation by θ is Equation 21.

Once we have the rotation matrix in the form of Equation 21, we can see why the generator of rotations is written the way it is. Since we're working in the $|\pm z\rangle$ basis, we know that the angular momentum of a $|+z\rangle$ state must be $\frac{\hbar}{2}$, and the angular momentum of a $|-z\rangle$ state must be $-\frac{\hbar}{2}$. If we have an operator that gives us the angular momentum along the $z-$axis, we can produce Equation 21 just by cancelling the $\hbar$. That gives:

$$R_z(\theta) = \begin{bmatrix} e^{-i\hat{J}_z/\hbar} & 0 \\ 0 & e^{-i\hat{J}_z/\hbar} \end{bmatrix} \tag{22}$$

3. Hermitianness and Rotation Operators

We've seemingly randomly introduced a factor of $\frac{i}{\hbar}$ in our definition of rotations. Of course, it's not random, but you only know that after you've arrived at your destination. What we can say robustly, though, is that once you have a rotation matrix of the form of Equation 10, we can say something important about $\hat{J}_z$. Since our operations obey the rules of linear algebra, an operator has to behave the same way when operating to the right on a ket/column vector as when it operates to the left on a bra/row vector. To conserve probability, we absolutely need that

$$\langle\Psi|\, \hat{R}^\dagger(\phi)\hat{R}(\phi)\, |\Psi\rangle = 1 \tag{23}$$

since the state $\hat{R}(\phi)\,|\Psi\rangle$ has to be normalized. That means

$$\hat{R}^\dagger(\phi)\hat{R}(\phi) = \mathrm{I} \tag{24}$$

We know that if we rotate through an angle ϕ about an axis, to undo the rotation we must rotate through an angle of $-\phi$ about the same axis. That means that our $\hat{R}^\dagger(\phi)$ must be equal to $\hat{R}(-\phi)$. We can plug that into Equation 10 where $\phi \to -\phi$, so

$$\hat{R}^\dagger(\phi) = e^{\frac{i}{\hbar}\hat{J}_z\phi} \tag{25}$$

Of course, we know that we need to take the complex conjugate when going from a ket to a bra, so we also know that

$$\hat{R}^\dagger(\phi) = e^{\frac{i}{\hbar}\hat{J}_z^\dagger\phi} \tag{26}$$

The only way that Equations 25 and 26 can be true is if

$$\hat{J}_z = \hat{J}_z^\dagger \tag{27}$$

In other words, the operator $\hat{J}_z$ is its own conjugate-transpose. We'll see that this is a general property of any physical observable. Any operator that gives us a physical observable *must* equal its own conjugate-transpose when written as a matrix. We call such operators/matrices *Hermitian*. In this case, $\hat{J}_z$ gives us the angular momentum about the $z-$axis. The rotation operator does not produce a physical quantity, and so it does not have to be Hermitian. The entire point of the rather odd way we wrote the generator of rotations was to express it in terms of a Hermitian operator.

Supplementary Notes for lecture 9 for McGill Phys 357.

1. Orthogonal Eigenvectors

We know that operators that produce an observable must be Hermitian (equal to their own conjugate-transpose). The eigenvalues must be real since for eigenvector v

$$v^\dagger \mathrm{A} v = v^\dagger \lambda v = \lambda \tag{1}$$

but if I take the conjugate transpose, I still get λ since $\mathrm{A}^\dagger$=A. Before moving to the general case, take the case of a symmetric matrix, which is just a Hermitian matrix with no imaginary part. If we have eigenvectors v_1 and v_2 with eigenvalues λ_1 and λ_2, then starting with the symmetmry of A, we know

$$v_1^T \mathrm{A} v_2 = v_2^T \mathrm{A} v_1 \tag{2}$$

$$v_1^T \lambda_2 v_2 = v_2^T \lambda_1 v_1 \tag{3}$$

$$(\lambda_2 - \lambda_1)(v_2^T v_1) = 0 \tag{4}$$

where we have used that the transpose of a number is itself, so $v_2^T v_1 = v_1^T v_2$. The only way this is true if the eigenvalues are different is if $v_2^T v_1 = 0$, so the eigenvectors of a symmetric matrix are orthogonal.

We need to be slightly (but only slightly) more careful. Keeping track of daggers, and using the fact that the eigenvalues are real, we end up with the very similar equation

$$\lambda_2 v_1^\dagger v_2 = \lambda_1 v_2^\dagger v_1 = \lambda_1 (v_1^\dagger v_2)^\dagger \tag{5}$$

The magnitude of $v_1^\dagger v_2$ doesn't change when you take its dagger, so the only way Equation refeqn:evecdag can hold when $\lambda_1 \neq \lambda_2$ is if $v_1^\dagger v_2 = 0$. Therefore all the eigenvectors of a Hermitian matrix must be orthogonal, just like the eigenvectors of a real symmetric matrix.

2. Expectation and Variance

Now we'll try to figure out the uncertainty on a measurement if our state is an eigenvector of an observable operator. Recall that

$$Var(x) = \left\langle x^2 \right\rangle - \langle x \rangle^2 \tag{6}$$

In quantum mechanics, if x is the observable we get by taking an operator $\hat{A}$ operating on wave function Ψ, then we have

$$\langle x \rangle \, \Psi^\dagger \hat{A} \Psi \tag{7}$$

We can always feed the output of an operator into another operator [1]. Since we want to get $\langle x^2 \rangle$, we can get that by operating by $\hat{A}$ twice:

$$\langle x^2 \rangle \, \Psi^\dagger \hat{A} \hat{A} \Psi \tag{8}$$

We can see this makes sense in the context of our familiar J_Z operator:

$$J_z = \begin{bmatrix} \hbar/2 & 0 \\ 0 & -\hbar/2 \end{bmatrix} \tag{9}$$

$$J_z^2 = J_z J_z = \begin{bmatrix} \hbar^2/4 & 0 \\ 0 & \hbar^2/4 \end{bmatrix} \tag{10}$$

$$\tag{11}$$

Since the operator for J_z^2 is just $\hbar/4$ times the identity matrix, we know the square of the angular momentum of any state will give us $\hbar^2/4$. Every measurement gives us either $\hbar/2$ or $-\hbar/2$, so the square of every measurement will give us $\hbar^2/4$, just like the J_z^2 operator tells us.

Now what happens if Ψ is an eigenvector of $\hat{A}$ with eigenvalue λ? We need to make

$$\langle \Psi | \, \hat{A} \, | \Psi \rangle = \langle \Psi | \Psi \rangle \, \lambda = \lambda \tag{12}$$

$$\langle \Psi | \, \hat{A} \hat{A} \, | \Psi \rangle = \lambda \, \langle \Psi | \, \hat{A} \, | \Psi \rangle = \lambda^2 \tag{13}$$

We have $\langle x \rangle = \lambda$ and $\langle x^2 \rangle = \lambda^2$, so we know that the uncertainty is $\langle x^2 \rangle - \langle x \rangle^2 = \lambda^2 - \lambda^2 = 0$. If we have a state that is an eigenvector of $\hat{A}$, then when we measure the value of the corresponding observable, the mean value is the eigenvalue, and the uncertainty is zero. Because of the zero uncertainty, we don't just get λ on average, we always get exactly λ.

In summary, if we have an observable operator $\hat{A}$ then the eigenvectors of $\hat{A}$ are *pure* states. Whenever we measure the value associated with $\hat{A}$, a pure state always returns its corresponding eigenvalue. Since the eigenvectors are orthogonal, we can always describe an arbitrary state as a combination of eigenvectors. That means when we make a measurement, the only possible results are the eigenvalues of $\hat{A}$. When we make a measurement of the observable, we collapse the wave function into the pure state corresponding to the value we measured.

3. Hermitian to Unitary

We'll often see objects of the form $exp(\hat{A})$ where $\hat{A}$ is a Hermitian operator. There are lots of ways to derive this, but let's just accept that by $\exp(\hat{A})$ we mean:

$$\exp(\hat{A}) = I + \hat{A} + \hat{A}^2/2! + \hat{A}^3/3! + ... \tag{14}$$

[1] After all, they are both just matrix multiplies. Townsend goes through a lot of effort to show explicitly that the result of a pair of operators is the product of their matrix representations, but once you know an operator is a matrix multiply, you do actually know the result of a sequence of operators has to be a sequence of matrix multiplies.

where $\hat{A}^2 = \hat{A}\hat{A}$. If we decompose $\hat{A}$ into eigenvalues and eigenvectors, then

$$\hat{A}^2 = \mathrm{V}\Lambda\mathrm{V}^{-1}\mathrm{V}\Lambda\mathrm{V}^{-1} = \mathrm{V}\Lambda^2\mathrm{V}^{-1} \tag{15}$$

Carrying out the sum, we're left with

$$\exp(\hat{A}) = \mathrm{V}\left(I + \Lambda + \Lambda^2/2! + \Lambda^3/3!\right)\mathrm{V}^{-1} = \mathrm{V}\exp(\Lambda)\mathrm{V}^{-1} \tag{16}$$

If I take $\hat{A}$ to be Hermitian, then

$$\exp(i\hat{A}) = \mathrm{V}\exp(i\Lambda)\mathrm{V}^{\dagger} \tag{17}$$

Since the eigenvalues of a Hermitian matrix are real, then the magnitude of the eigenvalues of $\hat{A}$ is one (this would *not* be true if there were an imaginary component to Λ), and we see that $\exp(i\hat{A})$ must be unitary. We can map every Hermitian operator to a unitary rotation matrix by taking $\exp(i\hat{A})$.

Supplementary Notes for lecture 12 for McGill Phys 357.

1. Raising/Lowering Operators and Defective Matrices

Let's say we have a multi-state system, and we've ordered the states by some quantum number m of an operator. For spin-1/2 systems, the allowed values of m would be $(-1/2, 1/2)$, for spin-1 they would be $(-1, 0, 1)$ and so on. If the basis we're working in are the eigenvectors of the operator, with the highest values of m first, then the raising/lowering operators are trivial to write down. Up to constants, the raising has to be:

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

If you multiply by a vector on the right, then the first value (corresponding to the highest value of m) goes away, and all the remaining values move up by one spot, with a zero in the final value. If we put the ones along the diagonal, we'd get the same thing out we started with, but because we shifted them up by one spot in the matrix, we end up shifting up the values in our vector by one. In math:

$$\mathrm{A}_+ = (a, b, c, d, e) \rightarrow (b, c, d, e, 0)$$

If we're already at the maximum m, then we can't raise that state to a higher value, so it goes away. Similarly, there is no state below the smallest value of m, so when we raise our state, we need to get a zero for the smallest m. If you start with a state $(0, 0, 1, 0, 0)$ say, then we get out $(0, 1, 0, 0, 0)$, and we can see that if we apply the raising operator to our starting pure state, the output is a pure state with the next higher m. The lowering operator is the same, but now we want to shift the ones down by one from the diagonal. That leaves us with

$$\mathrm{A}_- = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

and (a, b, c, d, e) goes to $(0, a, b, c, d)$. Depending on how the raising/lowering operators are defined, there may be multiplicative constants (*e.g.* the usual definitions of the angular momentum raising/lowering operators include factors of $\hbar$ and numerical constants). The basic idea, that you have an operator with numbers one above or one below the diagonal, does not change.

1.1. Defective Matrices and SVD

The raising/lowering operators are a bit peculiar from a linear algbera viewpoint. If we have a triangular matrix, the eigenvalues are just the diagonal elements since the characteristic equation

is the product of the diagonal[1]. We can see that the raising/lowering operators are triangular, because everything below/above the diagonal is zero. Of course, the diagonal is also zero, so all of our eigenvalues are zero. You might think "if all the eigenvalues are zero, the matrix A must be zero" because A times any eigenvector gives zero. Of course, we can see that not all vectors go away, so we must have vectors that *aren't* eigenvectors. That means we can't write the raising/lowering operators as

$$\mathrm{A} = \mathrm{V}\Lambda\mathrm{V}^T$$

. We do get the right number of eigenvalues, but the corresponding eigenvectors don't form a set of basis vectors. These matrices are called *defective*, and much of our intuition about linear algebra doesn't apply.

In this sort of situation, the singular value decomposition (SVD) of a matrix can be very useful. SVD factors a matrix A like this:

$$\mathrm{A} = \mathrm{USV}^{T?}$$

where U and V are orthogonal (if A is square), and S is diagonal. If A is complex, then U and V are unitary, while S is still real[2]. The values of S are the singular values, and behave similar to eigenvalues. We note that the SVD reduces to eigendecomposition for square symmetric/Hermitian matrices. The question mark on the transpose reflects that some packages take the transpose of V, while others don't. You should check before blindly using (or not) the tranpose. Python does not use the transpose, so we have $\mathrm{A} = \mathrm{USV}$. Happily, the SVD exists for all matrices, even when the eigendecomposition fails. When we call SVD on a raising/lowering operator, something interesting happens. We find there are $n-1$ singular values of 1 and, one singular value with a value of 0. The columns of U are just the columns of V shifted up by one (unless you're using python, in which case it's the rows of V).

What this is telling us is that if we multiply a vector by the SVD, V^T picks out the amplitude of the state for each of the pure states and moves them to the next state, except for the state with the zero singular value. For the simple case of working in your own basis, this is overkill. However, if someone gave you the matrix for the raising operator for a different operator, you could figure out the pure states of that operator just by taking the SVD of the raising operator. We can go even further, and order the pure states, because we know that each time we apply it, the states move up by one. We can find the state with the largest quantum number, because that corresponds to the zero singular value (this behavior is due to the fact that the raising operator has to kill off the state with the largest quantum number). So then I look for the state that turns into the highest state,

---

[1]If you aren't familiar with this, you can show it using the determinant. One way of calculating a determinant is to take every element in a column (or row), block out the row/column of that element, and take the determinant of the sub-matrix. The total determinant is then the sum over the row of the row times the sub-matrix determinant, with some signs. For a triangular matrix, we only have one non-zero element in the first column, then when we block out that row/column, we're left with a single non-zero entry in the first row/column of the sub-matrix. That process continues until the determinant is the product of the diagonal elements. If we subtract λ from the diagonal, we get a zero whenever λ is equal to an element along the diagonal, so the eigenvalues are just the diagonal entries.

[2]If A is rectangular, then we can zero-pad S to be rectangular, and then U and V are unitary. The more usual thing to do, though, is keep S square and just keep the corresponding columns of U and V. Is say, A has more rows than columns, then V will be unitary and U will have orthogonal columns, but not be fully unitary

which much be the state with the next-highest quantum number. If you keep multiplying by A, then you can rank-order the states by watching how many multiplies of A are needed to zero out a state. And of course, everything works with lowering operators as well, only it's the bottom-most state that disappears first.

In summary, raising/lowering operators are straightforward to write down if you're working in the basis of the operator that is getting raised/lowered. We won't always have that luxury, though, and in that case SVD can tell us a lot about the behavior of the operator that our usual eigenvalue decomposition can't.

2. Some Questions About Matrices

Are real symmetric matrices Hermitian?
Is the product of Hermitian matrices Hermitian?
Is the square of a Hermitian matrix Hermitian?
what can you say about eigenvalues?
If all the eigenvalues of a matrix are zero, is the matrix zero?
If [AB]=0 and [BC]=0, does [AC]=0?
how about if eigenvalues are distinct for (A,B,C)?
Does a matrix commute with itself?
Does a matrix commute with powers of itself?
If A is Hermitian, what about iA?
Express a general real matrix A as the sum of symmetric and antisymmetric matrices
Express a general complex matrix as sum of Hermitian and anti-Hermitian.
If B=-$B^\dagger$, how about iB Hermitianness?
[A,BC]=?

Supplementary Notes for lecture 17 for McGill Phys 357.

1. Unitary Operators and Generators of Translation

Unitary operators play an incredibly important role in quantum mechanics. If I want to turn a state into another state, and keep those states normalized, then the operator must be unitary. That follows because if $|\Phi\rangle = \mathrm{A}\,|\Psi\rangle$, then

$$\langle\Phi|\Phi\rangle = \langle\Psi|\mathrm{A}^\dagger\mathrm{A}|\Psi\rangle = 1 \tag{1}$$

That is only true if $\mathrm{A}^\dagger\mathrm{A} = \mathrm{I}$, or $\mathrm{A}^{-1} = \mathrm{A}^\dagger$. From now on, if a matrix is unitary, we'll call it U.

Right away, we know several properties of a unitary operator. The columns of U must be orthogonal $U[:,i]^\dagger U[:,j] = \delta_{ij}$, because the rows of the inverse must be orthogonal to the columns of the original matrix. Since inverses commute, the rows of A are also orthogonal, because the columns of $\mathrm{U}^\dagger$ have to be orthogonal. The eigenvectors also have to be orthogonal, because if $\mathrm{U} = \mathrm{V}\Lambda\mathrm{V}^{-1}$ then because $\mathrm{U}^{-1} = \mathrm{U}^\dagger$ we know

$$\mathrm{V}\Lambda^{-1}\mathrm{V}^{-1} = \mathrm{V}^{-1\dagger}\Lambda^\dagger\mathrm{V}^\dagger \tag{2}$$

For that to hold generally requires that $\mathrm{V}^{-1} = \mathrm{V}^\dagger$ so the eigenvectors themselves form a unitary matrix. We also know that $\Lambda^{-1} = \Lambda^\dagger$, or $\Lambda^\dagger\Lambda = \mathrm{I}$, that is the absolute value of each eigenvalue is 1.

We can write any complex number down as $c\exp(i\theta)$ where c is the magnitude of the number, and θ is a real angle. Since the magnitude of the eigenvalues is one for a unitary matrix, we can write them as $\exp(i\theta)$ and so the general form for a unitary matrix is

$$\mathrm{U} = \mathrm{V}\exp(i\Theta)\mathrm{V}^\dagger \tag{3}$$

for real-valued diagonal matrix Θ and orthogonal eigenvector V. Because Hermitian matrices have real eigenvalues and orthogonal eigenvectors, there's a unique mapping from every Hermitian matrix A (we aren't calling it H so as to not be confused with the Hamiltonian) to a unitary matrix:

$$\mathrm{U} = \exp(i\mathrm{A}) \tag{4}$$

Similarly, modulo branch cuts, there's a unique mapping from every unitary matrix to a Hermitian matrix.

One final piece really locks in the form of evolution operators. Let's say I want to evolve a wave function that is a function of some value x. We then want to be able to go from $|\Psi(x)\rangle$ to $|\Psi(x+dx)\rangle$. If we zoom in far enough, so nothing else in the system is changing, then whatever operator U takes $|\Psi(x)\rangle$ to $|\Psi x + dx\rangle$ also will take $|\Psi(x+dx)\rangle$ to $|\Psi(x+2dx)\rangle$. Intuitively this makes sense - two rotations about the same axis by angle $d\theta$ is the same as a rotation by angle $2d\theta$, or moving forward by dt twice is the same as moving forward by $2dt$. That leaves us with $\mathrm{U}(ndx) = \mathrm{U}(dx)^n$. That works if the eigenvalues of U are proportional to dx. That leaves us with the general form of a unitary operator that translates a wave function by some small amount dx:

$$\mathrm{U}_x(dx) = \mathrm{V}\exp(i\Lambda_x dx)\mathrm{V}^\dagger \tag{5}$$

I've added the subscript x to emphasize that over different ranges where we might want to carry out translations, the eigenvalues can change, hence the translation operator changes. We'll indeed see cases where this happens, but if we make dx small enough, we can always treat Λ as constant over a region. Since $\mathrm{V}\Lambda_x\mathrm{V}^\dagger$ is a Hermitian matrix, we can write

$$\mathrm{U}_x(dx) = \exp(i\mathrm{A}_x dx) \tag{6}$$

If I've zoomed in even further, I can expand the right-hand side of Equation 6 to first order to get:

$$\mathrm{U}_x(dx) = 1 + i\mathrm{A}_x dx \tag{7}$$

This is the fundamental reason why the generator of rotations looked the way it did, and that the generator of time translations looks the same way. If we have a unitary operator that continuously transforms a state into another state, you *have* to be able to write it down as a generator of translations, as in Equation 7. We can often give Equation 7 more physical meaning by introducing constants. For rotations, we set $\mathrm{A} = -\mathrm{J}/\hbar$ because it turned out that writing things that way meant we could generate rotations with the angular momentum operator. For time evolution, we'll write the generator of time translations as

$$\mathrm{U}_t(dt) = \mathrm{I} - i\mathrm{H}dt/\hbar \tag{8}$$

because it will turn out the operator H is the Hamiltonian of the system when we write things this way. We haven't shown that, of course, but at this point we do know that there exists some operator H that has the *units* of energy that will translate a wave function forward in time.

2. Conservation and Eigenstates

Let's consider the class of translation operators that don't explicitly depend on the particular spot we're doing the translation, *i.e.*

$$\mathrm{U}_x(dx) = \mathrm{U}(dx) \tag{9}$$

In the case of time translations, that means that U does not *explicitly* depend on time. Right away, from Equation 6, we see that we can generate large translations as well as small ones. If we know that our starting wave function is $|\Psi(x=0)\rangle \equiv |\Psi_0\rangle$, then

$$|\Psi(x)\rangle = \exp(i\mathrm{A}x)\,|\Psi_0\rangle \tag{10}$$

If pick our starting state to be an eigenvector of A, with eigenvalue λ, then we know that

$$\exp(i\mathrm{A}x)\,|\Psi_0\rangle = \exp(i\lambda x)\,|\Psi_0\rangle \tag{11}$$

since the we saw that the eigenvalues of unitary matrix $\mathrm{U} = \exp(i\mathrm{A})$ for Hermitian matrix A are $\exp(i\lambda)$, and the eigenvectors stay the same. The important thing to take away is that the left side of Equation 11 has an operator, but in the right side that has turned into a scalar. Because it's an eigenstate, our wave function has picked up an overall phase, but remains an eigenstate. That

means if we go out and measure our Hermitian operator A in the new state, we have to get the same answer. This follows from:

$$\langle\Psi(x)|\,\mathrm{A}\,|\Psi(x)\rangle = \langle\Psi_0|\exp(-i\lambda x)\mathrm{A}\exp(i\lambda x)\,|\Psi_0\rangle = \lambda \tag{12}$$

Whatever value of A we measure *does not change* as we evolve our system. It is *conserved*. Even more generally, if we let $|\Psi\rangle$ be a superposition of states, the expectation value of our operator won't change. If we have a mixed state, each component picks up a different phase, but the expectation of an operator doesn't depend on the relative phases of the eigenstates of that operator (they all just drop out when we form *e.g.* $c_+^* c_+$ and $c_-^* c_-$ for a spin-1/2 system). Equivalently, since U and A commute (because U consists of powers of A), $\langle\Psi_0|\,\mathrm{U}^\dagger\mathrm{AU}\,|\Psi_0\rangle = \langle\Psi_0|\,\mathrm{AU}^\dagger\mathrm{U}\,|\Psi_0\rangle = \langle\Psi_0|\,\mathrm{A}\,|\Psi_0\rangle$.

This is really a remarkable conclusion. Simply by requiring that probability is conserved, we see that if we want to evolve a wave function, there's a unitary operator that does that. That unitary operator has an associated Hermitian operator, and the expectation of that operator does not depend on the wave function's evolution, as long as the evolution does not depend explicitly on our variable. For the particular case of translations in time, there's an associated Hermitian operator with units of energy, the expectation of which does not change with time (again, as long as there is no explicit time dependence, so H is not $\mathrm{H}(t)$). It's a small (but important!) leap to say that operator *is* the energy of the system, but that is indeed true and so the time evolution operator is the Hamiltonian H.

3. The Schrodinger Equation(s)

Let's go back to Equation 6, and insert a wave function on the right. That leaves us with

$$|\Psi(x_0+x)\rangle = \mathrm{U}_{x_0}(x)\,|\Psi(x_0)\rangle = \exp(i\mathrm{A}_{x_0}x)\,|\Psi(x_0)\rangle \tag{13}$$

where we've changed variables to say we're starting at some value x_0 and shifting by x (I'm doing this mainly to avoid writing ddx). If we differentiate this with respsect to x, we have

$$\frac{d\Psi}{dx} = i\mathrm{A}\,|\Psi\rangle \tag{14}$$

I've dropped the explicit dependence on x_0 because we're now assuming you're evaluating A for the same value of x where you're evaluating your wave function. Of course, A may well be constant (and indeed will be for most of the examples we look at), but now it doesn't *have* to be. We're now left with the quite general theorem that for a wave function that depends on parameter x, any continuous[1] unitary transformation of our wave function can be described by Equation 14. The operator A is Hermitian, and is related to our unitary transform by $\mathrm{U} = \exp(i\mathrm{A})$. If we take the special case of evolving with time, and letting $\mathrm{A} = -\mathrm{H}/\hbar$, Equation 14 turns into the Schrodinger equation:

$$\frac{d\,|\Psi\rangle}{dt} = -i\mathrm{H}\,|\Psi\rangle\,/\hbar \tag{15}$$

[1]Because we do absolutely need that $\mathrm{A}(2dx) = \mathrm{A}(dx)\mathrm{A}(dx)$.

Moving the i and $\hbar$ to the left side of the equation, we're left with the Schrodinger equation as it's normally written:

$$i\hbar\frac{d\,|\Psi\rangle}{dt} = \mathrm{H}\,|\Psi\rangle \tag{16}$$

Unitarity guarantees time evolution has to look like the Schrodinger equation, and for now since H is conserved and has units of energy, we'll take the leap of faith that it is the Hamiltonian. While we'll concentrate on solving Schrodinger's equation for much of the rest of the course, do please remember that there is a Schrodinger-like equation for *any* unitary evolution. We'll see this for several cases. If we replaced dt by $d\theta$ in Equation 16, then the associated operator has the units of angular momentum, and indeed is J. The generator of rotations is simply the first-order expansion of Equation 16. If we replace dt by dx, then the associated operator has units of momentum, and once again turns out to be the momentum operator (with a sign flip). If we replace dt by dp, then the associated operator is position (without a sign flip)[2].

[2]It doesn't actually matter which of the momentum and position operators get the minus sign, but it's does very much matter that they have opposite signs. We'll see that position and momentum representations are forward/inverse Fourier transforms of each other. If you want to end up back where you started after a forward/inverse transform, then the operators have to have opposite signs.