Introduction

These notes are meant to accompany a short class on quantum mechanics at Canada/USA Mathcamp 2016. It might be possible to read through them outside the context of that class, but that would probably not be the best way to learn the subject for the first time; it will certainly be harder to follow without the benefit of the in-class discussion. Still, if you're reading this outside of Mathcamp and have a question, you can e-mail me at `njmford@gmail.com`.

I'll assume that the reader is familiar with a few topics in linear algebra — in particular, we'll be using basic ideas like linear independence, spanning, bases, and dimension; eigenvectors and eigenvalues; and the notion of an inner product on a finite-dimensional $\mathbb{C}$-vector space. It's also helpful, but not technically required, if you've seen a little bit of classical Newtonian mechanics, at least up through the definition of momentum. For the material at the end, it will be necessary to know a small amount of calculus; if you know what a derivative is you're probably fine, and if you know what a partial derivative is you're definitely fine.

1 States and Observables

Any well-defined physical theory should have some notion of "states" and "observables." A *state* is just what it sounds like — some mathematical object that represents the state of the physical system being modeled. An *observable* is some physical quantity that can be measured by someone observing the system. For example, in Newtonian mechanics, the state might consist of the position and momentum of all the particles under consideration, and an observable might be something like the distance between two of the particles, or the velocity of one of them with respect to the center of mass. It should always be possible, given a state, to "measure" an observable, producing the value the corresponding measurable quantity takes in that state.

1.1 Classical States and Observables

It's possible to frame Newtonian mechanics in this language. Suppose we're modeling a system with a finite number of particles. (It's also possible to do this for infinite classical systems, like a continuous rigid body, but it's more complicated.) As mentioned above, you can form a state by writing down the position and momentum of each of the particles — since the position and momentum can each be described by three numbers, such a state for an n-particle system can be described by a point in $\mathbb{R}^{6n}$. We'll sometimes call this the *phase space* of the system. (Sometimes it's helpful to only allow the state to live on a subset of the phase space: if, for example, two of the particles are connected by a rigid rod of a fixed length, then we might only allow those states in which those particles are the corresponding distance apart.)

A classical observable, then, is just a real-valued function on phase space. Given a point in phase space and such a function, you can measure the observable just by plugging that point into the function and get a number.

Many popular descriptions of quantum mechanics say (or at least imply) that quantum states are what you get by taking this picture and adding in what we'll call "mixed states," that is, a state that somehow combines aspects of multiple pure states. (Sometimes these accounts will call these "superpositions.") For example, suppose s and s' are two pure states, and f is an observable which takes the value 1 in state s and 5 in state s'. If you measure f in the mixed

state $\frac{1}{3}s + \frac{2}{3}s'$, you should get 1 with probability $\frac{1}{3}$ and 5 with probability $\frac{2}{3}$ — in particular, when you measure an observable in a mixed state, the result is not a single, definite number but a probability distribution on the possible outcomes.

This idea — that you might find your physical system in a mixed state — turns out to not be nearly weird enough to capture what's commonly referred to as quantum behavior. It's certainly possible and useful to build a physical theory this way, but *this theory is not quantum mechanics*. Indeed, this perspective is very useful for doing statistical mechanics; when you're studying large-scale physical systems made out of enormous numbers of particles, it's very natural to work with mixed states, since there's no reason to assume you know the state of every particle.

Indeed, if this was all that was going on, a lot of the "quantum paradoxes" that these popular accounts discuss would not be so mysterious: even if you have a mixed state in front of you, it's easy to assume that there is a true pure state that the universe is actually in but that you just don't know what it is. Then the fact that measuring observables produces probability distributions shouldn't surprise you at all. It's a direct reflection of your ignorance about the true state, and if you had more information the uncertainty would go away.

As we'll see, it's not possible to do away with the probabilities in quantum mechanics so easily. But to see why, we have to first understand what we mean by quantum states and quantum observables. The definitions will seem strange at first; while I can't promise to make them seem less strange, I will at least do my best to explain what it is they're telling you about how quantum measurement works.

1.2 Quantum States and Observables

We're ready to talk about where the states live in quantum mechanics. Again, the definitions will probably look pretty arbitrary, but I'll tell you all of them first, and then I'll explain how they're usually interpreted.

To every quantum system, we'll associate a vector space over the complex numbers. We'll call this the "state space," and we'll often call it H. We'll also give H a *complex inner product*. As a reminder, a complex inner product is a function from $H \times H \to \mathbb{C}$ satisfying the following properties. (We'll write $\langle x, y\rangle$ for the image of the pair (x, y) under this function.)

- The inner product is linear in its second argument, that is, for vectors $x, y, z \in H$ and any $\alpha \in \mathbb{C}$, we have
$$\langle x, y+z\rangle = \langle x, y\rangle + \langle x, z\rangle$$
and
$$\langle x, \alpha y\rangle = \alpha\langle x, y\rangle.$$
- The inner product is *conjugate symmetric*, that is,
$$\langle x, y\rangle = \overline{\langle y, x\rangle},$$
where the bar denotes the complex conjugate.
- The inner product is *positive definite*, that is,
$$\langle x, x\rangle \geq 0,$$
(in particular it's a real number) and we get 0 if and only if $x = 0$.

(Note that this implies that scalars pull out of the first argument as complex conjugates, that is, $\langle \alpha x, y\rangle = \overline{\alpha}\langle x, y\rangle$.)

Our pure states will be the vectors in H of norm 1, that is, the vectors v for which $|v| = \sqrt{\langle v, v\rangle} = 1$. (There is also a way to talk about mixed states in this context, but we won't be using them here, so we're not going to talk about it.) In general — and indeed in many physical applications — H can be infinite-dimensional. This setting is quite a bit more complicated mathematically, though, and we won't worry about it until the end of these notes. So *until the last section, we'll assume that H is finite-dimensional.*

There are a couple ways to define quantum observables; the one I've picked here is not the one that most physicists work with, but I'm using it because it requires a little less linear algebra machinery to define: we'll simply say that an observable is an orthonormal basis of H.

What happens when you want to measure the value of an observable in some state? Suppose the system is in the state corresponding to some vector v, and we want to measure an observable corresponding to the basis $v_1, \ldots, v_n$. When you perform the measurement, the result is one of the v_i's — that is, there are $\dim H$ possible results you can see. You get some particular v_i with probability $|\langle v_i, v\rangle|^2$. After the measurement has happened, the state becomes whichever v_i resulted from the measurement.[1]

We should pause to say a bit about what's going on here. First, it's tempting to think of the state space H as being roughly the same object as the phase space we discussed in the classical context. After all, they're both vector spaces with elements that are supposed to stand for our states. So it's worth pointing out that this interpretation is very misleading. Indeed, if H is finite-dimensional, then when I measure any observable there are only $\dim H$ different outcomes I can get. So the coordinates of the vector that represents a state shouldn't be thought of as the position or momentum of anything.

Instead, in this finite-dimensional context, we should think of the system as modelling some object for which any observable can only take on finitely many values. There are physical systems that behave this way — for example the spin of an electron lives in a two-dimensional space — but a particle with position and momentum isn't one of them. It's possible to model a system like that with an infinite-dimensional vector space, and we'll consider that at the very end of these notes.

For some observable corresponding to the basis $v_1, \ldots, v_n$, what happens when you measure it when the state is one of the v_i's? Since the v's form an orthonormal basis, we have that $\langle v_i, v_j\rangle = 0$ for $i \neq j$ and $\langle v_i, v_i\rangle = 1$. So we get v_i as the result of our measurement with probability 1, and it's impossible to see any of the other v. So we see that for each observable, the corresponding basis consists of states in which that observable has a definite value, and the state doesn't change as a result of the measurement.

Let's look at an example. Say H is $\mathbb{C}^2$, with the basis $e_1 = (1,0)$ and $e_2 = (0,1)$, and with the standard inner product. That is,

$$\langle a e_1 + b e_2, c e_1 + d e_2\rangle = \overline{a}c + \overline{b}d.$$

Let's consider two states and two observables: the states will be

$$s_1 = e_1; \qquad s_2 = \frac{1}{\sqrt{2}}(e_1 + e_2)$$

[1] A lot can be said about this idea that the measurement changes the state of the particle being measured. If you'd like to make the fewest assumptions necessary to make predictions about experiments, you can just say that all it means is that all subsequent measures will have the same result as they would have if the state were represented by the basis vector you obtained. Like most people who've spent a lot of time thinking about these ideas, I certainly have a favorite interpretation, but I also recognize the existence of intelligent people who have different interpretations from mine, and while this is definitely an interesting question we won't spend much more time on it in these notes.

and the observables will be A, corresponding to the basis

$$v_1 = e_1, \quad v_2 = e_2,$$

and A', corresponding to the basis

$$v_1' = \frac{1}{\sqrt{2}}(e_1 + e_2), \quad v_2' = \frac{1}{\sqrt{2}}(e_1 - e_2).$$

From our previous discussion, we can see that if we measure A in the state s_1 or measure A' in the state s_2, we get the corresponding result with probability 1. What about the other way? If we measure A in state s_2, then we see e_1 with probability

$$\left|\left\langle e_1, \frac{1}{\sqrt{2}}(e_1 + e_2)\right\rangle\right|^2 = \left|\frac{1}{\sqrt{2}}\right|^2 |\langle e_1, e_1 + e_2\rangle|^2 = \frac{1}{2} \cdot 1 = \frac{1}{2},$$

and similarly we also get e_2 with probability $\frac{1}{2}$. A similar computation also shows that if we measure A' in state s_1, each outcome again happens with probability $\frac{1}{2}$.

This example demonstrates something important about quantum states that doesn't happen in classical physics: since different observables correspond to different bases, it's impossible to have a state in which every observable takes on a definite value. Moreover, this isn't a question of "pure states" versus "mixed states": the same state will behave deterministically with respect to one observable and nondeterministically with respect to many others. (Indeed, while there is a notion of mixed states in quantum mechanics, all the states we've discussed so far are pure states.)

Still, you'd be excused for thinking that this description is hiding part of the story: perhaps, like we mentioned when we discussed statistical mechanics at the beginning of this section, the probabilities discussed here are merely a reflection of our ignorance of the true, underlying state of the physical system we're measuring. There are so-called "hidden-variable theories" of quantum mechanics, but in the next section we'll see that there are a lot more limitations to such a theory than you might think.

Exercises

1. Let s be a state, and say α is a complex number with absolute value 1. Show that the state αs behaves identically to s when measuring any observable. (Because of this, we often think of s and αs as representing the same state.)

2. Why do we insist that states only correspond to vectors in H of norm 1? [Hint: what does the norm of a state have to do with the probabilities we get when we measure an observable in that state?]

3. Electrons have a propety called *spin* which behaves in some ways like angular momentum — in particular, the spin depends on which axis you measure it around — but when you measure spin around some axis there are only two possible results you can get. (They're usually called "up" and "down.") That is, the spin around a specified axis is an observable, and when you measure it you always get one of two possible results.

   The spin of an electron can be represented by a vector in $\mathbb{C}^2$. We'll write $(\uparrow_x)$ for the state in which the electron has spin up around the x-axis. There is a basis in which we have:

$$(\uparrow_x) = \frac{1}{\sqrt{2}}(e_1 + e_2); \quad (\downarrow_x) = \frac{1}{\sqrt{2}}(e_1 - e_2);$$

$$(\uparrow_y) = \frac{1}{\sqrt{2}}(e_1 + i\,e_2); \quad (\downarrow_y) = \frac{1}{\sqrt{2}}(e_1 - i\,e_2);$$
$$(\uparrow_z) = e_1; \quad (\downarrow_z) = e_2.$$

(a) Suppose I measure an electron's spin around the y-axis and find that it's spin up, then I measure its spin around the x-axis. What is the probability of getting each outcome? What if I used the z-axis at the end instead of the x-axis?

(b) What if, instead, I measure the electron's spin around the y-axis, find it's spin up, and then measure its spin around the y-axis again?

(c) The laws of physics don't change when you apply a rotation, so the relationship between the spin states for the x, y, and z axes should hold for any choice of three perpendicular axes. Suppose we pick an axis A in the xy-plane. What does this rotational symmetry tell us about the relationship between the states $(\uparrow_A)$, $(\downarrow_A)$, $(\uparrow_z)$, and $(\downarrow_z)$?

(d) Suppose A is one of the two axes that makes a 45° angle with both the x and y axes. What new information can we conclude about $(\uparrow_A)$ and $(\downarrow_A)$?

2 Evolution and Entanglement

2.1 How Quantum States Evolve

There is a pretty big hole in our description of quantum mechanics so far: we've described what states look like and what it means to measure an observable, but we haven't described how states evolve in time. In the last section, we'll analyze a particular physical system in which we'll be able to answer the question in detail, but at this level of generality, the answer is, of course, that it depends on which system you're talking about.

But there is one constraint that the evolution of any quantum system will always satisfy. Recall from the exercises to the previous section that the fact that states correspond to vectors of norm 1 is what makes the probabilities of the outcomes of a measurement sum to 1. However a quantum state evolves through time, then, it seems important to preserve the norm of the corresponding vector.

This is all a way of making the following rule seem a bit more believable: quantum states evolve in time according to a *unitary* linear map. Recall that unitarity is the complex version of orthogonality: a linear map $U : H \to H$ is unitary if, for each $v, w \in H$, we have

$$\langle U(v), U(w)\rangle = \langle v, w\rangle.$$

There is an equivalent condition in terms of matrices: U is a unitary matrix if $UU^* = 1$, where U^* is the *conjugate transpose* of U, i.e. the matrix you get by taking the transpose and then conjugating each entry. Unitary maps are precisely the linear maps on H that take vectors of norm 1 to vectors of norm 1.[2]

Let's look at an example. Consider the two-dimensional space H from the previous section, with the observable A corresponding to e_1 and e_2 we used before. Suppose there's some physical process which acts on states via the matrix

$$U = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

So if our system starts in the state e_1, then when this process is done it's in the state

$$s = \frac{1}{\sqrt{2}}(e_1 + e_2).$$

As we saw before, measuring A in this state will give either e_1 or e_2 each with probability $\frac{1}{2}$.

A very tempting interpretation of this is that this simply means that what we actually have in front of us at this point is an object that is either in state e_1 or e_2, but we just don't know which. But you run into trouble with this interpretation when you want to apply U again: you can check that U takes s to e_2, that is, a state in which A has a definite value. If we're thinking of s as just representing a state which has some possibility of secretly being e_1, then we would conclude that applying U again should leave some chance of having e_1 again. But that's not what we see: after applying U to s, it's become impossible to get e_1.

[2] This, of course, leaves open the question of why time evolution ought to be linear at all. I don't have as good a reason for this — and anyway, my reasoning for insisting on preserving the norm is of course predicated on believing the picture of states and observables laid out in the previous section. If you look for a while, you can find various attempts to justify the postulates of quantum mechanics on the Internet from more reasonable-sounding assumptions, but I'm not going to worry too hard about this here.

This is meant to push us into a different interpretation of these states: we can't interpret our state in the way that we thought we were going to — we've done something more than just introduce some probabilities in the way one does in statistical physics. If one insists on interpreting states in this way, one has to accept the possibility that *adding* a possible outcome to a measurement (like adding the possibility of getting e_2 to the s example) can make it *less likely* to observe some other outcome from a different observation (as happened to the possibility of getting e_1 after applying U the second time).

2.2 Entanglement

We'll start our discussion of entanglement by exploring how to describe states and observables for a physical system with more than one particle in it. That is, suppose we have some particle whose states live in H and another particle whose states live in H'. What should we use as the state space for the system consisting of the two particles together?

One way to motivate the answer to this question is to think about what should happen when we measure some observables corresponding to some properties these particles could have. So suppose A is an observable corresponding to the basis $e_1, \dots, e_n$ of H, and A' is an observable corresponding to the basis $e'_1, \dots, e'_m$ of H'. If we measure A for our first particle and A' for our second particle, we should get some pair of outcomes, e_i and e'_j, so this pair needs to give us a state for our combined system.

That is, our states should live in a new vector space, which we'll write as $H \otimes H'$, which has an orthonormal basis consisting of all of these "paired up" vectors. We'll write such a basis vector $e_i \otimes e'_j$. That is, this vector describes a state in which, if we measure A for the first particle and A' for the second, we get the outcomes e_i and e'_j respectively. If $\dim H = n$ and $\dim H' = m$, then $\dim(H \otimes H') = nm$.

There's one more question to answer here: if we have a state in $H \otimes H'$ and I measure some observable on just one of the particles? As an example, say that $H = H'$, and H is two-dimensional, so a general state looks like

$$\alpha_{11} e_1 \otimes e_1 + \alpha_{12} e_1 \otimes e_2 + \alpha_{21} e_2 \otimes e_1 + \alpha_{22} e_2 \otimes e_2,$$

with $|\alpha_{11}|^2 + |\alpha_{12}|^2 + |\alpha_{21}|^2 + |\alpha_{22}|^2 = 1$. When you measure our observable A from before, you see that that particle is either in state e_1 or e_2. What does this tell you about the state of the whole system? If you see e_1 from this measurement, all you've learned is that the resulting state ought to belong to the two-dimensional subspace S_1 of $H \otimes H'$ consisting of vectors of the form $\sum_i \alpha_i e_1 \otimes e_i$ — that is, ones where we know the first component is in state e_1.

This is all to motivate the following stipulation: when we measure A, we take our state and project it onto this subspace S_1. So if we get e_1 from our measurement of A, our state should become

$$s_1 = \frac{1}{\sqrt{|\alpha_{11}|^2 + |\alpha_{12}|^2}} (\alpha_{11} e_1 \otimes e_1 + \alpha_{12} e_1 \otimes e_2),$$

and this should happen with probability $|\alpha_{11}|^2 + |\alpha_{12}|^2$.

(Why $|\alpha_{11}|^2 + |\alpha_{12}|^2$? One way to make this feel more like our other examples is to notice that our original state can be written as

$$\sqrt{|\alpha_{11}|^2 + |\alpha_{12}|^2}\, s_1 + \sqrt{|\alpha_{21}|^2 + |\alpha_{22}|^2}\, s_2,$$

where $s_1 \in S_1$ and $s_2 \in S_2$; s_1 is the expression given above, and s_2 is the analogous expression with the roles of e_1 and e_2 switched. That is, s_1 is a state in which measuring A would always

yield e_1 and s_2 is in a state where it would yield e_2. These square-root factors are exactly the ones we need to make s_1 and s_2 have norm 1.)

Note also that this requires a slight extension of our earlier definition of observables: we'd like to allow an observable that corresponds not just to an orthonormal basis of our state space but also to a decomposition of the state space into orthogonal subspaces, like our S_1 and S_2 above. (This is the same as taking an orthonormal basis, but then grouping some sets of vectors in the basis together and considering the subspaces spanned by the vectors in each group. In this language, S_1 would come from the group $\{e_1 \otimes e_1, e_1 \otimes e_2\}$ and S_2 from $\{e_2 \otimes e_1, e_2 \otimes e_2\}$.)

Some states in $H \otimes H$ come from taking two states s and s' from H and considering them together — that is, they describe a system in which the first particle is in state s and the second is in state s'. For example, the state $e_1 \otimes e_1$ we considered before is like this, as is the state

$$\left(\frac{1}{\sqrt{2}}(e_1 + e_2)\right) \otimes \left(\frac{1}{\sqrt{2}}(e_1 - e_2)\right) = \frac{1}{2}(e_1 \otimes e_1 - e_1 \otimes e_2 + e_2 \otimes e_1 - e_2 \otimes e_2).$$

States like this are called *separable*. (In particular, the procedure we used in this last example is the general way to combine a state from H and a state from H' to produce a state from $H \otimes H'$: allowing the $\otimes$ symbol to distribute over addition in this way is exactly the thing that reproduces the rules for measurement in $H \otimes H'$ we just discussed.)

States that aren't separable are called *entangled*. For example, the state

$$e = \frac{1}{\sqrt{2}}(e_1 \otimes e_1 + e_2 \otimes e_2)$$

isn't separable: if we could write it in the form

$$(\alpha e_1 + \beta e_2) \otimes (\alpha' e_1 + \beta' e_2),$$

then we get that one of α or β' has to be 0, which is inconsistent with having a nonzero coefficient on both $e_1 \otimes e_1$ and $e_2 \otimes e_2$.

There is a famous example, called the *CHSH game*, that illustrates a strange property of entangled states. Alice and Bob will play a game. Each of them is assigned a separate room containing a coin and a button, but they have a chance to agree on a strategy before they're separated. When they're ready to play, they'll go to their appointed rooms and flip the coin. After that, they'll have a chance to either press the button or not.

The goal is as follows: they want exactly one of them to press the button if and only if both coins land heads. (So if at least one coin lands tails, then they want either to both press the button or both not press the button.) Notice that each player has only four strategies to choose between (the only choices are what to do if the coin lands heads and what to do if it's tails) so there are 16 strategies total. It's not difficult to check that no matter what they do, they can't do better than a $\frac{3}{4}$ chance of winning. Furthermore, choosing their strategy randomly can't help: this just amounts to randomly choosing one of the 16 possible strategies according to some probability distribution, and randomly choosing among strategies that can't win more than $\frac{3}{4}$ of the time can't result in a strategy that wins more than $\frac{3}{4}$ of the time.

But what if Alice and Bob have a pair of particles that they've placed into the state e we discussed above? Then if Alice takes one of the particles and Bob takes the other, they have one more thing they can do after the coin has been flipped: each of them can choose which basis to use to measure the particle they have. Let's see what happens when they do this.

It will be useful to have some notation to talk about the bases they can use. Write $\{e_1^\alpha, e_2^\alpha\}$ for the bases you get by rotating $\{e_1, e_2\}$ counter-clockwise by α, so

$$e_1^\alpha = (\cos\alpha)e_1 + (\sin\alpha)e_2,$$

and

$$e_2^\alpha = (-\sin\alpha)e_1 + (\cos\alpha)e_2.$$

This notation makes it easy to take inner products: we get that

$$|\langle e_1^\alpha, e_1^\beta\rangle|^2 = |\langle e_2^\alpha, e_2^\beta\rangle|^2 = \cos^2(\alpha-\beta)$$

and

$$|\langle e_1^\alpha, e_2^\beta\rangle|^2 = |\langle e_2^\alpha, e_1^\beta\rangle|^2 = \sin^2(\alpha-\beta);$$

one quick way to see this is to rotate both bases by $-\beta$ so that one of them is $\{e_1, e_2\}$ and the other is $\{e_1^{\alpha-\beta}, e_2^{\alpha-\beta}\}$.

Say Alice has the first particle and Bob has the second. If Alice's coin lands heads, she'll measure her particle in the basis $\{e_1^{\pi/4}, e_2^{\pi/4}\}$. If she gets tails, she'll use $\{e_1, e_2\}$. Bob will use $\{e_1^{-\pi/8}, e_2^{-\pi/8}\}$ for heads and $\{e_1^{\pi/8}, e_2^{\pi/8}\}$ for tails. In each case, they'll press the button if and only if the measurement resulted in the first basis vector.

You'll check the following two facts in the exercises:

- If you use one of these rotated bases to construct the state e under discussion, the result is the same. That is,

$$\frac{1}{\sqrt{2}}(e_1\otimes e_1 + e_2\otimes e_2) = \frac{1}{\sqrt{2}}(e_1^\alpha\otimes e_1^\alpha + e_2^\alpha\otimes e_2^\alpha).$$

- As a consequence of this, if Alice measures her particle in some basis $\{e_1^\alpha, e_2^\alpha\}$ and sees, say, e_2^α, then the two particles together end up in the state $e_2^\alpha\otimes e_2^\alpha$. In particular, after this happens, the second particle is in the state e_2^α, so if Bob now measures his particle in the basis $\{e_1^\beta, e_2^\beta\}$, the probability that he sees e_1^β is $|\langle e_2^\alpha, e_1^\beta\rangle|^2$, and likewise for e_2^β.

So here are the possible outcomes:

Alice's coin	Bob's coin	P(same)	P(different)
H	H	$\cos^2(3\pi/8)$	$\sin^2(3\pi/8)$
H	T	$\cos^2(\pi/8)$	$\sin^2(\pi/8)$
T	H	$\cos^2(\pi/8)$	$\sin^2(\pi/8)$
T	T	$\cos^2(-\pi/8)$	$\sin^2(-\pi/8)$

The winning outcome is "different" in the top row and "same" everywhere else, so in every case, they have the same probability of winning, namely $\cos^2(\pi/8) = \sin^2(3\pi/8) = \frac{1}{4}(2+\sqrt{2}) \approx 0.8536$. Note, in particular, that this beats the upper bound of $\frac{3}{4}$ we came up with from before!

It's worth emphasizing how strange this is: it's tempting to look at an entangled state and say that all we've done is randomly choose one of e_1 or e_2 to assign to both particles, but we don't know which one. But if that were all that was happening, then Alice and Bob couldn't have improved their chances of winning any more than they could with any other random strategy. Once again, we're forced to conclude that something much stranger is at work than "classical physics plus probabilities."

Exercises

1. Let
$$e = \frac{1}{\sqrt{2}}(e_1 \otimes e_1 + e_2 \otimes e_2)$$
be the state we considered in the final example. Suppose $\{f_1, f_2\}$ is a different orthonormal basis for H, and that we have
$$f_1 = \alpha e_1 + \beta e_2 \qquad f_2 = \gamma e_1 + \delta e_2$$
with $\alpha, \beta, \gamma, \delta$ real. We can form a different entangled state from this basis, say
$$f = \frac{1}{\sqrt{2}}(f_1 \otimes f_1 + f_2 \otimes f_2).$$
Show that in fact $e = f$.

2. Verify the consequence of this mentioned in the section: if Alice measures her particle in some basis $\{f_1, f_2\}$ and sees, say, f_2, then the two particles end up in the state $f_2 \otimes f_2$.

3. Verify that the numbers in the table of outcomes from the CHSH game are correct.

4. In this problem we'll describe another coordination game like the CHSH game, except that we'll use entanglement to win the game all the time instead of just with higher probability than expected. We'll make use of the electron spin states discussed in Problem 3 from Section 1, so you should do that problem before this one.

   The game is played as follows. Alice, Bob, and Charlie are allowed to decide on a strategy and are then sent to separate rooms, each of which has a card and a button. The card has either a 1 or a 2 on it; after looking at the card, each player can decide to either press the button or not.

   There are only two possibilities that will occur for the distribution of cards. The first is that one of the cards has a 1 and the other two cards have 2's. In this case, their goal is for an *even* number of players to press their button. (That is, either two of them press it or nobody does.) The other possibility is that all the cards are 1's. In this case, their goal is for an *odd* number of players to press their button. (So either one of them presses it or they all do.)

   (a) Show that, at least without deploying some fancy trick involving entanglement, there's no way for the players to always win this game.

   (b) On the other hand, suppose they share three entangled electrons in the state
   $$\frac{1}{\sqrt{2}}((\uparrow_z)\otimes(\uparrow_z)\otimes(\uparrow_z)+(\downarrow_z)\otimes(\downarrow_z)\otimes(\downarrow_z)).$$
   Show that Alice, Bob, and Charlie can use this to win the game all the time. [Hint: there's a way to do it in which each player has exactly the same strategy.]

3 Statistics

3.1 Observables as Operators and Expected Value

Given a state and an observable, we get a probability distribution on the possible results of measuring that observable in that state, so we can ask statistical questions about this probability distribution, things like the expected value or the standard deviation. In this section we'll develop the machinery that's necessary to answer questions like this.

We'll start with the question about expected value. Suppose we're measuring an observable corresponding to the orthonormal basis $e_1, \ldots, e_n$. In order to talk about the expected value of our measurement, we need to have some numbers attached to each of these outcomes so we can take their average, so let's attach some real number λ_1 to e_1, λ_2 to e_2, and so on.

Let's write our state in the same basis, so in the form

$$s = \alpha_1 e_1 + \cdots + \alpha_n e_n.$$

The probability of getting any e_i is just $|\alpha_i|^2$, so the expected value is

$$\sum_{i=1}^{n} \lambda_i |\alpha_i|^2.$$

Now, suppose A is a diagonal matrix with diagonal entries $\lambda_1, \ldots, \lambda_n$. Then observe that

$$\langle As, s \rangle = \sum_i \lambda_i \overline{\alpha_i} \alpha_i = \sum_i \lambda_i |\alpha_i|^2.$$

That is, we can recover the expected value by taking the inner product of As with s.

This is a much more powerful idea than it might first appear: this inner product doesn't depend on the basis that we write As and s in. Any time we have a linear map $A : H \to H$ with an orthonormal basis of eigenvectors $e_1, \ldots, e_n$ with real eigenvalues $\lambda_1, \ldots, \lambda_n$, then $\langle As, s \rangle$ is the expected value of the result of measuring the observable corresponding to that basis, where we attach the number λ_i to each basis vector e_i.

Let's do a quick example. Consider the 2×2 matrix

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

This has an orthonormal basis of eigenvectors: $f_1 = \frac{1}{\sqrt{2}}(e_1 + e_2)$ with eigenvalue $\lambda_1 = 1$, and $f_2 = \frac{1}{\sqrt{2}}(-e_1 + e_2)$ with eigenvalue $\lambda_2 = -1$. Consider the state $s = e_1$. Say we're measuring s in the basis f_1, f_2, and we record a 1 when we get f_1 and a -1 when we get f_2, Our new result says we can compute the expected value of this measurement as $\langle Ae_1, e_1 \rangle = \langle e_2, e_1 \rangle = 0$. We can verify this directly: The probability of getting f_1 is then $|\langle f_1, e_1 \rangle|^2 = \frac{1}{2}$, as is the probability of getting f_2, so the overall expected value is $1 \cdot \frac{1}{2} + (-1) \cdot \frac{1}{2} = 0$.

In practice, because it makes it easier to do statistical computations like this one, we often identify an observable with an operator like this. (Note that the operator contains strictly more information than the basis: it also records the numbers we've attached to each basis vector.) It turns out that the matrices of the form we want — the ones with an orthonormal basis of eigenvectors with real eigenvalues — can be identified rather easily:

Theorem (Spectral Theorem)**.** *Take a linear map $A : H \to H$ and write its matrix with respect to some orthonormal basis. Write A^* for the conjugate transpose of A, that is, the matrix you get by taking the complex conjugate of every entry in the tranpose of A.*

Then $A = A^*$ if and only if there is an orthnormal basis of eigenvectors of A, each of which has a real eigenvalue.

Matrices of this form are sometimes called *self-adjoint* or *Hermitian*. Having this condition is nice for a couple reasons: First, finding all the eigenvectors of a matrix is hard but checking whether a matrix is self-adjoint is easy. Second, we'll see in the next section that this self-adjointness property is easier to generalize to the infinite-dimensional setting than the orthonormal basis version of the same condition. In fact, most texts use this as the first definition of a quantum observable rather than doing what we did.

3.2 Variance and Uncertainty

One of the results from quantum mechanics that's managed to make it out into the world at large is something called the "Heisenberg uncertainty principle." It's often stated as a fact about the position and momentum of a particle, but a similar statement can actually be made about any pair of observables. We'll derive the general version here, applying it to position and momentum in the next section.

We saw in the last subsection how to compute the expected value of an observable in a given state. In order to write down the uncertainty principle, we'll have to discuss how to compute another statistical quantity, called the *variance* from a quantum observable and a state. The definition of variance comes from probability and statistics — there's nothing uniquely quantum about the concept — and it measures how far, on average, the measured value is from the expected value. Since it turns out to be easier to work with mathematically, the definition of the variance is given in terms of the square of the difference between the measured value and the expected value. That is, if X is a random variable with expected value μ, then the variance of X is the expected value of $(X-\mu)^2$.

If we're measuring an observable A in a state s, then as we saw before the expected value μ is given by $\langle As, s\rangle$. The observable corresponding to the square of the difference between μ and the result of measuring A is just $(A-\mu)^2$; you can see this by working in the orthonormal basis where A is diagonal. (In this basis, subtracting μ times the identity and squaring doesn't affect the orthonormal basis of eigenvectors, and it has the effect of subtracting μ from each eigenvalue and squaring the result.) So the variance is given by

$$\langle (A-\mu)^2 s, s\rangle.$$

It's often convenient to rewrite this as

$$\langle A^2 s, s\rangle - 2\mu\langle As, s\rangle + \mu^2\langle s, s\rangle;$$

since $\mu = \langle As, s\rangle$ and $\langle s, s\rangle = 1$, this is

$$\langle A^2 s, s\rangle - \langle As, s\rangle^2.$$

Now we're ready to begin deriving our uncertainty principle; we'll sketch out the proof here, leaving some details to the exercises.

Theorem (Uncertainty Principle)**.** *Let A and B be observables and let s be a state. We'll write $[A,B] = AB - BA$; this is called the* commutator *of A and B. Then if V_A and V_B are the variances of A and B respectively, we have*

$$V_A V_B \geq \frac{1}{4}|\langle [A,B]s, s\rangle|^2.$$

Proof. If μ_A and μ_B are the expected values of A and B in the state s, let's write $\bar{A} = A - \mu_A$ and $\bar{B} = B - \mu_B$. Note that then $V_A = \langle \bar{A}^2 s, s\rangle$ and similarly for V_B.

We'll look at the quantity

$$\langle (\bar{A} + i\,x\,\bar{B})s, (\bar{A} + i\,x\,\bar{B})s\rangle$$

where x is any real number. This is ≥ 0, being the inner product of a vector with itself. We can expand the left side to get

$$\langle \bar{A}s, \bar{A}s\rangle + x^2\langle \bar{B}s, \bar{B}s\rangle + i\,x(\langle \bar{A}s, \bar{B}s\rangle - \langle \bar{B}s, \bar{A}s\rangle) \geq 0.$$

In particular, even though it isn't otherwise obvious, the last term is real. (The minus sign on the last term comes from the fact that the i pulls out of the left half of the inner product as a $-i$.)

In one of the exercises, you'll prove that for any observable C and any vectors $v, w \in H$, we have $\langle Cv, w\rangle = \langle v, Cw\rangle$. Using this, we get that

$$\langle \bar{B}^2 s, s\rangle x^2 - i\langle (\bar{A}\bar{B} - \bar{B}\bar{A})s, s\rangle x + \langle \bar{A}^2 s, s\rangle \geq 0.$$

(Again, we get from this that the middle term is real.)

Treat this expression as a quadratic polynomial in x. Since it's always nonnegative, we see that its discriminant has to be less than or equal to zero. But this gives us that

$$(-i\langle [\bar{A}, \bar{B}]s, s\rangle)^2 - 4\langle \bar{A}^2 s, s\rangle\langle \bar{B}^2 s, s\rangle \leq 0.$$

Since $-i\langle [\bar{A}, \bar{B}]s, s\rangle$ is real, the first term is equal to $|\langle [\bar{A}, \bar{B}]s, s\rangle|^2$. You'll check in the exercises that $[\bar{A}, \bar{B}] = [A, B]$; once we know this, we have

$$|\langle [A,B]s, s\rangle|^2 - 4V_A V_B \geq 0,$$

so the proof is complete. □

What good is this fact, and what does it have to do with the uncertainty principle you might have heard of? The content of the uncertainty principle is that it gives a lower bound on the variances of A and B, so that in a state where one variance is small the other has to be correspondingly larger, since their product has to end up larger than $|\langle [A,B]s, s\rangle|^2$.

As you'll verify in the exercises, though, there's a limit to this interpretation in our finite-dimensional setting: the variance will be zero whenever s is an eigenvector of our observable, and in this case the theorem doesn't tell us much of anything about the variance of the other observable. But we'll see in the next section that in a lot of interesting cases the commutator $[A,B]$ ends up being a scalar, which makes the quantity $|\langle [A,B]s, s\rangle|^2$ constant. (In paricular, we'll see that the position and momentum of a particle behave this way.) In this case it's possible to take the "compensating variances" idea more seriously — it really is the case that the less variance a state has in position, the more it must have in momentum and vice versa.

Exercises

1. The definition of "self-adjoint" given in most linear algebra texts is the following: we say that A is self-adjoint if for any vectors $v, w \in H$ we have $\langle Av, w\rangle = \langle v, Aw\rangle$. Show that if A has an orthonormal basis of eigenvectors with real eigenvalues, then it is self-adjoint in this sense. (The reverse implication is also true, but requires the Spectral Theorem.)

2. Write down operators to represent the spin of an electron around the x-, y-, and z-axes, treating spin up as a 1 and spin down as a -1.

3. Prove that, for a state s and an observable A, s has a definite value — that is, an eigenvector of the matrix for A — if and only if the variance of A in the state s is zero.

4. Verify that the commutator of two matrices doesn't change when you subtract a constant from either one. That is, if A and B are matrices and α and β are numbers, then

$$[A-\alpha, B-\beta] = [A, B].$$

4 Infinite-dimensional State Spaces

We're finally ready to extend the ideas we've developed to a setting where it's possible to talk about position and momentum. A warning: this section will be less rigorous than the preceding ones; doing this all carefully would require worrying a great deal about various questions of convergence, and in my judgment this would only obscure the main ideas.

4.1 Wavefunctions

Let's meet our first infinite-dimensional state space, which is the one that will let us talk about positions and momenta of particles. In particular, our states will describe particles living in a one-dimensional space. The elements of our state space are going to be complex-valued functions on $\mathbb{R}$. Given such a function ψ, we'll think of $|\psi(x)|^2$ as giving us something like the probability that the particle is at x.

Since knowing that the particle is at some position x means that it definitely isn't at any other position, we know that states in which the particle has definite position should all be orthogonal to each other. So the values of the function at different points serve the same role as the coefficients α_i in an expansion like

$$s = \alpha_1 s_1 + \alpha_2 s_2 + \cdots + \alpha_n s_n,$$

it's just that instead of needing to pick a number α_i for each $i \in \{1, 2, \ldots, n\}$, we now need to pick a number $\psi(x)$ for each $x \in \mathbb{R}$.

With this motivation, it shouldn't be too surprising that we'll want to define our inner product by multiplying the values that two different functions take at the same point and integrating — integrals are, after all, the continuous analogue of sums. That is, we'll define

$$\langle \phi, \psi \rangle = \int \overline{\phi(x)} \psi(x) dx.$$

The functions that we want to have representing our states are the ones with norm 1, that is, the functions ψ for which

$$\langle \psi, \psi \rangle = \int |\psi(x)|^2 dx = 1.$$

Functions like this will be called *normalized*.

Right away there are a couple differences between this and the finite-dimensional version. First, in a finite-dimensional space, it's always possible to divide any nonzero vector by its norm to get a vector of norm 1. But it's not always possible to do this with an arbitrary function: the norm, which is defined to be $\sqrt{\int |\psi|^2}$, might be infinite! We usually solve this by disallowing functions for which this is true: we'll restrict membership in our space H to those functions whose norm is finite.[3]

The next difference shows up when we try to identify the states in which the particle has a definite position. First, let's identify the linear function $Q : H \to H$ that's supposed to correspond

[3] There's one more annoying detail here: the inner product as we've defined it isn't nondegenerate, that is, it's possible to have $\langle \psi, \psi \rangle = 0$ without ψ being the zero function. This happens, for example, if $\psi(x)$ is defined to be 0 unless $x = 0$, in which case it's 1. Normally this is resolved by simply considering two functions f and g to be the same if $\int |f - g|^2 = 0$, so that function ψ just described would be identified with 0; we won't worry about it any further here.

to position. Using the metaphor we came up with for interpreting the values of our functions — that the values of the functions are like the coefficients we get when expanding in an orthonormal basis — I claim that the most natural choice for Q is to make $(Q\psi)(x) = x\psi(x)$. In the finite-dimensional case, multiplying some vector $v = \alpha_1 e_1 + \cdots + \alpha_n e_n$ by the matrix A corresponding to some observable with orthonormal basis $e_1, \ldots, e_n$ has the effect of multiplying each coefficient α_i by the eigenvalue λ_i corresponding to e_i. So here we ought to do the same thing: if $\psi(x)$ is to serve the same role as the coefficient of a basis vector corresponding to being at position x, then we want to multiply it by x when we apply Q.

But what could this basis be? If α_q is the function corresponding to a state which is definitely in position q, then we'd need $\alpha_q(x) = 0$ for $x \neq q$, but still $\int |\alpha_q(x)|^2 dx = 1$. It's pretty easy to see that there's no value for $\alpha_q(q)$ — which is the only number we have left to pick — that can make that integral nonzero. So there's actually no element of H that does what we want. That is, Q doesn't have any eigenvectors.

(Physicists often find it convenient to introduce an imaginary function with the two properties just laid out so that they can write equations the sorts of equations we had in the previous sections but involving eigenvectors of Q. The eigenvector for $q = 0$ is called the "Dirac delta function" and written $\delta(x)$; then $\alpha_q(x) = \delta(x - q)$. It is possible to introduce a mathematical object called a *distribution* that behaves the way one would like an eigenvector of Q to behave, but it's certainly not a function, and it's not an element of H.)

If the position observable is represented by the operator Q, how do we represent momentum? It is at this point that I run out of ways to justify the answers to these questions in an intuitive way without going beyond what can be done in a four-day class. It is possible for the make the facts I'm about to relate sound like they're coming from somewhere, but I won't be able to present it in these notes.

The momentum observable is represented by the operator $P = -i\hbar\frac{d}{dx}$. (The quantity $\hbar$ appearing here is a physical constant, equal to about $1.055 \times 10^{-34}\,\mathrm{kg\,m^2\,s^{-1}}$. It's called "Planck's constant" and it's often pronounced "h-bar.") What happens when we try to apply the uncertainty principle we derived in the last section to Q and P? Recall that it says that in some state ψ,

$$V_Q V_P \geq \frac{1}{4}|\langle [Q,P]\psi, \psi\rangle|^2,$$

where V_P and V_Q are the variances of position and momentum respectively. So what's $[Q,P]$? We can compute

$$([Q,P]\psi)(x) = -i\hbar\left(x\frac{d\psi}{dx} - \frac{d}{dx}(x\psi(x))\right) = -i\hbar\left(x\frac{d\psi}{dx} - x\frac{d\psi}{dx} - \psi(x)\right) = i\hbar\psi(x).$$

So $[Q,P]$ is just multiplication by $i\hbar$. This lets us conclude the form of the Heisenberg uncertainty principle you might have seen before:

$$V_Q V_P \geq \frac{\hbar^2}{4}.$$

Unlike the examples in the last section, in this case the right side of the inequality is a constant, so the interpretation we were reaching for before applies more precisely now. The less variance there is in a state's position, the more variance there has to be in momentum.

The final piece of the puzzle I'd like to present here is how states evolve through time. We discussed before that the time evolution ought to be unitary, but we didn't describe exactly which unitary functions to use, since that choice depends on the physics. But in this setting we can give more of an answer (again without much justification.)

In any quantum system describing something in the physical world, there is an observable corresponding to the total energy of the system. This is usually called the *Hamiltonian* and written as H. (There is an unfortunate clash of notation between this and our state space.) The exact form the Hamiltonian depends on the physics of the system you're trying to model. This is also true in classical physics: if your physics is taking place in a gravitational field, for example, then there has to be a term for gravitational potential energy.

Once a Hamiltonian has been given, a state ψ evolves through time as follows:

$$i\hbar \frac{\partial \psi}{\partial t} = H\psi.$$

This is called the *Schrödinger equation*. In the exercises, we'll work through a simple example of the Schrödinger equation for a particular physical system and draw some conclusions about how it behaves.

Exercises

1. What would the wavefunction (that is, the function ψ we described in the section) for a state with definite momentum look like? Is it actually an element of H?

2. In this problem we'll describe how to model the state of a single, one-dimensional free particle of mass m. We'll just be scratching the surface here; for more details, check any book on quantum mechanics.

   (a) Since there are no forces acting on the particle, its total energy is just its kinetic energy. Show that the (classical) kinetic energy of a particle of mass m is equal to $p^2/2m$, where p is its momentum.

   (b) Given this, our Hamiltonian for this system will be $H = P^2/2m$, so the Schrödinger equation says that
   $$i\hbar \frac{\partial \psi}{\partial t} = \frac{1}{2m} P^2 \psi.$$
   Use the definiton of P from the section to write this as a differential equation involving derivatives of ψ with respect to both t and x.

   (c) Suppose ψ were an eigenvector of P with eigenvalue p. What would this equation imply about ψ? Solve the equation in this case.