APPPHYS225 - Tuesday 21 October 2008

- The essential function of physical theory is to make quantitative predictions about experiments.

preparation	*representation* $\longrightarrow$	$\|\Psi(t_1)\rangle$
$\downarrow$		$\downarrow$
evolution	*specification of dynamics*	$i\hbar\frac{d}{dt}\|\Psi\rangle = \mathbf{H}\|\Psi\rangle$
$\downarrow$		$\downarrow$
measurement	*prediction* $\longleftarrow$	$\langle\Psi(t_2)\|\mathbf{O}\|\Psi(t_2)\rangle$

- State vectors $|\Psi\rangle$ and density operators ρ are mathematical objects that summarize our knowledge about the state (preparation+subsequent evolution) of a physical system.
- The use of representing knowledge in this way is that it efficiently allows us to make statistical predictions about the outcome of **any** measurement that could be made on the system.

Given that we fix a complete orthonormal basis for the Hilbert space, kets can be represented as column vectors

$$|\Psi\rangle = c_0|0\rangle + c_1|1\rangle \leftrightarrow \begin{pmatrix} c_0 \\ c_1 \end{pmatrix},$$

bras can be represented as row vectors

$$\langle\Psi| = c_0^*\langle0| + c_1^*\langle1| \leftrightarrow \begin{pmatrix} c_0^* & c_1^* \end{pmatrix},$$

and density operators can be represented as square matrices

$$\rho = |\Psi\rangle\langle\Psi| = (c_0|0\rangle + c_1|1\rangle)(c_0^*\langle0| + c_1^*\langle1|)$$

$$\leftrightarrow \begin{pmatrix} c_0 \\ c_1 \end{pmatrix} \begin{pmatrix} c_0^* & c_1^* \end{pmatrix} = \begin{pmatrix} |c_0|^2 & c_0 c_1^* \\ c_0^* c_1 & |c_1|^2 \end{pmatrix}.$$

There is a distinct standard measurement *procedure* corresponding to every complete set of orthogonal projectors. We say that measurement A is 'distinct' from measurement B if there exists some preparation of the system such that A and B have different measurement statistics. There are an uncountable number of distinct measurements (even in two Hilbert-space dimensions), so $|\Psi\rangle$ and ρ are pretty compact ways to summarize the statistics of *all* of them!

On the other hand, there are important consequences of this 'many-to-one-ness.' To illustrate this concretely let's work with a simple two-dimensional Hilbert space, with basis kets $|x\rangle$ and $|y\rangle$. Given an arbitrary pair of states

$$|\Psi_c\rangle = c_x|x\rangle + c_y|y\rangle,$$

$$|\Psi_d\rangle = d_x|x\rangle + d_y|y\rangle,$$

we ask under what conditions is it possible to find a (standard) measurement that can distinguish them with zero probability of error. For our purposes today, a standard measurement is specified by a complete set of orthogonal projectors. Since for now we are working in just two dimensions, we actually only need to specify a single ket $|\phi\rangle$. Then unambiguously,

$$\mathbf{P}_1 = |\phi\rangle\langle\phi|,$$

$$\mathbf{P}_2 = \mathbf{1} - \mathbf{P}_1.$$

In the current scenario, we are trying to pick $|\phi\rangle$ such that

$$\Pr(1|c) = \langle\Psi_c|\mathbf{P}_1|\Psi_c\rangle = 1,$$

$$\Pr(1|d) = \langle\Psi_d|\mathbf{P}_1|\Psi_d\rangle = 0,$$

$$\Pr(2|c) = \langle\Psi_c|\mathbf{P}_2|\Psi_c\rangle = 0,$$

$$\Pr(2|d) = \langle\Psi_d|\mathbf{P}_2|\Psi_d\rangle = 1.$$

In order to satisfy the first condition, we clearly need to choose

$$|\phi\rangle = |\Psi_c\rangle.$$

As a consequence

$$\Pr(1|d) = \langle\Psi_d|\mathbf{P}_1|\Psi_d\rangle = |c_x^* d_x + c_y^* d_y|^2,$$

and we find that our two states Ψ_c and Ψ_d are not perfectly distinguishable **unless** they have zero overlap:

$$\Pr(1|d) = 0 \quad \text{iff} \quad \langle\Psi_c|\Psi_d\rangle = 0.$$

Hopefully it should be clear that among all possible pairs of vectors in a Hilbert space, only a **vanishing** fraction are orthogonal! Furthermore, as we'll see below, the overlap of quantum states is preserved by physical dynamics according to the Schrödinger Equation.

And yet, according to our quantum representation rule, every vector in the Hilbert space corresponds to a distinct physical state of the system – that is, to a distinct *preparation* procedure.

Even though quantum measurement theory allows for non-projective measurements (as we have already briefly discussed), it is nonetheless a theorem that no measurement can distinguish nonorthogonal states, with zero probability of error, in a single trial. It is a mystery that quantum mechanics presents us with such a huge space of possible physical states without allowing us perfectly to distinguish between them. Why is it that we can "put more information into" the preparation of a quantum system than we can "get back out" in a single measurement of that very same system?

Random-basis quantum cryptography
Over the next few lectures we'll be talking a lot about entangment and its role in

enabling quantum information/communication protocols. Before we do that however we should note that some quantum information protocols don't even need entanglement!

As an example of this let's look at the crytographic key distribution protocol of C. H. Bennett and G. Brassard (*Proceedings of the IEEE International Conference on Computers, Systems, and Signal Processing, Bangalore, India* (IEEE, New York, 1984), p. 175; also, C. H. Bennett, G. Brassard, and N. D. Mermin, Phys. Rev. Lett. **68**, 557 (1992)). (In another lecture we'll look at a quantum cryptographic key distribution protocol that does require entanglement.)

A common way to encrypt data (string of meaningful bits) is the so-called "one-time pad," which requires a secret key (string of random bits) with length equal to that of the data. The encrypted version of the data is obtained simply by performing a bitwise XOR between the plain data and the secret key:

$data$:	$01011011101111\ldots$
key :	$01101010001101\ldots$
$encrypted$:	$00110001100010\ldots$

Recall that the XOR operation is defined by the following truth table:

$\mathrm{XOR}(0,0) = 0,$	$\mathrm{XOR}(0,1) = 1,$
$\mathrm{XOR}(1,0) = 1,$	$\mathrm{XOR}(1,1) = 0.$

If the key is truly random, the encrypted data is now statistically independent of the plain data. The encrypted data can only be decrypted by someone who knows the secret key. The procedure is then quite simple, corresponding to another XOR:

$encrypted$:	$00110001100010\ldots$
key :	$01101010001101\ldots$
$data$:	$01011011101111\ldots$

The drawback of one-time pad is that the communicators always need to have lots of secret key available (it's no good if you re-use it). In general, when all the available key has been used up the two parties have to either get together to decide on some new key bits, or they must use a trusted courier to "carry" new key information from one party to the other. Quantum cryptographic key distrubtion seeks to use quantum phenomena to allow generation of shared random key over public communication channels.

The basic idea is simple. Let the two communicators be named Alice and Bob (a common convention in the field of cryptography). Alice will send Bob a sequence of two-level quantum systems, each of which is randomly prepared in one of four possible states:

$$|0\rangle, \quad |1\rangle, \quad |x\rangle \equiv \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \quad |y\rangle \equiv \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle).$$

Alice doesn't initially say anything to anybody about the sequence of states. Every time Bob receives one of these quantum systems, he randomly chooses to make one of two standard measurements:

$$\{|0\rangle\langle 0|, |1\rangle\langle 1|\} \quad \text{or} \quad \{|x\rangle\langle x|, |y\rangle\langle y|\}.$$

3

After the entire transmission is done, Alice publicly announces for each transmission whether the state was selected from the $0, 1$ basis or the x, y basis (but without actually revealing which of the two basis states was sent). Bob also announces whether he measured in the $0, 1$ basis or the x, y basis, but without revealing his result. In cases where they happened to choose the same basis, Bob should have been able to perfectly recover the identity of the state that Alice sent. They can check this by using up a subset of the transmissions.

- State vectors can be used when the preparation and evolution of a system are as sharply defined as possible – in such cases, we have only **intrinsic** uncertainties regarding the outcome of future measurements.
- Density operators must be used when there are **extrinsic** uncertainties in the problem, or when the system of interest is entangled with an inaccessible system.
- Intrinsic uncertainties are associated with coherent superposition of eigenstates; extrinsic uncertainties arise (among other reasons) when we are not sure which of several preparations or dynamics has actually occured.
- Coherent superpositions can lead to interference effects; probabilistic mixtures cannot.

Recall that for essentially all states in a Hilbert space,

$$\Delta \mathbf{O}_q = \langle |\mathbf{O}_q - \langle \mathbf{O}_q \rangle| \rangle = \sqrt{\langle \mathbf{O}_q^2 \rangle - \langle \mathbf{O}_q \rangle^2} > 0$$

for *every* observable (physical quantity) $\mathbf{O}_q$ except those that are proportional to the identity operator. In orther words, for any given state in the Hilbert space there are plenty of questions one can ask to which there is no definite answer! Contrast this to the classical situation – if you know the position and momentum of a classical point particle, there's no measurement I could perform for which you can't predict the exact answer. The only way to have $\Delta \mathbf{O}_q = 0$ is to prepare the system in an eigenstate of $\mathbf{O}_q$. Any time we have a *superposition* of eigenstates corresponding to distinct eigenvalues of $\mathbf{O}_q$, we have **intrinsic** quantum-mechanical uncertainty.

Stern-Gerlach example

Consider the following experimental schematic, which depicts a sequential version of the famous 1922 experiment by Stern and Gerlach:
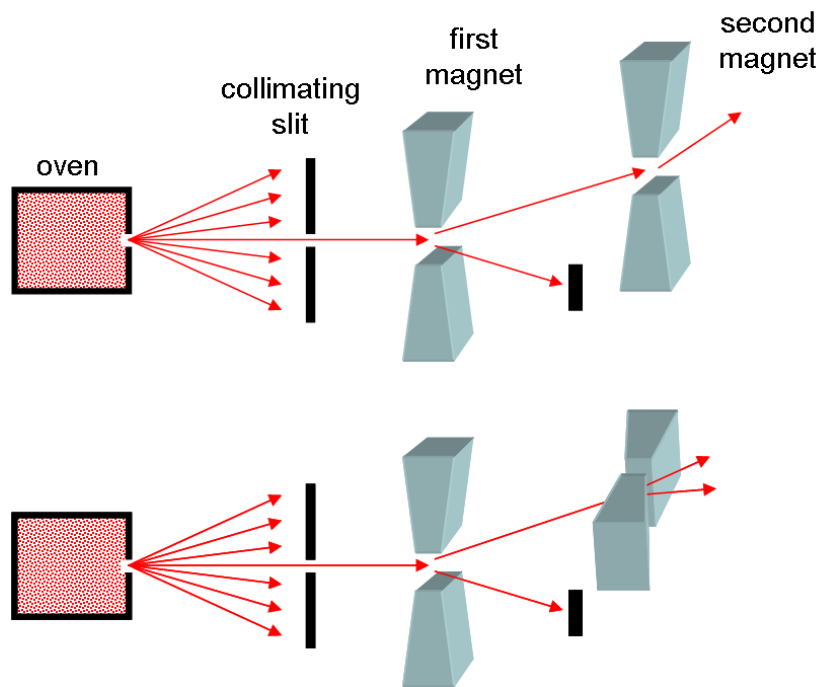
The oven and collimating slit produce a beam of silver atoms that passes through a specially-configured magnet; that magnet splits the beam in two (for reasons we'll discuss below) and then one of the branches is directed through a second specially configured magnet. If the second magnet is oriented in the same way as the first (upper diagram), one observes that it does not further split the atomic beam. But if the orientation of the second magnet is perpendicular to that of the first (lower diagram), one observes that the atomic beam is once again split into two branches.

The manipulation of the atomic beam by the specially-configured magnets is actually based on the internal spin of each atom (since each atom's magnetic moment is proportional to its spin) and must therefore be modeled quantum mechanically in order to make accurate predictions. Silver atoms happen to have an overall spin angular momentum of $\frac{1}{2}\hbar$, so the corresponding quantum probability model involves a simple two dimensional Hilbert space. The observables we'll need correspond to the z- and x-components of the atomic spin:

$$\mathbf{S}_z = \frac{\hbar}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \mathbf{S}_x = \frac{\hbar}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{S}_y = \frac{\hbar}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}.$$

(For completeness we have also written the observable corresponding to the y-component of spin, but we won't need to use it here.)

The action of the first magnet is to split the atomic beam into two components, one of which contains atoms whose internal spins are prepared in the state $|\Psi_{z+}\rangle$ while the other contains atoms whose internal states are perpared in the state $|\Psi_{z-}\rangle$. In terms of the matrix notation we have used for specifying the observables,

$$|\Psi_{z+}\rangle \to \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |\Psi_{z-}\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

It is easy to see that $|\Psi_{z+}\rangle$ is the eigenstate of $\mathbf{S}_z$ with eigenvalue $\hbar/2$, while $|\Psi_{z-}\rangle$ is the eigenstate of $\mathbf{S}_z$ with eigenvalue $-\hbar/2$. This splitting occurs because the magnetic field between the two pole pieces of the specially-configured magnet is oriented in the z direction and the flux density has a gradient along z. Atoms whose magnetic moments point downwards thus experience an impulse in one direction as they pass through the magnet, while atoms whose magnetic moments point up experience an opposite impulse. Since magnetic moment is proportional to spin, it follows that we can actually think of the specially-configured magnet as performing a measurement of the observable $\mathbf{S}_z$. As long as we are just doing theory, it is reasonable to model this as an ideal measurement in the sense that we can use the projection postulate for conditioning the state of each atom based on this measurement of $\mathbf{S}_z$. Hence we can assign each atom a spin state that is simply determined by which branch of the atomic beam it ends up in, in the manner stated above. (Note that it would be hard to do this perfectly in practice since the magnet would have some fringe fields, the atoms would be travelling at different velocities, *et cetera*.)

According to the schematic diagram, we dump the branch of the atomic beam in which the spins have state $|\Psi_{z-}\rangle$ and direct the atoms in state $|\Psi_{z+}\rangle$ towards the second magnet. If the second magnet has the same orientation as the first magnet, it will perform exactly the same splitting action. That is, it should re-measure $\mathbf{S}_z$ and divide up the beam according to the result. It should be clear from the basic rules we have already discussed that we should expect not to see any branching of the beam in this configuration.

If the orientation of the second magnet is perpendicular to that of the first, however, we need to do a little calculation. Let the second magnet be oriented along the x axis, so that it splits the beam according to a measurement of $\mathbf{S}_x$. We note that

$$\langle \Psi_{z+}|\mathbf{S}_x|\Psi_{z+}\rangle = \begin{pmatrix} 1 & 0 \end{pmatrix} \frac{\hbar}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0.$$

What does this mean? Getting a little more sophisticated, we note the spectral decomposition of $\mathbf{S}_x$:

$$\mathbf{P}_{x+} = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

$$\mathbf{P}_{x-} = \frac{1}{2} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \begin{pmatrix} 1 & -1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix},$$

$$\mathbf{S}_x = \frac{\hbar}{2}\mathbf{P}_{x+} - \frac{\hbar}{2}\mathbf{P}_{x-}.$$

It follows that the probabilities for each atom (starting in the state $|\Psi_{z+}\rangle$) to end up in either branch after the $\mathbf{S}_x$ measurement are equal:

$$\langle \Psi_{z+} | \mathbf{P}_{x+} | \Psi_{z+} \rangle = \begin{pmatrix} 1 & 0 \end{pmatrix} \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{2},$$

$$\langle \Psi_{z+} | \mathbf{P}_{x-} | \Psi_{z+} \rangle = \begin{pmatrix} 1 & 0 \end{pmatrix} \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{2}.$$

So in the end it is easy to see why the atomic beam is split in half if the second magnet is oriented along x. Hence we have intrinsic uncertainty in predicting the result of a measurement of the observable $\mathbf{S}_x$ for atoms prepared in the pure spin state $|\Psi_{z+}\rangle$.

Mixed ensembles of quantum states

Let's say we're working in a nice, simple two-dimensional Hilbert space and that we've chosen orthonormal basis kets $|x\rangle$ and $|y\rangle$. For the following discussion, we'll need to define two state vectors

$$|\Psi_A\rangle = a_x|x\rangle + a_y|y\rangle,$$
$$|\Psi_B\rangle = b_x|x\rangle + b_y|y\rangle.$$

Suppose I ask you to perform a series of N measurements on this system (here N is just a large integer), corresponding to the projectors

$$\mathbf{P}_x = |x\rangle\langle x|,$$
$$\mathbf{P}_y = |y\rangle\langle y|.$$

The trick is, in this series of measurements I will sometimes prepare the initial state $|\Psi_A\rangle$ (with probability p) and sometimes $|\Psi_B\rangle$ (with probability $1-p$). That is, I will be giving you a ***mixed*** ensemble of quantum states.

How shall we predict the overall number of times n_x we expect to obtain the measurement outcome x? According to simple probability theory,

$$n_x = N[\Pr(\Psi_A)\Pr(x|\Psi_A) + \Pr(\Psi_B)\Pr(x|\Psi_B)]$$
$$= N[p\langle\Psi_A|\mathbf{P}_x|\Psi_A\rangle + (1-p)\langle\Psi_B|\mathbf{P}_x|\Psi_B\rangle]$$
$$= N[p|a_x|^2 + (1-p)|b_x|^2].$$

Note that since $0 \le p \le 1$, the quantity n_x/N is bounded from below by the smaller of $|a_x|^2$ and $|b_x|^2$. In particular, if both of these quantities are nonzero than n_x must also be greater than zero.

A very different expression for n_x would be obtained if, instead of a mixed ensemble, I were to present you with a 'coherent superposition' of the states $|\Psi_A\rangle$ and $|\Psi_B\rangle$, corresponding to the 'pure' state $|\Psi(p_A,p_B)\rangle$

$$|\Psi(p_A,p_B)\rangle = p_A|\Psi_A\rangle + p_B|\Psi_B\rangle$$
$$= (p_Aa_x + p_Bb_x)|x\rangle + (p_Aa_y + p_Bb_y)|y\rangle.$$

(Note that we should be careful in choosing p_A and p_B such that $|\Psi(p_A,p_B)\rangle$ is normalized.) In this case, we would predict

$$n_x = N\langle \Psi(p_A,p_B) | \mathbf{P}_x | \Psi(p_A,p_B) \rangle$$
$$= N|p_A a_x + p_B b_x|^2.$$

Note that in certain cases, *e.g.* $p_A a_x = -p_B b_x \neq 0$, it is possible to choose p_A, p_B such that $n_x = 0 < |a_x|^2, |b_x|^2$ through the phenomenon of *destructive interference*. This is the truly important distinction between coherent superpositions (of the type that produce a single pure state) and incoherent admixtures (of the type that produce a mixed ensemble of quantum states).

To provide a really simple example, consider the following two states:

$$|\Psi_A\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle),$$
$$|\Psi_B\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle).$$

Clearly each of these states has $\langle \mathbf{P}_1 \rangle = \frac{1}{2}$, where $\mathbf{P}_1 \equiv |1\rangle\langle 1|$. Even in a situation where we are not sure which of these two quantum states has been prepared, we are sure that $\langle \mathbf{P}_1 \rangle$ should still be $\frac{1}{2}$. In other words, the two possible scenarios should not 'interfere' with one another when we make predictions about measurement results. Suppose we know that $|\Psi_A\rangle$ and $|\Psi_B\rangle$ are prepared with equal probability. Let's convince ourselves that this does not correspond to a coherent superposition of $|\Psi_A\rangle$ and $|\Psi_B\rangle$:

$$|\Psi_{coh}\rangle = \frac{1}{\sqrt{2}}(|\Psi_A\rangle + |\Psi_B\rangle)$$
$$= \frac{1}{2}(|0\rangle + |1\rangle + |0\rangle - |1\rangle)$$
$$= |0\rangle.$$

Coherent superpositions *do* exhibit interference, and this is why we can't use them to represent *extrinsic* uncertainty about the preparation of a quantum system.

If we want to combine the list of kets and probabilities in a mixed ensemble of quantum states in order to form a more compact representation, we turn to the density matrix. For the scenario we are considering,

$$\boldsymbol{\rho} = p|\Psi_A\rangle\langle\Psi_A| + (1-p)|\Psi_B\rangle\langle\Psi_B|.$$

We can conveniently represent ensemble-averaged quantities such as $\Pr(x)$ in terms of $\boldsymbol{\rho}$:

$$\Pr(x) = \mathrm{Tr}[\boldsymbol{\rho}\mathbf{P}_x]$$
$$= \mathrm{Tr}[\mathbf{P}_x\boldsymbol{\rho}].$$

To see that this is true we simply write out the trace,

$$\mathrm{Tr}[\boldsymbol{\rho}\mathbf{P}_x] = \langle x|\boldsymbol{\rho}\mathbf{P}_x|x\rangle + \langle y|\boldsymbol{\rho}\mathbf{P}_x|y\rangle$$
$$= \langle x|\boldsymbol{\rho}|x\rangle$$
$$= p\langle x||\Psi_A\rangle\langle\Psi_A||x\rangle + (1-p)\langle x||\Psi_B\rangle\langle\Psi_B||x\rangle$$
$$= p|a_x|^2 + (1-p)|b_x|^2.$$

In general, the probability of obtaining the outcome x in a measurement of $\{\mathbf{P}_x, \dots\}$ on

a mixed ensemble with density operator $\rho \leftrightarrow \{p_i, |\Psi_i\rangle\}$ is given by

$$\Pr(x) = \sum_i \Pr(i)\Pr(x|i)$$

$$= \sum_i p_i \langle \Psi_i | \mathbf{P}_x | \Psi_i \rangle$$

$$= \sum_{ij} p_i \langle \Psi_i | \mathbf{P}_x | j \rangle\langle j \| \Psi_i \rangle$$

$$= \sum_{ij} p_i \langle j \| \Psi_i \rangle\langle \Psi_i | \mathbf{P}_x | j \rangle$$

$$= \sum_j \langle j | \sum_i p_i | \Psi_i \rangle\langle \Psi_i | \mathbf{P}_x | j \rangle$$

$$= \mathrm{Tr}[\rho \mathbf{P}_x],$$

where

$$\rho \to \sum_i p_i |\Psi_i\rangle\langle\Psi_i|.$$

(Note that ρ is technically an operator on the Hilbert space, and has the form of a linear combination of projection operators. However, the ensemble representation used to define a density operator is not necessarily also a spectral decomposition, as the various $|\Psi_i\rangle$ that constitute an ensemble do not need to be mutually orthogonal.) For a two-member ensemble it is not entirely clear that ρ is any more convenient than working directly with a $\{p_i, |\Psi_i\rangle\}$. But what if we have, *e.g.*, a fifty-member ensemble? We can either compute fifty expectation values every time someone gives us a projection operator, or we can simply take the trace of that projector with the 2×2 density operator

$$\rho = \sum_{i=1}^{50} p_i |\Psi_i\rangle\langle\Psi_i|.$$

Surely ρ is a more compact representation of our knowledge about the system state! We need only compute it once, after all, to simplify an arbitrarily large number of statistical predictions. The same advantage applies when we want to evolve the system state dynamically according to the Schrödinger Equation (see below).

What about the post-measurement state? For pure quantum states the projection postulate reads

$$|\Psi_i\rangle \mapsto \frac{\mathbf{P}_x|\Psi_i\rangle}{\sqrt{\langle\Psi_i|\mathbf{P}_x|\Psi_i\rangle}},$$

and for density operators

$$\rho \mapsto \frac{\mathbf{P}_x\rho\mathbf{P}_x}{\mathrm{Tr}[\rho\mathbf{P}_x]}.$$

Once again, the role of the projection operator is to enforce consistency of the post-measurement state with the actual outcome of the measurement; the denominator ensures that the post-measurement state is still normalized (i.e.

$\mathrm{Tr}[\boldsymbol{\rho}] = 1$). Note that any of the states $|\Psi_i\rangle$ in the original list that are not consistent with the measurement result will be eliminated, and others will have their probabilities renormalized.

Many different ensembles have the same density operator. For example, both

$$p_1 = \frac{1}{2}, \quad |\Psi_1\rangle \leftrightarrow \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad p_2 = \frac{1}{2}, \quad |\Psi_2\rangle \leftrightarrow \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

and

$$p_1 = \frac{1}{2}, \quad |\Psi_1\rangle \leftrightarrow \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad p_2 = \frac{1}{2}, \quad |\Psi_2\rangle \leftrightarrow \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix},$$

have the same density operator

$$\boldsymbol{\rho} \leftrightarrow \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

What this tells us is that there is no measurement that can distinguish between two mixed ensembles of quantum states that have exactly the same density operator. In fact the density operator is a *sufficient statistic* for computing the probabilities of the outcomes for any POVM, because of the linear structure of the probability rule $\mathrm{Pr}(i) = \mathrm{Tr}[\boldsymbol{\rho} E_i]$. It likewise determines the expectation value for all observables and therefore specifies the *state* in the technical sense we discussed earlier in the term. This is not to say that there are not some situations in which it helps to know not only the ensemble density matrix but also the full list of states and probabilities - see for example C. A. Fuchs and C. M. Caves, "Ensemble-Dependent Bounds for Accessible Information in Quantum Mechanics," Phys. Rev. Lett. **73**, 3047 (1994).

Density matrix representation of mixed quantum states

At this point we should note that there are actually a number of fairly distinct reasons why density operators appear in quantum probability models, and mixed preparations such as the one we discussed above are only one of them. Another context in which density operators must be used is the representation of *thermal* states such as the spin state of the Ag atoms emerging from the oven in our Stern-Gerlach example from last week. Recall that the observables representing Cartesian components of the atomic spin are given by

$$\mathbf{S}_z = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \mathbf{S}_x = \frac{\hbar}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{S}_y = \frac{\hbar}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix},$$

and that one empirically finds that half of the atoms have eigenvalue $\pm\hbar/2$ for $\mathbf{S}_z$ after emerging from the oven. If we insist on representing the spin state of the atoms as a state vector, we must therefore choose

$$|\Psi\rangle \rightarrow \begin{pmatrix} u \\ l \end{pmatrix}, \quad |u|^2 = \langle \mathbf{P}_{z+} \rangle = |l|^2 = \langle \mathbf{P}_{z-} \rangle = \frac{1}{2}.$$

We can write any such vector as

$$|\Psi\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} \exp(i\phi_u) \\ \exp(i\phi_l) \end{pmatrix}, \quad \phi_u, \phi_l \in R,$$

with the consequence that

$$\langle \mathbf{S}_x \rangle = \frac{\hbar}{4} \begin{pmatrix} \exp(-i\phi_u) & \exp(-i\phi_l) \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \exp(i\phi_u) \\ \exp(i\phi_l) \end{pmatrix}$$

$$= \frac{\hbar}{4} \{ e^{i(\phi_l - \phi_u)} + e^{-i(\phi_l - \phi_u)} \} = \frac{\hbar}{2} \cos(\phi_l - \phi_u),$$

$$\langle \mathbf{S}_y \rangle = \frac{\hbar}{4} \begin{pmatrix} \exp(-i\phi_u) & \exp(-i\phi_l) \end{pmatrix} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \begin{pmatrix} \exp(i\phi_u) \\ \exp(i\phi_l) \end{pmatrix}$$

$$= \frac{\hbar}{4} \{ -ie^{i(\phi_l - \phi_u)} + ie^{-i(\phi_l - \phi_u)} \} = \frac{\hbar}{2} \sin(\phi_l - \phi_u).$$

However one can empirically determine $\langle \mathbf{S}_x \rangle = \langle \mathbf{S}_y \rangle = 0$ for atoms emerging from the oven (meaning that Stern-Gerlach-type measurements reveal 50/50 distributions for the $\pm\hbar/2$ eigenvalues of both $\mathbf{S}_x$ and $\mathbf{S}_y$), which cannot be reproduced by any possible assignment of $\phi_l - \phi_u$ in the state vector considered above!

We conclude that the thermal spin state cannot be represented by a state vector. Although we won't get to talk about it this term, basic considerations of quantum statistical mechanics lead naturally to representations of thermal states via density operators.

Generally speaking, valid density operators must be Hermitian and have the property

$$\mathrm{Tr}[\rho] = 1.$$

Here Tr denotes the 'trace' operation

$$\mathrm{Tr}[\rho] = \sum_k \langle k | \rho | k \rangle,$$

where $\{|k\rangle\}$ is **any** orthonormal basis for the Hilbert space – the numerical result is independent of choice of basis. In particular, since ρ is Hermitian we can chose to take the trace in its own eigenbasis, which makes it clear that

$$\mathrm{Tr}[\rho] = \sum_i \lambda_i^\rho,$$

where λ_i^ρ are the eigenvalues of ρ. Note that if ρ happens to be available in matrix form, we can further make use of the fact that the sum of the eigenvalues of a matrix is equal to the sum of its diagonal elements.

We can easily derive some of the important properties of the eigenvalues and trace of a density operator in the case where it happens to represent a given mixed

ensemble,

$$\rho = \sum_k p_k |\Psi_k\rangle\langle\Psi_k| = \sum_k p_k \mathbf{P}_k.$$

(Note however that the properties must hold for density matrices of arbitrary 'origin.')
Here $\mathbf{P}_k$ is the projector onto $|\Psi_k\rangle$, which itself is a normalized pure state that occurs in the ensemble with relative probability p_k. For any arbitrary normalized state $|\Phi\rangle$ in the entire Hilbert space,

$$\langle\Phi|\rho|\Phi\rangle = \sum_k p_k \langle\Phi|\mathbf{P}_k|\Phi\rangle.$$

Since all of the p_k and $\langle\Phi|\mathbf{P}_k|\Phi\rangle$ must be between 0 and 1, it follows that

$$0 \le \langle\Phi|\rho|\Phi\rangle \le 1$$

for absolutely any (normalized) choice of the state $|\Phi\rangle$. This includes, in particular, the eigenstates of ρ, for which we have the deduction

$$\rho|\psi_j\rangle = \lambda_j|\psi_j\rangle,$$

$$\langle\psi_j|\rho|\psi_j\rangle = \lambda_j,$$

$$0 \le \lambda_j \le 1.$$

If we furthermore consider the trace for a density matrix representing a mixed ensemble,

$$\mathrm{Tr}[\rho] = \sum_k \langle k|\rho|k\rangle = \sum_k \langle k|\left(\sum_j p_j|\Psi_j\rangle\langle\Psi_j|\right)|k\rangle$$

$$= \sum_k \sum_j p_j \langle k|\Psi_j\rangle\langle\Psi_j|k\rangle$$

$$= \sum_k \sum_j p_j \langle\Psi_j|k\rangle\langle k|\Psi_j\rangle$$

$$= \sum_j p_j \langle\Psi_j|\left(\sum_k |k\rangle\langle k|\right)|\Psi_j\rangle = \sum_j p_j \langle\Psi_j|\Psi_j\rangle = \sum_j p_j = 1.$$

Density operators can represent either pure states,

$$\rho = |\Psi\rangle\langle\Psi|,$$

or mixed states

$$\rho = \sum_i p_i|\Psi_i\rangle\langle\Psi_i|$$

where there is more than one $p_i > 0$. Note that in the former (pure state) case ρ is a true projection operator, so

$$\text{pure}: \quad \rho^2 = \rho.$$

In particular, $\mathrm{Tr}\,\rho^2 = 1$ for a pure state.

For a mixed state, however, we can use the spectral decomposition to show that $\mathrm{Tr}\,\rho^2 < 1$. We start by writing

12

$$\boldsymbol{\rho}^2 = \left(\sum_i \lambda_i^\rho \mathbf{P}_i^\rho \right)^2 = \sum_i (\lambda_i^\rho)^2 \mathbf{P}_i^\rho,$$

and note that since $\mathrm{Tr}\,\boldsymbol{\rho} = \sum_i \lambda_i^\rho = 1$, each of the λ_i^ρ must be strictly less than one for a mixed state. Hence the eigenvalues of $\boldsymbol{\rho}^2$, which are equal to the $(\lambda_i^\rho)^2$, must add up to *less than* one.

Cyclic property of the Trace operation

For any pair of operators $\mathbf{A}$ and $\mathbf{B}$,

$$\mathrm{Tr}[\mathbf{AB}] = \mathrm{Tr}[\mathbf{BA}].$$

To see this, let's pick a complete orthonormal basis for the relevant Hilbert space and write matrix representations:

$$\mathbf{A} = \sum_{ij} a_{ij}|i\rangle\langle j|, \quad \mathbf{B} = \sum_{ij} b_{ij}|i\rangle\langle j|.$$

By definition of the trace,

$$\mathrm{Tr}[\mathbf{AB}] = \sum_i \langle i|\mathbf{AB}|i\rangle,$$

but we can insert a 'closure relation'

$$\sum_j |j\rangle\langle j| = \mathbf{1}$$

in the middle to obtain

$$\mathrm{Tr}[\mathbf{AB}] = \sum_{ij} \langle i|\mathbf{A}|j\rangle\langle j|\mathbf{B}|i\rangle = \sum_{ij} a_{ij} b_{ji} = \sum_{ji} b_{ij} a_{ji} = \mathrm{Tr}[\mathbf{BA}].$$

Just for amusement, let's look at an alternative proof that holds when either $\mathbf{A}$ or $\mathbf{B}$ is Hermitian. Let's say that $\mathbf{A}$ is Hermitian, and therefore has real eigenvalues and a complete set of orthogonal eigenvectors:

$$\mathbf{A}|m\rangle = \alpha_m|m\rangle, \quad \langle m|\mathbf{A} = \alpha_m\langle m|.$$

We can therefore choose to take our trace in the eigenbasis of $\mathbf{A}$, so

$$\mathrm{Tr}[\mathbf{AB}] = \sum_m \langle m|\mathbf{AB}|m\rangle = \sum_m \alpha_m\langle m|\mathbf{B}|m\rangle = \sum_m \langle m|\mathbf{BA}|m\rangle = \mathrm{Tr}[\mathbf{BA}].$$

An analogous proof holds when $\mathbf{B}$ is Hermitian. Since the product of operators is still an operator,

$$\mathrm{Tr}[\mathbf{ABC\,D}] = \mathrm{Tr}[\mathbf{D\,ABC}],$$

$$\mathrm{Tr}[\mathbf{A\,BCD}] = \mathrm{Tr}[\mathbf{BCD\,A}],$$

and so on and so forth... Hence one says that the trace operation is cyclic.

The Schrödinger Equation

The physical dynamics of a quantum system are specified by a Hermitian operator $\mathbf{H}$, called the Hamiltonian. Just as the (initial) state vector is the mathematical representation of what we know about the preparation of a system, the Hamiltonian is the mathematical representation of what we know about its physical dynamics. Time-evolution of quantum states under such dynamics is then given by the Schrödinger Equation,

$$i\hbar \frac{d}{dt}|\Psi\rangle = \mathbf{H}|\Psi\rangle,$$

where $\hbar = h/2\pi$ and $h \simeq 6.6261 \times 10^{-34}$ [J sec] is Planck's constant. For finite-dimensional systems this is just a system of coupled linear ordinary differential equations. If the physical system is truly isolated (autonomous), then $\mathbf{H}$ must be constant and we may write the formal solution

$$|\Psi(t)\rangle = \exp\left[\frac{-i}{\hbar}\mathbf{H}t\right]|\Psi(0)\rangle.$$

(Something similar can be done even f the Hamiltonian is time-dependent, but one has to be careful - we'll return to this topic later.) In some cases it is actually possible analytically to compute the operator exponential, which is defined (as usual) via Taylor expansion:

$$\exp[i\alpha\mathbf{O}] = \mathbf{1} + i\alpha\mathbf{O} - \frac{\alpha^2}{2}\mathbf{O}^2 - i\frac{\alpha^3}{3!}\mathbf{O}^3 + \frac{\alpha^4}{4!}\mathbf{O}^4 + \cdots.$$

Here α is an arbitrary (real) scalar. Note that if $\mathbf{O}$ is a Hermitian operator,

$$(\exp[i\alpha\mathbf{O}])^\dagger = \mathbf{1} - i\alpha\mathbf{O} - \frac{\alpha^2}{2}\mathbf{O}^2 + i\frac{\alpha^3}{3!}\mathbf{O}^3 + \frac{\alpha^4}{4!}\mathbf{O}^4 + \cdots$$

and

$$\exp[i\alpha\mathbf{O}](\exp[i\alpha\mathbf{O}])^\dagger = (\exp[i\alpha\mathbf{O}])^\dagger \exp[i\alpha\mathbf{O}] = \mathbf{1}.$$

That is, $\exp[i\alpha\mathbf{O}]$ is a **_unitary_** operator.

In the case of the Schrödinger Equation, we write

$$\mathbf{T}(t_2, t_1) = \exp\left[\frac{-i}{\hbar}\mathbf{H}(t_2 - t_1)\right]$$

and refer to $\mathbf{T}(t_2, t_1)$ as the system's unitary "time development operator" from time t_1 to t_2. Note that

$$(\mathbf{T}(t_2, t_1))^{-1} = (\mathbf{T}(t_2, t_1))^\dagger \sim \mathbf{T}(t_1, t_2)$$

can be thought of as an operator that evolves a state backwards in time.

Recall that unitary operators may be thought of as the complex generalization of rotation operators in a real vector space. Hence **quantum evolution for an isolated system corresponds to a** "**rigid rotation**" **of the state space**. As a consequence, time evolution preserves the norms of individual state vectors, and preserves the inner product (angle) between arbitrary pairs of state vectors.

Note that by taking the Hermitian conjugate of the entire Schrödinger Equation, we get a time evolution equation for bras:

$$-i\hbar \frac{d}{dt}\langle\Psi| = \langle\Psi|\mathbf{H},$$

$$\langle\Psi(t_2)| = \langle\Psi(t_1)|\mathbf{T}(t_1, t_2).$$

Accordingly,

$$\langle\Psi_a(t_2)|\Psi_b(t_2)\rangle = \langle\Psi_a(t_1)|\mathbf{T}(t_1,t_2)\mathbf{T}(t_2,t_1)|\Psi_b(t_1)\rangle = \langle\Psi_a(t_1)|\Psi_b(t_1)\rangle,$$

as noted above.

For a density matrix representing an ensemble of pure states,

$$\boldsymbol{\rho}(t_1) = \sum_i p_i |\Psi_i(t_1)\rangle\langle\Psi_i(t_1)|,$$

we note that

$$|\Psi_i(t_2)\rangle\langle\Psi_i(t_2)| = \mathbf{T}(t_2,t_1)|\Psi_i(t_1)\rangle\langle\Psi_i(t_1)|\mathbf{T}(t_1,t_2),$$

$$\boldsymbol{\rho}(t_2) = \sum_i p_i \mathbf{T}(t_2,t_1)|\Psi_i(t_1)\rangle\langle\Psi_i(t_1)|\mathbf{T}(t_1,t_2)$$

$$= \mathbf{T}(t_2,t_1)\left(\sum_i p_i |\Psi_i(t_1)\rangle\langle\Psi_i(t_1)|\right)\mathbf{T}(t_1,t_2)$$

$$= \mathbf{T}(t_2,t_1)\boldsymbol{\rho}(t_1)\mathbf{T}(t_1,t_2).$$