8.06 Spring 2016 Lecture Notes

3. Entanglement, density matrices and decoherence

Aram W. Harrow

Last updated: May 18, 2016

Contents

1 Axioms of Quantum Mechanics

We begin with a (very) quick review of some concepts from 8.04 and 8.05.

1.1 One system

States are given by unit vectors $|\psi\rangle \in V$ for some vector space V.

Observables are Hermitian operators $\hat{A} \in \mathcal{L}(V)$.

Measurements

Suppose $\hat{A} = \sum_{i=1}^{d} \lambda_i |v_i\rangle\langle v_i|$ for $\{|v_1\rangle \ldots, |v_d\rangle\}$ an orthonormal basis of eigenvalues and (for simplicity) each λ_i distinct. If we measure observable $\hat{A}$ on state $|\psi\rangle$ then the outcomes are distributed according to $\Pr[\lambda_i] = |\langle\psi|v_i\rangle|^2$.

Time evolution is given by Schrödinger's equation: $i\hbar\frac{\partial}{\partial t}|\psi\rangle = H|\psi\rangle$ where H is the Hamiltonian.

Heisenberg picture We can instead evolve operators in time using

$$i\hbar\frac{\partial}{\partial t}\hat{A}_H = [\hat{A}_H, H_H]. \tag{1}$$

Time-independent solution

If the Hamiltonian does not change in time, then the time evolution operator for time t is the unitary operator $\mathcal{U} = e^{-\frac{iHt}{\hbar}}$. The state evolves according to $|\psi(t)\rangle = \mathcal{U}|\psi(0)\rangle$ in the Schrödinger picture or the operator evolves according to $\hat{A}_H(t) = \mathcal{U}^\dagger A_H(0)\mathcal{U}$ in the Heisenberg picture.

Systems are described by a pair (V, H).

1.2 Two systems

Let's see how things change when we have two quantum systems: (V_1, H_1) and (V_2, H_2).

States are given by unit vectors $|\psi\rangle \in V$ where

$$V = V_1 \otimes V_2 \equiv \text{span}\{|\psi_1\rangle \otimes |\psi_2\rangle : |\psi_1\rangle \in V_1, |\psi_2\rangle \in V_2\}.$$

A special case are the *product states* of the form $|\psi_1\rangle \otimes |\psi_2\rangle$. States that are not product are called *entangled.*

Observables are still Hermitian operators $\hat{A} \in \mathcal{L}(V)$. A general observable may involve interactions between the two systems. Local observables are of the form $\hat{A} \otimes I$, $I \otimes \hat{B}$ or more generally $\hat{A}\otimes I + I \otimes \hat{B}$, and correspond to properties that can be measured without interacting the two systems.

Measurements

The usual measurement rule still holds for collective measurements. But when only one system is measured, we need a way to explain what happens to the other system. Suppose we measure the first system using the orthonormal basis $\{|v_1\rangle, \ldots, |v_d\rangle\}$. (Equivalently, we measure an operator with distinct eigenvalues and with eigenvectors $|v_1\rangle, \ldots, |v_d\rangle$.) If the overall system is in state $|\psi\rangle$, then the first step is to write $|\psi\rangle$ as

$$|\psi\rangle = \sum_{i=1}^{d} \sqrt{p_i}|v_i\rangle \otimes |w_i\rangle,$$

for some unit vectors $|w_1\rangle, \ldots, |w_d\rangle$ (not necessarily orthogonal) and some $p_1, \ldots, p_d$ such that $p_i \geq 0$ and $\sum_{i=1}^{d} p_i = 1$.

Then the probability of outcome i is p_i and the residual state in this case is $|v_i\rangle \otimes |w_i\rangle$.

Time evolution is still given by Schrödinger's equation, but now the joint Hamiltonian of two non-interacting systems is

$$H = H_1 \otimes I + I \otimes H_2. \tag{2}$$

Interactions can add more terms, such as the $\frac{q_1 q_2}{|\vec{r}_1 - \vec{r}_2|}$ Coulomb interaction, which generally cannot be written in this way. Note that a Hamiltonian term of the form $\hat{A}_1 \otimes \hat{A}_2$ *does* represent an interaction; e.g. $\sigma_z \otimes \sigma_z$ has energy ±1 depending on whether the two spins have Z components pointing in the same or opposite directions.

Time-independent solution

For a Hamiltonian of the form in (2), the time evolution operator is

$$\mathcal{U} = e^{-\frac{iHt}{\hbar}} = e^{-\frac{iH_1 t}{\hbar}} \otimes e^{-\frac{iH_2 t}{\hbar}}.$$

You should convince your that this second equality is true. Of course if the Hamiltonian contains interactions then $\mathcal{U}$ will generally not be of this form.

These principles are actually profoundly different from anything we have seen before. For example, consider the number of degrees of freedom. One d-level system needs d complex numbers to describe (neglecting normalization and the overall phase ambiguity) but N d-level systems need d^N complex numbers to describe, instead of dN. This exponential extravagance is behind the power of quantum computers, which will be discussed briefly at the end of the course, if time permits. It also seemed intuitively wrong to many physicists in the early 20th century, most notably including Einstein. The objections of EPR [A. Einstein, B. Podelsky and N. Rosen, *Physical Review*, **47** 777–780 (1935)] led to Bell's theorem, which we saw in 8.05 and will review on pset 6. Here, though, we will consider a simpler problem.

1.3 The problem of partial measurement

Let us revisit the scenario where we measure part of an entangled state. Suppose that Alice and Bob each have a spin-1/2 particle in the singlet state

$$|\psi\rangle = \frac{|+\rangle \otimes |-\rangle - |-\rangle \otimes |+\rangle}{\sqrt{2}} \tag{3}$$

(The singlet is an arbitrary but nice choice. The argument would be essentially the same for any entangled state.) Imagine that Alice and Bob are far apart so that they should not be able to quickly send messages to one another.

Now suppose that Alice decides to measure her state in the $\{|+\rangle, |-\rangle\}$ basis. Using the above rules we find that the outcomes are as described in Table 1.

Alice's outcome	joint state	Bob's state
$\Pr[+] = \frac{1}{2}$	$\lvert+\rangle \otimes \lvert-\rangle$	$\lvert-\rangle$
$\Pr[-] = \frac{1}{2}$	$\lvert-\rangle \otimes \lvert+\rangle$	$\lvert+\rangle$

Table 1: Outcomes when Alice measures her half of the singlet state (3) in the $\{|+\rangle, |-\rangle\}$ basis.

What can we say about Bob's state after such a measurement? It is not a deterministic object, but rather an *ensemble* of states, each with an associated probability. For this we use the notation

$$\{(p_1, |\psi_1\rangle), \ldots, (p_m, |\psi_m\rangle)\} \tag{4}$$

to indicate that state $|\psi_i\rangle$ occurs with probability p_i. The numbers $p_1, \ldots, p_m$ should form a probability distribution, meaning that they are nonnegative reals that sum to one. The states $|\psi_i\rangle$ should be unit vectors but do not have to be orthogonal. In fact, the number m could be much larger than the dimension d, and could even be infinite; e.g. we could imagine a state with some coefficients that are given by a Gaussian distribution. We generally consider m to be finite because it keeps the notation simple and doesn't sacrifice any important generality.

In the example where Alice measures in the $\{|+\rangle, |-\rangle\}$ basis, Bob is left with the ensemble

$$\left\{ \left(\frac{1}{2}, |+\rangle\right), \left(\frac{1}{2}, |-\rangle\right) \right\}. \tag{5}$$

What if Alice chooses a different basis? Recall from 8.05 that if $\vec{n} \in \mathbb{R}^3$ is a unit vector then a spin-1/2 particle pointing in that direction has state

$$|\vec{n}\rangle \equiv |\mathbf{n}; +\rangle = \cos\frac{\theta}{2}|+\rangle + \sin\frac{\theta}{2}e^{i\phi}|-\rangle. \tag{6}$$

Here $(1, \theta, \phi)$ are the polar coordinates for $\vec{n}$; i.e. $n_x = \sin\theta\,\cos\phi$, $n_y = \sin\theta\,\sin\phi$, and $n_y = \cos\theta$. The notation $|\mathbf{n}; +\rangle$ was what we used in 8.05 and $|\vec{n}\rangle$ will be the notation used in 8.06 in contexts where it is clear that we are talking about spin states. The orthonormal basis $\{|\mathbf{n}; +\rangle, |\mathbf{n}; -\rangle\}$ in our new notation is denoted $\{|\vec{n}\rangle, |-\vec{n}\rangle\}$.

Suppose that Alice measures in the $\{|\vec{n}\rangle, |-\vec{n}\rangle\}$ basis. It can be shown (see 8.05 notes or Griffiths §12.2) that for any $\vec{n}$,

$$|\psi\rangle = \frac{|\vec{n}\rangle \otimes |-\vec{n}\rangle - |-\vec{n}\rangle \otimes |\vec{n}\rangle}{\sqrt{2}}. \tag{7}$$

Thus, for any choice of $\vec{n}$, the two outcomes are equally likely and in each case Bob is left with a spin pointing in the opposite direction, as described in Table 2.

Alice's outcome	joint state	Bob's state
$\Pr[\vec{n}] = \frac{1}{2}$	$\lvert\vec{n}\rangle \otimes \lvert-\vec{n}\rangle$	$\lvert-\vec{n}\rangle$
$\Pr[-\vec{n}] = \frac{1}{2}$	$\lvert-\vec{n}\rangle \otimes \lvert\vec{n}\rangle$	$\lvert\vec{n}\rangle$

Table 2: Outcomes when Alice measures her half of the singlet state (3) in the $\{|\vec{n}\rangle, |-\vec{n}\rangle\}$ basis.

This leaves Bob with the ensemble

$$\left\{ \left(\frac{1}{2}, |\vec{n}\rangle\right), \left(\frac{1}{2}, |-\vec{n}\rangle\right) \right\}. \tag{8}$$

Uh-oh! At this point, our elegant theories of quantum mechanics have run into a number of problems.

- **Theory isn't closed.** When we combine two systems with tensor product we get a new system, meaning a new vector space and a new Hamiltonian. It still fits the definition of a quantum system. But when we look at the state of a subsystem, we do not get a single quantum state, we get an ensemble. Thus, if we start with states being represented by unit vectors, we are inevitably forced into having to use *ensembles* of vectors instead.

- **Ensembles aren't unique.** Any choice of $\vec{n}$ will give Bob a different ensemble. We expect our physical theories to give us unique answers, but here we cannot uniquely determine which ensemble is the right one for Bob. Note that other choices of measurement can leave Bob with different ensembles as well; e.g. if Alice flips a coin and uses that to choose between two measurements settings, then Bob will have a distribution over four states, each occurring with probability 1/4.

- **Time travel?!** If Bob *could* distinguish between these different ensembles (including the case in which Alice does nothing and he still holds half of an entangled state), then Alice could instantaneously communicate to Bob with her choice of measurement basis (or perhaps her choice of whether to measure at all or not). According to special relativity, there is a different inertial frame in which this process looks like Alice sending a message backwards in time. This rapidly leads to trouble...

Fortunately density operators solve all three problems! As a bonus, they are far more elegant than ensembles.

2 Density operators

2.1 Introduction and definition

We would like to develop a theory of states that combines *randomness* and *quantum mechanics.* So it is worth reviewing how both randomness and quantum mechanics can be viewed as two different ways of generalizing classical states. For simplicity, consider a classical system which can be in d different states labelled $1, 2, \ldots d$. The quantum mechanical generalization of this would be to consider complex d-dimensional unit vectors while the probabilistic generalization would be nonnegative real d-dimensional vectors whose entries sum to one. These can be thought of as two incomparable generalizations of the classical picture. We are interested in considering both generalizations at once so that we consider state spaces that are both probabilistic and quantum. We summarize these different choices of state spaces in Table 3.

	classical	**quantum**
deterministic	$\{1, \ldots, d\}$	$\lvert\psi\rangle \in \mathbb{C}^d$ s.t. $\langle\psi\lvert\psi\rangle = 1$
probabilistic	$p_1, \ldots, p_d \geq 0$ s.t. $p_1 + \ldots + p_d = 1$	ensembles? density operators?

Table 3: Different theories yield different state spaces.

What do we put in the fourth box (probabilistic quantum) of Table 3? One possibility is to put ensembles of quantum states, as defined in (4). Besides the drawbacks mentioned in the previous section, these also have the flaw that of involving an unbounded number of degrees of freedom. For example, let's take a spin-1/2 particle (i.e. $d = 2$), so our quantum states are of the form $c_+|+\rangle + c_-|-\rangle$. Then one such probability distribution is $|+\rangle$ with probability 1/3, $|-\rangle$ with probability 1/2 and $\frac{|+\rangle - i|-\rangle}{\sqrt{2}}$ with probability 1/6. Another distribution is $\cos(\theta)|+\rangle + \sin(\theta)|-\rangle$ where θ is distributed according to a Gaussian with mean 0 and variance σ^2.

This works, but there are an infinite number of degrees of freedom, even if we start with a single lousy electron spin! Surely nature would not be so cruel.

Another drawback with this approach is that different distributions can give the same measurement statistics for all possible measurements. As a result, many of these infinite degrees of freedom turn out to be simply redundant.

To see how this works, suppose that we have a discrete distribution where state $|\psi_a\rangle$ occurs with probability p_a, for $a = 1, \ldots, m$. Consider an observable $\hat{A}$. The expectation of $\hat{A}$ with respect to this ensemble is:

$$\begin{aligned}
\sum_{a=1}^{m} p_a \langle\psi_a|\hat{A}|\psi_a\rangle &= \sum_{a=1}^{m} p_a \mathrm{tr}\left[\langle\psi_a|\hat{A}|\psi_a\rangle\right] && \text{since tr has no effect on } 1 \times 1 \text{ matrices} \\
&= \sum_{a=1}^{m} p_a \mathrm{tr}\left[\hat{A}|\psi_a\rangle\langle\psi_a|\right] && \text{cyclic property of the trace} \\
&= \mathrm{tr}\left[\hat{A} \underbrace{\sum_{a=1}^{m} p_a |\psi_a\rangle\langle\psi_a|}_{\text{density matrix } \rho}\right] && \text{linearity of the trace} \\
&= \mathrm{tr}[\hat{A}\rho] && \text{nice simple formula} \\
&= \langle \hat{A}, \rho \rangle && \text{alternate interpretation}
\end{aligned}$$

We see that the measurement statistics are a function of the distribution only via the *density operator* $\rho = \sum_{a=1}^{m} p_a |\psi_a\rangle\langle\psi_a|$. This has about d^2 degrees of freedom, by contrast with $O(d)$ degrees of freedom for known quantum states and with the ∞ degrees of freedom associated with ensembles. This already solves one of the problems of ensembles, namely their use of unlimited amounts of redundant information.

This argument used only discrete distributions over $\mathbb{C}^d$ but the extension to continuous distributions and/or infinite-dimensional states is straightforward.

Facts about traces: Let X be a matrix of dimension $m \times n$ and Y a matrix of dimension $n \times m$. In general these will not be square, but XY and YX both are, so their traces are well-defined. In fact, they are equal! A quick calculation shows

$$\mathrm{tr}[XY] = \sum_{i=1}^{m} \sum_{j=1}^{n} X_{i,j} Y_{j,i} = \mathrm{tr}[YX]. \tag{9}$$

This is called the *cyclic property of the trace* because it is often applied to traces of long strings of matrices. For example, we can repeatedly apply (9) (using curly braces to indicate which blocks of matrices we are calling X and Y) to obtain

$$\mathrm{tr}[\underbrace{ABC}_{X}\underbrace{D}_{Y}] = \mathrm{tr}[\underbrace{DAB}_{X}\underbrace{C}_{Y}] = \mathrm{tr}[\underbrace{CDA}_{X}\underbrace{B}_{Y}] = \mathrm{tr}[BCDA] \tag{10}$$

The trace can also be used to define an inner product on operators. Define

$$\langle X, Y\rangle \equiv \mathrm{tr}[X^\dagger Y] = \sum_{i,j} X^*_{i,j} Y_{i,j}. \tag{11}$$

From this last expression we see that $\langle X, Y\rangle$ is equivalent to turning X and Y into vectors in the natural way (just listing all the elements in order) and taking the conventional inner product between those vectors.

2.2 Examples

2.2.1 Pure states

If we know the state is $|\psi\rangle$, then the density matrix is $|\psi\rangle\langle\psi|$. Observe that there is no phase ambiguity ($|\psi\rangle \mapsto e^{i\phi}|\psi\rangle$ leaves the density matrix unchanged) and each $|\psi\rangle$ gives rise to a distinct density matrix. Such density matrices are called *pure states*, and sometimes this terminology is also used when talking about wavefunctions, to justify not using the density matrix formalism. By contrast, all other density matrices are called *mixed states*.

2.2.2 Spin-1/2 pure states

Let us consider the special case of pure states when $d = 2$, corresponding to a spin-1/2 particle. If $|\vec{n}\rangle = \cos(\frac{\theta}{2})|+\rangle + \sin(\frac{\theta}{2})e^{i\phi}|-\rangle$ then

$$\begin{aligned}
|\vec{n}\rangle\langle\vec{n}| &= \begin{pmatrix} \cos^2(\frac{\theta}{2}) & \cos(\frac{\theta}{2})\sin(\frac{\theta}{2})e^{-i\phi} \\ \cos(\frac{\theta}{2})\sin(\frac{\theta}{2})e^{i\phi} & \sin^2(\frac{\theta}{2}) \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} + \begin{pmatrix} \cos^2(\frac{\theta}{2}) - \frac{1}{2} & 0 \\ 0 & \sin^2(\frac{\theta}{2}) - \frac{1}{2} \end{pmatrix} + \cos\frac{\theta}{2}\sin\frac{\theta}{2}\begin{pmatrix} 0 & e^{-i\phi} \\ e^{i\phi} & 0 \end{pmatrix} \\
&= \frac{I}{2} + \frac{\cos(\theta)}{2}\sigma_z + \frac{\sin(\theta)}{2}(\cos(\phi)\sigma_x + \sin(\phi)\sigma_y) \\
&= \frac{I + \vec{n}\cdot\sigma}{2}
\end{aligned}$$

This result is beautiful enough to frame.

$$\boxed{|\vec{n}\rangle\langle\vec{n}| = \frac{I + \vec{n}\cdot\vec{\sigma}}{2}} \tag{12}$$

With this in hand we can return to the example of Alice measuring half of a singlet state. Whatever her choice of $\vec{n}$, Bob's density matrix is

$$\frac{1}{2}|\vec{n}\rangle\langle\vec{n}| + \frac{1}{2}|-\vec{n}\rangle\langle-\vec{n}| = \frac{1}{2}\quad\frac{I + \vec{n}\cdot\sigma}{2}\quad + \frac{1}{2}\quad\frac{I - \vec{n}\cdot\sigma}{2}\quad = \frac{I}{2}. \tag{13}$$

This rules out their earlier attempts at instantaneous signaling (and later we will prove this in more generality). Bob's density matrix fully determines the results of any measurement he makes, and it is independent of Alice's choice of $\vec{n}$.

2.2.3 The maximally mixed state

If $\{|v_1\rangle, \dots, |v_d\rangle\}$ are an orthonormal basis, and each occurs with probability $1/d$, then the resulting density matrix is

$$\rho = \frac{1}{d}\sum_{i=1}^{d} |v_i\rangle\langle v_i| = \frac{I}{d}, \tag{14}$$

independent of the choice of basis. This is called the *maximally mixed state.* The previous example was the $d = 2$ case of this: a 1/2 probability of spin-up and 1/2 probability of spin-down results in the same density matrix, *no matter which direction "up" refers to.*

The continuous distribution over all unit vectors in $\mathbb{C}^d$ also yields the same density matrix, although this is a harder calculation.

2.2.4 Multiple decompositions

Consider the distribution where $|+\rangle$ occurs with probability 2/3 and $|-\rangle$ with probability 1/3. The density matrix is

$$\frac{2}{3}|+\rangle\langle +| + \frac{1}{3}|-\rangle\langle -| = \begin{pmatrix} \frac{2}{3} & 0 \\ 0 & \frac{1}{3} \end{pmatrix}. \tag{15}$$

Now consider the distribution

$$\begin{aligned} |\psi_1\rangle &\equiv \sqrt{\frac{2}{3}}|+\rangle + \sqrt{\frac{1}{3}}|-\rangle \quad \text{with probability } 1/2 \\ |\psi_2\rangle &\equiv \sqrt{\frac{2}{3}}|+\rangle - \sqrt{\frac{1}{3}}|-\rangle \quad \text{with probability } 1/2 \end{aligned}$$

The density matrix is now

$$\frac{1}{2}|\psi_+\rangle\langle\psi_+| + \frac{1}{2}|\psi_-\rangle\langle\psi_-| = \frac{1}{2}\begin{pmatrix} \frac{2}{3} & \frac{\sqrt{2}}{3} \\ \frac{\sqrt{2}}{3} & \frac{1}{3} \end{pmatrix} + \frac{1}{2}\begin{pmatrix} \frac{2}{3} & -\frac{\sqrt{2}}{3} \\ -\frac{\sqrt{2}}{3} & \frac{1}{3} \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & 0 \\ 0 & \frac{1}{3} \end{pmatrix}. \tag{16}$$

One lesson is that we shouldn't take the probabilities p_a too seriously; i.e. they are not uniquely determined by the density matrix. Neither is the property of the states in the ensemble being orthogonal.

2.2.5 Thermal states

Introduce a Hamiltonian $H = {}_{i=1}^{d} E_i|i\rangle\langle i|$. We can think of this classically, as saying that state i has energy E_i. In this case, the Boltzmann distribution at temperature T is $p_i = e^{-\beta E_i}/Z$, where $Z = {}_{i=1}^{d} e^{-\beta E_i}$, $\beta = 1/k_B T$ and $k_B = 1.380688 \cdot 10^{-23} J/K$ is Boltzmann's constant. In the quantum setting, "state i" is replaced by $|i\rangle$. The resulting density matrix is

$$\rho = \sum_{i=1}^{d} p_i|i\rangle\langle i| = \frac{{}_{i=1}^{d} e^{-\beta E_i}|i\rangle\langle i|}{Z} = \frac{e^{-\beta H}}{Z} = \frac{e^{-\beta H}}{\mathrm{tr} e^{-\beta H}} \tag{17}$$

This is known as the *Gibbs state* or the *thermal state.* It describes the state of a quantum system at thermal equilibrium.

One specific example comes from NMR. Consider a proton spin in a magnetic field, say a 11.74 Tesla field in the $\hat{z}$ direction. At this field strength, the proton spin will experience the Hamiltonian $H = - \ \omega_0 \sigma_z$ where $\omega_0 \approx 500\,\mathrm{MHz}$. (In fact, if you buy a 11.74T superconducting magnet, the vendor will probably call it a "500 MHz" magnet for this reason. It could also reasonably be called a 500K magnet because of its price.) The thermal state is

$$\begin{aligned} \rho &= \frac{e^{-\beta H}}{\mathrm{tr} e^{-\beta H}} = \frac{e^{\beta \ \omega_0}|+\rangle\langle+| + e^{-\beta \ \omega_0}|-\rangle\langle-|}{e^{\beta \ \omega_0} + e^{-\beta \ \omega_0}} \\ &\approx \frac{(1+\beta \ \omega_0)|+\rangle\langle+| + (1-\beta \ \omega_0)|-\rangle\langle-|}{2} \\ &= \frac{I}{2} + \frac{\beta \ \omega_0}{2}\sigma_z. \end{aligned}$$

Let $T = 300\mathrm{K}$, which is close to room temperature. Then $1/\beta \approx 6.25\mathrm{THz}$ and so $\beta \ \omega_0 \approx 10^{-4}$. This is the "polarization" of the state. In an NMR experiment we can think of a $\frac{1}{2} + 10^{-4}$ fraction of spins aligning with the external field and a $\frac{1}{2} - 10^{-4}$ fraction of the spins anti-aligning with the external field at thermal equilibrium. This means that observables are effectively attenuated by a factor of (in this case) 10^{-4}.

3 The general rule for density operators

We know what makes a vector a valid probability distribution or a valid quantum wave-vector. What makes an operator a valid density operator? One answer is "ρ is a valid density operator if there exists $p_1, \ldots, p_m, |\psi_1\rangle, \ldots, |\psi_m\rangle$ such that $\rho = \ _a p_a |\psi_a\rangle\langle\psi_a|$, $p_1, \ldots, p_m \geq 0$, $p_1 + \ldots + p_m = 1$ and $\langle\psi_a| =\rangle 1$ for each i." This is somewhat unsatisfactory if we want to build a theory where the density operators are the fundamental objects.

Fortunately, there is a simple answer.

Theorem 1. *If a $d \times d$ matrix ρ is a density matrix for some ensemble of quantum states then*

1. $\mathrm{tr}\rho = 1$.
2. $\rho \succeq 0$.

Conversely, for any $d \times d$ matrix ρ satisfying these two conditions, there exists an ensemble $\{p_a, |\psi_a\rangle\}_{1\leq a\leq m}$ such that $\rho = \ _{a=1}^{m} p_a|\psi_a\rangle\langle\psi_a|$. Here m can be taken to be the rank of ρ.

The inequality $\rho \succeq 0$ means that ρ is *positive semidefinite*, which is defined to mean that $\langle\psi|\rho|\psi\rangle \geq 0$ for all $|\psi\rangle$. It is the matrix analogue of being nonnegative.

3.1 Positive semidefinite matrices

We say that a square matrix A is positive semidefinite if $\langle\psi|A|\psi\rangle \geq 0$ for all $|\psi\rangle$. Physically, A might be an observable that takes on only nonnegative values. Or it might be a density matrix. If furthermore A is Hermitian, then there are three equivalent ways to characterise the condition of being positive semidefinite.

Theorem 2. *If $A = A^\dagger$ then TFAE (the following are equivalent)*

1. *For all $|\psi\rangle$, $\langle\psi|A|\psi\rangle \geq 0$.*
2. *All eigenvalues of A are nonnegative.*

3. There exists a matrix B such that $A = B^\dagger B$. (This is called a Cholesky factorization.)

To get intuition for this last condition, observe that for 1×1 matrices, it is the statement that a real number $x \geq 0$ iff $x = z^* z$ for some complex z.

Proof of Theorem 2. Since A is Hermitian, we can write $A = {}^{d}_{i=1} \lambda_i |e_i\rangle\langle e_i|$ for some orthonormal basis $\{|e_1\rangle, \ldots, |e_d\rangle\}$ and some real $\lambda_1, \ldots, \lambda_d$.

$(1 \to 2)$: Take $|\psi\rangle = |e_i\rangle$. Then $0 \leq \langle\psi|A|\psi\rangle = \langle e_i|A|e_i\rangle = \lambda_i$.

$(2 \to 3)$: Let $B = {}^{d}_{i=1} \sqrt{\lambda_i} |e_i\rangle\langle e_i|$. As an aside, one can show that B satisfies $A = B^\dagger B$ if and only if $B = {}^{d}_{i=1} \sqrt{\lambda_i} |e_i\rangle\langle f_i|$ for some orthonormal basis $\{|f_1\rangle, \ldots, |f_d\rangle\}$. Sometimes we say that $B = \sqrt{A}$, by analogy to the scalar case.

$(3 \to 1)$: For any $|\psi\rangle$, let $|\varphi\rangle = B|\psi\rangle$. Then $\langle\psi|A|\psi\rangle = \langle\psi|B^\dagger B|\psi\rangle = \langle\varphi|\varphi\rangle \geq 0$. □

3.2 Proof of the density-matrix conditions

Here we prove Theorem 1. Start with an ensemble $\{p_i, |\psi_i\rangle\}_{1 \leq i \leq m}$. Define $\rho = {}^{m}_{i=1} p_i |\psi_i\rangle\langle\psi_i|$.

- Then $\rho^\dagger = {}^{m}_{i=1} p_i^* |\psi_i\rangle\langle\psi_i| = \rho$, since $p_i = p_i^*$. Thus ρ is Hermitian.
- Next, define $B = {}^{m}_{i=1} \sqrt{p_i} |\psi_i\rangle\langle i|$. Then $\rho = B^\dagger B$, implying that $\rho \succeq 0$.
- Finally, $\mathrm{tr}\rho = {}^{m}_{i=1} p_i \mathrm{tr}|\psi_i\rangle\langle\psi_i| = {}^{m}_{i=1} p_i = 1$.

To prove the other direction, suppose that $\mathrm{tr}\rho = 1$ and $\rho \succeq 0$. By Theorem 2, $\rho = {}^{d}_{i=1} \lambda_i |e_i\rangle\langle e_i|$ for $\{|e_1\rangle, \ldots, |e_d\rangle\}$ an orthonormal basis and each $\lambda_i \geq 0$. Additionally $\mathrm{tr}\rho = {}^{d}_{i=1} \lambda_i = 1$. Thus we can take $p_i = \lambda_i$ and now ρ is the density matrix corresponding to the ensemble $\{p_i, |e_i\rangle\}_{1 \leq i \leq d}$. If $\mathrm{rank}\,\rho < d$, then the sum only needs $\mathrm{rank}\,\rho$ terms.

3.3 Application to spin-1/2 particles: the Bloch ball

The geometry of the set of density matrices is unfortunately not quite as simple as the state spaces we have encountered previously. Pure quantum states form a ball (modulo the phase ambiguity) and probability distributions form a simplex. Intersections of planes (such as $\mathrm{tr}[\rho] = 1$) with the set of positive semidefinite matrices are called *spectrahedra*, and apart from the wonderful name, I will not explore their general properties here.

However, the case of $d = 2$ is indeed simple and elegant. Given a Hermitian 2×2 matrix A, when is it a valid density matrix? First, if it is Hermitian then we can express it as

$$A = \frac{a_0 I + a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3}{2}$$

for some real numbers a_0, a_1, a_2, a_3. The factor of 2 in the denominator is arbitrary, but we will see later that it simplifies things. If A is a density matrix, then $1 = \mathrm{tr}[A] = a_0$. Thus, $A = \frac{I + \vec{a}\cdot\vec{\sigma}}{2}$ with $\vec{a} \equiv (a_1, a_2, a_3)$. We saw in 8.05 that $\mathrm{eig}(\vec{a}\cdot\vec{\sigma}) = \pm|\vec{a}|$. Thus $\mathrm{eig}(A) = \frac{1 \pm |\vec{a}|}{2}$. A is psd iff these are both nonnegative, which is true iff $|\vec{a}| \leq 1$.

This proves that the set of two-dimensional density matrices is precisely equal to the set

$$\frac{I + \vec{a}\cdot\vec{\sigma}}{2} : |\vec{a}| \leq 1 \quad . \tag{18}$$

Geometrically this looks like the unit ball in $\mathbb{R}^3$. The pure states form the surface of the ball, corresponding to the case $|\vec{a}| = 1$. The maximally mixed state $I/2$ corresponds to $\vec{a} = 0$. In general, $|\vec{a}|$ can be thought of as the "purity" of a state.

This set is called the *Bloch ball.* The unit vectors at the surface are called the *Bloch sphere.* These have nothing to do with Bloch states or Bloch's theorem (which arise in the solution of periodic potentials) except for the name of the inventor.

Beware also that for $d > 2$, the set of density matrices is no longer a ball and there is no longer a canonical way to quantify "purity." However, notions of entropy do exist and are used in fields such as quantum statistical mechanics.

4 Dynamics of density matrices

4.1 Schrödinger equation

The Schrödinger equation states that

$$\frac{\partial}{\partial t}|\psi\rangle = -\frac{i}{}H|\psi\rangle \tag{19a}$$

$$\frac{\partial}{\partial t}\langle\psi| = \frac{i}{}\langle\psi|H \tag{19b}$$

$$\frac{\partial}{\partial t}|\psi\rangle\langle\psi| = -\frac{i}{}(H|\psi\rangle\langle\psi| - |\psi\rangle\langle\psi|H) = -\frac{i}{}[H, |\psi\rangle\langle\psi|] \tag{19c}$$

$$\frac{\partial}{\partial t}\sum_{a=1}^{m} p_a|\psi_a\rangle\langle\psi_a| = -\frac{i}{}[H, \sum_{a=1}^{m} p_a|\psi_a\rangle\langle\psi_a|] \tag{19d}$$

We conclude that the density matrix evolves in time according to

$$\boxed{i\ \frac{\partial}{\partial t}\rho = [H, \rho]}. \tag{20}$$

This is reminiscent of the Heisenberg equation of motion for operators, but with the opposite sign

$$i\ \frac{\partial}{\partial t}\hat{A}_H = [\hat{A}_H, H_H]. \tag{21}$$

One way to explain the different signs is that states and observables are dual to each other, in the sense that they appear in the expectation value as $\langle \hat{A}, \rho\rangle$.

Another way to talk about quantum dynamics is in terms of unitary transformations. If a system undergoes Hamiltonian evolution for a finite time then this evolution can be described by a unitary operator $\mathcal{U}$, so that state $|\psi\rangle$ gets mapped to $\mathcal{U}|\psi\rangle$. In this case $|\psi\rangle\langle\psi|$ is mapped to $\mathcal{U}|\psi\rangle\langle\psi|\mathcal{U}^\dagger$. By linearity, a general density matrix ρ is then mapped to $\mathcal{U}\rho\mathcal{U}^\dagger$.

4.2 Measurement

A similar argument shows that if we measure ρ in the orthonormal basis $\{|v_1\rangle, \ldots, |v_d\rangle\}$, then the probability of outcome j is $\langle v_j|\rho|v_j\rangle$ and the post-measurement state is $|v_j\rangle\langle v_j|$. The fastest way to see this is to consider the observable $|v_j\rangle\langle v_j|$ which has eigenvalue 1 (corresponding to obtaining outcome $|v_j\rangle$) and eigenvalue 0 repeated $d - 1$ times (corresponding to the orthogonal outcomes). Then we use the fact that $\langle\hat{A}\rangle = \mathrm{tr}[\hat{A}\rho]$ and set $\hat{A} = |v_j\rangle\langle v_j|$.

An alternate derivation is to decompose $\rho = \sum_{a=1}^{m} p_a|\psi_a\rangle\langle\psi_a|$. Then $\Pr[j|a] = |\langle v_j|\psi_a\rangle|^2$ and

$$\begin{aligned}
\Pr[j] &= \sum_{a=1}^{m} p_a \Pr[j|a] \\
&= \sum_{a=1}^{m} p_a|\langle v_j|\psi_a\rangle|^2 \\
&= \sum_{a=1}^{m} p_a\langle v_j|\psi_a\rangle\langle\psi_a|v_j\rangle \\
&= \langle v_j|\left(\sum_{a=1}^{m} p_a|\psi_a\rangle\langle\psi_a|\right)|v_j\rangle \\
&= \langle v_j|\rho|v_j\rangle
\end{aligned}$$

It should be reassuring that, even though we used the ensemble decomposition in this derivation, the final probability we obtained depends only on ρ.

What if we forget the measurement outcome, or never knew it (e.g. someone else measures the state while our back is turned)? Then ρ is mapped to

$$\sum_{j=1}^{d}\langle v_j|\rho|v_j\rangle|v_j\rangle\langle v_j| = \sum_{j=1}^{d}|v_j\rangle\langle v_j|\rho|v_j\rangle\langle v_j|. \tag{22}$$

Here it is important to note that density matrices, like probability distributions, represent not only objective states of the world but also subjective states; in other words, they describe our knowledge about a state. So subjective uncertainty (i.e. the state "really is" something definite but we don't know what it is) will have implications for the density matrix.

If we now write ρ from (22) as a matrix in the $|v_1\rangle, \ldots, |v_d\rangle$ basis, this looks like

$$\begin{pmatrix}
\rho_{1,1} & 0 & \ldots & 0 \\
0 & \rho_{2,2} & & \vdots \\
\vdots & & \ddots & 0 \\
0 & \ldots & 0 & \rho_{d,d}
\end{pmatrix}$$

Can we unify measurement and unitary evolution the way that we have unified the probabilistic and quantum pictures of states? For example, how should we model an atom in an excited state undergoing fluorescence? We will return to this topic later when we discuss *open quantum systems* and *quantum operations*. However, already we are equipped to handle the phenomenon of decoherence, which is the monster lurking in the closet of every quantum mechanical experiment.

4.3 Decoherence

Unitary operators correspond to reversible operations: if U is a valid unitary time evolution then so is $U^\dagger$. In terms of Hamiltonians, evolution according to $-H$ will reverse evolution according to H. But other quantum processes cause an irreversible loss of information. Irreversible quantum processes are generally called "decoherence." This somewhat imprecise term refers to the fact that this information loss is always associated with a loss of "coherence" and with quantum systems becoming more like classical systems. In what follows we will illustrate it via a series of examples, but will not give a general definition.

Let's warm up with the concept of a *mixture*. If state $|\psi_a\rangle$ occurs with probability p_a, then the density matrix is ${}_a\, p_a|\psi_a\rangle\langle\psi_a|$. But what if we have an ensemble of density matrices? e.g. $\{(p_1,\rho_1),\dots,(p_m,\rho_m)\}$ Then the "average" density matrix is

$$\rho = \sum_{a=1}^{m} p_a \rho_a. \tag{23}$$

We can use this to model *random unitary evolution*. Suppose that our state experiences a random Hamiltonian. Model this by saying that unitary $\mathcal{U}_a$ occurs with probability p_a for $a = 1,\dots,m$. This corresponds to the map

$$\rho \mapsto \sum_{a=1}^{m} p_a \mathcal{U}_a \rho \mathcal{U}_a^\dagger. \tag{24}$$

Let's see how this can explain how coherence is lost in simple quantum systems. Suppose we start with the density matrix

$$\rho = \begin{pmatrix} \rho_{+,+} & \rho_{+,-} \\ \rho_{-,+} & \rho_{-,-} \end{pmatrix}$$

and choose a random unitary to perform as follows: with probability $1-p$ we do nothing and with probability p we perform a unitary transformation equal to σ_z. This corresponds to the ensemble of unitary transformations $\{(1-p, I), (p, \sigma_z)\}$. The density matrix is then mapped to

$$\begin{aligned}
\rho' &\equiv (1-p) I \rho I^\dagger + p \sigma_z \rho \sigma_z^\dagger \\
&= (1-p)\begin{pmatrix} \rho_{+,+} & \rho_{+,-} \\ \rho_{-,+} & \rho_{-,-} \end{pmatrix} + p \begin{pmatrix} \rho_{+,+} & -\rho_{+,-} \\ -\rho_{-,+} & \rho_{-,-} \end{pmatrix} \\
&= \begin{pmatrix} \rho_{+,+} & (1-2p)\rho_{+,-} \\ (1-2p)\rho_{-,+} & \rho_{-,-} \end{pmatrix}
\end{aligned}$$

If $p = 0$ then this of course corresponds to doing nothing, and if $p = 1$, we simply have $\rho' = \sigma_z \rho \sigma_z$. In between we see that the diagonal terms remain the same, but the off-diagonal terms are reduced in absolute value. The diagonal terms correspond to the probability of outcomes we would observe if we measured in the $\hat{z}$ basis, and so it is not surprising that a $\hat{z}$ rotation would not affect these. However, the off-diagonal terms reduce just as we would expect for a vector that is averaged with a rotated version of itself. If $p = 1/2$, then the off-diagonal terms are completely eliminated, meaning that all polarization in the $\hat{x}$ and $\hat{y}$ directions has been eliminated. One way to see this is that the $\hat{x}$ and $\hat{y}$ polarization of $\sigma_z\rho\sigma_z$ is opposite to that of ρ. Thus averaging ρ and $\sigma_z\rho\sigma_z$ leaves zero polarization in the $\hat{x}$-$\hat{y}$ plane.

With a series of examples, I will illustrate that:

- Decoherence can be achieved in several ways that look different but have the same results.
- Decoherence destroys some quantum/wave-like effects, such as interference.
- This also involves the loss of information, often of phase information.

5 Examples of decoherence

5.1 Looking inside a Mach-Zehnder interferometer

This example is physically unrealistic (in one place) but makes the decoherence phenomenon clearest to see.

A Mach-Zehnder interferometer is depicted in Fig. 1.

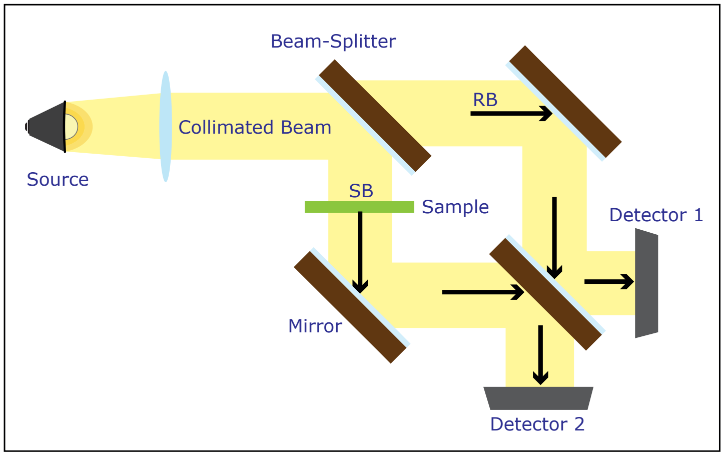

Image by MIT OpenCourseWare, adapted from the wikipedia article on Mach-Zehnder interferometers.

Figure 1: Mach-Zehnder interferometer. Image taken from the wikipedia article with this name.

At each point the photon can take one of two possible paths, which we denote by the states $|1\rangle$ and $|2\rangle$. Technically $|1\rangle$ means photon number in one mode and zero in the other modes, and similarly for $|2\rangle$. Also, we use $|1\rangle, |2\rangle$ to first denote the two inputs to the first beam splitter, then the two possible paths through the interferometer, and finally the two outputs of the second beam splitter leading to the detectors.

Each beam splitter can be modeled as a unitary operator. If they are “50-50” beam splitters, then this operator is

$$U_{\text{bs}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Thus, a photon entering in state $|1\rangle$ will go through the first beam splitter and be transformed into the state $\frac{|1\rangle+|2\rangle}{\sqrt{2}}$, corresponding to an even superposition of both paths. Assuming the paths have the same length and refractive index, it will have the same state when it reaches the second beam splitter. At this point the state will be mapped to

$$U_{\text{bs}} \frac{|1\rangle + |2\rangle}{\sqrt{2}} = \frac{|1\rangle + |2\rangle + |1\rangle - |2\rangle}{2} = |1\rangle$$

and the first detector will click with probability 1.

This is very different from what we’d observe if a particle entering a 50-50 beam splitter chose randomly which path to take. In that case, both detectors would click half the time.

The usual reason to build a Mach-Zehnder experiment, though, is not only to demonstrate the wave nature of light, but to measure something. Suppose we put some object in one of the paths so that light passing through it experiences a phase shift of θ. This corresponds to the unitary transformation

$$U_{\text{ph}} \equiv \begin{pmatrix} e^{i\theta} & 0 \\ 0 & 1 \end{pmatrix}. \tag{25}$$

Our modified experiment now corresponds to the sequence $U_{\text{bs}}U_{\text{ph}}U_{\text{bs}}$, which maps $|1\rangle$ to

$$U_{\text{bs}}U_{\text{ph}}U_{\text{bs}}|1\rangle = U_{\text{bs}}U_{\text{ph}}\frac{|1\rangle + |2\rangle}{\sqrt{2}} = U_{\text{bs}}\frac{e^{i\theta}|1\rangle + |2\rangle}{\sqrt{2}} = \frac{e^{i\theta}|1\rangle + e^{i\theta}|2\rangle + |1\rangle - |2\rangle}{2}.$$

The probability of the first detector clicking is now $|\frac{1+e^{i\theta}}{2}|^2 = \cos^2(\theta/2)$.

Now add decoherence. Suppose you find a way to look at which branch the photon is in without destroying the photon. (This part is a bit unrealistic, but if we use larger objects, then it becomes more reasonable. See the readings for a description of a two-slit experiment conducted with C_{60} molecules.) If we observe it then we will find that regardless of the phase shift θ:

- the photon is equally likely to be in each path; and
- each detector is equally likely to click.

Our measurement has caused decoherence that has destroyed the phase information in θ.

5.2 Spin rotations in NMR

Start with a spin-1/2 particle in the $|+\rangle$ state. Apply $H = S_y$ for time $t = \pi/2$, so that

$$U = e^{-\frac{iHt}{\hbar}} = e^{-i\frac{\pi}{4}\sigma_y} = \frac{I - i\sigma_y}{\sqrt{2}} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

Applying U once yields $U|+\rangle = \frac{|+\rangle+|-\rangle}{\sqrt{2}}$ and applying U a second time yields $|-\rangle$ (calculation omitted).

Suppose that we measure in the $\{|+\rangle, |-\rangle\}$ basis after applying the first U. Then each outcome occurs with probability 1/2 and the resulting density matrix is

$$\frac{1}{2}|+\rangle\langle+| + \frac{1}{2}|-\rangle\langle-| = \frac{I}{2}.$$

Applying U again leaves the density matrix unchanged. Decoherence has destroyed the polarization of the spin.

In actual NMR experiments, we have a test tube with 10^{20} water molecules at room temperature and we are not going to measure their individual spins. Instead, suppose that two nuclear spins get close to each other and interact briefly. Suppose that the first spin is in state ρ and the second spin is maximally mixed (i.e. density matrix $I/2$). Suppose that they interact for a time T according to the Hamiltonian

$$H = \frac{2}{}\lambda S_z \otimes S_z.$$

(Why not $\vec{S}^{(1)} \cdot \vec{S}^{(2)} \equiv \sum_{i=1}^{3} S_i \otimes S_i$? This is a consequence of perturbation theory: if there is a large $S_z \otimes I + I \otimes S_z$ term in the Hamiltonian, then the $S_x \otimes S_x$ and $S_y \otimes S_y$ terms are suppressed but the $S_z \otimes S_z$ term is not.) This is equivalent to the first spin experiencing a Hamiltonian λS_z if the second spin is in a $|+\rangle$ state. and experiencing $-\lambda S_z$ if the second spin is in a $|-\rangle$ state.

Averaging over these, the first spin is mapped to the state

$$\begin{aligned}
\rho' &= \frac{1}{2} e^{-\frac{i}{\hbar} t\lambda S_z} \rho e^{\frac{i}{\hbar} t\lambda S_z} + \frac{1}{2} e^{\frac{i}{\hbar} t\lambda S_z} \rho e^{-\frac{i}{\hbar} t\lambda S_z} \\
&= \frac{1}{2} \begin{pmatrix} e^{-i\lambda t/2} & 0 \\ 0 & e^{i\lambda t/2} \end{pmatrix} \rho \begin{pmatrix} e^{i\lambda t/2} & 0 \\ 0 & e^{-i\lambda t/2} \end{pmatrix} + \frac{1}{2} \begin{pmatrix} e^{i\lambda t/2} & 0 \\ 0 & e^{-i\lambda t/2} \end{pmatrix} \rho \begin{pmatrix} e^{-i\lambda t/2} & 0 \\ 0 & e^{i\lambda t/2} \end{pmatrix} \\
&= \begin{pmatrix} \rho_{++} & \cos(\lambda t/2)\rho_{+-} \\ \cos(\lambda t/2)\rho_{-+} & \rho_{--} \end{pmatrix}
\end{aligned}$$

This doesn't complete destroy the off-diagonal terms, but attenuates them. Here we should think of λt as usually small.

If we average over many such interactions, then this might (skipping many steps, which you will explore on the pset) result in a process that looks like

$$\dot{\rho} = -\frac{1}{T_2} \begin{pmatrix} 0 & \rho_{+-} \\ \rho_{-+} & 0 \end{pmatrix}, \tag{26}$$

where T_2 is the *decoherence time*, sometime also called the *dephasing time* for this kind of decoherence.

If this is T_2, then is there also a T_1? Yes, T_1 refers to a different kind of decoherence. In NMR, there is typically a static magnetic field in the $\hat{z}$ direction, which gives rise to a Hamiltonian of the form $H = -\gamma B S_z$. From this (together with the temperature) we obtain a thermal state ρ_{thermal} described in Section 2.2.5. The process of thermalization is challenging to rigorously derive from the Schrödinger equation but it is usually sufficient to model it phenomenologically. Suppose that according to a Poisson process with rate $1/T_1$, the spin is discarded and replaced with a fresh spin in the state ρ_{thermal}. Then we would obtain the differential equation

$$\dot{\rho} = -\frac{1}{T_1}(\rho - \rho_{\text{thermal}}). \tag{27}$$

Of course, there is another source of dynamics, which is the natural time evolution from the Schrödinger equation: $\dot{\rho} = -\frac{i}{\hbar}[H, \rho]$. Putting this together, we obtain the *Bloch equation*:

$$\dot{\rho} = -\frac{i}{\hbar}[H, \rho] - \frac{1}{T_1}(\rho - \rho_{\text{thermal}}) - \frac{1}{T_2} \begin{pmatrix} 0 & \rho_{+-} \\ \rho_{-+} & 0 \end{pmatrix}. \tag{28}$$

If we write $\rho = \frac{I + \vec{a} \cdot \vec{\sigma}}{2}$, then (28) becomes

$$\frac{\partial \vec{a}}{\partial t} = \hat{M}\vec{a} + \vec{b}, \tag{29}$$

for $\hat{M}, \vec{b}$ to be determined on a pset.

(Why do I keep talking about NMR, and not ESR (electron spin resonance)? The electron's gyromagnetic ratio is about 657 times higher than the proton's so its room-temperature polarization is larger by about this amount, and signals from it are easier to detect. However, it also interacts more promiscuously and thus often decoheres quickly, with T_2 on the order of microseconds or worse in most cases. So when you get a knee injury, your diagnosis will be made via your nuclei and not your electrons.)

5.3 Spontaneous emission

Consider an atom with states $|g\rangle$ and $|e\rangle$, corresponding to "ground" and "excited." We will also consider a photon mode, i.e. a harmonic oscillator. Suppose the initial state of the system is $|\psi\rangle_{\text{atom}} \otimes |0\rangle_{\text{photon}}$ with $|\psi\rangle = c_1|g\rangle + c_2|e\rangle$. These will interact via the Jaynes-Cummings Hamiltonian

$$H = \ \Omega(|g\rangle\langle e| \otimes \hat{a}^\dagger + |e\rangle\langle g| \otimes \hat{a}). \tag{30}$$

(For simplicity we have left out some terms that are usually in this Hamiltonian. This Hamiltonian can be derived using perturbation theory, as we discussed on a pset.) Suppose that the atom and photon field interact via this Hamiltonian for a time t. Assume that $\delta \equiv \Omega t$ is small and expand the state of the system in powers of δ:

$$e^{-\frac{iHt}{\hbar}}|\psi\rangle \otimes |0\rangle = (c_1|g\rangle + c_2|e\rangle) \otimes |0\rangle - i\delta c_2|g\rangle \otimes |1\rangle - \frac{\delta^2}{2}c_2|e\rangle \otimes |0\rangle + O(\delta^3). \tag{31}$$

Now measure and we see that with probability $|c_2|^2\delta^2$ the photon number is 1 and the atom is in the state $|g\rangle$. In this case, we observe an emitted photon and can conclude that the atom must currently be in the state $|g\rangle$. (It is tempting to conclude that we know it was previously in the state $|e\rangle$. This sort of reasoning about the past can be dangerous. In fact, all we can conclude is that c_2 must have been nonzero.)

With probability $|c_1|^2 + (1-\delta^2)|c_2|^2 = 1 - |c_2|^2\delta^2$ we observe 0 photons and the state is

$$\frac{c_1|g\rangle + (1-\delta^2)c_2|e\rangle}{\sqrt{1-|c_2|^2\delta^2}},$$

again, up to $O(\delta^3)$ corrections. If we repeat this for long enough then we *also* end up in the state $|g\rangle$. This is because if we watch an atom for a long time and it never emits a photon we can conclude that it's probably in the ground state.

6 Multipartite density matrices

In Section 1.3 I complained that pure-state quantum mechanics is not closed under discarding subsystems. Here we will see that density matrices solve this problem. More generally we will extend the formalism of density matrices to handle composite quantum systems.

6.1 Product states

Suppose that we have two systems, called A and B, with density matrices ρ and σ respectively. I claim that their joint state is $\rho \otimes \sigma$.

One way to see this is by explicitly decomposing $\rho = \quad {}_i\, p_i|\alpha_i\rangle\langle\alpha_i|, \sigma = \quad {}_j\, q_j|\beta_j\rangle\langle\beta_j|$ and considering these as *independent* ensembles $\{(p_i, |\alpha_i\rangle)\}$ and $\{(q_i, |\beta_i\rangle)\}$. According to the rule for independent probability distributions the probability of finding system A in state $|\alpha_i\rangle$ and system B in state $|\beta_j\rangle$ is $p_i \cdot q_j$. In this case the joint state is $|\alpha_i\rangle \otimes |\beta_j\rangle$. This corresponds to the ensemble $\{(p_i q_j, |\alpha_i\rangle \otimes |\beta_j\rangle)\}$ which has density matrix

$$\sum_{i,j} p_i q_j (|\alpha_i\rangle \otimes |\beta_j\rangle)(\langle\alpha_i| \otimes \langle\beta_j|) = \sum_{i,j} p_i q_j |\alpha_i\rangle\langle\alpha_i| \otimes |\beta_j\rangle\langle\beta_j| = \rho \otimes \sigma. \tag{32}$$

Therefore the product rule for density matrices can be inferred from the product rule for pure states.

We can also derive it from observables. If we measure observable $\hat{A}$ on the first system then this corresponds to the observable $\hat{A} \otimes I$ on the composite system; likewise $\hat{B}$ on the second system corresponds to $I \otimes \hat{B}$ on the joint system. Their product is $\hat{A} \otimes \hat{B}$. This arises for example when the dipole moments of two spins are coupled and the Hamiltonian gets a term proportional to $\vec{S}_1 \cdot \vec{S}_2 = S_x \otimes S_x + S_y \otimes S_y + S_z \otimes S_z$. Let ω be the joint state of a system where the first particle is in state ρ and the second is in state σ. The expectation of $\hat{A} \otimes \hat{B}$ with respect to ω should be

$$\mathrm{tr}[\rho \hat{A}]\mathrm{tr}[\sigma \hat{B}] = \mathrm{tr}[(\rho \otimes \sigma)(\hat{A} \otimes \hat{B})]. \tag{33}$$

Since this is equal to $\mathrm{tr}[\omega(\hat{A} \otimes \hat{B})]$ for all choices of $\hat{A}, \hat{B}$, we must have that $\omega = \rho \otimes \sigma$.

6.2 Partial measurement and partial trace

Density matrices were introduced by the fact that measuring one part of a larger system leaves the rest in a random state. As a result, pure-state quantum mechanics is not a closed theory; if the state of the joint system AB is pure, then it is possible that the states of A and B are not themselves pure. However, if there is a density matrix ρ^{AB} describing the joint state of systems A and B then we should be able to define density matrices for the individual systems. Indeed any observable X on system A should have a well-defined expectation value, and these can be used to define a *reduced density matrix* for system A. Mathematically we denote the density matrix of system A by ρ^A and define it to be the unique density matrix satisfying

$$\mathrm{tr}[\rho^A X] = \mathrm{tr}[\rho^{AB}(X \otimes I)], \tag{34}$$

for all observables X. (It is an instructive exercise to verify that there is always a solution to (34) and that it is unique.) Similarly we can define the state of system B to be ρ^B satisfying $\mathrm{tr}[\rho^B X] = \mathrm{tr}[\rho^{AB}(I \otimes X)]$.

Expanding (34) in terms of matrix elements yields

$$\sum_{a,a'} \rho^A_{a,a'} X_{a,a'} = \sum_{a,b,a',b'} \rho^{AB}_{ab,a'b'} X_{a,a'} \delta_{b,b'} = \sum_{a,a',b} \rho^{AB}_{ab,a'b} X_{a,a'}. \tag{35}$$

Since this must hold for any X, we have

$$\rho^A_{a,a'} = (\mathrm{tr}_B[\rho])_{a,a'} = \sum_b \rho^{AB}_{ab,a'b}. \tag{36}$$

This looks like taking a trace over the B subsystem (i.e. summing over the $b = b'$ entries) while leaving the A system alone. For this reason we call the map from $\rho^{AB} \mapsto \rho^A$ the "partial trace" and denote it tr_B; i.e. $\rho^A = \mathrm{tr}_B[\rho^{AB}]$. The partial trace is the quantum analogue of the rule for marginals of probability distributions: $p^X(x) = \quad {}_y\, p^{XY}(x, y)$.

A similar equation holds for $\rho^B \equiv \mathrm{tr}_A[\rho^{AB}]$ which can be expressed in terms of matrix elements as

$$\rho^B_{b,b'} = (\mathrm{tr}_A[\rho])_{b,b'} = \sum_a \rho^{AB}_{ab,ab'}. \tag{37}$$

If A and B have dimensions d_A and d_B respectively and M_d denotes the set of $d \times d$ matrices, then $\mathrm{tr}_A : M_{d_A d_B} \to M_{d_B}$ and $\mathrm{tr}_B : M_{d_A d_B} \to M_{d_A}$ are linear maps defined by

$$\begin{aligned}
\mathrm{tr}_A[|\alpha\rangle\langle\beta| \otimes |\gamma\rangle\langle\delta|] &= \langle\alpha|\beta\rangle \cdot |\gamma\rangle\langle\delta| \\
\mathrm{tr}_B[|\alpha\rangle\langle\beta| \otimes |\gamma\rangle\langle\delta|] &= \langle\gamma|\delta\rangle \cdot |\alpha\rangle\langle\beta|
\end{aligned}$$

If $\{|a\rangle\}$ and $\{|b\rangle\}$ are orthonormal bases then $\mathrm{tr}_A[|a\rangle\langle a'| \otimes |b\rangle\langle b'|] = \delta_{a,a'}|b\rangle\langle b'|$ and $\mathrm{tr}_B[|a\rangle\langle a'| \otimes |b\rangle\langle b'|] = \delta_{b,b'}|a\rangle\langle a'|$.

Let's illustrate this by revisiting the example of spontaneous emission from Section 5.3. Suppose

$$|\psi\rangle = e^{-iHt/} \quad \frac{|g\rangle + |e\rangle}{\sqrt{2}} \quad \otimes |0\rangle,$$

where $H = \ \Omega(|g\rangle\langle e| \otimes a^\dagger + |e\rangle\langle g| \otimes a)$. $H|g,0\rangle = 0$ and H acts on the $\{|e,0\rangle, |g,1\rangle\}$ subspace as a rotation. Thus

$$|\psi\rangle = \frac{1}{\sqrt{2}}|g,0\rangle + \frac{1}{\sqrt{2}}(\cos(\theta)|e,0\rangle - i\sin(\theta)|g,1\rangle).$$

The corresponding density matrix is

$$|\psi\rangle\langle\psi| = \begin{array}{c} \\ |g,0\rangle \\ |e,0\rangle \\ |g,1\rangle \\ |e,1\rangle \end{array} \!\!\begin{array}{c} \begin{array}{cccc} \langle g,0| & \langle e,0| & \langle g,1| & \langle e,1| \end{array} \\ \begin{pmatrix} \frac{1}{2} & \frac{1}{2}\cos(\theta) & \frac{i}{2}\sin(\theta) & 0 \\ \frac{1}{2}\cos(\theta) & \frac{1}{2}\cos^2(\theta) & \frac{i}{2}\cos(\theta)\sin(\theta) & 0 \\ \frac{-i}{2}\sin(\theta) & \frac{-i}{2}\sin(\theta)\cos(\theta) & \frac{1}{2}\sin^2(\theta) & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{array} \tag{38}$$

The reduced state of the atom is

$$\mathrm{tr}_{\text{photon}}|\psi\rangle\langle\psi| = \begin{array}{c|cc} & \langle g| & \langle e| \\ |g\rangle & \frac{1}{2} + \frac{1}{2}\sin^2(\theta) & \frac{1}{2}\cos(\theta) \\ |e\rangle & \frac{1}{2}\cos(\theta) & \frac{1}{2}\cos^2(\theta) \end{array} \tag{39}$$

and the reduced state of the photon is

$$\mathrm{tr}_{\text{atom}}|\psi\rangle\langle\psi| = \begin{array}{c|cc} & \langle 0| & \langle 1| \\ |0\rangle & \frac{1}{2} + \frac{1}{2}\cos^2(\theta) & \frac{i}{2}\sin(\theta) \\ |1\rangle & \frac{-i}{2}\sin(\theta) & \frac{1}{2}\sin^2(\theta) \end{array} . \tag{40}$$

(The decorations surrounding the above matrices are meant as reminders of which basis elements the rows and columns correspond to.)

6.3 Purifications

One way density matrices can arise is via subjective uncertainty; i.e. we don't know what the state is, but it "really" is pure. If so, we might imagine that density matrices would be useful for a quantum theory of statistics or information, but are not essential to quantum physics. However, density matrices also arise in settings where the overall state is known exactly. We saw this earlier where Bob could not distinguish his half of a singlet from a uniformly random state. Conversely, a uniformly random state cannot be distinguished from half of a singlet, with the other half in an unknown location. This is in fact only a representative example of the general rule that *any* density matrix could arise by being part of an entangled state.

First, let $|\psi\rangle = \ {}_{i=1}^{d_A} \ {}_{j=1}^{d_B} \alpha_{i,j}|i\rangle \otimes |j\rangle$. If Bob measures his system, he obtains outcome j with probability $p_j \equiv \ {}_i |\alpha_{i,j}|^2$ and the residual state for Alice is $\ {}_i \alpha_{i,j}|i\rangle/\sqrt{p_j}$. Her density matrix is

$$\sum_{j=1}^{d_B} p_j \frac{{}_{i=1}^{d_A} \alpha_{i,j}|i\rangle}{\sqrt{p_j}} \frac{{}_{i'=1}^{d_A} \alpha^*_{i',j}\langle i'|}{\sqrt{p_j}} = \sum_{i=1}^{d_A}\sum_{i'=1}^{d_A} \underbrace{\sum_{j=1}^{d_B} \alpha_{i,j}\alpha^*_{i',j}}_{=(\alpha\alpha^\dagger)_{i,i'}} |i\rangle\langle i'| = \alpha\alpha^\dagger.$$

What if Alice measures? Working this out is a good exercise. The answer is $(\alpha^\dagger\alpha)^T$.

By Theorem 2 any density matrix ρ can be written as $\alpha\alpha^\dagger$ for some matrix α. It remains only to check the normalization to ensure that $|\psi\rangle$ is a valid state:

$$1 = \mathrm{tr}\rho = \mathrm{tr}\alpha\alpha^\dagger = \sum_{i,j} |\alpha_{i,j}|^2.$$

This means that if we produce ρ in the lab, we can never know whether the state is mixed because of uncertainty about which pure state it is, or because it is entangled with a particle that is out of our control.

MIT OpenCourseWare
http://ocw.mit.edu

8.06 Quantum Physics III
Spring 2016

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.