PHY256 Lecture notes: Introducing Quantum Information

A. C. Quillen

April 30, 2021

Contents

1 The Density operator

The density operator encapsulates all that can be learned about a system through measurement.

1.1 Density operator for a pure state

The density operator for a pure state $|\psi\rangle = \sum_i c_i |i\rangle$ is a matrix (or operator)

$$\rho = |\psi\rangle \langle\psi|$$

$$\rho = \sum_{ij} c_i c_j^* |i\rangle \langle j|$$

Since ψ is normalized, (it has length 1; $\langle\psi|\psi\rangle = \sum_i c_i c_i^* = 1$),

$$\operatorname{tr} \rho = \sum_i c_i c_i^* = 1.$$

It is useful to know that the trace of a product

$$\operatorname{tr}(\mathbf{AB}) = \operatorname{tr}(\mathbf{BA}).$$

We show that this is true

$$\begin{aligned}
\mathbf{C} &= \mathbf{AB} \\
C_{ij} &= \sum_k A_{ik} B_{kj} \\
\operatorname{tr}\ \mathbf{C} &= \sum_i C_{ii} = \sum_{ik} A_{ik} B_{ki} \\
&= \sum_{ki} B_{ki} A_{ik} \\
&= \operatorname{tr}(\mathbf{BA}).
\end{aligned}$$

The trace of a matrix is independent of basis. A unitary transformation transforms from one orthonormal basis to another. In other words $\operatorname{tr} UAU^\dagger = \operatorname{tr} A$, where U is a unitary matrix that allows you to change basis from one orthonormal basis to another. This follows because we can change the order $\operatorname{tr} UAU^\dagger = \operatorname{tr} AU^\dagger U = \operatorname{tr} A$.

Eigenvalues are independent of basis and this means that

$$\operatorname{tr} \rho = \sum_i \lambda_i = 1$$

for a density matrix ρ, where λ_i are its eigenvalues.

We can always find an orthonormal basis that has our state vector as a basis vector. In this basis $|\psi\rangle = |0\rangle$ where $|0\rangle$ is that basis vector. In this basis the density matrix has a single 1 on its diagonal and the rest of the matrix contains zeros. That means that $\mathrm{tr}\,\rho^2 = 1$ and the density matrix has a single non-zero eigenvalue that is 1. This is only true for density matrices that are derived from a single pure state.

We can write the expectation of an observable as

$$\langle \mathbf{A} \rangle = \langle \psi | \, \mathbf{A} \, | \psi \rangle$$

Since observables are real, $\mathbf{A}$ should be Hermitian. Writing this out in a basis with $|\psi\rangle = \sum_i c_i \, |i\rangle$

$$\langle \mathbf{A} \rangle = \sum_{ij} c_i^* A_{ij} c_j$$

We can show that this is the same thing as

$$\langle \mathbf{A} \rangle = \mathrm{tr}(\rho \mathbf{A}) = \sum_{ij} c_i c_j^* A_{ji}$$

(Since A is Hermitian $A_{ij} = A_{ji}^*$).

While there is a complex phase ambiguity in a state vector, there is none in the density matrix.

1.2 The density operator for mixtures

We can consider a system which has probability of being in various pure states,

$$|\psi\rangle = \sum_i p_i \, |\psi_i\rangle$$

where p_i are probabilities with $\sum p_i = 1$. Call ρ_i the density matrix for each pure state.

$$\rho = \sum_i p_i \rho_i$$

and

$$\langle \mathbf{A} \rangle = \sum_i p_i \, \mathrm{tr}(\rho_i \mathbf{A}) = \sum_i \mathrm{tr}(p_i \rho_i \mathbf{A}) = \mathrm{tr}(\rho \mathbf{A})$$

The trace $\mathrm{tr}\,\rho = 1$, as before for a pure state. Using the Schwartz inequality one can show that $\mathrm{tr}\,\rho^2 \leq 1$. If $\mathrm{tr}\,\rho^2 = 1$ then the system is in a *pure* state.

For a non-pure state, the off-diagonal elements are called *coherences*. Because ρ is Hermitian one can always find a basis that makes it diagonal. Eigenvalues are real and positive (or zero). If the system is a pure state, then in that basis, the density matrix has a single 1 on the diagonal.

Summary.

- **Classical states** An object is either in one or not. We can have a linear vector space describing all the possible states $|a_i\rangle$.
- **Pure quantum states** These are complex superpositions of classical states; $|\psi\rangle = \sum c_i |a_i\rangle$. The probability p_i that an object is in a state $|a_i\rangle$ depends on the amplitude; $p_i = c_i c_i^*$.
- **Mixed states** Classical probability distribution over pure states.

$$\rho = \sum p_j |\psi_j\rangle \langle\psi_j|$$

1.2.1 Examples of density matrices for pure states and mixtures

Let's compute the density matrix of a pure state with $|\psi\rangle_0 = |0\rangle$. The density matrix is $\rho_0 = |0\rangle \langle 0|$. If the system is a two state system, in matrix form

$$\rho_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Now consider a composite state

$$|\psi\rangle_+ = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle).$$

The density matrix is

$$\begin{aligned} \rho_+ &= \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)\frac{1}{\sqrt{2}}(\langle 0| + \langle 1|) \\ &= \frac{1}{2}(|0\rangle \langle 0| + |1\rangle \langle 0| + |0\rangle \langle 1| + |1\rangle \langle 0|). \end{aligned}$$

In matrix form

$$\rho_+ = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}. \tag{1}$$

Now consider a mixture where the probability is 1/2 that the state is in $|0\rangle$ and 1/2 that the state is in $|1\rangle$. The density matrix is

$$\rho_2 = \frac{1}{2} |0\rangle \langle 0| + \frac{1}{2} |1\rangle \langle 1|.$$

In matrix form

$$\rho_2 = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}. \tag{2}$$

1.3 Transformations of the density matrix

If the wave-function $|\psi'\rangle = U|\psi\rangle$ where U is a unitary operater, then the density matrix transforms as

$$\rho' = U\rho U^\dagger$$

This is true even for mixed states.

If we have a measurement operator M_m the probability of getting value m is

$$p(m) = \mathrm{tr}\Big(M_m^\dagger M_m \rho\Big)$$

and the density matrix after measurement becomes

$$\rho'_m = \frac{M_m \rho M_m^\dagger}{\mathrm{tr}\Big(M_m^\dagger M_m \rho\Big)}$$

If we don't record the result of the measurement then the wavefunction is not collapsed and we sum over all the probabilities for each possible measurement

$$\begin{aligned}
\rho' &= \sum_m p(m)\rho'_m \\
\rho' &= \sum_m \mathrm{tr}\Big(M_m^\dagger M_p \rho\Big) \frac{M_m \rho M_m^\dagger}{\mathrm{tr}\Big(M_m^\dagger M_m \rho\Big)} \\
\rho' &= \sum_m M_m \rho M_m^\dagger. \qquad (3)
\end{aligned}$$

This type of transformation is usually not unitary. However it is deterministic as a measurement value is not chosen with a probability.

This gives two types of types of quantum simulations involving a series of measurements.

- We can start with a density matrix and transform it via each measurement, conserving probability as we go and using equation 3 each time a measurement is done. This method is deterministic.
- Or we can start with a wave-vector, and simulate the measurements by randomly choosing measured values consistent with the probability of measurement. After each measurement the wave function is collapsed. Each measurement requires a monte-Carlo simulated random event.

In the first case we evolve probability. In the second case we follow a path of possible measurements. An ensemble of paths should be consistent with the probability distribution predicted via the first method. In other words, if we carry out the second method a number

of times and average the density matrices of the results at the end, we should get the density matrix estimated from the first method.

The usefulness of the first of these methods partially explains why the density matrix is important. As we will discuss below, the density matrix can also be used to quantitative measure the amount of information available in a quantum system.

1.4 The density matrix for a product space

Suppose we have a two component system with quantum state in a product of two Hilbert spaces, $H_{AB} = H_A \otimes H_B$. An example is two qubits.

We can consider a partial measurement. For example, a measurement single of a single qubit when there are two.

For example, if we have a pure state

$$|\psi\rangle_{AB} = a\,|00\rangle + b\,|01\rangle + c\,|10\rangle + d\,|11\rangle \tag{4}$$

the probability that the first qubit has spin up would be $aa^* + bb^*$. If spin up is measured in the first qubit, the full state becomes

$$|\psi\rangle'_{AB} = \frac{a}{aa^* + bb^*}\,|00\rangle + \frac{b}{aa^* + bb^*}\,|01\rangle\,.$$

The probability that spin down is measured in the first qubit is $cc^* + dd^*$ and afterwards the full state becomes

$$|\psi\rangle'_{AB} = \frac{c}{cc^* + dd^*}\,|10\rangle + \frac{d}{cc^* + dd^*}\,|11\rangle\,.$$

Another way to describe probabilities for measurement is to use the density matrix. For a pure state $|\psi\rangle$

$$\rho_{AB} = |\psi\rangle_{AB}\,\langle\psi|_{AB}$$

where AB means we are taking the full system using both qubits. If the state is that of equation 4, the complex conjugate or bra state vector

$$\langle\psi|_{AB} = \langle 00|\,a^* + \langle 01|\,b^* + \langle 10|\,c^* + \langle 11|\,d^*.$$

The density matrix

$$\rho_{AB} = \begin{pmatrix} aa^* & ba^* & ca^* & da^* \\ ab^* & bb^* & cb^* & db^* \\ ac^* & bc^* & cc^* & dc^* \\ ad^* & bd^* & cd^* & dd^* \end{pmatrix}$$

More generally if pure state

$$|\psi\rangle_{AB} = \sum_{ij} a_{ij} |i\rangle_A |j\rangle_B = \sum_{ij} a_{ij} |ij\rangle$$
$$\langle\psi|_{AB} = \sum_{kl} \langle k|_A \langle l|_B a^*_{kl} = \sum_{kl} \langle kl| a^*_{kl}$$
$$\rho_{AB} = \sum_{ijkl} a_{ij} a^*_{kl} |i\rangle_A |j\rangle_B \langle k|_A \langle l|_B = \sum_{ijkl} a_{ij} a^*_{kl} |ij\rangle \langle kl| .$$

We can also consider mixtures, giving density matrices in the H_{AB} product space. For a mixture

$$\rho_{AB} = \sum_{ijkl} \rho_{ijkl} |ij\rangle \langle kl|$$

where ρ_{AB} is Hermitian, positive definite and has trace 1. The condition that the trace is 1 is

$$\operatorname{tr} \rho_{AB} = \sum_{ij} \rho_{ijij} = 1$$

where we need to sum over both indices.

1.5 The partial traces – tracing out one part of a product space

The density matrix should let us find out about all possible measurements of a system. If we have a product space of two subsystems, an experimenter may have access to to one subsystem. How can we describe the density matrix for that subsystem alone?

Consider a density matrix for a product space of two qubits

$$\rho_{AB} = \sum_{ijkl} \rho_{ijkl} |ij\rangle \langle kl| = \sum_{ijkl} \rho_{ijkl} |i\rangle_A |j\rangle_B \langle k|_A \langle l|_B .$$

We can define a density matrix for the first qubit ρ_A with

$$\begin{aligned}
\rho_A &= \operatorname{tr}_B(\rho_{AB}) \\
&= \sum_{ik} c_{ik} |i\rangle_A \langle k|_A = \sum_{ik} \left(\sum_j \rho_{ijkj} \right) |i\rangle_A \langle k|_A \\
c_{ik} &= \sum_j \rho_{ijkj}.
\end{aligned}$$

Here tr_B means that we are taking the trace of values in the second Hilbert space and $|i\rangle$ and $\langle k|$ are states in the first quantum system. Likewise we can take the trace over values

in the first system giving

$$\begin{aligned}\rho_B &= \mathrm{tr}_A(\rho_{AB}) \\ &= \sum_{jl} c'_{jl} \left|j\right\rangle_B \left\langle l\right|_B = \sum_{jl} \left(\sum_i \rho_{ijil} \right) \left|j\right\rangle_B \left\langle l\right|_B \\ c'_{jl} &= \sum_i \rho_{ijil}.\end{aligned}$$

The matrices ρ_A, ρ_B are positive definite, Hermitian and have trace 1. They are density matrices or operators for the sub-systems.

With a two qubit system the density matrices ρ_A, ρ_B are 2x2 matrices.

Consider ρ_{AB}, the density matrix of a pure state, with $\mathrm{tr}\,\rho_{AB}^2 = 1$ and eigenvalues 1,0. The partially traced matrices $\rho_A = \mathrm{tr}_B\,\rho_{AB}$ and $\rho_B = \mathrm{tr}_A\,\rho_{AB}$ need not satisfy these relations. In other words $\mathrm{tr}\,\rho_A^2$ may not be 1, and $\mathrm{tr}\,\rho_B^2$ may not be 1. Their eigenvalues need not be 1 or 0.

The von Neumann entropy of a pure state is 0. All information is known. However, not all information may be available about the pure state if one can only measure quantities in one of the subsystems, so the von Neumann entropy of ρ_A and ρ_B can be greater than 0.

1.5.1 Examples of density matrices for bipartite qubit systems

Consider a pure state

$$\left|\psi\right\rangle_{AB} = \left|0\right\rangle_A \otimes \left|0\right\rangle_B = \left|00\right\rangle.$$

The density matrices are

$$\begin{aligned}\rho_{AB} &= \left|00\right\rangle \left\langle 00\right| \\ \rho_A &= \left|0\right\rangle_A \left\langle 0\right|_A \\ \rho_B &= \left|0\right\rangle_B \left\langle 0\right|_B.\end{aligned}$$

The eigenvalues of all three density matrices are 1,0.

Let's consider again a pure state that is a product state

$$\left|\psi\right\rangle_{AB} = \frac{1}{\sqrt{2}}(\left|0\right\rangle_A + \left|1\right\rangle_A) \otimes \left|0\right\rangle_B = \frac{1}{\sqrt{2}}(\left|00\right\rangle + \left|10\right\rangle).$$

The density matrices are

$$\begin{aligned}
\rho_{AB} &= \frac{1}{2}(|00\rangle + |10\rangle)(\langle 00| + \langle 10|) \\
&= \frac{1}{2}(|00\rangle\langle 00| + |10\rangle\langle 00| + |00\rangle\langle 10| + |10\rangle\langle 10|) \\
\rho_A &= \frac{1}{2}(|0\rangle_A\langle 0|_A + |0\rangle_A\langle 1|_A + |1\rangle_A\langle 0|_A + |1\rangle_A\langle 1|_A) \\
\rho_B &= |0\rangle_B\langle 0|_B .
\end{aligned}$$

These are consistent with density matrices formed from $\frac{1}{\sqrt{2}}(|0\rangle_A + |1\rangle_A)$ and $|0\rangle_B$ alone, which is expected since the wave vector was a product state! The eigenvalues of all three density matrices are again 1,0.

Now let's consider a pure but entangled state, the Bell pair,

$$|\psi\rangle_{\text{Bell}} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle).$$

The density matrices are

$$\begin{aligned}
\rho_{AB} &= \frac{1}{2}(|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|) \\
\rho_A &= \frac{1}{2}(|0\rangle_A\langle 0|_A + |1\rangle_A\langle 1|_A) \\
\rho_B &= \frac{1}{2}(|0\rangle_B\langle 0|_B + |1\rangle_B\langle 1|_B). \qquad (5)
\end{aligned}$$

The eigenvalues of ρ_{AB} are 1,0 because it is derived from a pure state. The eigenvalues of ρ_A, ρ_B are 1/2,1/2. Notice that ρ_A, ρ_B resemble a density matrix of a mixture for a single qubit (equation 2) even though $|\psi\rangle_{\text{Bell}}$ is a pure state!

If Alice is separate from Bob then she would measure $|0\rangle$ half the time and $|1\rangle$ half the time. However, Alice and Bob have strongly correlated measurements.

Notice that ρ_A is not the same as that for ρ_+ (see equation 1) which comes from a superposition state $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$.

1.6 Positivity of the density matrix

A density operator ρ

- is Hermitian ($\rho^\dagger = \rho$).
- has trace 1 ($\text{tr}\rho = 1$).
- is **positive**.

The density matrix has to be Hermitian as we could construct hypothetical measurements using projection operators in any orthonormal basis and these should give real values. The operator has to have trace 1 because it is comprised of a weighted sum of density matrices formed from pure states and these have trace 1.

Positivity: What does it mean for an operator or matrix to be positive? If ρ is positive, then for any wave vector $|v\rangle$, the expectation

$$\langle v| \rho |v\rangle \geq 0.$$

Positivity also implies that means the eigenvalues of ρ must be real and non-negative.

Why should the density matrix be positive? For a pure state, we can find a basis that gives a diagonal density matrix that has a single 1 on the diagonal. This matrix is positive. A weighted sum of positive matrices where all the weights are positive should also be positive. So density matrices that are derived from mixtures should also be positive.

1.7 Acronyms and definitions!

- Positive operator value measurement **POVM**. A generalization of measurement that describes measurements with a set of complete, positive, but not necessarily orthogonal matrices. Superoperators or quantum channels can be decomposed into such a set of operators.
- Local operations and classical communication **LOCC**. Local operations include unitary transformations and measurements on portions of a quantum system. An unentangled state cannot be converted to an entangled one using only LOCC. Two states can be considered to be **LOCC equivalent** if one can be converted into the other using LOCC.
- **TPCP** trace-preserving (TP) completely positive (CP) map.
- A **quantum channel** is a communication channel which can transmit quantum information. A **quantum channel** is a completely positive (CP) trace-preserving (TP) map between spaces of operators.
- A **superoperator** is a linear operator acting on a vector space of linear operators. In quantum computing it is a completely positive map which preserves the trace of its argument.

In quantum computing and quantum information theory I think people tend to use use quantum channel and superoperator interchangeably.

1.8 The no communication theorem

- A density matrix encodes all and only what is physically observable. This is also true for the reduced or traced density matrix of a subsystem.
- **No communication theorem**. If Alice and Bob share an entangled state (where each of them has a piece of tensor product space), nothing Alice chooses to do (applying unitary transformations or carrying out measurements) will have an effect on Bob's density matrix.

Let's check that this is true for the Bell pair. Suppose that Alice performs a Pauli-X on her qubit. The wave vector becomes

$$\begin{aligned} |\psi\rangle' &= \sigma_x \otimes \mathbf{I} \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \\ &= \frac{1}{\sqrt{2}}(|10\rangle + |01\rangle). \end{aligned}$$

The density matrices become

$$\begin{aligned} \rho'_{AB} &= \frac{1}{2}(|10\rangle\langle 10| + |10\rangle\langle 01| + |01\rangle\langle 10| + |01\rangle\langle 01|) \\ \rho'_A &= \frac{1}{2}(|0\rangle_A\langle 0|_A + |1\rangle_A\langle 1|)_A \\ \rho'_B &= \frac{1}{2}(|0\rangle_B\langle 0|_B + |1\rangle_B\langle 1|_B). \end{aligned} \tag{6}$$

Bob's traced density matrices is not changed (see that listed in equation 5).

Suppose Alice makes a measurement of her qubit. She has a probability of 1/2 to measure spin up. If this happens the wave vector becomes $|00\rangle$ and Bob's density matrix would become $\rho_B = |0\rangle_B\langle 0|_B$. Alice has a probability of 1/2 to measure spin down and the wave vector would become $|11\rangle$. Because there is 1/2 probability for these outcomes Bob's density matrix is

$$\rho_B = \frac{1}{2}|0\rangle_B\langle 0|_B + \frac{1}{2}|1\rangle_B\langle 1|_B$$

which is equivalent to that given in equation 5.

1.9 Schmidt decomposition of a bipartite pure state

We consider a product space of Hilbert spaces $H_A \otimes H_B$ and a wave vector ψ in $H_A \otimes H_B$. There is always an orthonormal basis with basis vectors $|i\rangle_A$ in H_A and an orthonormal basis with basis vectors $|i\rangle_B$ in H_B and a list of non-negative real numbers p_i such that

$$\psi = \sum_i \sqrt{p_i}\,|i\rangle_A \otimes |i\rangle_B\,. \tag{7}$$

This is known as the Schmidt decomposition for a pure state.

We can always diagonalize $\rho_A = \mathrm{tr}_B \, \rho_{AB}$ where $\rho_{AB} = |\psi\rangle \langle\psi|$. We chose $|i\rangle_A$ to be a basis in which ρ_A is diagonal so

$$\rho_A = \sum_i p_i \, |i\rangle_A \, \langle i|_A$$

and p_i are its non negative eigenvalues. In an orthogonal basis for H_B described with basis vectors $\left|\hat{j}\right\rangle_B$

$$\psi = \sum_{ij} \psi_{ij} \, |i\rangle_A \otimes \left|\hat{j}\right\rangle_B = \sum_i |i\rangle_A \otimes \sum_j \psi_{ij} \left|\hat{j}\right\rangle_B .$$

This gives

$$\rho_{AB} = \left[\sum_i |i\rangle_A \otimes (\sum_j \psi_{ij} \left|\hat{j}\right\rangle_B) \right] \left[\sum_l \langle l|_A \otimes (\sum_k \psi^*_{lk} \left\langle\hat{k}\right|_B) \right]$$

Let's take the trace of this over H_B

$$\rho_A = \sum_{il} |i\rangle_A \, \langle l|_A \sum_j \psi_{ij} \psi^*_{lj}$$

We assumed that ρ_A is diagonal, and has eigenvalues p_i so

$$\sum_j \psi_{ij} \psi^*_{lj} = p_i \delta_{il} \tag{8}$$

This is a dot product.

We can define a series of orthonormal vectors in H_B

$$|i\rangle_B \equiv \sum_j \frac{\psi_{ij}}{\sqrt{p_i}} \left|\hat{j}\right\rangle_B$$

These vectors are orthonormal because of equation 8. Using them we can write

$$\psi = \sum_i \sqrt{p_i} \, |i\rangle_A \otimes |i\rangle_B ,$$

as in equation 7.

Among the consequences of the Schmidt decomposition (for a pure state) is that the non-zero eigenvalues of the partially traced density matrices ρ_A and ρ_B must be the same. If the dimensions of the Hilbert spaces H_A and H_B differ, then the partially traced density matrices may have a different number of zero eigenvalues. If there are degenerate eigenvalues then it is more work to find the Schmidt decomposition (and there is an ambiguity in it).

The Schmidt decomposition gives us the eigenvalues p_i of the partially traced density matrices ρ_A, ρ_B. The full density matrix ρ_{AB} has eigenvalues 1,0 because it is derived from a pure state. The von Neumann entropy is 0.

The density matrices ρ_A and ρ_B can have eigenvalues other than 1,0. The subsystems would have non-zero entropy (and less information!). While the full system is fully specified (and everything is known about it), this is not true in each of the subsystems.

Question: Any pure wave vector in a bipartite system can be decomposed via Schmidt decomposition. Can this be done for a system comprised of a tripartite system, with three subsystems, $H_{ABC} = H_A \otimes H_B \otimes H_C$? In other words is it possible to find orthonormal bases in H_A, H_B and H_C such that

$$|\psi\rangle_{ABC} = \sum_i \sqrt{p_i} \, |i\rangle_A \, |i\rangle_B \, |i\rangle_C$$

for any $|\psi\rangle_{ABC}$?

Answer: No, Schmidt decomposition is not possible for *any* wave vector in a tripartite system but is possible for some wave vectors in a tripartite system.

1.10 Entanglement and the Schmidt number

We consider a pure state $|\psi\rangle_{AB}$ in bipartite system $H_A \otimes H_B$. The **Schmidt number** of a bipartite system is the number of non-zero eigenvalues of ρ_A or ρ_B.

When can $|\psi\rangle_{AB}$ can be written as a product state? In other words when can we write

$$|\psi\rangle_{AB} = |\phi\rangle_A \otimes |\xi\rangle_B$$

where $|\phi\rangle_A$ is a wave vector in H_A and $|\xi\rangle_B$ is a wave vector in H_B?

Suppose we can write $|\psi\rangle$ as a product state. Then $\rho_A = |\phi\rangle_A \langle\phi|_A$ and $\rho_B = |\xi\rangle_B \langle\xi|_B$. The traced density matrices are those of pure states and the Schmidt number is 1.

If $|\psi\rangle_{AB}$ can be written as a product state then any unitary transformation in the form $\mathbf{U} \otimes \mathbf{I}$ or $\mathbf{I} \otimes \mathbf{U}$ would give another product state. These are **local** transformations as they only involve a transformation of one of the subsystems. Similarly a partial measurement in one of the subsystems would give another product state. Local transformations (either unitary or measurements) do not change the Schmidt number. These types of transformations are in the class of **Local Operations and Classical Communication** or LOCC.

1.11 Generalized Measurements: A positive operator value measure (POVM)

We first review the conventional definition of quantum measurement. Previously we discussed measurement as derived from an observable **A** which is a Hermitian operator. We can use the eigenvectors of the observable to form an orthonormal basis which spans the

Hilbert space. We can construct a compete set of Hermitian orthogonal projection operators from this orthonormal basis. For each basis vector $|i_a\rangle$ in the orthonormal basis we can form a projection operator

$$\mathbf{E}_a = |i_a\rangle \langle i_a| .$$

The projection operators satisfy

$$\mathbf{E}_a = \mathbf{E}_a^\dagger \qquad \mathbf{E}_a \mathbf{E}_b = \delta_{ab} \mathbf{E}_a.$$

We can see that they are orthogonal. If the basis spans the Hilbert space, then the projection operator set is complete and

$$\sum_a \mathbf{E}_a = \mathbf{I}$$

is the identity matrix. If a is the eigenvalue associated with $|i_a\rangle$ then

$$\mathbf{A} = \sum_a a \mathbf{E}_a.$$

This is known as a spectral decomposition. The probability that a measurement of wave vector $|\psi\rangle$ gives value a is

$$\langle\psi| \mathbf{E}_a |\psi\rangle ,$$

and the post measurement wave vector is

$$\frac{\mathbf{E}_a |\psi\rangle}{\langle\psi| \mathbf{E}_a |\psi\rangle}.$$

If the outcome of the measurement is not known, then post measurement the system is in a mixed state. If the original density matrix is $\boldsymbol{\rho}$, then the post measurement the density matrix

$$\boldsymbol{\rho} \to \sum_a \mathbf{E}_a \boldsymbol{\rho} \mathbf{E}_a$$

which is equivalent to expressing $\boldsymbol{\rho}$ as a sum of probabilities of pure states.

What if we are only doing measurements only in a subspace? For example in one part of a bipartite system that is described with a tensor product? What if unitary evolution takes place in the full system and we would like to know how the density matrix evolves in a subsystem? We can make a generalization of our measurements by relaxing the orthogonality requirement.

Definition: A **positive operator value measure** (POVM) is a **complete** collection of **positive and Hermitian** operators $\mathbf{E}_m$. The summation property

$$\sum_m \mathbf{E}_m = \mathbf{I}$$

where $\mathbf{I}$ is the identity is known as **completeness**. Positivity means that $\langle\psi|\,\mathbf{E}_m\,|\psi\rangle \geq 0$ for all possible wave vectors $|\psi\rangle$. The operators do **not** necessarily need to be **orthogonal**. The $\mathbf{E}_m$ are can be considered elements associated with partial measurements and they can determine probabilities of outcomes.

We can specify the density matrix after partial measurements or after unitary evolution. If we are tracking a subspace, then it may not be possible to know the full wave function. However we should be able to track the density matrix of a subsystem and we can describe evolution of the subsystem density matrix in terms of the operators $\mathbf{E}_m$.

A non negative Hermitian operator has a non-negative square root $\mathbf{M}_m = \sqrt{\mathbf{E}_m}$. The new density matrix after evolution or partial measurement becomes

$$\boldsymbol{\rho}' = \sum_m \mathbf{M}_m \boldsymbol{\rho} \mathbf{M}_m^\dagger. \tag{9}$$

1.12 Evolution of density matrices – quantum channels and superoperators

A mixture can be described by a density matrix ρ. At a later time, either due to measurements or interaction with another quantum system, the density matrix may change to ρ'.

We construct a map from ρ to ρ'. The map must maintain the trace and positivity. It should keep ρ Hermitian as its eigenvalues must remain real. Such a **trace preserving completely positive map** (TPCP map) is sometimes called a **quantum channel**.

A **quantum channel** $\epsilon()$ is a map from a density matrix to another density matrix $\rho \to \rho'$ or

$$\rho' = \epsilon(\rho)$$

that preserves

- Linearity $\epsilon(\alpha\rho_1 + \beta\rho_2) = \alpha\epsilon(\rho_1) + \beta\epsilon(\rho_2)$ where α, β are real numbers.
- Hermiticity. $\rho = \rho^\dagger$ implies that $\epsilon(\rho) = \epsilon(\rho)^\dagger$.
- Preserves positivity.
- Preserves the trace. $\mathrm{tr}(\epsilon(\rho)) = \mathrm{tr}(\rho)$.

An example: We start with a pure state $|\psi\rangle_{AB}$ in a bipartite system. The full system can evolve via unitary evolution $|\psi\rangle'_{AB} = U\,|\psi\rangle_{AB}$. The density matrix ρ_A can evolve according to equation 9.

$$\begin{aligned}
\rho_{AB} &\to U\rho_{AB}U^\dagger \\
\rho_A &= \mathrm{tr}_B\,\rho_{AB} \\
\rho_A &\to \rho'_A = \mathrm{tr}_B\,U\rho_{AB}U^\dagger.
\end{aligned}$$

The evolution of ρ_A is not necessarily unitary. The map of $\rho_A \to \rho'_A$ giving the traced density matrix at a later time would be a quantum channel $\epsilon()$. If ρ_{AB} varies due to unitary transformation, the map $\rho_A \to \rho'_A$ is a quantum channel and sometimes it is called a **superoperator**.

$$\begin{array}{ccc} \rho_{AB} & \xrightarrow{U} & \rho'_{AB} = U\rho U^\dagger \\ \downarrow{\scriptstyle \mathrm{tr}_B} & & \downarrow{\scriptstyle \mathrm{tr}_B} \\ \rho_A & \xrightarrow{\epsilon} & \rho'_A \end{array}$$

Question What is a simple example of a POVM that is not orthogonal?

Question Is it possible to write every possible quantum channel in terms of operators M_m such that $\rho' = \sum_m M_m \rho M_m^\dagger$, the operators are complete; $\sum_m M_m M_m^\dagger = \mathbf{I}$, and $M_m M_m^\dagger$ are positive? The answer I think is yes.

The **operator sum decomposition** I think is exactly this. Every superoperator $\epsilon(\rho)$ can be written as a sum

$$\epsilon(\rho) = \sum_{i=0^N} A_i \rho A_i^\dagger$$

The A_i are linear operators. The operator sum decomposition is not necessarily unique.

2 Entropy and Information

2.1 Entropy in statistical physics

We consider a system of a total energy but a number of different ways g to distribute the energy in the system. The number of different ways to distribute the energy is called the multiplicity, g. The entropy of the system is log of the multiplicity,

$$\sigma = \ln g.$$

With multiplicity g, the system is equally likely to be in any of the g possible configurations. The probability that it is in one of them is $p = 1/g$. The entropy can be written as

$$\sigma = -\ln p.$$

If we have two systems, one with multiplicity g_1 and the other with multiplicity g_2 (and they are not exchanging anything), then the multiplicity of the combined system would be $g_1 g_2$. The entropy is then

$$\sigma = \ln g_1 + \ln g_2.$$

For the first system the probability that it is in any single configuration is $1/p_1$. The probability that the first system is in a single configuration and the second sytem is in a particular configuration is $1/(p_1 p_2)$ and the entropy can be written as

$$\sigma = -\ln p_1 - \ln p_2.$$

Why is this related to information? We would need a number of size g_1 to specify exactly which state the first system is in and we would need a number of size g_2 to specify exactly the state that the second system is in. So the entropy is related to the information needed to specify the exact configuration of the system.

If we have multiple systems, each with multiplicity g_i and probability $1/p_i$ to be in a particular configuration, the entropy

$$\sigma = \sum_i \ln g_i = -\sum_i \ln p_i. \tag{10}$$

2.2 Shannon entropy

In the realm of classical communication, Shannon answered the following question: What is the maximal compression that can be applied to a message?

Consider a message that is sent with only 4 letters, a, b, c, d. We could encode each letter with two bits

$$a = 00 \qquad b = 01 \qquad c = 10 \qquad d = 11.$$

A message that is n letters long would require 2n bits.

Suppose that the letters occur with different probabilities. This would not be surprising if our letters were in an alphabet and we were transmitting text. For example, In English 'e' is the most frequent letter. Suppose the letter a has a probability $1/2$, the letter b has a probability $1/4$ and c, d both have probabilities of $1/8$. The following encoding would be more efficient

$$a = 0 \qquad b = 10 \qquad c = 110 \qquad d = 111.$$

The most common letter a is given 1 bit, the next most common b is given 2 bits and the least common letters, c, d, are given 3 bits. The average length of a message that is n letters long is

$$n\left(\frac{1}{2}\times 1 + \frac{1}{4}\times 2 + \frac{1}{8}\times 3 + \frac{1}{8}\times 3\right) = \frac{7}{4}n$$

This is less than 2n so this encoding system is more efficient than using 2 bits for each letter.

More generally, assume that we have k alphabet letters and each letter a_i has a probability p_i with the sum of the probabilities $\sum_{i=1}^{k} p_i = 1$.

In a typical message that is n letters long, what is the expected number of alphabet letter a_i?

The answer depends on the Shannon entropy

$$S \equiv -\sum_{i=1}^{k} p_i \log_2 p_i. \tag{11}$$

Stirling's approximation is used to estimate this. Notice that the log is base 2, which is convenient if we are encoding information digitally with bits that are 0 or 1.

Compare the Shannon entropy to that we estimated from conventional entropy in equation 10. Instead of have equal probability states and keeping track of the multiplicity of states we are choosing states with a set of probabilities. Consider a single letter in our alphabet sequence. If the letter is a with probability $1/2$ then the information content would be $-\log_2 1/2$. But the probability of that happening would be $1/2$. This gives an estimate for the information content of the first letter, $-\frac{1}{2}\log_2 \frac{1}{2}$. The total information content would be $\sum_i p_i \log_2 p_i$ which is, in fact, the Shannon entropy. You can think of the Shannon entropy as a probability weighted sum of information in each possible alphabet choice.

2.3 von Neumann entropy and quantum information

Consider a system that has a various states described with probabilities p_i. The **Shannon** entropy is

$$S_{Shannon} \equiv -\sum_i p_i \ln p_i.$$

For a quantum system, the analogous entropy is known as the **von Neumann** entropy and it is computed from the density matrix

$$s \equiv -\operatorname{tr}(\rho \ln \rho). \tag{12}$$

The exponential of a matrix can be defined in terms of a series. The log of a matrix can be defined as consistent from that definition. If we diagonalize ρ and write it as a diagonal matrix with eigenvalues λ_i on its diagonal then

$$s \equiv -\sum_i \lambda_i \ln \lambda_i$$

A pure state has density matrix that has a single eigenvalue of 1 so its entropy $s = 0$. A mixed state has a positive entropy as the eigenvalues are positive and they are all smaller than 1.

2.4 Quantifying the extent of entanglement

Consider a bipartite system that is a tensor product of two Hilbert spaces. If the density matrix is constructed from a pure state in the full product space ρ_{AB}, then the von-Neumann entropy of ρ_{AB} is 0. An entangled state is one that has non zero entropy in the traced density matrices ρ_A and ρ_B. We could try to measure the extent of entanglement from the von-Neuman entropy of the traced density matrices or the trace of their squares.

Given a state vector $|\psi\rangle$ in $H_A \otimes H_B$ is it possible to write it as a product of two states $|\psi\rangle = |\phi\rangle_A \otimes |\phi\rangle_B$. There is discussion of using negative off diagonal matrices in Presskill's notes to try to partially answer the question of whether a state vector can be written as a product and so is *not* entangled.

2.5 Distance measures for quantum information

The distance between two classical probability distributions p_i, q_i can be described in terms of a trace distance or $\sum_i |p_i - q_i$ where the indices for the events are matched.

An analogy for two density matrices ρ, σ is the trace distance

$$D(\rho, \sigma) = \frac{1}{2}\text{tr}|\rho - \sigma|.$$

Related is called **fidelity**

$$F(\rho, \sigma) = \text{tr}\sqrt{\rho^{\frac{1}{2}}\sigma\rho^{\frac{1}{2}}}.$$

3 Systems with three qubits

In this section we consider a tensor product of 3 Hilbert spaces $H_A \otimes H_B \otimes H_C$ where each one describes a single qubit.

3.1 The GHZ state and the monogamy of entanglement

Consider

$$|\psi\rangle_{Cat} = \frac{1}{\sqrt{2}}\left(|000\rangle + |111\rangle\right)$$

which is known as the *cat* state for 3 qubits. Each qubit in the entangled state is hosted by a different person. Alice gets the first one, Bob gets the second and Charlie gets the third.

Let's look at some density matrices. The density matrix of the pure state given by $|\psi\rangle$ is

$$\rho_{ABC,Cat} = \frac{1}{2}\left(|000\rangle\langle 000| + |111\rangle\langle 111| + |000\rangle\langle 111| + |111\rangle\langle 000|\right)$$

We trace out the first qubit giving the reduced density matrix

$$\rho_{BC,Cat} = \text{tr}_A\rho_{ABC} = \frac{1}{2}\left(|00\rangle\langle 00| + |11\rangle\langle 11|\right) = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}.$$

Similarly

$$\rho_{AC,Cat} = \mathrm{tr}_B \rho_{ABC} = \frac{1}{2}\left(|00\rangle\langle 00| + |11\rangle\langle 11|\right) = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \tag{13}$$

$$\rho_{AB,Cat} = \mathrm{tr}_C \rho_{ABC} = \frac{1}{2}\left(|00\rangle\langle 00| + |11\rangle\langle 11|\right).$$

We take the trace over pairs of two qubits

$$\rho_{A,Cat} = \mathrm{tr}_{BC} \rho_{ABC} = \frac{1}{2}\left(|0\rangle\langle 0| + |1\rangle\langle 1|\right)$$

$$\rho_{B,Cat} = \mathrm{tr}_{AC} \rho_{ABC} = \frac{1}{2}\left(|0\rangle\langle 0| + |1\rangle\langle 1|\right)$$

$$\rho_{C,Cat} = \mathrm{tr}_{AB} \rho_{ABC} = \frac{1}{2}\left(|0\rangle\langle 0| + |1\rangle\langle 1|\right).$$

The non-zero eigenvalues of $\rho_{A,Cat}$, $\rho_{B,Cat}$, $\rho_{C,Cat}$, $\rho_{AB,Cat}$, $\rho_{BC,Cat}$, $\rho_{AC,Cat}$ are 1/2,1/2. There is the same amount of information in a single qubit as in the other two.

Consider Alice and Bob who together have information present in $\rho_{AB,Cat}$ and Charlie who has information present in ρ_C. Nothing Charlie does should affect Alice and Bob's measurements but if Charlie makes a measurement of the 3-bit Cat state then there is no longer any entanglement between Alice and Bob.

Compare $\rho_{AB,Cat}$ to that of a two bit 2 Cat state or Bell pair state. With

$$|\psi\rangle_{\mathrm{Bell}} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$$

the density matrix

$$\rho_{\mathrm{Bell}} = \frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 11|) = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

This is not the same as $\rho_{AB,Cat}$ (equation 13).

This is illustrates the principle called **monogamy of entanglement**. If Alice has a qubit that is maximally entangled with one that Bob has, (like the Bell pair state) then Alice's qubit cannot be maximally entangled with Charlie's.

An example of a 3 qubit state in which Alice and Bob share a **maximally** entangled state is $|\psi\rangle = |\psi\rangle_{\mathrm{Bell}} \otimes |0\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |110\rangle)$. For this state $\rho_{AB} = \rho_{Bell}$ and $\rho_C = |0\rangle\langle 0|$.

Let's compare the situation in what is known as the 'W-state'

$$|\psi\rangle_W = \frac{1}{\sqrt{3}}\left(|001\rangle + |010\rangle + |100\rangle\right)$$

$$\rho_{ABC,W} = \frac{1}{3}(|100\rangle\langle 100| + |010\rangle\langle 010| + |001\rangle\langle 001| + |001\rangle\langle 010| + |010\rangle\langle 001| + |100\rangle\langle 010| + |010\rangle\langle 100| + |001\rangle\langle 100| + |100\rangle\langle 001|)$$

$$\rho_{AB,W} = \frac{1}{3}(|10\rangle\langle 10| + |01\rangle\langle 01| + |00\rangle\langle 00| + |10\rangle\langle 01| + |01\rangle\langle 10|)$$

$$\rho_A = \frac{2}{3}|0\rangle\langle 0| + \frac{1}{3}|1\rangle\langle 1|$$

The eigenvalues of $\rho_{AB,W}$ are 0,0,1/3, 2/3. The eigenvalues of ρ_A are 1/3, 2/3. Again the amount of information present in any two bits is equivalent to that in any single bit. However, notice that there is more information content coming from the $|1\rangle$ state than in the $|0\rangle$ state. There's some entanglement between Alice and Charlie, and between Alice and Bob, but neither pair is maximally entangled.