QI Lecture 9

Quantum Information I

Mixed States ↔ Pure States

Purification

The partial trace is an incredibly powerful tool; it allows us to take into account the fact that we don't have access to some subspace of a quantum system which, technically, is the only situation we're ever in. As isolated as a quantum system is, there's always at least a little interaction with the environment, and the environment is a subspace we can never entirely take into account. As we saw in the last example, taking a partial trace can turn pure states into mixed states.

The "**purification**" of a mixed state is a way of writing it in a larger space such that in the larger space it's a pure state. That is, if

$$\rho_a = \sum_{k=1}^{N} p_k |\psi_k\rangle_a \langle\psi_k|_a$$

where $|\psi_k\rangle_a \in A$ with $dim(A) = N$, then we call $|\Phi\rangle_{ab} \in A \otimes B$ with $dim(B) \geq N$ a "purification" of ρ_a if

$$Tr_b \left[|\Phi\rangle\langle\Phi| \right] = \rho_a$$

Define the basis of the space B to be $\{|j\rangle_b\}$. Then

$$\boxed{|\Phi\rangle = \sum_k \sqrt{p_k} |\psi_k\rangle_a |k\rangle_b}$$

is a purification of ρ_a because

$$
\begin{aligned}
& Tr_b\left[|\Phi\rangle\langle\Phi|\right] \\
= \;& \textstyle\sum_j \langle j|_b \left(\sum_i \sqrt{p_i}|\psi_i\rangle_a|i\rangle_b\right)\left(\sum_k \sqrt{p_k}\langle\psi_k|_a\langle k|_b\right)|j\rangle_b \\
= \;& \textstyle\sum_j \left(\sum_i \sqrt{p_i}|\psi_i\rangle_a\langle j|i\rangle_b\right)\left(\sum_k \sqrt{p_k}\langle\psi_k|_a\langle k|j\rangle_b\right) \\
= \;& \textstyle\sum_j \left(\sum_i \sqrt{p_i}|\psi_i\rangle_a\delta_{ij}\right)\left(\sum_k \sqrt{p_k}\langle\psi_k|_a\delta_{jk}\right) \\
= \;& \textstyle\sum_j \left(\sqrt{p_j}|\psi_j\rangle_a\right)\left(\sqrt{p_j}\langle\psi_j|_a\right) \\
= \;& \textstyle\sum_j p_j|\psi_j\rangle_a\langle\psi_j|_a \\
= \;& \rho_a
\end{aligned}
$$

The purification is not unique. Not only did we not specify which space we're using for B, but we could have done the same derivation with any basis for B.

The purification is rather profound. Is every mixed state a pure state in a larger space that we don't have complete access to? Is the universe in a pure state? What if it's in a mixed state?

Fun stuff.

Example Find and verify a purification for

$$\rho_a = \frac{3}{4}|1\rangle\langle 1| + \frac{1}{4}\left(\frac{3|1\rangle + 4|2\rangle}{5}\right)\left(\frac{3\langle 1| + 4\langle 2|}{5}\right) \in A$$

Following the procedure above, the purification is $\rho = |\Phi\rangle\langle\Phi| \in A \otimes B$ where

$$|\Phi\rangle = \frac{\sqrt{3}}{2}|1\rangle_a|1\rangle_b + \frac{1}{2}\left(\frac{3|1\rangle_a + 4|2\rangle_a}{5}\right)|2\rangle_b = \frac{\sqrt{3}}{2}|1\rangle_a|1\rangle_b + \frac{3}{10}|1\rangle_a|2\rangle_b + \frac{2}{5}|2\rangle_a|2\rangle_b$$

and $|1\rangle_b$, $|2\rangle_b$ are basis states in space B. We can pick absolutely any orthogonal states in B, so we could just as easily switch $|1\rangle_b \leftrightarrow |2\rangle_b$ or, what the heck, use $|\square\rangle_b$ and $|\circ\rangle_b$.

$$\boxed{\rho = |\Phi\rangle\langle\Phi| = \left(\tfrac{\sqrt{3}}{2}|1\rangle_a|1\rangle_b + \tfrac{3}{10}|1\rangle_a|2\rangle_b + \tfrac{2}{5}|2\rangle_a|2\rangle_b\right)\left(\tfrac{\sqrt{3}}{2}\langle 1|_a\langle 1|_b + \tfrac{3}{10}\langle 1|_a\langle 2|_b + \tfrac{2}{5}\langle 2|_a\langle 2|_b\right)}$$

As ugly as $|\Phi\rangle$ is, it's still a single state in $A \otimes B$ and that means that $\rho = |\Phi\rangle\langle\Phi|$ is a pure state.

The partial trace of a purification is the original density matrix. So to verify that ρ is the purification of ρ_a, we need to take the trace over B of ρ.

$$
\begin{aligned}
\rho_a &= Tr_b[\rho] \\
&= \langle 1|_b \rho |1\rangle_b + \langle 2|_b \rho |2\rangle_b \\
&= \left\{ \begin{array}{l} \quad\ \langle 1|_b \left(\frac{\sqrt{3}}{2}|1\rangle_a|1\rangle_b + \frac{3}{10}|1\rangle_a|2\rangle_b + \frac{2}{5}|2\rangle_a|2\rangle_b \right) \left(\frac{\sqrt{3}}{2}\langle 1|_a\langle 1|_b + \frac{3}{10}\langle 1|_a\langle 2|_b + \frac{2}{5}\langle 2|_a\langle 2|_b \right) |1\rangle_b \\ + \ \langle 2|_b \left(\frac{\sqrt{3}}{2}|1\rangle_a|1\rangle_b + \frac{3}{10}|1\rangle_a|2\rangle_b + \frac{2}{5}|2\rangle_a|2\rangle_b \right) \left(\frac{\sqrt{3}}{2}\langle 1|_a\langle 1|_b + \frac{3}{10}\langle 1|_a\langle 2|_b + \frac{2}{5}\langle 2|_a\langle 2|_b \right) |2\rangle_b \end{array} \right. \\
&= \left(\frac{\sqrt{3}}{2}|1\rangle_a + 0 + 0 \right) \left(\frac{\sqrt{3}}{2}\langle 1|_a + 0 + 0 \right) + \left(0 + \frac{3}{10}|1\rangle_a + \frac{2}{5}|2\rangle_a \right) \left(0 + \frac{3}{10}\langle 1|_a + \frac{2}{5}\langle 2|_a \right) \\
&= \frac{3}{4}|1\rangle_a\langle 1|_a + \left(\frac{3}{10}|1\rangle_a + \frac{2}{5}|2\rangle_a \right) \left(\frac{3}{10}\langle 1|_a + \frac{2}{5}\langle 2|_a \right) \\
&= \frac{3}{4}|1\rangle_a\langle 1|_a + \frac{1}{4} \left(\frac{3|1\rangle_a + 4|2\rangle_a}{5} \right) \left(\frac{3\langle 1|_a + 4\langle 2|_a}{5} \right)
\end{aligned}
$$

The system works!

■

Undisclosed Measurements

An observable is a set of projective operations, P_k, with corresponding probabilities p_k. Given the state $|\psi\rangle$ or density matrix ρ the probability of each result and the effect are

$$p_k = \langle \psi | P_k | \psi \rangle = Tr\left[P_k \rho \right] \qquad |\psi\rangle \xrightarrow{k} |\psi_k\rangle = \frac{P_k|\psi\rangle}{\sqrt{p_k}} \qquad \rho \xrightarrow{k} \rho_k = \frac{P_k \rho P_k}{p_k}$$

But what happens if a measurement is made, but not reported? The best you can say is that the state or density matrix is one of $\{|\psi_k\rangle\}$ or $\{\rho_k\}$ and that the probability of that result is p_k. Fortunately, we already have a way to describe probability distributions over quantum states: another density matrix!

The effect of an unrevealed observable, with projection operators $\{P_k\}$, is

$$\boxed{\rho \longrightarrow \rho' = \sum_k p_k \rho_k = \sum_k P_k \rho P_k}$$

It may occur to you that for projective measurements $\sum_k P_k = I$ and therefore the right side should be equal to ρ. That would be the case if the right side were $\left(\sum_j P_j\right) \rho \left(\sum_k P_k\right) = \sum_{jk} P_j \rho P_k$. As it is, the right side describes a “chopping up” of the density matrix that destroys any coherence between the ranges of the different projection operations.

Example The state $|\psi\rangle = \frac{3|0\rangle+4i|1\rangle}{5}$ is measured in the computational basis, but the result is not reported. What is the resulting density matrix?

The projection operators for a measurement in the computational basis are $P_0 = |0\rangle\langle 0|$ and $P_1 = |1\rangle\langle 1|$. The pure state density matrix for $|\psi\rangle = \frac{3|0\rangle+4i|1\rangle}{5}$ is

$$\rho = \left(\frac{3|0\rangle + 4i|1\rangle}{5}\right)\left(\frac{3\langle 0| - 4i\langle 1|}{5}\right)$$

The result of the unreported measurement is

$$\begin{aligned}
\rho' &= P_0\rho P_0 + P_1\rho P_1 \\
&= |0\rangle\langle 0|\left(\tfrac{3|0\rangle+4i|1\rangle}{5}\right)\left(\tfrac{3\langle 0|-4i\langle 1|}{5}\right)|0\rangle\langle 0| + |1\rangle\langle 1|\left(\tfrac{3|0\rangle+4i|1\rangle}{5}\right)\left(\tfrac{3\langle 0|-4i\langle 1|}{5}\right)|1\rangle\langle 1| \\
&= |0\rangle\left(\tfrac{3}{5}\right)\left(\tfrac{3}{5}\right)\langle 0| + |1\rangle\left(\tfrac{4i}{5}\right)\left(\tfrac{-4i}{5}\right)\langle 1| \\
&= \tfrac{9}{25}|0\rangle\langle 0| + \tfrac{16}{25}|1\rangle\langle 1|
\end{aligned}$$

■

Von Neumann Entropy

The Shannon entropy, which describes the entropy of classical probability distributions, $\{p_1, p_2, \ldots\} = \{p_k\}$, is given by

$$H\left[\{p_k\}\right] = -\sum_k p_k \log(p_k)$$

We can describe a classical probability distribution by using a set of orthogonal states, $\{|k\rangle\}$. This describes a classical situation because we think of classical states as being distinguishable from each other. For example, if you roll a die, the probability of having a 5 but observing a 2 is zero, which is exactly what we expect from orthogonal states: $p = |\langle 2|5\rangle|^2 = 0$.

Given a classical probability distribution, $\{p_k\}$, we can use the ensemble $\{p_k, |k\rangle\}$ and the associated density matrix

$$\rho = \sum_k p_k |k\rangle\langle k|$$

The trace[1] sums over diagonal terms and it's easy to apply functions to diagonal matrices.

[1]Here we use the shortcut mentioned in lecture 6: "The trace turns outer products into inner products."

$$-Tr\left[\rho\log(\rho)\right] = -Tr\left[\sum_k p_k \log\left(p_k\right)|k\rangle\langle k|\right] = -\sum_k p_k \log\left(p_k\right)\langle k|k\rangle = -\sum_k p_k \log\left(p_k\right) = H\left[\{p_k\}\right]$$

So we have a formula for density matrices that recreates the Shannon entropy for classical probability distributions. Since density matrices are always hermitian, an arbitrary $\rho = \sum_k p_k|\psi_k\rangle\langle\psi_k|$ has a spectral decomposition

$$\rho = \sum_j p_j|\psi_j\rangle\langle\psi_j| = \sum_k \lambda_k|v_k\rangle\langle v_k|$$

where λ_k are the eigenvalues and $\{|v_k\rangle\}$ are orthonormal eigenstates. So ρ is also the density matrix for the ensemble $\{p(|v_k\rangle) = \lambda_k\}_k$, which is composed of distinguishable (orthogonal) states.

We define the Von Neumann Entropy as

$$\boxed{S[\rho] \equiv -Tr\left[\rho\log[\rho]\right] = -\sum_k \lambda_k \log\left(\lambda_k\right)}$$

Notice that, because the trace permutes, it doesn't matter what basis we use.[2]

$$\begin{aligned}
S[U\rho U^\dagger] &= -Tr\left[U\rho U^\dagger \log[U\rho U^\dagger]\right] \\
&= -Tr\left[U\rho U^\dagger U \log[\rho] U^\dagger\right] \\
&= -Tr\left[U^\dagger U\rho U^\dagger U \log[\rho]\right] \\
&= -Tr\left[\rho\log[\rho]\right] \\
&= S[\rho]
\end{aligned}$$

Example Find the Von Neumann Entropy for the ensemble

$$\left\{p(|1\rangle) = \frac{1}{4},\, p\left(\frac{|2\rangle + i|3\rangle}{\sqrt{2}}\right) = \frac{1}{2},\, p\left(\frac{|2\rangle - i|3\rangle}{\sqrt{2}}\right) = \frac{1}{4}\right\}$$

The density matrix is

[2] Here we a clever trick for Taylor polynomials and similarity transforms. Because $\left(UMU^\dagger\right)^n = UMU^\dagger UMU^\dagger UMU^\dagger \ldots UMU^\dagger = UM^nU^\dagger$, wherever the Taylor polynomial converges $f\left(UMU^\dagger\right) = \sum_n a_n\left(UMU^\dagger\right)^n = U\left(\sum_n a_n M^n\right)U^\dagger = Uf(M)U^\dagger$.

$$
\begin{aligned}
\rho &= \tfrac{1}{4}|1\rangle\langle 1| + \tfrac{1}{2}\left(\tfrac{|2\rangle+i|3\rangle}{\sqrt{2}}\right)\left(\tfrac{\langle 2|-i\langle 3|}{\sqrt{2}}\right) + \tfrac{1}{4}\left(\tfrac{|2\rangle-i|3\rangle}{\sqrt{2}}\right)\left(\tfrac{\langle 2|+i\langle 3|}{\sqrt{2}}\right) \\
&= \tfrac{1}{4}|1\rangle\langle 1| + \tfrac{1}{4}\left(|2\rangle\langle 2| - i|2\rangle\langle 3| + i|3\rangle\langle 2| + |3\rangle\langle 3|\right) + \tfrac{1}{8}\left(|2\rangle\langle 2| + i|2\rangle\langle 3| - i|3\rangle\langle 2| + |3\rangle\langle 3|\right) \\
&= \tfrac{1}{4}|1\rangle\langle 1| + \tfrac{3}{8}|2\rangle\langle 2| - \tfrac{1}{8}i|2\rangle\langle 3| + \tfrac{1}{8}i|3\rangle\langle 2| + \tfrac{3}{8}|3\rangle\langle 3| \\
&\sim \begin{bmatrix} \frac{1}{4} & 0 & 0 \\ 0 & \frac{3}{8} & -\frac{1}{8}i \\ 0 & \frac{1}{8}i & \frac{3}{8} \end{bmatrix}
\end{aligned}
$$

Since $S[\rho] \equiv -Tr\left[\rho \log[\rho]\right] = -\sum_k \lambda_k \log\left(\lambda_k\right)$, we need to find the eigenvalues of this matrix.

$$
\begin{aligned}
C(\lambda) &= \begin{vmatrix} \frac{1}{4}-\lambda & 0 & 0 \\ 0 & \frac{3}{8}-\lambda & -\frac{1}{8}i \\ 0 & \frac{1}{8}i & \frac{3}{8}-\lambda \end{vmatrix} \\
&= \left(\tfrac{1}{4}-\lambda\right)\begin{vmatrix} \frac{3}{8}-\lambda & -\frac{1}{8}i \\ \frac{1}{8}i & \frac{3}{8}-\lambda \end{vmatrix} \\
&= \left(\tfrac{1}{4}-\lambda\right)\left[\left(\tfrac{3}{8}-\lambda\right)\left(\tfrac{3}{8}-\lambda\right) - \left(\tfrac{1}{8}i\right)\left(-\tfrac{1}{8}i\right)\right] \\
&= -\left(\lambda-\tfrac{1}{4}\right)\left[\lambda^2 - \tfrac{3}{4}\lambda + \tfrac{9}{64} - \tfrac{1}{64}\right] \\
&= -\left(\lambda-\tfrac{1}{4}\right)\left[\lambda^2 - \tfrac{3}{4}\lambda + \tfrac{1}{8}\right] \\
&= -\left(\lambda-\tfrac{1}{4}\right)\left(\lambda-\tfrac{1}{4}\right)\left(\lambda-\tfrac{1}{2}\right)
\end{aligned}
$$

We have that the eigenvalues are $\lambda = \frac{1}{2}, \frac{1}{4}, \frac{1}{4}$ and therefore

$$
S[\rho] = -\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) = \frac{1}{2} + \frac{2}{4} + \frac{2}{4} = \frac{3}{2}
$$

■

Properties of Von Neumann Entropy

The "**relative entropy**" is a quantity which, on its own, isn't terribly useful, but which is very useful for proofs.

$$
\boxed{S[\rho\|\sigma] \equiv Tr[\rho\log[\rho]] - Tr[\rho\log[\sigma]]}
$$

Klein's inequality[3] says that for any density matrices, ρ and σ, defined on the same space

$$\boxed{0 \le S[\rho \| \sigma]}$$

where equality only occurs for $\rho = \sigma$ (the relative entropy between a state and itself is zero).

Entropy is bound from below by pure states and from above by completely mixed states.

$$\boxed{0 \le S[\rho] \le \log(d)}$$

The entropy is positive and only zero for pure states. Like Shannon entropy, Von Neumann entropy is maximized by equally probable orthogonal states, a "**completely mixed state**", $\rho = \frac{1}{d}I$. In other words,

$$0 = S[|\psi\rangle\langle\psi|] \le S[\rho] \le S\left[\frac{1}{d}I\right] = \log(d)$$

The lower bound is trivial; the probability distribution for a pure state is $p = 1$ and so $S[|\psi\rangle\langle\psi|] = -1\log_2(1) = 0$.

We can prove the (fairly reasonable) upper bound using Klein's inequality. Remember that $Tr[\rho] = 1$.

$$\begin{aligned}
0 &\le S\left[\rho \| \tfrac{1}{d}I\right] \\
0 &\le Tr[\rho\log[\rho]] - Tr\left[\rho\log\left[\tfrac{1}{d}I\right]\right] \\
0 &\le Tr[\rho\log[\rho]] - Tr\left[\rho\log\left(\tfrac{1}{d}\right)I\right] \\
0 &\le Tr[\rho\log[\rho]] - \log\left(\tfrac{1}{d}\right)Tr\left[\rho\right] \\
0 &\le Tr[\rho\log[\rho]] - \log\left(\tfrac{1}{d}\right) \\
-Tr[\rho\log[\rho]] &\le \log(d) \\
S[\rho] &\le \log(d)
\end{aligned}$$

[3] I don't include the proof of Klein's inequality here because it's long, algebraic, and unenlightening. It essentially boils down to "log is concave" and then a lot of work.

Klein's inequality is an equality if and only if the arguments are equal, so this maximum occurs for $\rho = \frac{1}{d} I$.

QED

Von Neumann Entropy is Subadditive.

$$\boxed{S[\rho_{ab}] \leq S[\rho_a] + S[\rho_b]}$$

We can prove this quickly using Klein's inequality, $0 \leq S[\rho \| \sigma]$, with $\rho = \rho_{ab}$ and $\sigma = \rho_a \otimes \rho_b$:

$$\begin{aligned}
0 &\leq S[\rho\|\sigma] \\
0 &\leq Tr[\rho_{ab}\log[\rho_{ab}]] - Tr[\rho_{ab}\log[\rho_a \otimes \rho_b]] \\
-Tr[\rho_{ab}\log[\rho_{ab}]] &\leq -Tr[\rho_{ab}\log[\rho_a \otimes \rho_b]] \\
S[\rho_{ab}] &\leq -Tr[\rho_{ab}\log[\rho_a \otimes \rho_b]] \\
S[\rho_{ab}] &\leq -Tr\left[\rho_{ab}\log[\rho_a] \otimes I_b + \rho_{ab} I_a \otimes \log[\rho_b]\right] \\
S[\rho_{ab}] &\leq -Tr\left[\rho_{ab}\log[\rho_a] \otimes I_b\right] - Tr\left[\rho_{ab} I_a \otimes \log[\rho_b]\right] \\
S[\rho_{ab}] &\leq -Tr_a\left[\rho_a \log[\rho_a]\right] - Tr_b\left[\rho_b \log[\rho_b]\right] \\
S[\rho_{ab}] &\leq S[\rho_a] + S[\rho_b]
\end{aligned}$$

The trace without a subscript is a trace over all spaces, so when $Tr \to Tr_a$ we've taken a trace over B, but haven't yet taken the trace over A. Remember that whenever you apply a function to a diagonalized matrix (the spectral decomposition), you can just apply it to each diagonal term. This helps explain the biggest algebraic jump, expanded here:

$$\begin{aligned}
\log\left[\rho_a \otimes \rho_b\right] &= \log\left[\left(\sum_j p_j |j\rangle_a \langle j|_a\right)\left(\sum_k q_k |k\rangle_b \langle k|_b\right)\right] \\
&= \log\left[\sum_{jk} p_j q_k |j\rangle_a |k\rangle_b \langle j|_a \langle k|_b\right] \\
&= \sum_{jk} \log(p_j q_k) |j\rangle_a |k\rangle_b \langle j|_a \langle k|_b \\
&= \sum_{jk} \left(\log(p_j) + \log(q_k)\right) |j\rangle_a |k\rangle_b \langle j|_a \langle k|_b \\
&= \sum_{jk} \log(p_j) |j\rangle_a |k\rangle_b \langle j|_a \langle k|_b + \sum_{jk} \log(q_k) |j\rangle_a |k\rangle_b \langle j|_a \langle k|_b \\
&= \left(\sum_j \log(p_j) |j\rangle_a \langle j|_a\right)\left(\sum_k |k\rangle_b \langle k|_b\right) + \left(\sum_j |j\rangle_a \langle j|_a\right)\left(\sum_k \log(q_k) |k\rangle_b \langle k|_b\right) \\
&= \log[\rho_a] \otimes I_b + I_a \otimes \log[\rho_b]
\end{aligned}$$

QED

Knowing vs. not knowing which state you're looking at. Suppose that you want to deal with randomly chosen density matrices, where $p(\rho_k) = p_k$. A good way to declare that "someone has kept track of which density matrix is which" is to use $S\left[\sum_k p_k |k\rangle\langle k| \otimes \rho_k\right]$. This describes, for example, the situation of Alice picking states among $\{\rho_k\}$ and sending them to Bob. *Alice knows which state she sent*, so Alice's system includes a "**pointer state**", $|k\rangle$, indicating which ρ_k was sent to Bob. So the system as a whole isn't in the state $\sum_k p_k \rho_k$, it's in $\sum_k p_k |k\rangle\langle k| \otimes \rho_k$, because that "which state" information is accessible. If you were to make a measurement like asking Alice which state she sent, you'd find that p_k is the probability of seeing $|k\rangle$ and the state would "collapse" to

$$\sum_k p_k |k\rangle\langle k| \otimes \rho_k \longrightarrow |k\rangle\langle k| \otimes \rho_k$$

Here, a measurement of first space resulting in $|k\rangle\langle k|$ reveals which density matrix was selected, and the entropy of this selection is $H(\{p_k\})$. After that selection we then have the entropy of the density matrix itself, $S[\rho_k]$. We find that the entropy of this situation is:

$$\boxed{S\left[\sum_k p_k |k\rangle\langle k| \otimes \rho_k\right] = H(\{p_k\}) + \sum_k p_k S[\rho_k]}$$

where the right side should look a little familiar from lecture 7; the entropy of selecting one of many scenarios plus the average entropy of those scenarios.

Each density matrix has its own spectral decomposition, $\rho_k = \lambda_{kj} |v_{kj}\rangle_b \langle v_{kj}|_b$, so we'll use $\left\{|k\rangle_a |v_{kj}\rangle_b\right\}_{jk}$ as a basis for $A \otimes B$. Despite being funny looking, this is still an orthonormal basis.

$$\begin{aligned}
S\left[\sum_k p_k |k\rangle\langle k| \otimes \rho_k\right] &= -\sum_{jk} p_k \lambda_{jk} \log\left(p_k \lambda_{jk}\right) \\
&= -\sum_{jk} p_k \lambda_{jk} \log\left(p_k\right) - \sum_{jk} p_k \lambda_{jk} \log\left(\lambda_{jk}\right) \\
&= -\sum_k \left(\sum_j \lambda_{jk}\right) p_k \log\left(p_k\right) - \sum_k p_k \left(\sum_j \lambda_{jk} \log\left(\lambda_{jk}\right)\right) \\
&= -\sum_k p_k \log\left(p_k\right) + \sum_k p_k S[\rho_k] \\
&= H\left(\{p_k\}\right) + \sum_k p_k S[\rho_k]
\end{aligned}$$

QED

If on the other hand there's no way to keep track of which density matrix is being selected, we're left looking at $S\left[\sum_k p_k \rho_k\right]$ instead of $S\left[\sum_k p_k |k\rangle\langle k| \otimes \rho_k\right]$.

Defining $\rho_{ab} = \sum_k p_k \rho_k \otimes |k\rangle\langle k|$ we find that the reduced density matrices are

$$\rho_{ab} = \sum_k p_k \rho_k \otimes |k\rangle\langle k| \qquad \Rightarrow \qquad \rho_a = \sum_k p_k \rho_k \qquad \rho_b = \sum_k p_k |k\rangle\langle k|$$

and applying subadditivity,

$$S\left[\sum_k p_k \rho_k \otimes |k\rangle\langle k|\right] \leq S\left[\sum_k p_k \rho_k\right] + S\left[\sum_k p_k |k\rangle\langle k|\right] = S\left[\sum_k p_k \rho_k\right] + H[\{p_k\}]$$

then plugging in the result from above

$$H(\{p_k\}) + \sum_k p_k S[\rho_k] \leq S\left[\sum_k p_k \rho_k\right] + H[\{p_k\}]$$

we find that

$$\boxed{\sum_k p_k S[\rho_k] \leq S\left[\sum_k p_k \rho_k\right]}$$

Exercises

#1) The New Entropy.

a) Find the Von Neumann Entropy in bits for this ensemble.

$$p\left(|1\rangle\right) = \frac{1}{2} \qquad p\left(|2\rangle\right) = \frac{1}{6} \qquad p\left(\frac{-3i|2\rangle + 4|3\rangle}{5}\right) = \frac{1}{3}$$

b) In the computational basis, $\rho \sim \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$. What is $S[\rho]$?

c) What does the answer to part b tell you about ρ? Can you find an ensemble for this density matrix?

#3) Keeping Track And Then Not.

$$\rho_1 = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| \qquad \rho_2 = |\frac{\pi}{6}\rangle\langle\frac{\pi}{6}| = \left(\frac{\sqrt{3}|0\rangle + |1\rangle}{2}\right)\left(\frac{\sqrt{3}\langle 0| + \langle 1|}{2}\right)$$

Alice is sending photons to Bob in one of these two density matrices, where $p(\rho_1) = p(\rho_2) = \frac{1}{2}$.

Verify that the two equations

$$S\left[\sum_k p_k |k\rangle\langle k| \otimes \rho_k\right] = H(\{p_k\}) + \sum_k p_k S[\rho_k]$$

and

$$S\left[\sum_k p_k \rho_k\right] \geq \sum_k p_k S[\rho_k]$$

hold in this situation.

#4) Unmixing.

Find a purification for the density matrix

$$\rho = \frac{1}{3}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| + \frac{1}{6}|-\rangle\langle -|$$

#4) Lots of Ways to be Pure.

Assume that $|\Phi\rangle = \sum_k \sqrt{p_k}|\psi_k\rangle_a|k\rangle_b$ is a purification for $\rho_a = \sum_k p_k|\psi_k\rangle_a\langle\psi_k|_a$. Here B is assumed to have the basis $\{|k\rangle_b\}$. Show that if another basis, $\{|k'\rangle_b\}$, for the space B is related to the first by a unitary operation on B, $|k'\rangle_b = U_b|k\rangle_b$, then

$$|\Phi'\rangle = (I_a \otimes U_b)\,|\Phi\rangle$$

is also a purification for ρ_a.

This problem is about forcing you to carefully apply operators on a composite space and to know the properties of unitary operators.