Introduction to Quantum Information

MATH 667
University of Calgary
Winter semester 2016

Lecturer:
Gilad Gour
Notes taken by Mark Girard

April 7, 2016

Contents

6 Lecture 6

(**4 February 2016**)

6.1 Theory of Entanglement

This type of material is usually encountered at the end of textbooks on quantum information, but we will cover it earlier.

The main idea is that entanglement can be viewed as a resource. It can be used to perform quantum teleportation, superdense coding, and other protocols that we will discuss later. Resources are always viewed in terms of what the limitations are. In this case, the operations that we are restricted to are the LOCC operations that Alice and Bob can perform.

The most general thing that we can do on any quantum system. It is not just *observing* the system, but a measurement is also *changing* the system. Recall that a measurement is associated with a collection of operators $\{M_j\}$ satisfying

$$\sum_j M_j^* M_j = I.$$

A ***unitary*** operator U can then be viewed as a measurement with one outcome, with associated operator U, since $U^*U = I$. If Alice performs a measurement with more than one possible outcome, she can tell Bob which outcome she received.

6.1.1 LOCC with one round of communication

An arbitrary LOCC operation can be performed as follows. Alice performs a measurement and sends the result to Bob. Based on the message received from Alice, Bob performs another operation on his system. He performs a measurement and sends the result to Alice. This can go on back-and-forth meany many times indefinitely. At the end of the protocol, they will end up with some final state. It turns out that any arbitrary of this form can be *simulated* by an LOCC operation with one round of communication.

Claim 1. LOCC van be simulated with the following protocol:

1. Alice makes a measurement and communicates to Bob.
2. Bob performs a unitary operations on his system.

Consider a composite system consisting of two subsystems (Alice and Bob). Suppose the state that Alice and Bob initially shared is $|\psi\rangle \in \mathbb{C}^n \otimes \mathbb{C}^m$. Usually when Alice makes a measurement, there are many possible outcomes. If she obtained outcome x, the associated measurement operator is M_x, which is an operator that acts on Alice's space $\mathbb{C}^n$. Then the resulting (unnormalized) state is

$$M_x \otimes I|\psi\rangle.$$

Alice tells Bob that she obtained x. For each different x, Bob performs a unitary U_x that depends on Alice's outcome. The resulting (unnormalized) state is now

$$M_x \otimes U_x|\psi\rangle.$$

Quantum mechanics is not deterministic, so the outcomes are usually different depending on x. But suppose there were some measurement that Alice can perform and some set of unitaries Bob could perform such that the outcome was the same no matter what Alice's result x was. Let's say this outcome is $|\varphi\rangle$. That is, for all measurement outcomes x,

$$\frac{1}{\|M_x \otimes U_x|\psi\rangle\|} M_x \otimes U_x|\psi\rangle = |\varphi\rangle.$$

If there exist such a measurement with corresponding unitaries, then we say that $|\psi\rangle$ can be converted into $|\varphi\rangle$ by LOCC with one measurement and one round of communication.

Notation. We use the notation

$$|\psi\rangle \xrightarrow{\text{LOCC}} |\varphi\rangle$$

as shorthand for the statement that "$|\psi\rangle$ can be converted into $|\varphi\rangle$ by LOCC".

Without loss of generality, we write $|\psi\rangle$ is Schmidt form as

$$|\psi\rangle = \sum_{y=1}^{n} \sqrt{p_y}|y\rangle|y\rangle$$

and $|\varphi\rangle$ as

$$|\varphi\rangle = \sum_{y=1}^{n} \sqrt{q_y}|y\rangle|y\rangle.$$

The corresponding reduced density matrices are

$$\rho_\psi = \begin{pmatrix} p_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & p_n \end{pmatrix} \qquad \text{and} \qquad \rho_\varphi = \begin{pmatrix} q_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & q_n \end{pmatrix}.$$

Recall that $|\psi\rangle \xrightarrow{\text{LOCC}} |\varphi\rangle$ if and only if there exist unitary matrices U_x and probabilities $\{r_x\}_x$ such that

$$\rho_\psi = \sum_x r_x U_x^* \rho_\varphi U_x.$$

Lemma 6.1. *With $|\psi\rangle$, $|\varphi\rangle$, $\{p_j\}$, and $\{q_j\}$ as above, it holds that there exist unitaries U_x and probabilities $\{r_x\}_x$ such that*

$$\rho_\psi = \sum_x r_x U_x^* \rho_\varphi U_x$$

if and only if there exists a doubly stochastic matrix D such that $\vec{p} = D\vec{q}$.

We use an important and famous result (that we will prove in the assignments).

Lemma 6.2 (Birkoff-von Neumann Theorem)**.** *Every $n \times n$ doubly stochastic cam be written as a convex combination of $n \times n$ permutation matrices,*

$$D = \sum_x r_x \Pi_x.$$

Furthermore, the permutation matrices are exactly the extreme points of the convex set of doubly stochastic matrices.

We denote the $n \times n$ doubly stochastic matrices as $\mathscr{D}_{n\times n}$ and the $n \times n$ permutation matrices as $\mathscr{P}$. It holds that

$$\mathscr{D}_{n\times n} = \text{conv}(\mathscr{P}_{n\times n}) \qquad \text{and} \qquad \text{ext}(\mathscr{D}_{n\times n}) = \mathscr{P}_{n\times n}$$

where conv and ext denote the convex hull and set of extremal points respectively.

Suppose that there exists a doubly stochastic matrix D such that $\vec{p} = D\vec{q}$. By Birkhoff's theorem, we can write D as $D = \sum_x r_x \Pi_x$. Hence

$$\vec{p} = \sum_x r_x \Pi_x \vec{q} \Leftrightarrow \begin{pmatrix} p_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & p_n \end{pmatrix} = \sum_x r_x \Pi_x \begin{pmatrix} q_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & q_n \end{pmatrix} \Pi_x^*.$$

Given a vector of probabilities $\vec{p}$, we can order the probabilities in decreasing order $\vec{p}^{\downarrow}$ so that the entries are

$$p_0 \geq p_1 \geq \cdots \geq p_{n-1}.$$

For now, assume without loss of generality that $\vec{p} = \vec{p}^{\downarrow}$ for all our probability vectors.

Lemma 6.3. *Given probability vectors $\vec{p}$ and $\vec{q}$ (with entries in decreasing order), there exists a doubly stochastic matrix D satisfying $\vec{p} = D\vec{q}$ if and only if the following hold:*

$$\begin{aligned} p_0 &\leq q_0 \\ p_0 + p_1 &\leq q_0 + q_1 \\ &\vdots \\ p_0 + p_1 \cdots + p_{n-1} &\leq q_0 + q_1 + \cdots + q_{n-1}. \end{aligned}$$

Definition 6.4. We say that $\vec{p}$ is ***majorized*** by $\vec{q}$ (or $\vec{q}$ majorizes $\vec{p}$) if all of the inequalities above hold. We write $\vec{p} \prec q$ or $\vec{q} \succ p$.

Example 6.5. For all probability vectors $\vec{p}$, it holds that

$$(1, 0, \ldots, 0) \succ (p_0, p_1, \ldots, p_{n-1}) \succ \left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right).$$

Indeed, we always have $1 \geq p_0$, $1 + 0 \geq p_0 + p_1$, etc. and $p_0 \geq \frac{1}{n}$, $p_0 + p_1 \geq \frac{2}{n}$, etc, since the values of $\vec{p}$ are in decreasing order.

Example 6.6. Not all vectors may be compared using this relation. In particular, consider the vectors

$$\vec{p} = \left(\frac{1}{2}, \frac{1}{2}, 0\right) \qquad \text{and} \qquad \vec{q} = \left(\frac{3}{4}, \frac{1}{8}, \frac{1}{8}\right).$$

Note that $\frac{1}{2} \leq \frac{3}{4}$ but $\frac{1}{2} + \frac{1}{2} \geq \frac{3}{4} + \frac{1}{8}$. So $\vec{p} \not\succ \vec{q}$ and $\vec{p} \not\prec \vec{q}$.

6.1.2 Majorization and LOCC conversion

Theorem 6.7. *Let $|\psi\rangle$ and $|\varphi\rangle$ be pure states of a bipartite system. Then it holds that $|\psi\rangle \xrightarrow{\text{LOCC}} |\varphi\rangle$ if and only if $\vec{q} \succ \vec{p}$, where $\vec{q}$ and $\vec{p}$ are the vectors of the Schmidt coefficients (probabilities) of $|\psi\rangle$ and $|\varphi\rangle$ respecitively.*

Remark 6.8 (Maximally entangled states)**.** If

$$|\psi\rangle = \frac{1}{\sqrt{n}}|00\rangle + \frac{1}{\sqrt{n}}|11\rangle + \cdots + \frac{1}{\sqrt{n}}|(n-1)(n-1)\rangle,$$

then $\vec{p} = \left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$. So $\vec{p} \prec \vec{q}$ for *every* other n-dimensional probability vector $\vec{q}$. Hence $|\psi\rangle$ can be converted into any other state $|\varphi\rangle$ by LOCC. We say that this state $|\psi\rangle$ is ***maximally entangled***.

Remark 6.9. Given any two states $|\psi\rangle, |\varphi\rangle \in \mathbb{C}^n \otimes \mathbb{C}^m$, determining whetther or not $|\psi\rangle \xrightarrow{\text{LOCC}} |\varphi\rangle$ is completely determined by the Schmidt coefficients of the state vector. Two state vectors with the same Schmidt coefficients are equivalent under LOCC.

Remark 6.10. Given probability vectors $\vec{p}$ and $\vec{q}$, the condition that $\vec{p} \prec \vec{q}$ ($\vec{p}$ is majorized by $\vec{q}$) is equivalent to the condition that

$$\sum_{j=k}^{n} p_j \geq \sum_{j=k}^{n} q_j$$

holds for all $k = 1, 2, \ldots, n$. (Note that when $k = 1$, this is trivial since $\sum_{j=1}^{n} p_j = 1$.) Define functions

$$E_k(\psi) = \sum_{j=k}^{n} p_j.$$

Note that $E_k(\psi) = 1 - \sum_{j=1}^{k-1} p_j$. Hence we have

$$\vec{p} \prec \vec{q} \Longleftrightarrow E_k(\psi) \geq E_k(\varphi) \quad \forall k = 1, \ldots, n.$$

Or equivalently

$$|\psi\rangle \xrightarrow{\text{LOCC}} |\varphi\rangle \Longleftrightarrow E_k(\psi) \geq E_k(\varphi) \quad \forall k = 1, \ldots, n.$$

So we can view the E_k's as ***entanglement measures***. That is, each E_k cannot increase under LOCC operations. Their values can only decrease.

6.1.3 Probabilistic LOCC conversion

Normally, quantum mechanics is not deterministic. It sould be that Alice and Bob perform some LOCC operation, but the outcome they achieve is not always the same and it depends on the measurement outcomes. An ***ensemble*** of states is a collection

$$\{(|\varphi_y\rangle, r_y)\}_y$$

of state vectors $|\varphi_y\rangle$ and probabilities r_y that sum to 1. The ensemble can be produced by LOCC if there is a protocol that outputs $|\varphi_y\rangle$ with probability r_y. What are the conditions for conversion $|\psi\rangle \xrightarrow{\text{LOCC}} \{(|\varphi_y\rangle, r_y)\}$?

Theorem 6.11. *Let $|\psi\rangle$ be a bipartite state and $\{(|\varphi_y\rangle, r_y)\}$ be an ensemble of bipartite states. Then*

$$|\psi\rangle \xrightarrow{\text{LOCC}} |\varphi_y\rangle \qquad \textit{with probability } r_y$$

if and only if

$$E_k(\psi) \geq \sum_y r_y E_k(\varphi_y) \qquad \forall k = 1, \ldots, n.$$

Corollary 6.12. *Let $P_{\max}\left(|\psi\rangle \xrightarrow{\text{LOCC}} |\varphi\rangle\right)$ denote the maximal probability over all possible LOCC protocols of obtaining $|\varphi\rangle$ from $|\psi\rangle$. It holds that*

$$P_{\max}\left(|\psi\rangle \xrightarrow{\text{LOCC}} |\varphi\rangle\right) = \min_{k=1,\ldots,n} \left\{ \frac{E_k(\psi)}{E_k(\varphi)} \right\}.$$

Sketch of proof of Corollary. Suppose that there was some LOCC protocol that produced $|\psi\rangle \xrightarrow{\text{LOCC}} |\varphi\rangle$ with some probability p and produced a separable state $|\psi\rangle \xrightarrow{\text{LOCC}} |00\rangle$ with probability $1 - p$. Then, from the theorem, it must hold that

$$E_k(\psi) \geq pE_k(\varphi) + (1-p)E_k(|00\rangle).$$

It is clear to see that

$$p \leq \min_{k=1,\ldots,n} \left\{ \frac{E_k(\psi)}{E_k(\varphi)} \right\}.$$

This shows that

$$P_{\max}\left(|\psi\rangle \xrightarrow{\text{LOCC}} |\varphi\rangle\right) \leq \min_{k=1,\ldots,n} \left\{ \frac{E_k(\psi)}{E_k(\varphi)} \right\}.$$

(Equality is harder to prove). □

7 Lecture 7

(**9 February 2016**)

The teleportation protocol is actually an LOCC transformation. Alice and Bob (with systems labelled A and B) share a Bell state $|\text{Bell}\rangle_{\mathsf{AB}}$. In addition, Alice has an extra system (labelled a) that is initialized in an extra state $|\psi\rangle_{\mathsf{a}}$. Then the teleportation protocol looks like

$$|\psi\rangle_{\mathsf{a}}|\text{Bell}\rangle_{\mathsf{AB}} \xrightarrow{\text{LOCC}} |\text{Bell}\rangle_{\mathsf{aA}}|\psi\rangle_{\mathsf{B}}$$

where the unknown state $|\psi\rangle$ has shifted to Bob's system. Note that, in the beginning, there is entanglement between the systems (aA) and (B), but there is no entanglement between Alice and Bob at the end.

7.1 The Density Operator

The density operator formalism is the most general way to do quantum mechanics.

Definition 7.1. Suppose a quantum system is in one of a number of states $|\psi_i\rangle$ where i is an index with respective probabilities p_i. An ***ensemble*** is a collection of pairs of probabilities and states

$$\{(p_i, |\psi_i\rangle)\}.$$

A ***density operator*** is a matrix constructed from an ensemble

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|.$$

7.1.1 Unitary evolution of an ensemble

If we have an ensemble of pure states $\{(p_i, |\psi_i\rangle)\}$, we don't know which of the $|\psi\rangle$ we actually have. If we perform a unitary operation U, we have

$$\{(p_i, |\psi_i\rangle)\} \mapsto \{(p_i, U|\psi_i\rangle)\}.$$

So the resulting density operator becomes

$$\rho \mapsto U\rho U^*.$$

7.1.2 Measurement of an ensemble

Again, suppose we have an ensemble of pure states $\{(p_i, |\psi_i\rangle)\}$. If the actual state was $|\psi_i\rangle$, the probability of obtaining result m after performing a measurement with measurement operators $\{K_x\}$ is

$$\text{Prob}(m|i) = \langle\psi_i|K_m K_m^*|\psi_i\rangle = \text{Tr}[K_m^* K_m |\psi_i\rangle\langle\psi_i|].$$

To obtain the probability of obtaining m if we don't know which i we have,

$$\text{Prob}(m) = \sum_i \text{Prob}(m|i)p_i = \sum_i p_i \,\text{Tr}[K_m^* K_m |\psi_i\rangle\langle\psi_i|] = \text{Tr}[K_m^* K_m \rho].$$

So ρ gives the correct measurement outcome probabilities if the initial state is unknown.

Note 7.2. Any vector $|\psi\rangle$ and any operator A, it holds that $\langle\psi|A|\psi\rangle = \text{Tr}[|\psi\rangle\langle\psi|A]$.

If the initial state was i, the state of the system after obtaining the result m is

$$|\psi_i^{(m)}\rangle = \frac{1}{\|K_m|\psi_i\rangle\|} K_m|\psi_i\rangle$$

where $\|K_m|\psi_i\rangle\| = \sqrt{\langle\psi_i|K_m^* K_m|\psi_i\rangle}$. Define $p_{i|m}$ as the probability that the initial state was i given that the outcome m was observed. Then the resulting ensemble is

$$\left\{\left(p_{i|m}, |\psi_i^{(m)}\rangle\right)\right\}.$$

How does this affect the density matrix? From conditional probabilities, we know that $p_i p_{m|i} = p_m p_{i|m}$, where $p_{i|m} = \text{Prob}(i|m)$ is the probability that we will observe outcome m given that the initial state was i. If we observed m but didn't know which i we began with, the resulting density operator is

$$\begin{aligned}
\rho_m &= \sum_i p_{i|m} |\psi_i^{(m)}\rangle\langle\psi_i^{(m)}| \\
&= \sum_i \frac{p_i p_{m|i}}{p_m} |\psi_i^{(m)}\rangle\langle\psi_i^{(m)}| \\
&= \sum_i \frac{K_m|\psi_i\rangle\langle\psi_i|K_m^*}{\text{Tr}\left[K_m^* K_m \rho\right]} \\
&= \frac{K_m \rho K_m^*}{\text{Tr}\left[K_m^* K_m \rho\right]}.
\end{aligned}$$

Suppose two ensembles are decompositions for the same density operator. I.e., consider two ensembles $\{p_i, |\psi_i\rangle\}$ and $\{q_j, |\phi_j\rangle\}$ such that

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i| = \sum_j p_j |\phi_j\rangle\langle\phi_j|.$$

They have the same measurement outcome probabilities for any possible measurement you could perform, and the reulting state after measurement only depends on ρ, not the specific decomposition. So the two ensembles are completely indisinguishable. All quantum mechanics can be done by considering only the density operator ρ.

7.1.3 Properties of density operators

Proposition 7.3. *For any density operator ρ arising from any ensemble decomposition satisfies the following properties:*

1. $\text{Tr}[\rho] = 1$
2. $\rho \geq 0$
3. $\text{Tr}[\rho^2] = 1 \iff \rho = |\psi\rangle\langle\psi|$

Proof. Note that $\text{Tr}[|\psi\rangle\langle\psi|] = 1$ for any normalized vector. Since all the $|\psi_i\rangle$ are normalized,

$$\text{Tr}\,\rho = \sum_i p_i \,\text{Tr}[|\psi_i\rangle\langle\psi_i|] = 1.$$

Note that ρ must have real eigenvalues, so suppose $|\psi\rangle$ is a normalized eigenvector of ρ. Then

$$\langle\psi|\rho|\psi\rangle = \sum_i p_i \langle\psi||\psi_i\rangle\langle\psi_i||\psi\rangle = \sum_i p_i |\langle\psi|\psi_i\rangle|^2 \geq 0.$$

The proof of 3 left as an exersize. □

Note 7.4. Given a quantum system with associated Hilbert space $\mathbb{C}^n$, states of the system are density operators $\rho\colon \mathbb{C}^n \to \mathbb{C}^n$ (i.e. they are linear operators, or matrices).

7.1.4 Reformulation of Quantum Mechanics

(See slides online.)

7.1.5 The Bloch Sphere

Denote by $\mathcal{H}_2$ the space of 2×2 hermitian matrices. This is a *real* 4-dimensional vector space. Indeed, for $A, B \in \mathcal{H}_2$ and any $a, b \in \mathbb{R}$, it holds that $aA + bB$ is also hermitian. This does not hold if a or b are complex. Furthermore, $\mathcal{H}_2$ is spanned by

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}.$$

Furthermore, these matrices are *orthogonal*. We can construct an inner product on $\mathcal{H}_2$ by

$$\langle A, B \rangle = \operatorname{Tr}[A^* B].$$

(This actually works as an inner product over any space of matrices. Note here that $A^* = A$ since A is hermitian.) The spanning set listed above is orthogonal in the sense that $\langle A, B \rangle = 0$ for any $A \neq B$. We denote these matrices as

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}.$$

These are also known as the *Pauli matrices* if we use the notation $\sigma_1 = Z$, $\sigma_2 = X$, $\sigma_3 = Y$, and $\sigma_0 = I$.

Let $\rho \in \mathcal{H}_2$ be a density matrix. Then $\operatorname{Tr} \rho = 1$. Note that we can write ρ uniquely as

$$\rho = r_0 I + r_1 \sigma_1 + r_2 \sigma_2 + r_3 \sigma_3.$$

Since $\operatorname{Tr} \sigma_i = 0$ unless $i = 0$, it follows that $1 = \operatorname{Tr} \rho = r_0 \operatorname{Tr} I = 2 r_0$ and thus $r_0 = \frac{1}{2}$.

(more...)

7.1.6 The reduced density matrix

(See slindes online...)

8 Lecture 8

(**11 February 2016**)

8.1 Density matrices (continued)

8.1.1 Separable density matrices

Consider a bipartite quantum system with associated Hilbert space $\mathcal{H} = \mathbb{C}^n \otimes \mathbb{C}^m$ (with subsystems corresponding to Alice A and Bob B). A state (i.e. a density operator) of this quantum system is a linear operator

$$\rho_{\mathsf{AB}} \colon \mathcal{H} \longrightarrow \mathcal{H}.$$

Suppose Alice can prepare her subsystem in one of two ways, either as state $\sigma_{\mathsf{A}}^{(1)}$ or $\sigma_{\mathsf{B}}^{(2)}$, which are linear operators on $\mathbb{C}^n$. Similarly, Bob can prepare his subsystem in one of two ways, either $\sigma_{\mathsf{B}}^{(1)}$ or $\sigma_{\mathsf{B}}^{(2)}$, which are linear operators on $\mathbb{C}^m$. As a state of the entire system, Alice and Bob together will either prepare

$$\sigma_{\mathsf{A}}^{(1)} \otimes \sigma_{\mathsf{B}}^{(1)} \qquad \text{or} \qquad \sigma_{\mathsf{A}}^{(2)} \otimes \sigma_{\mathsf{B}}^{(2)}$$

Alice and Bob can then prepare the ***mixture*** of these two states in the following manner.

- Alice flips a coin, which produces heads with probability p and tails with probability $1 - p$.
- Alice communicates the result to Bob.
- If heads was obtained, Alice and Bob prepare $\sigma_{\mathsf{A}}^{(1)} \otimes \sigma_{\mathsf{B}}^{(1)}$. If tails was obtained, Alice and Bob prepare $\sigma_{\mathsf{A}}^{(2)} \otimes \sigma_{\mathsf{B}}^{(2)}$.
- Alice and Bob subsequently 'forget' which result was obtained.
- The resulting state can be described as

$$\rho = p\, \sigma_{\mathsf{A}}^{(1)} \otimes \sigma_{\mathsf{B}}^{(1)} + (1-p) \sigma_{\mathsf{A}}^{(2)} \otimes \sigma_{\mathsf{B}}^{(2)}.$$

Likewise, we can consider any probability distribution $\{p_j\}$ and any set of states of the form $\{\sigma_{\mathsf{A}}^{(j)} \otimes \sigma_{\mathsf{B}}^{(j)}\}$ that Alice and Bob can prepare separately. Using the ensemble $\{(p_j, \sigma_{\mathsf{A}}^{(j)} \otimes \sigma_{\mathsf{B}}^{(j)})\}$, Alice and Bob can 'forget' which outcome j was obtained and the resulting state is the density operator

$$\rho_{\mathsf{AB}} = \sum_j p_j \sigma_{\mathsf{A}}^{(j)} \otimes \sigma_{\mathsf{B}}^{(j)}. \tag{8.1}$$

Definition 8.1. A density operator $\rho_{\mathsf{AB}} \colon \mathcal{H} \longrightarrow \mathcal{H}$ is ***separable*** if and only if it can be written in the form of (8.1).

Lemma 8.2. *A density operator ρ_{AB} is separable if and only if it can be written as*

$$\rho_{\mathsf{AB}} = \sum_j p_j |v_j\rangle\langle v_j| \otimes |w_j\rangle\langle w_j|$$

for some collections of vectors $\{|v_j\rangle\}$ and $\{|w_j\rangle\}$ and probabilities $\{p_j\}$.

Note that the $\{|v_j\rangle\}$ and $\{|w_j\rangle\}$ do not have to be orthogonal. They are just any collection of unit vectors in the appropriate space!

Proof. Suppose ρ_{AB} is separable. It follows that

$$\rho_{\mathsf{AB}} = \sum_x q_x \sigma_{\mathsf{A}}^{(x)} \otimes \sigma_{\mathsf{B}}^{(x)} = \sum_x q_x \underbrace{\left(\sum_y t_{y|x} |\psi_{y|x}\rangle\langle\psi_{y|x}| \right)}_{\sigma_{\mathsf{A}}^{(x)}} \otimes \underbrace{\left(\sum_z s_{z|x} |\phi_{z|x}\rangle\langle\phi_{z|x}| \right)}_{\sigma_{\mathsf{B}}^{(x)}}$$

$$= \sum_{x,y,z} q_x t_{y|x} s_{z|x} |\psi_{y|x}\rangle\langle\psi_{y|x}| \otimes |\phi_{z|x}\rangle\langle\phi_{z|x}|.$$

Define the index j to be the tuple $j = (x, y, z)$ where j runs over all possible values for x, y, and z. Then each p_j is defined as

$$p_j = q_x t_{y|x} s_{z|x} \geq 0$$

and it holds that $\sum_j p_j = \sum_{x,y,z} q_x t_{y|x} s_{z|x} = 1$. The vectors $|v_j\rangle$ and $|w_j\rangle$ are defined as

$$|v_j\rangle = |\psi_{y|x}\rangle \qquad \text{and} \qquad |w_j\rangle = |\phi_{z|x}\rangle,$$

so there are many js that have the same $|v_j\rangle$ and $|w_j\rangle$. □

Almost by definition, separable states are states that have no entanglement. (Or perhaps we can define entanglement this way).

8.1.2 Classical-quantum states

We now examine a certain subset of the separable states.

Definition 8.3. Consider a *classical* random variable X on a system X. The possible outcomes x are associated with a set of orthonormal vectors $\{|x\rangle\}$. The ***classical-quantum*** states are those that can be written as

$$\rho_{\mathsf{XB}} = \sum_x p_x |x\rangle\langle x|_{\mathsf{X}} \otimes \sigma_{\mathsf{B}}^{(x)}$$

where the states $\{\sigma_{\mathsf{B}}^{(x)}\}$ are arbitrary density operators of Bob's system.

Classical-quantum states are those that are only 'classically correlated'. The states of a classical system cannot be superimposed, because it is not quantum. The only allowable states are the orthonormal $|x\rangle$.

We can think of it as a die[1]. Alice rolls a die, and tells Bob what the outcome was. Bob prepares a corresponding state. Alice and Bob then 'forget' which outcome the die had. Alice can then later 're-examine' the die by looking at it. If she does, then she knows exactly what state Bob's system is in.

Definition 8.4. A ***classical-classical*** state is a density operator that can be written as

$$\rho = \sum_{x,y} p_{xy} |x\rangle\langle x| \otimes |y\rangle\langle y|$$

for some orthonormal sets of vectors $\{|x\rangle\}$ and $\{|y\rangle\}$.

We can imagine a classical-classical state simply as a pair of random variables X and Y that are correlated, where the joint probability distribution is $\{p_{xy}\}$.

[1] Gilad says 'a dice' here, but this is incorrect. The correct singular form is 'one die' and the plural form is 'many dice'. Look it up in any dictionary!

8.1.3 Entanglement of mixed states

Recall that, for pure states $|\psi\rangle$ and $|\phi\rangle$ in $\mathbb{C}^n \otimes \mathbb{C}^m$, it holds that

$$|\psi\rangle \xrightarrow{\text{LOCC}} |\phi\rangle \iff \vec{p}_\psi \prec \vec{p}_\phi.$$

What about necessary and sufficient conditions for converting arbitrary mixed states $\rho_{\mathsf{AB}} \xrightarrow{\text{LOCC}} \sigma_{\mathsf{AB}}$? The states ρ can be written in infinitely many different decompositions

$$\rho_{\mathsf{AB}} = \sum_x p_x |\psi_x\rangle\langle\psi_x|.$$

There are no simple ways of writing necessary and sufficient conditions for $\rho_{\mathsf{AB}} \xrightarrow{\text{LOCC}} \sigma_{\mathsf{AB}}$.

For certain cases, we do know. If σ_{AB} is separable, we can always convert any ρ_{AB} into σ_{AB} via LOCC. The method is simple: Alice and Bob can 'throw away' ρ_{AB} and just prepare σ_{AB} as discussed earlier.

But for arbitrary states that are not separable, the question is incredibly difficult to answer. In fact, the number of rounds of communication that are allowed changes whether or not the conversion can take palce. That is, if Alice and Bob can only perform *single*-round LOCC (that is, Alice performs a measurement and tells Bob what it is, then Bob performs a measurement, and that's it), then the states that Alice and Bob can obtain from ρ_{AB} might be different than the ones obtainable using an unlimited number of communication rounds (or two rounds, etc...).

8.1.4 Entanglement monotones – necessary conditions

Although it is impossible to have a simple necessary AND sufficient condition for determining if $\rho_{\mathsf{AB}} \xrightarrow{\text{LOCC}} \sigma_{\mathsf{AB}}$ is possible, we can find easy *necessary* conditions for $\rho_{\mathsf{AB}} \xrightarrow{\text{LOCC}} \sigma_{\mathsf{AB}}$ to be possible.

Consider (as always) a bipartite system $\mathbb{C}^n \otimes \mathbb{C}^m$ with subsystems labelled A and B. Let $\mathrm{H}^{\mathsf{AB}}_{nm,+,1}$ denote the set of all density operators on the system:

$$\mathrm{H}^{\mathsf{AB}}_{nm,+,1} = \left\{\rho \colon \mathbb{C}^n \otimes \mathbb{C}^m \to \mathbb{C}^n \otimes \mathbb{C}^m \,|\, \rho^* = \rho,\, \rho \geq 0,\, \operatorname{Tr}\rho = 1\right\}.$$

(This is the set of nm-dimensional hermitian matrices that are positive and have unit trace.) The set of *all* bipartite density operators in all dimensions will be denoted $\mathrm{H}^{\mathsf{AB}}_{+,1}$ (which I suppose we can think of as the disjoint union

$$\mathrm{H}^{\mathsf{AB}}_{+,1} = \coprod_{n,m=2}^{\infty} \mathrm{H}^{\mathsf{AB}}_{nm,+,1}$$

of all of these.)

Definition 8.5. A function $E \colon \mathrm{H}^{\mathsf{AB}}_{+,1} \longrightarrow \mathbb{R}^+$ is an ***entanglement monotone*** if

$$\rho_{\mathsf{AB}} \xrightarrow{\text{LOCC}} \sigma_{\mathsf{AB}} \implies E(\rho_{\mathsf{AB}}) \geq E(\sigma_{\mathsf{AB}}).$$

Example 8.6. For pure states, recall that $|\psi\rangle \xrightarrow{\text{LOCC}} |\phi\rangle$ if and only if $E_k(\vec{p}_\psi) \geq E_k(\vec{p}_\phi)$ holds for all $k = 1, \ldots, n$, where E_k are the functions

$$E_k(\vec{p}) := \sum_{j=k}^{n} \left(\vec{p}^{\downarrow}\right) = j.$$

Hence, for pure states, the E_k's are entanglement monotones.

Note 8.7. In fact, any function that is *Shur-concave* on vectors is an entanglement monotone for pure states. A function f of probability vectors is said to be ***Schur-concave*** if

$$\vec{p} \prec \vec{q} \iff f(\vec{p}) \geq f(\vec{q})$$

hodls for all probability vectors $\vec{p}$ and $\vec{q}$.

Why are they called Shur-*concave*? In fact, any function on probability vectors that is both *symmetric* and *concave* is also Schur-concave (but not vice versa). Recall also that $\vec{p} \prec \vec{q}$ holds if and only if $\vec{q} = D\vec{p}$ for some doubly stochastic matrix D, i.e.

$$\vec{q} = D\vec{p} = \sum_j t_j \Pi_j \vec{p}$$

for some probabilities $\{t_j\}$ and projection matrices Π_j. And this resembles the definition of a regular convex combination.

Perhaps the most well-known and most often used Schur-concave function in information theory is the Shannon entropy:

$$S(\vec{p}) = \sum_j p_j \log p_j.$$

Exercise 8.8. Find an example of a function that is Shur-concave but not concave.

Let f be a Schur-concave function and define an entanglement monotone on pure states as

$$E_f(\psi) = f(\vec{p}_\psi).$$

Then this entanglement monotone can be extended to mixed states in the following manner. For a density operator ρ_{AB}, we define

$$E(\rho_{\mathsf{AB}}) = \min_{\{(p_j, |\psi_j\rangle)\}} \sum_j p_j E_f(\psi_j)$$

where the minimum is taken over all ensembles that are decompositions $\{(p_j, |\psi_j\rangle)\}$ of ρ_{AB},

$$\rho_{\mathsf{AB}} = \sum_j p_j |\psi_j\rangle\langle\psi_j|.$$

This computation is unfortunately very difficult to perform on a computer in general. In fact, determining whether or not a state is separable at all is *as hard as* computing the value of $E(\rho)$ above.

8.2 Detecting if a state is separable

When we are 'given' a state, for example of a system of two qubits $\rho_{\mathsf{AB}} \colon \mathbb{C}^2 \otimes \mathbb{C}^2 \to \mathbb{C}^2 \otimes \mathbb{C}^2$, we are given a matrix with elements

$$\rho = \begin{bmatrix} \rho_{11} & \rho_{12} & \cdots \\ \vdots & \ddots & \\ \vdots & & \end{bmatrix}.$$

We want to determine if the state is separable. That is, we want to know if it can be written as

$$\rho = \sum_j p_j \sigma_{\mathsf{A}}^{(j)} \otimes \sigma_{\mathsf{B}}^{(j)}.$$

Can we come up with some necessary criterion for determining if such a state is separable?

The transpose operator is defined as $T(\sigma) = \sigma^T$ for matrices σ. Consider density operators on a bipartite system $\mathbb{C}^n \otimes \mathbb{C}^m$. Define the *partial transpose* operator in the following way:

$$I \otimes T\left(\sigma \otimes \tau\right) = \sigma \otimes \tau^T$$

and extended linearly to any linear combination of product operators. Suppose that ρ_{AB} is separable. It follows that $I \otimes T\left(\rho_{\mathsf{AB}}\right)$ is a positive operator:

$$I \otimes T(\rho_{\mathsf{AB}}) = I \otimes T\left(\sum_j p_j \sigma_{\mathsf{A}}^{(j)} \otimes \sigma_{\mathsf{B}}^{(j)}\right) = \sum_j p_j \sigma_{\mathsf{A}}^{(j)} \otimes \left(\sigma_{\mathsf{B}}^{(j)}\right)^T \geq 0.$$

Each $\sigma_{\mathsf{A}}^{(j)}$ and $\sigma_{\mathsf{B}}^{(j)}$ is a positive operator and $\sigma^T \geq 0$ holds whenever $\sigma \geq 0$. The tensor product of positive operators is positive, and the sum sum of positive operators is also positive. Therefore, the partial transpose applied to any separable state will always be a positive operator.

Let's look at the partial transpose applied to another operator $|\psi\rangle\langle\psi|$ when $|\psi\rangle$ is the Bell state

$$|\psi\rangle = \frac{1}{\sqrt{t}}\left(|00\rangle + |11\rangle\right).$$

Note that $I \otimes T\left(|ij\rangle\langle kl|\right) = |i\rangle\langle k| \otimes T(|j\rangle\langle l|) = |il\rangle\langle kj|$. Now

$$\begin{aligned} I \otimes T\left(|\psi\rangle\langle\psi|\right) &= \frac{1}{2} I \otimes T\Big(|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|\Big) \\ &= \frac{1}{2}\Big(|00\rangle\langle 00| + |01\rangle\langle 10| + |10\rangle\langle 01| + |11\rangle\langle 11|\Big). \end{aligned}$$

In matrix form, this is

$$I \otimes T\left(|\psi\rangle\langle\psi|\right) = \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \ngeq 0,$$

hence the partial transpose of $|\psi\rangle\langle\psi|$ is not positive definite (since the matrix has at least one negative eigenvalue). So we see that $|\psi\rangle$ is in fact not separable.

Note 8.9. For bipartite systems $\mathbb{C}^n \otimes \mathbb{C}^m$, the partial transpose condition is actually a necessary *and* sufficient condition for determining if a density operator is separable if and only if $nm \leq 6$. (Proving this will be one of the possible end-of-course projects.)

9 Lecture 9

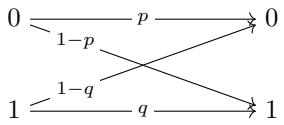
(**23 February 2016**)

9.1 Quantum channels

9.1.1 Classical noise

How do we model the transmission of classical information? Suppose we have some physical means of transmitting bits (a fiber optic, an antenna, etc) which we will call a *channel*. Consider an information *source* X that has some prior probability distribution that yields either 0 or 1. We want to transmit this bit of information using our channel, but the channel is *noisy* (it is not perfect). The *receiver* receives the bit Y, which may not be the same bit that the sender sent! We model it like this:

$$X \longrightarrow [\text{channel}] \longrightarrow Y$$



where p is the probability that a 0 is transmitted perfectly and q is the probability that a 1 is transmitted perfectly.

What is the probability that the receiver receives the correct information that the sender intended to send? Well, we can look at the following probabilities:

$$\begin{aligned}
\Pr(Y=0|X=0) &= p \\
\Pr(Y=1|X=0) &= 1-p \\
\Pr(Y=1|X=1) &= q \\
\Pr(Y=0|X=1) &= 1-q
\end{aligned}$$

How can we describe the probability distribution of the receiver? It looks like this:

$$\Pr(Y=y) = \sum_{x=0,1} \Pr(Y=y|X=x)\Pr(X=x).$$

If we write $s_y = \Pr(Y=y)$ and $t_x = \Pr(X=x)$, we have that the probability vectors corresponding to the distributions of X and Y can be written in matrix form as $\vec{s} = E\vec{t}$, or

$$\begin{pmatrix} s_0 \\ s_1 \end{pmatrix} = \begin{pmatrix} p & 1-q \\ 1-p & q \end{pmatrix} \begin{pmatrix} t_0 \\ t_1 \end{pmatrix},$$

where $E = \left(\begin{smallmatrix} p & 1-q \\ 1-p & q \end{smallmatrix}\right)$ is the *transition matrix*. The matrix elements of E are

$$E_{xy} = \Pr(Y=y|X=x)$$

Note that this matrix is *column* stochastic, since $\sum_y E_{xy} = 1$ for all x.

What is X? It is a random source. The sender can be thought of as sampling the distribution p_X over and over again, where each x occurs with probability $p_X(x) = \Pr(X=x)$ (where the outcomes don't have to be just 0 and 1; there could be more). The sender then sends this information through the channel, which may not transmit the information perfectly. We can think of transmitting a message (in English) by looking at the individual letters of the message. The letters may be thought of as occuring 'randomly'. For example, there is more probability that the letter 's' will be typed than the letter 'z' for example. In fact, we can assign a probability of each letter occuring in "standard English" and use this probability distribution on the letters as the distribution of the source.

9.1.2 The quantum case

Suppose we have a sender who wants to send a *quantum* state ρ to a receiver through a *quantum* channel. If $\rho \in \mathrm{L}(\mathbb{C}^n)$ is the input state and $\sigma \in \mathrm{L}(\mathbb{C}^m)$ is the output state, we have

$$\rho \xrightarrow{\mathcal{E}} \sigma$$

$$\sigma = \mathcal{E}(\rho)$$

One trivial example of a channel is just the identity mapping $\mathcal{I}(\rho) = \rho$. But there are other channels that are more interesting.

What are the conditions that the transformation $\mathcal{E}$ must satisfy? If ρ is a density operator, the output $\sigma = \mathcal{E}(\rho)$ must also be a density operator. Hence, if $\rho \geq 0$ with $\operatorname{Tr}\rho = 1$, then $\mathcal{E}(\rho) \geq 0$ and $\operatorname{Tr}\mathcal{E}(\rho) = 1$ must also hold. So the map $\mathcal{E}$ must be

- ***trace-preserving***: $\operatorname{Tr}[\mathcal{E}(\rho)] = \operatorname{Tr}\rho = 1$
- ***positive***: $\mathcal{E}(\rho) \geq 0$ if $\rho \geq 0$

It must also be the case that $\mathcal{E}$ is ***linear***. That is, if we have an ensemble $\{p_x, \rho_x\}$, then

$$\mathcal{E}\left(\sum_x p_x \rho_x\right) = \sum_x p_x \mathcal{E}(\rho_x)$$

must hold. But that's not all! The map $\mathcal{E}$ must also be ***completely positive***. That is, suppose we have the channel $\mathcal{E}$ that only acts on one part (A) of a composite system (AB), as in the following diagram:

$$\rho_{\mathsf{AB}} \left\{ \begin{array}{c} \xrightarrow{\mathsf{A}} \mathcal{E} \longrightarrow \\ \xrightarrow[\mathsf{B}]{\qquad} \end{array} \right\} \sigma_{\mathsf{AB}}$$

which we can write as $\sigma_{\mathsf{AB}} = \left(\mathcal{E}^{\mathsf{A}} \otimes \mathcal{I}^{\mathsf{B}\to\mathsf{B}}\right)(\rho_{\mathsf{AB}})$. For any state $\rho_{\mathsf{AB}} \geq 0$ of the composite system, the result

$$\left(\mathcal{E}^{\mathsf{A}} \otimes \mathcal{I}^{\mathsf{B}\to\mathsf{B}}\right)(\rho_{\mathsf{AB}}) \geq 0$$

must also be positive. And this must hold for *any* possible state on any possible extension of A to AB.

Example 9.1. Here we give an example of a mapping that is trace-preserving and positive, but not completely positive. Consider the transpose map $\mathcal{E}(\rho) = \rho^T$. Recall from a previous lecture that

$$\mathcal{E} \otimes I(|\phi^+\rangle\langle\phi^+|) \ngeq 0$$

where $|\phi^+\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$.

Definition 9.2. Let n and m be integers. Recall that H_n is the real vector space of $n \times n$ hermitian matrices. Let $\mathcal{E}$ be a linear mapping that acts on H_n. Then $\mathcal{E}$ is said to be m-***positive*** if the mapping $\mathcal{E} \otimes I_{\mathrm{H}_m}$ on $\mathrm{H}_n \otimes \mathrm{H}_m$ is a positive map, and $\mathcal{E}$ is ***completely positive*** if it is m-positive for every m.

Proposition 9.3. *It holds that $\mathcal{E} : \mathrm{H}_n \to \mathrm{H}_n$ is completely positive if and only if it is n-positive.*

Note that it's possible for a map to be m-positive for every $m < n$, even if the map is not completely positive. However, if $\mathcal{E} : \mathrm{H}_n \to \mathrm{H}_n$ is n-positive then it is m-positive for every $m \geq n$.

9.1.3 Definition and examples of quantum channels

Definition 9.4. A ***quantum channel*** is a linear map $\mathcal{E} : \mathbb{C}^{n\times n} \to \mathbb{C}^{m\times m}$ that satisfies the following properties:

(a) $\mathcal{E}$ is completely positive (CP)

(b) $\mathcal{E}$ is trace-preserving (TP)

(i.e. $\mathcal{E}$ is a linear CPTP map).

Example 9.5. Examples of quantum channels.

1. A *unitary channel* is a linear mapping of the form $\mathcal{E}(\rho) = U\rho U^*$ for some unitary operator U. This is clearly trace-preserving, since $\mathrm{Tr}[U\rho U^*] = \mathrm{Tr}[\rho U^* U] = \mathrm{Tr}\,\rho$, and also clearly positive, since the specturm of ρ will be the same as $U\rho U^*$ for any unitary U. (It's also clearly completely positive...)

2. A *replacement* channel is a mapping of the form

$$\mathcal{E}(A) = \mathrm{Tr}(A)\sigma$$

for all $A \in \mathbb{C}^{n\times n}$, where $\sigma \in \mathrm{H}_{m,+,1}$ is some density operator (on a possibly different space than the input space). This is clearly CPTP.

3. A *classical-quantum* channel takes a classical input and outputs a quantum state. For example,

$$\mathcal{E}(x) = \rho_x, \qquad \forall x \in \{1, 2, \ldots, n\}.$$

We can think of this as a honest-to-goodness quantum channel in the following way. Consider the space $\mathcal{K} = \mathrm{span}\{|x\rangle\langle x|\}$ where $\{|x\rangle\}$ is an orthonormal basis of $\mathbb{C}^n$, and define a map $\mathcal{E} : \mathcal{K} \to \mathrm{H}_m$ by $\mathcal{E}(|x\rangle\langle x|) = \rho_x$. By linearity, for $\sum_x p_x |x\rangle\langle x| \in \mathcal{K}$, we have

$$\mathcal{E}\left(\sum_x p_x |x\rangle\langle x|\right) = \sum_x p_x \mathcal{E}(|x\rangle\langle x|) = \sum_x p_x \rho_x.$$

9.2 Representation and characterizations of CPTP maps

9.2.1 The natural representation

(This is the least useful representation that we will consider.) Consider a linear map $\mathcal{E} : \mathbb{C}^{n\times n} \to \mathbb{C}^{m\times m}$. Note that $\mathbb{C}^{n\times n}$ (the space of $n \times n$ matrices with complex entries) is isomorphic to $\mathbb{C}^{n^2}$ (the regular n^2-dimensional vector space). So there is an isomorphism

$$\mathrm{vec} : \mathbb{C}^{n\times n} \to \mathbb{C}^{n^2}$$

which 'vectorizes' $n \times n$ matrices into n^2-dimensional vectors. Recall that we can define an inner product on $\mathbb{C}^{n\times n}$ by

$$\langle A, B\rangle = \mathrm{Tr}[A^* B].$$

Let $\sigma = \mathcal{E}(\rho)$. Under the isomorphism vec, we get $\mathrm{vec}(\sigma) = \mathrm{vec}(\mathcal{E}(\rho))$. Since $\mathcal{E}$ is a linear map, there exists some matrix $A_{\mathcal{E}} \in \mathbb{C}^{m^2\times n^2}$ such that

$$\mathrm{vec}(\sigma) = A_{\mathcal{E}}\,\mathrm{vec}(\rho).$$

More explicitly, denote $E_{xy} = |x\rangle\langle y|$ for every $x, y \in \{1, \ldots, n\}$, where $|x\rangle$ and $|y\rangle$ are standard basis elements of $\mathbb{C}^n$. Then the E_{xy} span the entire space of $n \times n$ matrices. Similarly, define $F_{wz} = |w\rangle\langle z|$

for every $z, w \in \{1, \ldots, n\}$, where $|w\rangle$ and $|z\rangle$ are standard basis elements of $\mathbb{C}^m$. These matrices form an orthonormal basis of $\mathbb{C}^{n\times n}$ and $\mathbb{C}^{m\times m}$ respectively, since

$$\langle E_{xy}, E_{x'y'}\rangle = \mathrm{Tr}[E_{yx}E_{x'y'}] = \delta_{xx'}\delta_{yy'} \qquad \text{and} \qquad \langle F_{wz}, F_{w'z'}\rangle = \mathrm{Tr}[F_{zw}F_{w'z'}] = \delta_{ww'}\delta_{zz'}.$$

From the vec isomorphism, we also have that $\{\mathrm{vec}(E_{xy})\}$ and $\{\mathrm{vec}(F_{wz})\}$ are orthonormal bases of $\mathbb{C}^{n^2}$ and $\mathbb{C}^{m^2}$ respectively. We can find the matrix elements of $A_{\mathcal{E}}$ in these bases. Every $\rho \in \mathbb{C}^{n\times n}$ can be written as

$$\rho = \sum_{x,y} \rho_{xy} E_{xy}$$

where $\rho_{xy} = \langle E_{xy}, \rho\rangle$. Similarly, with $\sigma_{wz} = \langle F_{wz}, \sigma\rangle$, we can write

$$\sigma = \sum_{w,z} \sigma_{wz} F_{wz}.$$

This allows us to find the matrix elements of $A_{\mathcal{E}}$ in the following manner. The matrix elements of σ are

$$\sigma_{wz} = \langle F_{wz}, \mathcal{E}(\rho)\rangle = \sum_{x,y} \rho_{xy}\langle F_{wz}, \mathcal{E}(E_{xy})\rangle = \sum_{x,y} \rho_{xy}\,\mathrm{Tr}[F_{wz}^*\mathcal{E}(E_{xy})]$$

and so the matrix elements of $E_{\mathcal{E}}$ are

$$(A_{\mathcal{E}})_{xz,xy} = \langle F_{wz}, \mathcal{E}(E_{xy})\rangle = \mathrm{Tr}[F_{wz}^*\mathcal{E}(E_{xy})].$$

Exercise 9.6. Let $\mathcal{E} : \mathbb{C}^{n\times n} \to \mathbb{C}^{m\times m}$ be a linear map. Check that $A_{\mathcal{E}^*} = (A_{\mathcal{E}})^*$, where $\mathcal{E}^*$ is the *dual* map to $\mathcal{E}$. (It is the unique linear map such that

$$\langle A, \mathcal{E}(B)\rangle = \langle \mathcal{E}^*(A), B\rangle$$

holds for all $B \in \mathbb{C}^{n\times n}$ and $A \in \mathbb{C}^{m\times m}$.) It follows that $A_{\mathcal{E}}$ is a hermitian matrix if $\mathcal{E}^* = \mathcal{E}$.

Note 9.7. This representation helps us understand quantum channels as linear mappings (i.e. matrices). Why is this representation not useful? Because the matrix representation $A_{\mathcal{E}}$ does not help us determine if the mapping $\mathcal{E}$ is trace-preserving or completely positive. The other representations that we will discuss later will be more useful.

10 Lecture 10

(**25 February 2016**)

10.1 Representations of Quantum Channels (cont.)

Recall that a quantum channel is a linear CPTP map $\mathcal{E} : \mathrm{H}_n \longrightarrow \mathrm{H}_m$.

10.1.1 The Choi representation

Definition 10.1. Let $\mathcal{E} : \mathrm{H}_n^{\mathsf{A}'} \longrightarrow \mathrm{H}_m^{\mathsf{A}}$ be a CPTP map. Then the ***Choi matrix*** of $\mathcal{E}$ is $J(\mathcal{E}) \in \mathrm{H}_{mn}^{\mathsf{AR}}$ (or also written as $\sigma_{\mathsf{AR}} \in \mathrm{H}_{mn}^{\mathsf{AR}}$) and is defined as

$$J(\mathcal{E}) = \sigma_{\mathsf{AR}} = (\mathcal{E} \otimes \mathrm{id}_{\mathsf{R}}) \left(|\phi^+\rangle\langle\phi^+|_{\mathsf{A}'\mathsf{R}} \right)$$

where $\mathcal{E}$ is a mapping on systems $\mathsf{A}' \to \mathsf{A}$, id_{R} is the identity linear map on R, and

$$|\phi^+\rangle = \sum_{x=1}^{n} |x\rangle_{\mathsf{A}'} |x\rangle_{\mathsf{R}}$$

is an unnormalized vector. Using the notation $E_{xy} = |x\rangle\langle y|$, we have

$$|\phi^+\rangle\langle\phi^+| = \sum_{x,y} |xx\rangle\langle yy| = \sum_{x,y} E_{xy} \otimes E_{xy}$$

and

$$J(\mathcal{E}) = \sigma_{\mathsf{AB}} = \sum_{x,y} \mathcal{E}(E_{xy}) \otimes E_{xy}.$$

This map J is actually a bijection. The notation 'R' denotes a 'reference system' that is independent of the input space A' and output space A. The dimension of R must be n, the same as the dimension of the input space A'.

What happens when we take the partial trace of $J(\mathcal{E})$ with respect to A? We see that

$$\mathrm{Tr}_{\mathsf{A}}\, J(\mathcal{E}) = \mathrm{Tr}_{\mathsf{A}}\, \sigma_{\mathsf{AR}} = \sum_{x,y} \mathrm{Tr}\left(\mathcal{E}(E_{xy})\right) E_{xy} = \sum_{xy} \delta_{xy} E_{xy} = \sum_{x} E_{xx} = I_{\mathsf{R}},$$

where I_{R} here is the identity matrix on the space R.

Now we show that J is actually a bijection between channels $\mathcal{E} : \mathrm{H}_n^{\mathsf{A}'} \longrightarrow \mathrm{H}_m^{\mathsf{A}}$ and density operators on $\mathrm{H}_{mn}^{\mathsf{AR}}$. That is, to every density matrix $\sigma_{\mathsf{AR}} \in \mathrm{H}_{mn}^{\mathsf{AR}}$ we can define a channel in the following manner. For every $\rho \in \mathrm{H}_n^{\mathsf{A}}$ define

$$\mathcal{E}(\rho) = \mathrm{Tr}_{\mathsf{R}} \left[\sigma_{\mathsf{AR}} (I_{\mathsf{A}} \otimes \rho^T) \right],$$

where ρ^T can be considered as an operator on $\mathrm{H}_n^{\mathsf{R}}$ since $\dim \mathsf{R} = \dim \mathsf{A} = n$. Then

$$\begin{aligned}
\mathrm{Tr}_{\mathsf{A}}\left[\sigma_{\mathsf{AR}}(I_{\mathsf{A}} \otimes \rho^T)\right] &= \sum_{x,y} \mathrm{Tr}_{\mathsf{R}}\left[(\mathcal{E}(E_{xy}) \otimes E_{xy})\,(I \otimes \rho^T)\right] \\
&= \sum_{x,y} \mathrm{Tr}_{\mathsf{R}}\left[\mathcal{E}(E_{xy}) \otimes E_{xy}\rho^T\right] \\
&= \sum_{x,y} \mathcal{E}(E_{xy})\,\mathrm{Tr}\left(E_{xy}\rho^T\right) \\
&= \sum_{x,y} \mathcal{E}(E_{xy})\,\mathrm{Tr}\left(|x\rangle\langle y|\rho^T\right) \\
&= \sum_{x,y} \mathcal{E}(E_{xy})\langle y|\rho^T|x\rangle
\end{aligned}$$

where we note that $\mathrm{Tr}[|x\rangle\langle y|A] = \langle y|A|x\rangle$, since

$$\mathrm{Tr}[|x\rangle\langle y|A] = \sum_{x'} \langle x'|\,(|x\rangle\langle y|A)\,|x'\rangle = \sum_{x'} \delta_{xx'}\langle y|A|x'\rangle = \langle x|A|y\rangle.$$

Continuing, we see

$$\sum_{x,y} \mathcal{E}(E_{xy})\langle y|\rho^T|x\rangle = \sum_{x,y} \mathcal{E}(E_{xy})\,\underbrace{\langle x|\rho|y\rangle}_{\rho_{xy}} = \mathcal{E}\left(\sum_{x,y} \rho_{xy}E_{xy}\right) = \mathcal{E}(\rho).$$

Why is this representation useful? The Choi isomorphism J can help determine when a map $\mathcal{E}$ is completely positive and trace-preserving.

Theorem 10.2. *A linear map $\mathcal{E} : \mathrm{H}_n \longrightarrow \mathrm{H}_m$ is CPTP if and only if it holds that*

$$J(\mathcal{E}) \geq 0 \qquad \textit{and} \qquad \mathrm{Tr}_{\mathsf{A}}\, J(\mathcal{E}) = I_{\mathsf{R}}.$$

Proof. Suppose that $\mathcal{E}$ is CPTP. Then $\mathcal{E} \otimes \mathrm{id}\,(|\phi^+\rangle\langle\phi^+|) \geq 0$ because $\mathcal{E}$ is completely positive (and $|\phi^+\rangle\langle\phi^+|$ is a positive $n^2 \times n^2$ matrix[1]). Also note that $\mathrm{Tr}_{\mathsf{A}}\, J(\mathcal{E}) = I_{\mathsf{R}}$ (which we proved earlier).

Now suppose that $J(\mathcal{E}) \geq 0$ and that $\mathrm{Tr}_{\mathsf{A}}\, J(\mathcal{E}) = I_{\mathsf{R}}$. We need to show that $(\mathcal{E} \otimes \mathrm{id})(\rho_{\mathsf{AR}}) \geq 0$ holds for every $\rho_{\mathsf{AB}} \geq 0$. Every positive semi-definite operator ρ_{AB} can be written as some linear combination of rank-1 positive semi-definite operators

$$\rho_{\mathsf{AB}} = \sum_j |\psi_j\rangle\langle\psi_j|,$$

where $|\psi_j\rangle$ is a (not necessarily normalized) vector on the system $\mathsf{A'R}$. Furthermore, each such vector can always be written as

$$|\psi_j\rangle = \sum_{x,y} M_{xy}^{(j)}|x\rangle|y\rangle = \sum_x (I \otimes M_j|xx\rangle$$

for some $M_{xy}^{(j)} \in \mathbb{C}$, and we denote M_j as the $n \times n$ matrix whose entries are $M_{xy}^{(j)}$

$$M_j|x\rangle = \sum_{y=1}^{n} M_{xy}^{(j)}|y\rangle.$$

[1]What are the eigenvalues of $|\phi^+\rangle\langle\phi^+|$? It is a rank-1 matrix with one eigenvalue equal to $\||\phi^+\rangle\| = n$. The remaining $n^2 - 1$ eigenvalues are all equal to zero.

Now, what is $\mathcal{E} \otimes \mathrm{id}(\rho_{\mathsf{AB}})$, for any arbitrary positive semidefinite operator ρ_{AR}?

$$\begin{aligned}(\mathcal{E} \otimes \mathrm{id})(\rho_{\mathsf{AB}}) &= (\mathcal{E} \otimes \mathrm{id}) \left(\sum_j I \otimes M_j |\phi^+\rangle\langle\phi^+| I \otimes M_j^* \right) \\ &= \sum_j (I \otimes M_j) \Big(\underbrace{(\mathcal{E} \otimes \mathrm{id}) \left(|\phi^+\rangle\langle\phi^+| \right)}_{J(\mathcal{E})} \Big) (I \otimes M_j^*)\end{aligned}$$

where we note that the different parts of the above equation 'commute' in the following sense. Suppose we defined a map $\eta(X) = MXM^*$ for some M. Then it is clear that

$$(\mathcal{E} \otimes \mathrm{id})(\mathrm{id} \otimes \eta)(\sigma) = (\mathcal{E} \otimes \eta)(\sigma) = (\mathrm{id} \otimes \eta)(\mathcal{E} \otimes \mathrm{id})(\sigma)$$

holds for all σ.

Continuing, we see that

$$(\mathcal{E} \otimes \mathrm{id})(\rho_{\mathsf{AB}}) = \sum_j \tilde{M}_j J(\mathcal{E}) \tilde{M}_j^* \geq 0, \tag{10.1}$$

where we define the matrices $\tilde{M}_j = I \otimes M_j$. The positivity of the matrix in (10.1) is due to Lemma 10.3 below.

We now need to show that $\mathcal{E}$ is trace preserving. Let $\rho \in \mathrm{H}_n$ be an arbitrary density operator. From the definition of $\mathcal{E}$ obtained from σ_{AB}, we have

$$\begin{aligned}\mathrm{Tr}\, \mathcal{E}(\rho) &= \mathrm{Tr}_{\mathsf{A}} \Big(\underbrace{\mathrm{Tr}_{\mathsf{R}} \left[\sigma_{\mathsf{AR}} (I \otimes \rho^T) \right]}_{\mathcal{E}(\rho)} \Big) \\ &= \mathrm{Tr}_{\mathsf{R}} \left(\mathrm{Tr}_{\mathsf{A}} \left[\sigma_{\mathsf{AR}} (I \otimes \rho^T) \right] \right) \\ &= \mathrm{Tr}_{\mathsf{R}} \left(\mathrm{Tr}_{\mathsf{A}} \left[\sigma_{\mathsf{AR}} (I \otimes I) \right] \rho^T \right) \\ &= \mathrm{Tr}_{\mathsf{R}}\, \rho^T = \mathrm{Tr}\, \rho\end{aligned}$$

(I may have missed something here.... he kind of rushed at the end). □

Lemma 10.3. *Let A and B be matrices (of appropriate sizes). If $A \geq 0$, it holds that $BAB^* \geq 0$.*

Proof. Note that a matrix A is positive if and only if $\langle\psi|A|\psi\rangle \geq 0$ for all vectors $|\psi\rangle$. Let $|\psi\rangle$ be an arbitrary vector (of the appropriate size). Define $|\psi\rangle = B^*|\phi\rangle$. Then

$$\langle\psi|BAB^*|\psi\rangle = \langle\phi|A|\phi\rangle \geq 0$$

by positivity of A. □

11 Lecture 11

(**1 March 2016**)

11.1 Homework hint

In question 2 of assignment 3, we are asked to consider a state $|\psi\rangle$ and an ensemble $\{r_j, |\phi_j\rangle\}$. We must prove that $|\psi\rangle$ can be converted into $|\phi_j\rangle$ with probability r_j if and only if

$$E_k(\psi) \geq \sum_j r_j E_k(\phi_j)$$

for all $k = 1, \ldots, n$. (All states are vectors in $\mathbb{C}^n \otimes \mathbb{C}^n$.)

Let $|\phi\rangle = \sum_{i,j} \phi_{ij} |i\rangle|j\rangle$ be an arbitrary pure state and let ϕ be the $n \times n$ matrix whose entries are ϕ_{ij}. It holds that

$$E_k(\phi) = \sum_{l=k}^{d} \lambda_l(\phi\phi^*)$$

where $\lambda_l(\phi\phi^*)$ is the l^{th} eigenvalue of the matrix $\phi\phi^*$ (where the eigenvalues are considered in decreasing order). You may use the fact that, for any matrices σ_1 and σ_2,

$$\sum_{l=k}^{n} \lambda_l\big(t\sigma_1 + (1-t)\sigma_2\big) \leq t \sum_{l=k}^{n} \lambda_l(\sigma_1) + (1-t) \sum_{l=k}^{n} \lambda_l(\sigma_2)$$

(i.e. convexity of the Ky-Fan norms). We can then set

$$|\phi_j\rangle = \frac{1}{\|\cdot\|}\big(M_j \otimes U_j\big)|\psi\rangle = \frac{1}{\|\cdot\|}\big((M_j A) \otimes U_j\big)|\phi^+\rangle$$

for some matrices M_j and A (where $\|\cdot\|$ are the correct normalization constants).

11.2 Representations of quantum channels (cont.)

Consider Hilbert spaces A', A, and R, where these spaces have dimensions $\dim \mathsf{A} = \dim \mathsf{R} = n$ and $\dim \mathsf{A}' = m$. We denote the (unnormalized) operator of the maximally entanged state by $\Phi^+ = |\phi^+\rangle\langle\phi^+|$, where $|\phi^+\rangle$ is the unnornalized vector

$$|\phi^+\rangle = \sum_{x=1}^{n} |x\rangle_{\mathsf{A}} |x\rangle_{\mathsf{R}}.$$

Recall that a linear map $\mathcal{E} : \mathrm{H}_n^{\mathsf{A}'} \longrightarrow \mathrm{H}_m^{\mathsf{A}}$ is CPTP if and only if $J(\mathcal{E}) = (\mathcal{E}_{\mathsf{A}' \to A} \otimes \mathrm{id})(\Phi^+) \geq 0$ and $\mathrm{Tr}_A\, J(\mathcal{E}) = I_{\mathsf{R}}$. As we said last time, this J is an isomorphism. That is, given a operator σ_{AR}, we can also define a mapping by

$$\mathcal{E}(\rho_{\mathsf{R}}) = \mathrm{Tr}_{\mathsf{R}}[\sigma_{\mathsf{AR}}(I_{\mathsf{A}} \otimes \rho^T)].$$

Let $\sigma_{\mathsf{AR}} \in \mathrm{H}_{nm}$ be an arbitrary positive semi-definite operator. Then it can be cecomposed as

$$\sigma_{\mathsf{AR}} = \sum_j |\psi_j\rangle\langle\psi_j|$$

for some collection of (unnoralized) vectors $|\psi_j\rangle$. Furthermore, each $|\psi_j\rangle$ can be written as

$$|\psi_j\rangle = \sum_{x=1}^{m} \sum_{y=1}^{n} M_{xy}^{(j)} |x\rangle_{\mathsf{A}} |y\rangle_{\mathsf{R}} = (M_j \otimes I) \sum_{y=1}^{n} |y\rangle_{\mathsf{A}} |y\rangle_{\mathsf{R}} = (M_j \otimes I)|\phi^+\rangle$$

where each M_j is some $m \times n$ matrix with entries $M_{xy}^{(j)}$. Now we examamine the mapping $\mathcal{E}$. We have

$$\begin{aligned}
\mathcal{E}(\rho) &= \sum_j \mathrm{Tr}_{\mathsf{R}}[|\psi_j\rangle\langle\psi_j|(I_{\mathsf{A}} \otimes \rho^T)] \\
&= \sum_j \mathrm{Tr}_{\mathsf{R}}\big[(M_j \otimes I)|\phi^+\rangle\langle\phi^+|_{\mathsf{AR}}(M_j^+ \otimes I)(I_{\mathsf{A}} \otimes \rho^T)\big] \\
&= \sum_j M_j\big(\mathrm{Tr}_{\mathsf{R}}[|\phi^+\rangle\langle\phi^+|_{\mathsf{AR}}(I \otimes \rho^T)]\big)M_j^* \\
&= \sum_j M_j\big(\mathrm{Tr}_{\mathsf{R}}[|\phi^+\rangle\langle\phi^+|_{\mathsf{AR}}(\rho \otimes I)]\big)M_j^* \\
&= \sum_j M_j\big(\underbrace{\mathrm{Tr}_{\mathsf{R}}[|\phi^+\rangle\langle\phi^+|_{\mathsf{AR}}]}_{I}\big)\rho M_j^* \\
&= \sum_j M_j\rho M_j^*,
\end{aligned}$$

where we used the fact that $(A\otimes I)|\phi^+\rangle = (I\otimes A^T)|\phi^+\rangle$ holds for any matrix A. Note that the M_j's are not unique, because we could have chosen a different representation of σ_{AR}. What properties must the operators M_j satisfy? We have

$$\begin{aligned}
\sum_j M_j^* M_j = \sum_j M_j^* I M_j &= \sum_j M_j^* \,\mathrm{Tr}_{\mathsf{A}}[|\phi^+\rangle\langle\phi^+|]M_j \\
&= \sum_j \mathrm{Tr}_{\mathsf{A}}\big[(M_j^* \otimes I)|\phi^+\rangle\langle\phi^+|(M_j \otimes I)\big] \\
&= \overline{\sum_j \mathrm{Tr}_{\mathsf{A}}\big[(M_j \otimes I)|\phi^+\rangle\langle\phi^+|(M_j^* \otimes I)\big]} \quad \text{(since this is Hermitian)} \\
&= \overline{\sum_j \mathrm{Tr}_{\mathsf{A}}|\psi_j\rangle\langle\psi_j|} \\
&= \overline{\mathrm{Tr}_{\mathsf{A}}[\sigma_{\mathsf{AR}}]} \\
&= \overline{I} = I.
\end{aligned}$$

This the ***Kraus representation*** of the channel $\mathcal{E}$.

Theorem 11.1. *Every CPTP map $\mathcal{E}$ can be written as*

$$\mathcal{E}(\rho) = \sum_j M_j\rho M_j^*$$

for some matirces M_j that satisfy $\sum_j M_j^* M_j = I$.

This map is in fact trace-preserving, since[1]

$$\begin{aligned}\operatorname{Tr}\mathcal{E}(\rho) &= \sum_j \operatorname{Tr}[M_j\rho M_j^*] \\ &= \sum_j \operatorname{Tr}[M_j^* M_j\rho] \\ &= \operatorname{Tr}\Big[\underbrace{\sum_j M_j^* M_j}_{I}\,\rho\Big] = \operatorname{Tr}\rho.\end{aligned}$$

Theorem 11.2. *Let $\mathcal{E} : \mathrm{H}_n \longrightarrow \mathrm{H}_m$ be a CPTP map. It is always possible to find a Kraus representation $\mathcal{E}$ with at most nm Kraus operators.*

Proof. Let σ_{AR} be the Choi representation of $\mathcal{E}$. Note that σ_{AR} is an $nm \times nm$ matrix, so it can be decomposed into its eigenvectors

$$\sigma_{\mathsf{AR}} = \sum_{j=1}^{nm} |\psi_j\rangle\langle\psi_j|.$$

Taking $|\psi_j\rangle = M_j \otimes I|\phi^+\rangle$ in the derivation of the Kraus representation above, we get nm matrices M_j. □

Theorem 11.3. *Let $\{F_l\}$ and $\{M_j\}$ be families of Kraus operators. It holds that*

$$\sum_l F_l\rho F_l^* = \sum_j M_j\rho M_j^*$$

(i.e. they are Kraus representations of the same channel) if and only if there is a unitary matrix U with entries U_{jl} satisfying

$$M_j = \sum_l U_{jl}F_l.$$

Proof. Suppose there exists a unitary U satisfying $M_j = \sum_l U_{jl}F_l$. Now

$$\begin{aligned}\sum_j M_j\rho M_j^* &= \sum_j \sum_l \sum_{l'} U_{jl}\overline{U}_{jl'} F_l\rho F_{l'} \\ &= \sum_{l,l'} \delta_{ll'} F_l\rho F_{l'} \\ &= \sum_l F_l\rho F_l^*.\end{aligned}$$

The other direction is left as an exercise. (Hint: If they represent the same channel, then the channels have the same Choi matrix.) □

11.2.1 Kraus representations and measurements

One way of thinking about Kraus representations of channels is as measurements. Suppose we have a measurement with corresponding operators $\{M_j\}$. Recall that these operators must satisfy $\sum_j M_j^* M_j = I$. If we start with a state ρ and measure it with this measurement, we obtain outcome j with probability $p_j = \operatorname{Tr}[M_j\rho M_j^*]$ and the resulting state is

$$\sigma_j = \frac{M_j\rho M_j^*}{\operatorname{Tr}[M_j\rho M_j^*]} = \frac{1}{p_j} M_j\rho M_j^*.$$

[1] There is another way to show this. Suppose that $\langle\sum_j M_j^* M_j, \rho\rangle = \langle I, \rho\rangle$ holds for all ρ. This implies that $\langle I - \sum_j M_j^* M_j, \rho\rangle = 0$ holds for all ρ, which can only be true if $I - \sum_j M_j^* M_j = 0$ or $\sum_j M_j^* M_j = I$.

If we subsequently 'forget' which measurement outcome we obtained, we are left with the state

$$\sum_j p_j \sigma_j = M_j \rho M_j^* = \mathcal{E}(\rho),$$

which is exactly the way we defined the Kraus representation of a channel with Kraus operators $\{M_j\}$.

11.2.2 The Stinespring representation

Theorem 11.4. *Let $\mathcal{E} : \mathrm{H}_n^{\mathsf{A}} \longrightarrow \mathrm{H}_m^{\mathsf{B}}$ be a CPTP map. Then there is a space E and a unitary operator U on AE such that*

$$\begin{aligned}\mathcal{E}(\rho) &= \mathrm{Tr}_{\mathsf{E}}[U_{\mathsf{AE}}(\rho_{\mathsf{A}} \otimes |0\rangle\langle 0|_{\mathsf{E}})U_{\mathsf{AE}}^*] \\ &= \sum_x \langle x|_{\mathsf{E}} U_{\mathsf{AE}} |0\rangle_{\mathsf{E}} \rho^A \langle 0|_{\mathsf{E}} U_{\mathsf{AE}}^* |x\rangle_{\mathsf{E}}\end{aligned}$$

How is this related to the Kraus representation? Define a set of Kraus operators M_x on A by

$$M_x = \langle x|_{\mathsf{E}} U_{\mathsf{AE}} |0\rangle_{\mathsf{E}}$$

and we have the Kraus representation $\mathcal{E}(\rho) = \sum_x M_x \rho M_x^*$.

16 Lecture 16

(**17 March 2016**)

(Note: I was out-of-town for lectures 12-15.)

16.1 Distance mesaures

Recall the fidelity

$$F(\rho,\sigma) = \operatorname{Tr}\sqrt{\rho^{1/2}\sigma\rho^{1/2}}.$$

Some important theorems:

Theorem 16.1. *Let ρ and σ be density operators on the same space. It holds that*

$$F(\rho,\sigma) = \max_{|\psi\rangle,|\phi\rangle} |\langle\psi|\phi\rangle|,$$

where the maximum is taken over all purifications $|\psi\rangle$ and $|\phi\rangle$ of ρ and σ respectively.

For probability distributions $\{p_i\}$ and $\{q_i\}$, their fidelity is defined as $F(\{p_i\},\{q_i\}) = \sum_i \sqrt{p_i q_i}$.

Theorem 16.2. *Let ρ and σ be density operators on the same space. It holds that*

$$F(\rho,\sigma) = \min_{\{K_i\}} F\big(\{\operatorname{Tr}[K_i^* K_i \rho]\}, \{\operatorname{Tr}[K_i^* K_i \sigma]\}\big),$$

where the minimization is taken over all possible measurements $\{K_i\}$.

Theorem 16.3 (Monotonicity of fidelity)**.** *Let ρ and σ be density operators on the same space and let $\mathcal{E}$ be a CPTP map. It holds that*

$$F(\rho,\sigma) \leq F\big(\mathcal{E}(\rho),\mathcal{E}(\sigma)\big).$$

Theorem 16.4 (Strong concavity)**.** *For density operators $\{\rho_i\}$ and $\{\sigma_i\}$ with probabilities $\{p_i\}$ and $\{q_j\}$, it holds that*

$$F\left(\sum_i p_i\rho_i, \sum_i q_i\sigma_i\right) \geq \sum_i \sqrt{p_i q_i} F(\rho_i,\sigma_i).$$

Proof. Recall from Uhlmann's theorem that, for density matrices ρ_{A} and σ_{A} on system A, there exist purifications $|\psi\rangle_{\mathsf{AR}}$ and $|\phi\rangle_{\mathsf{AR}}$ of ρ and σ respectively such that $F(\rho,\sigma) = \langle\psi|\phi\rangle$.

For each ρ_i and σ_i, we may use Uhlmann's theorem to choose purifications $|\psi_i\rangle_{\mathsf{AR}}$ and $|\phi_i\rangle_{\mathsf{AR}}$ such that $F(\rho_i,\sigma_i) = \langle\psi_i|\phi_i\rangle$ holds for each i. We may define a new state $|\Psi\rangle_{ARR'}$ by

$$|\Psi\rangle = \sum_i \sqrt{p_i}|\psi_i\rangle \otimes |i\rangle.$$

States of this form are sometimes called *flag states*. It is straightforward to check that

$$\operatorname{Tr}_{\mathsf{RR}'}|\Psi\rangle\langle\Psi|_{\mathsf{ARR}'} = \sum_i p_i \operatorname{Tr}_{\mathsf{R}}|\psi_i\rangle\langle\psi_i|_{\mathsf{AR}} = \sum_i p_i\rho_i.$$

Similarly, we may define

$$|\Phi\rangle = \sum_i \sqrt{q_i}|\phi_i\rangle \otimes |i\rangle$$

such that $\mathrm{Tr}_{\mathsf{RR}'}|\Phi\rangle\langle\Phi|_{\mathsf{ARR}'} = \sum_i q_i\sigma_i$. So $|\Psi\rangle$ and $|\Phi\rangle$ are purifications of $\sum_i p_i\rho_i$ and $\sum_i q_i\sigma_i$ respectively. By Uhlmann's theorem, it holds that

$$\begin{aligned} F\left(\sum_i p_i\rho_i, \sum_i q_i\sigma_i\right) &\geq |\langle\Psi|\Phi\rangle| \\ &= \sum_i \sqrt{p_i q_i}\langle\psi_i|\phi_i\rangle \\ &= \sum_i \sqrt{p_i q_i} F(\rho_i, \sigma_i) \end{aligned}$$

as desired. □

Why are we studying distance measures? Because we want to know when a channel is 'close' to being an identity channel. In particular, we want compression schemes that send quantum information as accurately as possible.

16.1.1 Dynamic measures of distance

Let $\mathcal{E}$ be a quantum channel. We can define the following quantities:

$$F(\mathcal{E}) = \min_{|\psi\rangle} F\big(|\psi\rangle\langle\psi|, \mathcal{E}(|\psi\rangle\langle\psi|)\big)$$

and

$$D(\mathcal{E}) = \max_{|\psi\rangle} D\big(|\psi\rangle\langle\psi|, \mathcal{E}(|\psi\rangle\langle\psi|)\big)$$

where the minimum and maximum are taken over all pure states of the input system. Note that only the identity channel will give either $D(\mathcal{E}) = 0$ or $F(\mathcal{E}) = 1$.

Example 16.5. For the depolarizing channel $\mathcal{E}$, it holds that $F(\mathcal{E}) = \sqrt{1 - \frac{p}{2}}$.

Note that in the definition of $F(\mathcal{E})$ and $D(\mathcal{E})$ we only optimize over pure states. Why don't we define it as the optimization over *all* states? Well, it turns out that this would give us the same values! Indeed, if we instead define the quantity

$$f(\mathcal{E}) = \min_{\rho} F(\rho, \mathcal{E}(\rho)),$$

we see that

$$\begin{aligned} F(\rho, \mathcal{E}(\rho)) &= F\left(\sum_i p_i|v_i\rangle\langle v_i|, \sum_i p_i\mathcal{E}(|v_i\rangle\langle v_i|)\right) \\ &\geq \sum_i p_i F(|v_i\rangle\langle v_i|, \mathcal{E}(|v_i\rangle\langle v_i|)) \\ &\geq \min_i F(|v_i\rangle\langle v_i|, \mathcal{E}(|v_i\rangle\langle v_i|)) \\ &\geq F(\mathcal{E}). \end{aligned}$$

16.1.2 Preservation of Entanglement

Consider a quantum source $S = \{p_x, |\varphi_x\rangle\}$. That is, a device that generates the state $|\varphi_x\rangle$ with probability p_x. If we don't know the value of the classical variable x, we consider the state of the emitted from the source to be

$$\rho = \sum_x p_x|\varphi_x\rangle\langle\varphi_x|.$$

Given a state ρ_{A}, consider a purification $|\psi\rangle_{\mathsf{AR}}$ of ρ_{A}. Given a channel $\mathcal{E} : \mathsf{A} \longrightarrow \mathsf{A}'$, we define the ***entanglement fidelity*** of $\mathcal{E}$ with respect to ρ as

$$F_e(\rho, \mathcal{E}) = F(\mathsf{RA}, \mathsf{RA}') = \langle\psi|(\mathcal{E} \otimes \hat{I}_{\mathsf{R}})(|\psi\rangle\langle\psi|)|\psi\rangle.$$

It holds that

$$F\big(|\psi\rangle_{\mathsf{AR}}, (\mathcal{E} \otimes \hat{I})(|\psi\rangle\langle\psi|_{\mathsf{AR}})\big) \leq F(\rho_{\mathsf{A}}, \mathcal{E}(\rho_{\mathsf{A}})).$$

This follows directly from the monotonicity of fidelity, where we use the fact that the partial trace is a CPTP map:

$$F(\rho_{\mathsf{A}}, \mathcal{E}(\rho_{\mathsf{A}})) = F\left(\mathrm{Tr}_{\mathsf{R}}\left(|\psi\rangle_{\mathsf{AR}}\right), \mathrm{Tr}_{\mathsf{R}}\left((\mathcal{E} \otimes \hat{I})(|\psi\rangle\langle\psi|_{\mathsf{AR}})\right)\right) \geq F\big(|\psi\rangle_{\mathsf{AR}}, (\mathcal{E} \otimes \hat{I})(|\psi\rangle\langle\psi|_{\mathsf{AR}})\big)$$

(where we recall that, for any CPTP map Λ, it holds that $F(\Lambda(\rho_1), \Lambda(\rho_2)) \geq F(\rho_1, \rho_2)$ for all states states ρ_1 and ρ_2).

Why are we interested in this measure? Note that another way of writing the fidelity is as $F(\rho, \sigma) = \|\rho - \sigma\|_1$. Here are some important properties of the entanglement fidelity:

(1) The entanglement fidelity is independent of the purification $|\psi\rangle_{\mathsf{AR}}$.

(2) $F_e(\rho, \mathcal{E}) = |\mathrm{Tr}(\rho K_i)|^2$ where K_i are Kraus operators of $\mathcal{E}$.

(3) $F_e(\rho, \mathcal{E}) \leq \big(F(\rho, \mathcal{E}(\rho))\big)^2$.

(4) $F_e\big(\sum_i p_i \rho_i, \mathcal{E}\big) \leq \overline{F}(S)$.

(5) For pures tates, $F_e(|\varphi\rangle, \mathcal{E}) = \big(F(|\varphi\rangle, \mathcal{E}(|\varphi\rangle\langle\varphi|))\big)^2$.

Proof (of (2)).

$$\begin{aligned}
F_e(\rho, \mathcal{E}) &= \langle\psi|(\mathcal{E} \otimes \hat{I})(|\psi\rangle\langle\psi|)|\psi\rangle \\
&= \sum_j \langle\psi|K_j \otimes I|\psi\rangle\langle\psi|K_j^* \otimes I|\psi\rangle \\
&= \sum_j |\langle\psi|K_j \otimes I|\psi\rangle|^2 \\
&= \sum_j |\mathrm{Tr}\left[|\psi\rangle\langle\psi|(K_j \otimes I)\right]|^2 \\
&= \sum_j |\mathrm{Tr}_{\mathsf{A}}(\mathrm{Tr}_{\mathsf{R}}(|\psi\rangle\langle\psi|(K_j \otimes I)))|^2 \\
&= \sum_j |\mathrm{Tr}_{\mathsf{A}}(\rho_{\mathsf{A}} K_j)|^2.
\end{aligned}$$

□

16.2 Schumacher's quantum noiseless channel coding theorem

16.2.1 A compression scheme of rate r

$$\rho^{\otimes n} \xrightarrow{C} \sigma_{rn} \xrightarrow{\mathcal{E}=\hat{I}} \sigma_{rn} \xrightarrow{D} \tau_n$$

where C is a compression channel, D is a decompression channel, σ_{rn} is the intermediate state after compression, and τ_n is the state after decompression.

16.2.2 Typical subspaces

Let ρ be a density operator and suppose that it has spectral decomposition

$$\rho = \sum_x p_x |x\rangle\langle x|$$

where $|x\rangle$ are some orthonormal vectors. What does $\rho^{\otimes n}$ look like? It is

$$\sum_{x_1,x_2,\ldots,x_n} p_{x_1} p_{x_2} \cdots p_{x_n} |x_1\rangle\langle x_1| \otimes |x_2\rangle\langle x_2| \otimes \cdots \otimes |x_n\rangle\langle x_n|.$$

For every positive integer n and $\epsilon \geq 0$, the ϵ-typical subspace of $\rho^{\otimes n}$ is

$$T_q(n,\epsilon) = \text{span}\left\{|x_1\rangle|x_2\rangle\cdots|x_n\rangle \,|\, (x_1,x_2,\ldots,x_n) \text{ is an } \epsilon\text{-typical sequence of } \{p_x,x\}\right\}.$$

The projection onto the ϵ-typical subspace is

$$P(n,\epsilon) = \sum_{(x_1,\ldots,x_n) \text{ is } \epsilon\text{-typical}} |x_1\rangle\langle x_1| \otimes \cdots \otimes |x_n\rangle\langle x_n|$$

17 Lecture 17

(**22 March 2016**)

17.1 Homework hint

In the homework (problem 2 of assignment 4) we are asked to find necessary and sufficient conditions for when

$$S(\rho_{\mathsf{AB}}) = |S(\rho_{\mathsf{A}}) - S(\rho_{\mathsf{B}})|.$$

There is a condition given in Nielsen and Chuang, but it is too complicated. Look instead at the discussion of this problem in that book that appears just before that. In fact, it is sufficient to show that $\rho_{\mathsf{AR}} = \rho_{\mathsf{A}} \otimes \rho_{\mathsf{R}}$. (And you need to prove this.)

17.2 Typical Subspaces (cont.)

We'll see that, in a way, the quantum notion of typical subspaces is simpler than the classical notion of typical sequences.

As before, consider a density operator $\rho : \mathbb{C}^d \to \mathbb{C}^d$ and its spectral decomposition

$$\rho = \sum_y q_y |\psi_y\rangle\langle\psi_y| = \sum_x p_x |x\rangle\langle x|$$

for some orthonormal basis $\{|x\rangle\}$. Then the n-fold tensor product of ρ is

$$\rho^{\otimes n} = \sum_{\underbrace{x_1, x_2, \ldots, x_n}_{x^n}} \underbrace{p_{x_1} p_{x_2} \cdots p_{x_n}}_{p_{x^n}} \underbrace{|x_1\rangle\langle x_1| \otimes |x_2\rangle\langle x_2| \otimes \cdots \otimes |x_n\rangle\langle x_n|}_{|x^n\rangle\langle x^n|} = \sum_{x^n} p_{x^n} |x^n\rangle\langle x^n|$$

where we use the notation $x^n = (x_1, x_2, \ldots, x_n)$ to denote an arbitrary sequence. (Here we think about these sequences as indepentent results from as an i.i.d. classical source.)

The (quantum) ϵ-***typical subspace*** of $\rho^{\otimes n}$ is defined as

$$T_q(n, \epsilon) = \text{span}\{|x^n\rangle \,|\, x^n \text{ is } \epsilon\text{-typical}\}$$

where the sequences $x^n = (x_1, \ldots, x_n)$ are ϵ-typical in the classical sense. The projector onto the ϵ-typical subspace is defined by

$$P(n, \epsilon) = \sum_{x^n \text{ is } \epsilon\text{-typical}} |x^n\rangle\langle x^n|.$$

Theorem 17.1 (Theorem of typical subspaces)**.** *Let ρ be a density operator on $\mathcal{H} = \mathbb{C}^d$.*

(1) *Fix $\epsilon > 0$. For any $\delta > 0$, there exists an $n > 1$ such that*

$$\text{Tr}\big[P(n, \epsilon)\rho^{\otimes n}\big] > 1 - \delta.$$

(2) *Fix $\epsilon, \delta > 0$. Then there exists an $n > 1$ such that*

$$(1 - \delta)2^{n(S(\rho) - \epsilon)} \leq \dim T_q(n, \epsilon) \leq 2^{n(S(\rho) + \epsilon)}.$$

(3) *Let $r < S(\rho)$ be a nonnegative real number and for every postitive integer n let $S_n \subset \mathcal{H}^{\otimes n}$ be a subspace with $\dim S_n \leq 2^{nr}$. For any $\delta > 0$, there exists an $n > 1$ such that*

$$\mathrm{Tr}[\Pi_{S_n} \rho^{\otimes n}] \leq \delta$$

where Π_{S_n} is the projection operator onto the subspace S_n.

Proof. Parts (1) and (2) follow directly from the definitions. Only part (3) requires some work.

(1) From the theory of classical typical sequences, we have that

$$\begin{aligned}\mathrm{Tr}[P(n,\epsilon)\rho^{\otimes n}] &= \sum_{x^n \ \epsilon\text{-typical}} \langle x^n|\rho^{\otimes n}|x^n\rangle \\ &= \sum_{x^n \ \epsilon\text{-typical}} p_{x^n} = \Pr(\text{an arbitrary sequence is typical}) > 1-\delta,\end{aligned}$$

as desired.

(2) By definition, we have that

$$\dim T_q(n,\epsilon) = \mathrm{Tr}[P(n,\epsilon)] = |T_c(n,\epsilon)|$$

where $T_c(n,\epsilon)$ denotes the (classical) set of typical sequences and $|T_c(n,\epsilon)|$ denotes the number of elements in that set.

(3) Here, S^n is *any* subspace that may or may not have anything to do with typicality. We can break up the space into the typical and nontypical subspaces,

$$\mathrm{Tr}[\Pi_{S_n}\rho^{\otimes n}] = \mathrm{Tr}[\Pi_{S_n}\rho^{\otimes n}P(n,\epsilon)] + \mathrm{Tr}\big[\Pi_{S_n}\rho^{\otimes n}\big(I - P(n,\epsilon)\big)\big],$$

and show that both terms on the right-hand side go to zero. For the second term, from part (1) we know that there is some n large enough such that $\mathrm{Tr}\big[P(n,\epsilon)\rho^{\otimes n}\big] > 1-\delta$, and thus

$$\begin{aligned}\mathrm{Tr}\big[\Pi_{S_n}\rho^{\otimes n}\big(I-P(n,\epsilon)\big)\big] &\leq \mathrm{Tr}\big[\rho^{\otimes n}\big(I-P(n,\epsilon)\big)\big] \\ &< 1-(1-\delta) = \delta\end{aligned}$$

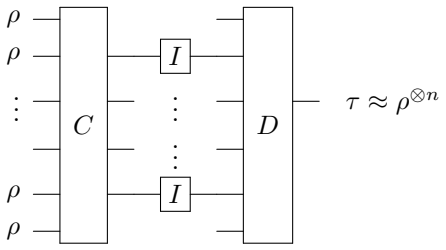
for that n. For the first term, note that

$$\begin{aligned}\mathrm{Tr}[\Pi_{S_n}\rho^{\otimes n}P(n,\epsilon)] &= \sum_{x^n \ \epsilon\text{-typical}} \mathrm{Tr}[\Pi_{S_n}\rho^n|x^n\rangle\langle x^n|] \\ &= \sum_{x^n \ \epsilon\text{-typical}} \langle x^n|\Pi_{S_n}\rho^n|x^n\rangle \\ &= \sum_{x^n \ \epsilon\text{-typical}} p_{x^n}\langle x^n|\Pi_{S_n}|x^n\rangle \\ &\leq 2^{-n(S(\rho)-\epsilon)} \sum_{x^n \ \epsilon\text{-typical}} \langle x^n|\Pi_{S_n}|x^n\rangle \\ &\leq 2^{-n(S(\rho)-\epsilon)} \underbrace{\sum_{x^n}\langle x^n|\Pi_{S_n}|x^n\rangle}_{\mathrm{Tr}\,\Pi_{S_n} < 2^{rn}} \\ &\leq 2^{-n(S(\rho)-\epsilon)}2^{rn} \\ &= 2^{-n(S(\rho)-r-\epsilon)} \xrightarrow{n\to\infty} 0,\end{aligned}$$

from the theory of classical typical sequences. Here we use the fact that, for any typical sequence x^n, we have $p_{x^n} \leq 2^{-\lfloor n(S(\rho)-\epsilon)\rfloor} \leq 2^{-n(S(\rho)-\epsilon)}$.

□

17.2.1 Compression-decompression schemes



Definition 17.2. Let ρ be a density operator. A compression-decompression scheme (C, D) of rate r is said to be ***reliable*** for ρ if

$$\lim_{n\to\infty} F_e(\rho^{\otimes n}, D \circ C) = 1.$$

Recall that the entanglement fidelity of a state to a channel is defined by

$$\begin{aligned} F_e(\rho_{\mathsf{A}}, \mathcal{E}) &= \langle\psi|(\hat{I}_{\mathsf{R}} \otimes \mathcal{E})(|\psi\rangle\langle\psi|)|\psi\rangle \\ &= \sum_j |\mathrm{Tr}(\rho_{\mathsf{A}} K_j)|^2 \end{aligned}$$

where $|\psi\rangle = |\psi\rangle_{\mathsf{AR}}$ is any purification of ρ_{A} to system AR and $\{K_j\}$ is any Kraus representation of $\mathcal{E}$.

Theorem 17.3 (Schumacher's noiseless channel coding theorem)**.** *Given an i.i.d. quantum information source $\rho : \mathbb{C}^d \otimes \mathbb{C}^d$ with entropy rate $S(\rho)$, a reliable compression-decompression scheme of rate r exists if and only if $r > S(\rho)$.*

Proof. Suppose $r > S(\rho)$ and let $\epsilon > 0$ such that $r > S(\rho) + \epsilon$. From the theory of typical subspaces, given any $\delta > 0$ there exists a sufficiently large n such that

$$\mathrm{Tr}[\rho^{\otimes n} P(n, \epsilon)] > 1 - \delta.$$

Moreover, it holds that

$$\dim T_q(n, \epsilon) = \mathrm{Tr}[P(n, \epsilon)] \leq 2^{\lfloor n(S(\rho)+\epsilon)\rfloor} \leq 2^{\lfloor nr \rfloor}.$$

It follows that there exists a subspace $W \subset \mathcal{H}^{\otimes n}$ such that $\dim W = 2^{\lfloor nr \rfloor}$ and $T_q(n, \epsilon) \subset W$. Define a compression map $C : \mathrm{H}_{d^n} \to \mathrm{H}_{2^{\lfloor nr \rfloor}}$ by

$$C(\gamma) = P(n, \epsilon)\gamma P(n, \epsilon) + \mathrm{Tr}\big[\big(I - P(n, \epsilon)\big)\gamma\big]|0\rangle\langle 0|$$

where $|0\rangle$ is some arbitrary pure state in $\mathbb{C}^{2^{\lfloor nr \rfloor}}$. Note that C is indeed a CPTP map, since it has a Kraus representation with Kraus operators.[1]

The decompression channel $D : \mathrm{H}_{2^{\lfloor nr \rfloor}} \to \mathrm{H}_{d^n}$ that we should use then is just an embedding channel

$$D(\sigma) = \sigma \in \mathrm{H}_{d^n}\,.$$

Finally, we compute the entanglement fidelity. We have

$$\begin{aligned} F_e(\rho^{\otimes n}, D \circ C) = |\mathrm{Tr}[P(n, \epsilon)\rho^{\otimes n}]|^2 + \sum_j |\mathrm{Tr}[K_j \rho^{\otimes n}]|^2 &\geq |\mathrm{Tr}[P(n, \epsilon)\rho^{\otimes n}]|^2 \\ &> |1 - \delta|^2 > 1 - 2\delta, \end{aligned}$$

where the Kraus operators $\{K_j\}$ are obtained from the footnote. □

[1] Indeed, the Kraus operators can be found in the following manner. Let $|\psi_i\rangle$ be any orthonormal basis such that $\sum_i |\psi_i\rangle\langle\psi_i| = I - P(n, \epsilon)$. Then we get Kraus operators

$$\mathrm{Tr}\big[(I - P(n, \epsilon))\gamma\big]|0\rangle\langle 0| = \sum_i \langle\psi_i|\gamma|\psi_i\rangle|0\rangle\langle 0| = \sum_i \underbrace{|0\rangle\langle\psi_i|}_{K_i}\,\gamma\,\underbrace{|\psi_i\rangle\langle 0|}_{K_i^*}.$$

So the Kraus operators for C are $\{P(n, \epsilon)\} \cup \{K_i\}$ such that $C(\gamma) = P(n, \epsilon)\gamma P(n, \epsilon) + \sum_i K_i \gamma K_i^*$.

18 Lecture 18

(**24 March 2016**)

18.1 What is a quantum source?

Something that we actually haven't discussed yet in class is what a quantum source actually is! Here are different ways of thinking about a 'quantum source'.

(1) Consider a pure state $|\psi\rangle_{\mathsf{AR}}$ on a joint system composed of A and a reference system R. Then a quantum source would be

$$\rho_{\mathsf{A}} = \mathrm{Tr}_{\mathsf{R}}|\psi\rangle\langle\psi|_{\mathsf{AR}}$$

where we only have access to the system A. We could have numerous identical copies of the same state $|\psi\rangle_{\mathsf{AR}}$ (where for each one we only have access to A), and we then 'compress' all of the A systems into something else and send what we have to Bob through a series of identity quantum channels.

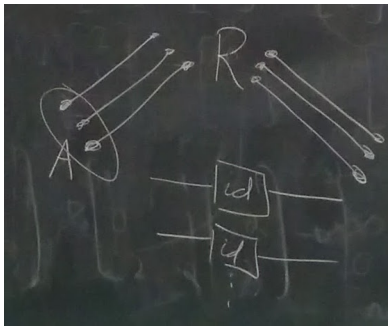


Figure 18.1: Depiction of a source as many copies of a pure state entangled with a resource system R.

The identity quantum channel is also called the *noiseless* quantum channel.

(2) Alternatively, we can consider a source as an ensemble $\{q_y, |\psi_y\rangle\}$ and have a state

$$\rho = \sum_y p_y|\psi_y\rangle\langle\psi_y| = \sum_x \lambda_x|x\rangle\langle x|$$

where λ_x are the eigenvalues of ρ. We could randomly sample from the source $\{q_y, |\psi_y\rangle\}$ where we obtain $|\psi_y\rangle$ with probability q_y. Given a random string of samples obtained (independently and identically distributed)

$$|\psi_{y_1}\rangle|\psi_{y_2}\rangle\cdots|\psi_{y_n}\rangle, \tag{18.1}$$

we would like to have a compression-decompression scheme such that the string in (18.1) can be retrieved as close as possible with high probability in the limit of large n. That is, we want

$$\sum_{x^n} p_{x^n}\left(F_e\left(|\psi_{x^n}\rangle, D\circ C(|\psi_{x^n}\rangle)\right)\right)^2 \xrightarrow{n\to\infty} 1$$

18.2 Back to Schumacher compression

We now want to prove the *converse* of Schumacher's compression theorem. Consider a density operator ρ. We want to show that there does **not** exist a valid compression-decompression scheme of rate r for **any** $r < S(\rho)$. Indeed, suppose $r < S(\rho)$ and suppose that we had some compression-decompression scheme of rate r with compression and decompression channels given by

$$C : \mathrm{H}_{d^n} \to W \qquad \text{and} \qquad D : W \to \mathrm{H}_{d^n}$$

where $W \subset \mathrm{H}_{d^n}$ is a subspace with $\dim W = 2^{\lfloor rn \rfloor}$. To show that this compression-decompression scheme is *not* reliable, we want to show that

$$\lim_{n\to\infty} F_e\left(\rho^{\otimes n}, D \circ C\right) = 0.$$

Let $\delta > 0$. We will show that $F_e\left(\rho^{\otimes n}, D \circ C\right) < \delta$ for n sufficiently large. Now

$$F_e\left(\rho^{\otimes n}, D \circ C\right) = \sum_{j,k} \left|\mathrm{Tr}\left(D_k C_j \rho^{\otimes n}\right)\right|^2 \tag{18.2}$$

where $\{D_k\}$ and $\{C_j\}$ are Kraus operators for the channels D and C

$$C(\gamma) = \sum_j C_j \gamma C_j^* \qquad \text{and} \qquad D(\sigma) = \sum_k D_k \sigma D_k^*$$

such that $\sum_j C_j^* C_j = I_{d^n}$ and $\sum_k D_k^* D_k = I_W$. Note that the matrices C_j are all the same size. They all have input dimension d^n and output dimension $2^{\lfloor rn \rfloor}$. Hence $\dim \mathrm{Range}(C_j) \leq 2^{nr}$ for each j and thus[1]

$$\dim \mathrm{Range}(D_k C_j) \leq 2^{rn}.$$

For each pair k, define the subspaces[2]

$$W_k = \mathrm{span}\left\{D_k v \,|\, v \in W\right\}$$

and let $S_k^{(n)}$ be the projection onto W_k. Then $\mathrm{Tr}[S_k^{(n)}] \leq 2^{\lfloor rn \rfloor} = \dim W$. Continuing from (18.2), we have

$$\begin{aligned}
F_e\left(\rho^{\otimes n}, D \circ C)\right) &= \sum_{j,k} \left|\mathrm{Tr}\left(D_k C_j \rho^{\otimes n}\right)\right|^2 \\
&= \sum_{j,k} \left|\mathrm{Tr}\left(S_k^{(n)} D_k C_j \rho^{\otimes n}\right)\right|^2 \\
&= \sum_{j,k} \left|\mathrm{Tr}\left(D_k C_j \sqrt{\rho^{\otimes n}} \sqrt{\rho^{\otimes n}} S_k^{(n)}\right)\right|^2 \\
&\leq \sum_{j,k} \mathrm{Tr}\left[D_k C_j \rho^{\otimes n} C_j^* D_k^*\right] \mathrm{Tr}\left[S_k^{(n)} \rho^{\otimes n} S_k^{(n)}\right]
\end{aligned}$$

where we make use of the Cauchy-Schwartz inequality for matrix inner products[3] in the last line. Hence we have proved that

$$F_e\left(\rho^{\otimes n}, D \circ C\right) \leq \sum_{j,k} \mathrm{Tr}\left[D_k C_j \rho^{\otimes n} C_j^* D_k^*\right] \mathrm{Tr}\left[S_k^{(n)} \rho^{\otimes n}\right].$$

[1] Recall that the range of an operator $A : V \to W$ is the subspace defined by $\mathrm{Range}(A) = \{Av \,|\, v \in V\} \subseteq W$.

[2] He erased this really quickly, so I'm not quite so sure this definition for W_k is corect... In the notes online, he just defines $S_k^{(n)}$ as the 'projector onto the range of $D_k S_n$,' where S_n is the projector onto W.

[3] The Cauchy-Schwartz inequality for matrix inner products states that $|\mathrm{Tr}(A^* B)|^2 \leq \mathrm{Tr}(A^* A)\,\mathrm{Tr}(B^* B)$ for all matrices A and B.

From the Theory of Typical Subspaces part (3), we know that

$$\mathrm{Tr}\left[S_k^{(n)}\rho^{\otimes n}\right] < \delta$$

for n sufficiently large. Furthermore, note that the result of the compression and decompression channels applied to $\rho^{\otimes n}$ can be written in term of their Kraus operators as

$$D \circ C(\rho^{\otimes n}) = \sum_{j,k} D_k C_j \rho^{\otimes n} C_j^* D_k^*.$$

Since the channels C and D are trace preserving, we have that $\mathrm{Tr}[D \circ C(\rho^{\otimes n})] = 1$ and thus

$$F_e\left(\rho^{\otimes n}, D \circ C\right) \leq \sum_{j,k} \mathrm{Tr}\left[D_k C_j \rho^{\otimes n} C_j^* D_k^*\right] \mathrm{Tr}\left[S_k^{(n)}\rho^{\otimes n}\right] \leq \delta\, \mathrm{Tr}[D \circ C(\rho^{\otimes n})] = \delta,$$

for sufficiently large n, where $\delta > 0$ is any positive constant. It follows that

$$\lim_{n\to\infty} F_e\left(\rho^{\otimes n}, D \circ C\right) = 0,$$

and thus this compression-decompression scheme is not reliable.

18.2.1 Mixed sources

Instead of a source composed of pure states, suppose instead we have the following source

$$\text{Source} = \{p_x, \rho_x\}.$$

In this case, the entropy of the source is *not* just $S\big(\sum_x p_x\rho_x\big)$.

Suppose we had an ensemble of the form $\{p_x, \rho_x\}$ for $x = 1, 2$ where ρ_1 and ρ_2 have orthogonal support. That is,

$$\rho_1\rho_2 = \rho_2\rho_1 = 0.$$

For example, we could have

$$\rho_1 = \lambda_1|1\rangle\langle 1| + \lambda_2|2\rangle\langle 2| \qquad \text{and} \qquad \rho_2 = \lambda_3|3\rangle\langle 3| + \lambda_4|4\rangle\langle 4|.$$

Then the state $\rho = \sum_x p_x\rho_x$ can really be written as a *direct* sum in block matrix form as

$$\rho = \bigoplus_x p_x\rho_x = \left(\begin{array}{c|c} p_1\rho_1 & 0 \\ \hline 0 & p_2\rho_2 \end{array}\right).$$

Then the entropy of ρ is

$$\begin{aligned} S(\rho) &= -p_1\lambda_1\log(p_1\lambda_1) - p_1\lambda_2\log(p_1\lambda_2) - p_2\lambda_3\log(p_2\lambda_3) - p_2\lambda_4\log(p_2\lambda_4) \\ &= H(p_1, p_2) + \sum_x p_x S(\rho_x). \end{aligned}$$

19 Lecture 19

(**29 March 2016**)

19.1 Compression (cont.)

Last time we were talking about Schumacher compression. We finished proving the Schumacher compression theorem, but now we will talk about different types of compression. Previously, we had considered sources of the form $\{p_x, |\psi_x\rangle\}$. The information that we wanted to compress was a random sequence of states

$$|\psi_{x_1}\rangle \otimes |\psi_{x_2}\rangle \otimes \cdots \otimes |\psi_{x_n}\rangle$$

such that any random such string of states can be reproduced as best as possible with high probability. We proved that compression with rate r is possible only if $r > S(\rho)$, where $\rho = \sum_x p_x |\psi_x\rangle\langle\psi_x|$.

Now suppose we have an arbitrary source of the form $\{p_x, \rho_x\}$ where the states ρ_x are arbitrary (not necessarily pure). The information that we want to compress is a random sequence of states

$$\rho_{x_1} \otimes \rho_{x_2} \otimes \cdots \otimes \rho_{x_n}$$

where the states are randomly and independently drawn from the source. Sometimes we can do even better than $r > S(\rho)$!

Suppose that the ρ_x are all orthogonal to each other. That is,

$$\rho_x \rho_{x'} = \rho_{x'} \rho_x = 0 \qquad \text{for all } x \neq x'.$$

Then there exist projection operators $\{P_x\}$ such that $P_{x'}\rho_x = \rho_x P_{x'} = \rho_x \delta_{xx'}$ for all x, x'. Without loss of generality we may then suppose that these projections sum to the identity

$$\sum_x P_x = I.$$

If Alice wants to communicate a random sequence of states $\rho_{x_1} \otimes \cdots \otimes \rho_{x_n}$, she can just perform the measurement corresponding to the measurement operators $\{P_x\}$ and obtain the string $(x_1, \ldots, x_n)$. She can then compress the message classically (at a rate $r > H(X)$) and send the result to Bob. Bob then decompresses the string and prepares the corresponding string of states (since Bob knows the source $\{p_x, \rho_x\}$).

If the states ρ_x are mixed (not pure), then $S(\rho) > H(X)$ where ρ is the state

$$\rho = \sum_x p_x \rho_x.$$

In the case when ρ_x are orthogonal, it holds that

$$S(\rho) = H(X) + \sum_x p_x \rho_x$$

If the states are really mixed, then we can compress the information even more. It is possible to show that a reliable compression scheme of rate r for a source $\{p_x, \rho_x\}$ is possible only if

$$r > S(\rho) - \sum_x p_x S(\rho_x) \geq 0.$$

If all of the ρ_x are pure, then $S(\rho_x) = 0$ for each x, and a bound for the best achievable rate is $r > S(\rho)$. This is exactly what we had already proved earlier in Schumacher's compression theorem.

19.2 Asymptotic Entanglement Theory

Consider the standard Bell state of two qubits held by Alice and Bob:

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

This state is in a sense the most *useful* state for quantum information processing. So if Alice and Bob share some arbitrary state ρ_{AB}, they would like to know if they can obtain $|\text{Bell}\rangle$ from ρ_{AB} via LOCC. In general, a single copy of an arbitrary state cannot be converted into a Bell state

$$\rho_{\mathsf{AB}} \xrightarrow{\text{LOCC}} |\text{Bell}\rangle.$$

Here, however, we will be concerned with asymptotic entanglement conversion.

Suppose Alice and Bob have a pure state $|\psi\rangle$, but they would like to have some other state $|\phi\rangle$. It is usually not possible to convert $|\psi\rangle \xrightarrow{\text{LOCC}} |\phi\rangle$, but it might be possible to convert many copies of $|\psi\rangle$ to many copies of $|\phi\rangle$

$$|\psi\rangle^{\otimes n} \xrightarrow{\text{LOCC}} |\phi\rangle^{\otimes m}$$

for some integers n and m.

The protocol of ***entanglement distillation*** is as follows. Suppose Alice and Bob have access to a pure state $|\psi\rangle_{\mathsf{AB}}$, and we write $\rho_{\mathsf{AB}} = |\psi\rangle\langle\psi|_{\mathsf{AB}}$. For a positive integer n, Alice and Bob want to know how many copies of the Bell state they can produce via LOCC

$$|\psi\rangle^{\otimes n} \xrightarrow{\text{LOCC}} |\text{Bell}\rangle^{\otimes m},$$

where m is the maximum number of Bell states that may be obtained from n copies of $|\psi\rangle$.

Figure 19.1: Depiction of $(|\psi\rangle_{\mathsf{AB}})^{\otimes n}$ being converted into $(|\text{Bell}\rangle_{\mathsf{AB}})^{\otimes m}$. (Sorry for the low quality)

The protocol of ***entanglement formation*** is the reverse. Suppose Alice and Bob have access to many copies of a Bell state, but they would like to have copies of some other state $|\psi\rangle$. For a positive integer n, Alice and Bob want to know how many copies of $|\psi\rangle$ they can produce via LOCC

$$|\text{Bell}\rangle^{\otimes n} \xrightarrow{\text{LOCC}} |\psi\rangle^{\otimes m},$$

where m is the maximum number copies of $|\psi\rangle$ that may be obtained from n copies of the Bell state. The entanglement of formation is sometimes called the *entanglement cost*.

19.2.1 Entanglement Distillation

Suppose Alice and Bob have access to the state $|\psi\rangle$. Without loss of generality, we may write $|\psi\rangle$ in Schmidt form as

$$|\psi\rangle = \sum_x \sqrt{p_x}|x_{\mathsf{A}}x_{\mathsf{B}}\rangle.$$

Define $\rho = \mathrm{Tr}_{\mathsf{B}}|\psi\rangle\langle\psi|_{\mathsf{AB}}$. If they share n copies of this state, then they have

$$\begin{aligned}|\psi\rangle^{\otimes n} &= \left(\sum_x \sqrt{p_x}|x_{\mathsf{A}}x_{\mathsf{B}}\rangle\right)^{\otimes n} \\ &= \sum_{x_1,x_2,\ldots,x_n} \sqrt{p_{x_1}p_{x_2}\cdots p_{x_n}}|\underbrace{x_1x_2\cdots x_n}_{x^n}\rangle_{\mathsf{A}} \otimes |\underbrace{x_1x_2\cdots x_n}_{x^n}\rangle_{\mathsf{B}},\end{aligned}$$

where x^n is a classical sequence, and we write $p_{x^n} = p_{x_1}p_{x_2}\cdots p_{x_n}$.

We'll show that the optimal achievable rate is $S(\rho)$ (where $\rho = \mathrm{Tr}_{\mathsf{B}}|\psi\rangle\langle\psi|_{\mathsf{AB}}$). Let $\epsilon > 0$. Alice can then perform the measurement corresponding to the operators

$$P(n,\epsilon) = \sum_{x^n \text{ is } \epsilon\text{-typical}} |x^n\rangle\langle x^n|_{\mathsf{A}} \qquad \text{and} \qquad I - P(n,\epsilon),$$

corresponding to outcomes "0" and "1" respectively. She obtains outcome 0 with probability

$$\langle\psi|^{\otimes n}\big(P(n,\epsilon)\otimes I_{\mathsf{B}}\big)|\psi\rangle^{\otimes n} = \sum_{x^n \text{ is } \epsilon\text{-typical}} p_{x^n} = \Pr(x^n \text{ is } \epsilon\text{-typical}) \to 1$$

which goes to 1 as $n \to \infty$. Alice's protocol 'succeeds' if she obtains outcome 0 and her protocol 'fails' if she obtains outcome 1. But she doesn't have to worry about obtaining outcome 1, since it will happen with vanishing probability as n gets large.

After obtaining outcome 0, the the resulting state after the measurement may be denoted

$$|\psi_n(\epsilon)\rangle = \frac{P(n,\epsilon)|\psi\rangle^{\otimes n}}{\sqrt{\Pr(x^n \text{ is } \epsilon\text{-typical})}}.$$

Let $\delta > 0$. Then for all n sufficiently large, the Schmidt coefficients of this state are

$$\frac{p_{x_n}}{\Pr(x^n \text{ is } \epsilon\text{-typical})} \le \frac{p_{x^n}}{1-\delta} \le \frac{2^{-n(S(\rho)-\epsilon)}}{1-\delta}$$

where x^n is ϵ-typical. Let $\vec{p}_n$ be the probability vector of these Schmidt coefficients.

Define $m(n)$ to be the largest integer such that $2^{-m} \ge \frac{1}{1-\delta}2^{-n(S(\rho)-\epsilon)}$. Consider the probability vector

$$\vec{q}_m = \begin{pmatrix} 2^{-m} \\ 2^{-m} \\ \vdots \\ 2^{-m} \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

These are exactly the Schmidt coefficients of the state $|\mathrm{Bell}\rangle^{\otimes m}$.[1] Since $\vec{p}_n \prec \vec{q}_m$, from Nielsen's theorem it follows that

$$|\psi_n(\epsilon)\rangle \xrightarrow{\text{LOCC}} |\mathrm{Bell}\rangle^{\otimes m}.$$

Now,

$$2^{-m} \ge \frac{1}{1-\delta}2^{-n(S(\rho)-\epsilon)} = 2^{-n\left(S(\rho)-\epsilon-\frac{\log(1-\delta)}{n}\right)}$$

[1] Indeed, we have

$$(|\mathrm{Bell}\rangle)^{\otimes m} = \left(\frac{1}{\sqrt{2}}\sum_{x\in\{0,1\}}|xx\rangle_{\mathsf{AB}}\right)^{\otimes m} = \frac{1}{\sqrt{2^m}}\sum_{x_1,\ldots,x_n}|x_1\cdots x_n\rangle_{\mathsf{A}}\otimes|x_1\cdots x_n\rangle_{\mathsf{B}}.$$

and thus $\frac{m}{n} \geq S(\rho) - \epsilon - \frac{\log(1-\delta)}{n}$. Since ϵ and δ can be chosen to be arbitrarily small, it follows that $\frac{m}{n} \geq S(\rho)$. Therefore the optimal limit must satisfy

$$\text{opt} \lim_{n\to\infty} \frac{m}{n} \geq S(\rho)$$

19.2.2 Entanglement of formation

(next time)

20 Lecture 20

(**31 March 2016**)

20.1 Homework hint

Need to show that, if there exists a CPTP map that is LOCC such that $\mathcal{E}(\rho_{\mathsf{AB}}) = \sigma_{\mathsf{AB}}$, then

$$E(\rho_{\mathsf{AB}}) \geq E(\sigma_{\mathsf{AB}})$$

For this particular measure (in the assignment), the only thing that you need to show is that, if ρ_{AB} is separable then $\sigma_{\mathsf{AB}} = \mathcal{E}(\rho_{\mathsf{AB}})$ must also be separable.

20.2 Asymptotic entanglement theory (cont)

Last time we discussed the asymptotic rates of entanglement distillation for bipartite pure states. Now we discuss the reverse protocol.

20.2.1 Entanglement of formation

(Sometimes also called the *entanglement cost*.)

Let $|\psi\rangle_{\mathsf{AB}}$ be a state on systems A and B (Alice and Bob). Without loss of generality we may assume that it is in Schmidt form

$$|\psi\rangle_{\mathsf{AB}} = \sum_x \sqrt{p_x}|x\rangle_{\mathsf{A}} \otimes |x\rangle_{\mathsf{B}}.$$

Suppose Alice and Bob have access to many copies of the Bell state,

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

If they want to obtain n copies of $|\psi\rangle$, Alice and Bob want to know the minimal number m of copies of $|\text{Bell}\rangle$ that are needed to obtain $|\psi\rangle^{\otimes n}$ by LOCC

$$|\text{Bell}\rangle^{\otimes m} \xrightarrow{\text{LOCC}} |\psi\rangle^{\otimes n}.$$

We are interested in the asymptotic limit of the ratio of m and n as n goes to infinity

$$\lim_{n\to\infty} \frac{m}{n}.$$

As before, for entanglement distillation, we can write the n-fold tensor product of $|\psi\rangle$ as

$$|\psi\rangle^{\otimes n} = \sum_{x^n} \sqrt{p_{x^n}}|x^n\rangle|x^n\rangle,$$

where the sum is taken over all sequences x^n of length n. For any $\epsilon > 0$ we define the state

$$|\psi_n(\epsilon)\rangle := \frac{1}{\sqrt{\Pr(x^n \text{ } \epsilon\text{-typical})}} \sum_{x^n \text{ is } \epsilon\text{-typical}} \sqrt{p_{x^n}}|x^n\rangle|x^n\rangle$$

where

$$\Pr(x^n \text{ } \epsilon\text{-typical}) = \sum_{x^n \text{ is } \epsilon\text{-typical}} p_{x^n}.$$

By definition of typicality, almost all of the probabilities p_{x^n} for ϵ-typical sequences x^n must be very close to $2^{-nS(\rho)}$. In fact, we know that

$$2^{-n(S(\rho)+\epsilon)} < p_{x^n} < 2^{-n(S(\rho)-\epsilon)}$$

holds for all ϵ-typical sequences x^n (where $\rho = \mathrm{Tr}_{\mathsf{B}}|\psi\rangle\langle\psi|_{\mathsf{AB}}$).

The original state $|\psi\rangle$ is a vector in the tensor product space $\mathbb{C}^d \otimes \mathbb{C}^d$ (where d is the dimension of the spaces held by Alice and Bob), and $|\psi\rangle^{\otimes n} \in \mathbb{C}^{d^n} \otimes \mathbb{C}^{d^n}$. Now the state $|\psi_n(\epsilon)\rangle$ 'lives' in a subspace

$$|\psi_n(\epsilon)\rangle \in \mathbb{C}^{d'_n} \otimes \mathbb{C}^{d'_n}$$

where $d'_n = |T(n,\epsilon)|$ is the number of ϵ-typical sequences (and d'_n is a number that depends on n). No matter what, the dimension satisfies

$$d'_n = |T(n,\epsilon)| \leq 2^{n(S(\rho)+\epsilon)} < d^n.$$

Furthermore, since all of the probabilities p_{x^n} for typical sequences must all be very close together, the state $|\psi_n(\epsilon)\rangle$ must be very close to a 'maximally entangled' state of dimension d'_n. Define m to be

$$m = \lceil n(S(\rho)+\varepsilon)\rceil.$$

Consider now the m-fold tensor product of the standard Bell state, which we can think of as

$$|\mathrm{Bell}\rangle^{\otimes m} = \frac{1}{\sqrt{2^m}} \sum_{j=1}^{2^m} |jj\rangle,$$

which is the maximally entangled state of dimension 2^m. All of the Schmidt coefficients of $|\mathrm{Bell}\rangle^{\otimes m}$ are just 2^{-m}, whereas the Schmidt coefficients of $|\psi_n(\epsilon)\rangle$ are

$$\frac{p_{x^n}}{\Pr(x^n \text{ } \epsilon\text{-typical})}$$

for every ϵ-typical sequence x^n. However $\Pr(x^n \text{ } \epsilon\text{-typical}) \to 1$ as $n \to \infty$. For n sufficiently large, it follows that

$$2^{-m} \leq 2^{-n(S(\rho)+\epsilon)} < p_{x^n}$$

holds for all ϵ-typical sequences x^n. Therefore the vector of Schmidt coefficients of $|\mathrm{Bell}\rangle^{\otimes m}$ is majorized by the vector of Schmidt coefficients of $|\psi_n(\epsilon)\rangle$, and thus

$$|\mathrm{Bell}\rangle^{\otimes m} \xrightarrow{\text{LOCC}} |\psi_n(\epsilon)\rangle.$$

Finally, we note that the state $|\psi_n(\epsilon)\rangle$ can be brought arbitrarily close to the state $|\psi\rangle^{\otimes n}$ if n is sufficiently large, since their fidelity is

$$\begin{aligned}[F(|\psi_n(\epsilon)\rangle, |\psi\rangle^{\otimes n})]^2 &= \langle\psi_n(\epsilon)|\psi\rangle^{\otimes n} \\ &= \frac{1}{\sqrt{\Pr(x^n \text{ } \epsilon\text{-typical})}} \sum_{x^n \text{ } \epsilon\text{-typical}} p_{x^n} = \sqrt{\Pr(x^n \text{ } \epsilon\text{-typical})}\end{aligned}$$

which approaches 1 as $n \to \infty$. Hence, if we choose $m = \lceil n(S(\rho)+\varepsilon)\rceil$, then $|\mathrm{Bell}\rangle^{\otimes m}$ can be brought arbitrarily close to $|\psi\rangle^{\otimes n}$ for n sufficiently large. So the optimal limit is at most

$$\lim_{n\to\infty} \frac{\lceil n(S(\rho)+\varepsilon)\rceil}{n} = S(\rho) + \epsilon.$$

Since ϵ can be chosen to be arbitrarily small, it follows that the optimal limit is

$$\text{opt} \lim_{n\to\infty} \frac{m}{n} \leq S(\rho). \tag{20.1}$$

What we really want to show is that

$$\text{opt} \lim_{n\to\infty} = S(\rho).$$

How can we show that $S(\rho)$ really is optimal?

20.2.2 Optimality of the rate $S(\rho)$

We showed last time that the optimal rate of entanglement *distillation* satisfies

$$\text{opt} \lim_{n\to\infty} \frac{m}{n} \geq S(\rho).$$

Compare this to the optimal rate of entanglement formation in (20.1). What does this mean?

- If Alice and Bob start with n copies of $|\psi\rangle$, there is a protocol for Alice and Bob to obtain at least $m = nS(\rho)$ copies of $|\text{Bell}\rangle$ (if n is large enough). That is, the optimal rate of distillation r_d is at least $r_d \geq S(\rho$.
- If Alice want n copies of $|\psi\rangle$, there is a protocol where Alice and Bob require at most $m = nS(\rho)$ copies of $|\text{Bell}\rangle$ (if n is large enough). That is, the optimal rate of formation r_f is at most $r_f \leq S(\rho)$.

Putting this together, we see that the asymptotic rates of distillation r_d and formation r_f for any state $|\psi\rangle$ must satisfy $r_f \leq S(\rho) \leq r_d$.

However, we will show that the optimal rates of distillation and formation are equal to each other, by showing that $r_f < r_d$ is impossible (i.e. it is impossible to distill more Bell states than from what you started with).

Consider a protocol where Alice and Bob start with some number of Bell states which they convert into n copies of $|\psi\rangle$ at an optimal rate of formation r_f, then convert those copies of $|\psi\rangle$ back into Bell states at an optimal rate of distillation r_d as depicted here:

$$|\text{Bell}\rangle^{\otimes nr_f} \xrightarrow{\text{LOCC}} |\psi\rangle^{\otimes n} \xrightarrow{\text{LOCC}} |\text{Bell}\rangle^{\otimes nr_d}.$$

If $r_d > r_f$, this would mean that we could actually end up with more Bell states than what we started with! This is clearly impossible. It follows that $r_f = r_d = S(\rho)$.

Reversibility The fact that the optimal rate of entanglement *distillation* is equal to the optimal rate of entanglement *formation* implies reversibility in the asymptotic limit. As n goes to infinity, the entanglement 'loss' for converting many copies of the Bell states $|\text{Bell}\rangle^{\otimes m} \xrightarrow{\text{LOCC}} |\psi\rangle^{\otimes n}$ to some number of copies of a state $|\psi\rangle$, and then back to the maximal number of copies of the Bell state will go to zero (assuming optimal conversion rates are performed at each step).

That is, for large n, we can convert $|\psi\rangle^{\otimes n}$ into $|\text{Bell}\rangle^{\otimes(nS(\rho))}$ and back with vanishing loss of entanglement as $n \to \infty$.

Interpretation of the rates of distillation and formation How do we interpret these numbers? Given some number n, what is the number of Bell states that are needed to produce n copies of $|\psi\rangle$? The answer is $nS(\rho)$ (in the limit of large n). Conversely, what is the number of Bell states that can be produced from n copies of $|\psi\rangle$? Again, in the limit of large n, the answer is $nS(\rho)$. So we say that $S(\rho)$ is the *conversion rate* in the asymptotic limit.

20.3 Some properties of entropy and information measures

Recall some notions from classical information. The *Shannon entropy* of a random variable X (whose probability distribution is $\{p_x\}$) is defined by

$$H(X) = H(p_1, p_2, \ldots, p_n) = -\sum_k p_k \log p_k.$$

We can think of the probability distribution as a vector $\vec{p}$ whose entries are p_k. If we have two random variables X and Y, the *joint probability* of the random variables is a probability distribution $\{p_{xy}\}$.

Some properties of the Shannon entropy include

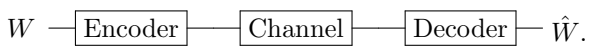
- **Concavity**:
$$H\Big(\sum_i t_i \vec{p}_i\Big) \geq \sum_i t_i H(\vec{p}_i)$$
for any collection of probability vectors $\{\vec{p}_i\}$ and any $t_i \geq 0$ satisfying $\sum_i t_i = 1$.
- **Symmetry**: $H(X,Y) = H(Y,X)$
- **Subadditivity**: $H(X,Y) \leq H(X) + H(Y)$ with equality if and only if X and Y are independent.
- **Strong subadditivity**: $H(X,Y,Z) + H(X) \leq H(X,Y) + H(Y,Z)$, with equality if and only if $Z \to Y \to X$ forms a Markov chain.

Given random variables X and Y with joint probability distribution $\{p_{xy}\}$, the ***joint entropy*** of X and Y is defined as
$$H(X,Y) = -\sum_{x,y} p_{xy} \log p_{xy}$$
If X and Y are *independent* random variables, then the joint probability distribution can be written as $\{p_{xy}\} = \{p_x q_y\}$ for individual probability distributions $\{p_x\}$ and $\{p_y\}$ of X and Y. It follows that
$$H(X,Y) = H\left(\{p_x q_y\}\right) = -\sum_{x,y} p_x q_y \log(p_x q_y) = -\sum_x p_x \log p_x - \sum_y q_y \log q_y.$$
Now consider some random variables X, Y, and Z, and their joint probability distribution $\{p_{xyz}\}$. We define the conditional probabilities
$$p_{x|yz} := \frac{p_{xyz}}{\sum_{x'} p_{x'yz}}.$$
These random variables form a ***Markov chain*** if it holds that
$$p_{x|yz} = p_{x|y}$$
for all x, y, z, where $p_{x|y}$ are the conditional probabilities of x given y supposing no information is known about Z.

The ***mutual information*** of random variables X and Y is defined as
$$H(X\colon Y) = H(X) + H(Y) - H(X,Y).$$
It holds that $H(X\colon Y) \geq 0$ for all random variables X and Y.

20.3.1 Message encoding

Suppose Alice and Bob have a channel between them for transmitting information (from Alice to Bob). Alive wants to transmit a message W by encoding it, sending it through the channel, and Bob decodes it to obtain the message $\hat{W}$ (after transmission, with possible errors):



The channel here is a classical channel on some alphabet $\mathcal{X}$. Given some input $x \in \mathcal{X}$ into the channel, the probability that the channel outputs some output $y \in \mathcal{X}$ is some probability $p_{y|x}$ (i.e. the probability that y is output given that x is the input). Alice can make multiple uses of this channel, but the goal is to use it as little as possible but still accurately send the intended message.

Alice samples a random source that outputs messages from the alphabet $i \in \{1, 2, \ldots, M\}$ (i.e. there are M different possible messages, each occurring with some probability). The source produces each message with some probability $\Pr(W = i)$. She converts the messages into a sequence of n

letters from the alphabet $\mathcal{X}$. Each message i is encoded into a distinct message or *codeword*. The codewords are denoted

$$\{x^n(1), x^n(2), \ldots, x^n(M)\}$$

where the set of these codewords is called the *codebook*. The encoding-decoding message-sending scheme works as follows:

1. Alice samples the source and obtains a message i that she wishes to send to Bob.
2. She then sends the corresponding codeword $x^n(i)$ through the channel, by sending each term of the sequence separately.
3. Bob then receives a sequence of letters from the alphabet $\mathcal{X}^n$ (which may or not be the same sequence sent by Alice if the channel is 'noisy', i.e. introduces errors). Hes must attempt to correctly decode it to obtain the message i.
4. Bob's decoder is a function $g : \mathcal{X}^n \to \{1, \ldots, M\}$ that takes sequences from the output of the channel and assigns them to messages.

The choice of codebook $\{x^n(1), \ldots, x^n(M)\}$ together with the function g is called an (M, n)-code for this source and channel.

Alice wants to make sure that the probability of Bob interpreting the message incorrectly to be as small as possible. For a given n, Alice and Bob want to choose an (M, n)-code that reduces the probability error. Define the probability of error for sending the message i to be

$$\lambda_i^{(n)} = \sum_{\{y^n \,:\, g(y^n) \neq i\}} \Pr(y^n | x^n(i)).$$

This is the probability that the message that Bob interprets a message different that i given that Alice wanted to send i. The maximal probability of error is

$$\lambda_{\max}^{(n)} = \max_i \lambda_i^{(n)}.$$

The *rate* of an (M, n)-code is defined as $\frac{\log M}{n}$ (i.e. the number of bits of information that can be sent per n uses of the channel).

Let $r > 0$. We say that the rate r is ***achievable*** if there exists a sequence of $(\lceil 2^{nr} \rceil, n)$-codes such that

$$\lambda_{\max}^{(n)} \to 0$$

as $n \to \infty$. The *capacity* C of a channel is defined as the supremum over all achievable rates r.

Theorem 20.1 (Shannon's capacity theorem for a classical channel)**.** *The maximum capacity of a channel is*

$$C = H(p)???$$

(This didn't make any sense... He also erased it way too fast.)

20.3.2 Conditional information

The ***conditional information*** of random variables X and Y is defined as

$$H(X|Y) = H(X) - H(X : Y) = H(X, Y) - H(Y).$$

It holds that $H(X|Y) \geq 0$ for all random variables X and Y.

Proof. Consider a probability distribution $\{p_{xy}\}$. Then

$$\begin{aligned} -\sum_{x,y} p_{xy} \log p_{xy} + \sum_{x,y} p_{xy} \log p_y &= -\sum_{x,y} p_{xy} \log \left(\frac{p_{xy}}{p_y} \right) \\ &= -\sum_{x,y} p_{xy} \log(p_{x|y}) \\ &\geq 0. \end{aligned}$$

□

The quantum version of the conditional entropy is *not* positive.

21 Lecture 21

(**5 April 2016**)

21.1 Homework hints

- For problem 3(c) of assignment 5: You only need to find a separable σ such that $S(\rho\|\sigma) = S(\rho_{\mathsf{A}})$. Proving that it is optimal is too hard.
- Problem 4 was already on assignment 4, but here we only need to prove the equality conditions.

21.2 Information measures

The ***conditional mutual information*** of random variables X, Y, and Z is defined by

$$H(X:Z|Y) = H(X,Y) + H(Y,Z) - H(Y) - H(X,Y,Z).$$

It holds that $H(X:Z|Y) \geq 0$ with equality if and only if $X \to Y \to Z$ forms a Markov chain.

21.2.1 The relative entropy

For probability distributions $\{p(x)\}$ and $\{q(x)\}$, their (classical) ***relative entropy*** is defined by

$$H(p\|q) = \sum_x p(x) \log \frac{p(x)}{q(x)} = -H(X) - \sum_x p(x) \log q(x)$$

where $H(x) = -\sum_x p(x) \log p(x)$. This is really the 'mother' of many other important quantities that arise in information theory.

Theorem 21.1 (Properties of the relative entropy)**.** *The following properties of the classical relative entropy hold.*

- ***Positivity****:* $H(p\|q) \geq 0$ *with equality if and only if* $p(x) = q(x)$ *for all* x*.*
- ***Sub-additivity****:* $H(X,Y) \leq H(X) + H(Y)$
- $H(X) \leq \log d$

These properties can be proven from the quantum case. For example, if ρ and σ are density operators that are simultaneously diagonalizable,

$$\rho = \sum_i p_i |i\rangle\langle i| \qquad \text{and} \qquad \sum_i q_i |i\rangle\langle i|,$$

then $S(\rho\|\sigma) = H(p\|q)$.

21.2.2 The von Neumann entropy

For a density operator ρ, the ***von Neumann entropy*** of ρ is

$$S(\rho) = -\operatorname{Tr}[\rho \log \rho].$$

Theorem 21.2. *Some important properties of the von Neumann entropy:*

(1) $S(\rho) \geq 0$ *with equality if and only if* ρ *is pure.*

(2) $S(\rho) \leq \log d$ *with equality if and only if* $\rho = \frac{1}{d} I$*.*

(3) If ρ_{AB} is pure, then $S(\mathsf{A}) = \mathsf{S}(\mathsf{B})$, where $S(\mathsf{A}) = S(\mathrm{Tr}_{\mathsf{B}}\,\rho_{\mathsf{AB}})$ and $S(\mathsf{B}) = S(\mathrm{Tr}_{\mathsf{A}}\,\rho_{\mathsf{AB}})$.

(4) ***Concavity****:*

$$\sum_i p_i S(\rho_i) \leq S\left(\sum_i p_i \rho_i\right) \leq \sum_i p_i S(\rho_i) + H(p_i).$$

(5) ***Additivity****:* $S(\rho \otimes \sigma) = S(\rho) + S(\sigma)$

(6) ***Sub-additivity****:* $|S(\mathsf{A}) - S(\mathsf{B})| \leq S(\mathsf{A},\mathsf{B}) \leq S(\mathsf{A}) + S(\mathsf{B})$

(7) ***Strong sub-additivity****:*

$$S(\mathsf{A},\mathsf{B},\mathsf{C}) + S(\mathsf{B}) \leq S(\mathsf{A},\mathsf{B}) + S(\mathsf{B},\mathsf{C})$$

and

$$S(\mathsf{A}) + S(\mathsf{B}) \leq S(\mathsf{A},\mathsf{B}) + S(\mathsf{B},\mathsf{C}).$$

The entropy gives a sense of how 'uniform' the eigenvalues of ρ are. If all of the eigenvalues are the same (i.e. $\frac{1}{d}$), then $S(\rho)$ is maximal.

Proof (sketch) of Theorem 21.2. Note that property (3) is actually a necessary and sufficient condition.

(4) Consider the so-called 'flag states'

$$\sum_i p_i |i\rangle\langle i| \otimes \rho_i = \bigoplus_i p_i \rho_i$$

(5) Note that $\log(\rho \otimes \sigma) = (\log \rho) \otimes I + I \otimes (\log \sigma)$. Then

$$\begin{aligned}\mathrm{Tr}[\rho \otimes \sigma \log(\rho \otimes \sigma)] &= \mathrm{Tr}[(\rho \otimes \sigma)(\log \rho \otimes I + I \otimes \log \sigma)] \\ &= \mathrm{Tr}[(\rho \log \rho) \otimes \sigma] + \mathrm{Tr}[\rho \otimes (\sigma \log \sigma)] \\ &= \mathrm{Tr}[\rho \log \rho]\,\mathrm{Tr}[\sigma] + \mathrm{Tr}[\rho]\,\mathrm{Tr}[\sigma \log \sigma] = \mathrm{Tr}[\rho \log \rho] + \mathrm{Tr}[\sigma \log \sigma].\end{aligned}$$

(7) The proof of this is hard. [1]

□

There are other areas there the von Neumann entropy occurs naturally, for example in thermodynamics. Many problems in thermodynamics are considered in the 'thermodynamic limit' (i.e. waiting a long time) in which the law of large numbers starts to kick in. 'Quantum thermodynamics' is a relatively new subfield of Quantum Information that combines the two.

There are some classical quantities that have quantum counterparts, and some don't. Here are some of the quantum counterparts.

21.2.3 Quantum mutual information

The ***joint entropy*** of two quantum systems A and B (in state ρ_{AB}) is $S(\mathsf{A},\mathsf{B}) = S(\rho_{\mathsf{AB}})$. The entropies of the subsystems are

$$S(\mathsf{A}) = S(\rho_{\mathsf{A}}) \quad \text{and} \quad S(\mathsf{B}) = S(\rho_{\mathsf{B}}).$$

The ***mutual information*** of systems A and B is defined as

$$S(\mathsf{A} : \mathsf{B}) = S(\mathsf{A}) + S(\mathsf{B}) - S(\mathsf{A},\mathsf{B}).$$

It holds that $S(\mathsf{A} : \mathsf{B}) \geq 0$ with equality if and only if $\rho_{\mathsf{AB}} = \rho_{\mathsf{A}} \otimes \rho_{\mathsf{B}}$.

[1] Originally proved by Lieb in the 1970s, and there is no known 'simple proof' of this. It is an open question in quantum information as to whether there is a simple proof or not

21.2.4 Quantum conditional information

The ***conditional information*** is defined as

$$S(\mathsf{A}|\mathsf{B}) = S(\mathsf{A}) - S(\mathsf{A}:\mathsf{B}) = S(\mathsf{A},\mathsf{B}) - S(\mathsf{B}).$$

If $\rho_{\mathsf{AB}} = |\psi\rangle\langle\psi|$ is pure, then $S(\mathsf{A},\mathsf{B}) = 0$. In this case, $S(\mathsf{B}) = 0$ if and only if $|\psi\rangle_{\mathsf{AB}}$ is separable. Hence $|\psi\rangle$ is entangled if and only if $S(\mathsf{A}|\mathsf{B}) < 0$.

What would it mean to have 'negative conditional entropy' mean[2]? This is David's research project.

21.2.5 Quantum relative entropy

The 'rock star' of all entropies. Many other important quantities are defined from this.

The ***quantum relative entropy*** of ρ and σ is defined as

$$S(\rho\|\sigma) = \mathrm{Tr}[\rho\log\rho] - \mathrm{Tr}[\rho\log\sigma] = -S(\rho) - \mathrm{Tr}[\rho\log\sigma].$$

The relative entropy of entanglement is a 'distance-like measure'. This fact is related to the ***relative entropy of entanglement***, which is defined as

$$E_R(\rho) = \min_{\sigma\in\mathrm{Sep}} S(\rho\|\sigma).$$

This gives us a somewhat 'geometric' picture of entanglement, by giving us a sense of how 'far' a state ρ is from the set of separable states. A picture can be found in Fig. 21.1.

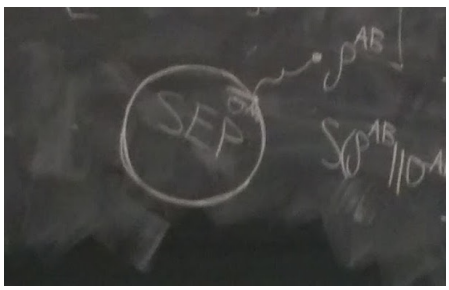


Figure 21.1: Depiction of the relative entropy of entanglement. Given a state ρ_{AB} outside of the set SEP of separable states, we can find the state $\sigma \in$ SEP that is 'closest' to ρ with respect to the 'distance' defined by the relative entropy.

Theorem 21.3 (Klein)**.** *Let ρ and σ be density operators. Then $S(\rho\|\sigma) \geq 0$ with equality if and only if $\rho = \sigma$.*

Note that proving this is on the assignment. The proof in Nielsen and Chuang [1] has an error! (Well, actually they just sweep something under the rug.)

Theorem 21.4 (Joint convexity)**.** *Let $\{p_i\}$ be probabilities and let ρ_i and σ_i be density operators for every i. It holds that*

$$S\left(\sum_i p_i\rho_i \middle\| \sum_i p_i\sigma_i\right) \leq \sum_i p_i S(\rho_i\|\sigma_i).$$

[2] M Horodecki, J Oppenheim, A Winter. *Quantum information can be negative.* Nature **436**, 673-676 (2005).

Theorem 21.5 (Monotonicity)**.** *Let ρ_{AB} and σ_{AB} be density operators. It holds that*

$$S(\rho_{\mathsf{AB}}\|\sigma_{\mathsf{AB}}) \geq S(\rho_{\mathsf{A}}\|\sigma_{\mathsf{A}}).$$

Monotonicity is related to the picture in Fig. 21.2. That is, *erasing* some of the information can only make the two states closer together.

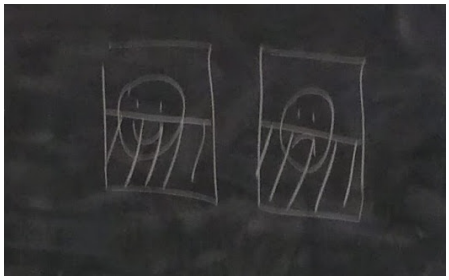


Figure 21.2: Smiley and frowny faces. Covering up the lower half make the two picture indistinguishable.

Theorem 21.6 (Monotonicity)**.** *Let ρ and σ be density operators and $\mathcal{E}$ be a CPTP map. It holds that*

$$S(\mathcal{E}(\rho)\|\mathcal{E}(\sigma) \leq S(\rho\|\sigma).$$

Proof. By Stinespring dilation, we can write the channel $\mathcal{E}$ as

$$\mathcal{E}(X) = \mathrm{Tr}_{\mathsf{E}}\left[U(X \otimes |0\rangle\langle 0|_{\mathsf{E}})U^*\right]$$

for all X. Now

$$\begin{aligned} S(\mathcal{E}(\rho)\|\mathcal{E}(\sigma) &= S\left(Tr_{\mathsf{E}}\left[U(\rho \otimes |0\rangle\langle 0|_{\mathsf{E}})U^*\right] \middle\| \mathrm{Tr}_{\mathsf{E}}\left[U(\sigma \otimes |0\rangle\langle 0|_{\mathsf{E}})U^*\right]\right) \\ &\leq S\left(U(\rho \otimes |0\rangle\langle 0|_{\mathsf{E}})U^* \,\middle\|\, U(\sigma \otimes |0\rangle\langle 0|_{\mathsf{E}})U^*\right) \\ &= S\left(\rho \otimes |0\rangle\langle 0|_{\mathsf{E}} \,\middle\|\, \sigma \otimes |0\rangle\langle 0|_{\mathsf{E}}\right) \\ &= S(\rho\|\sigma), \end{aligned}$$

where we note that $S(\rho \otimes \tau\|\sigma \otimes \gamma) = S(\rho\|\sigma) + S(\tau\|\gamma)$ holds for all $\rho, \sigma, \tau, \gamma$ (left as an exercise to the reader). □

Theorem 21.7. *The mutual information of quantum systems is positive.*

$$I(\mathsf{A} : \mathsf{B}) \geq 0$$

Proof. Note that

$$\begin{aligned} 0 \leq S(\rho_{\mathsf{AB}}\|\rho_{\mathsf{A}} \otimes \rho_{\mathsf{B}}) &= \mathrm{Tr}[\rho_{\mathsf{AB}} \log(\rho_{\mathsf{AB}})] - \mathrm{Tr}[\rho_{\mathsf{AB}} \log(\rho_{\mathsf{A}} \otimes \rho_{\mathsf{B}})] \\ &= -S(\rho_{\mathsf{AB}}) - \mathrm{Tr}[\rho_{\mathsf{AB}}(\log \rho_{\mathsf{A}} \otimes I + I \otimes \log \rho_{\mathsf{B}})] \\ &= -S(\rho_{\mathsf{AB}}) - \mathrm{Tr}[\rho_{\mathsf{A}} \log \rho_{\mathsf{A}}] - \mathrm{Tr}[\rho_{\mathsf{B}} \log \rho_{\mathsf{B}}], \end{aligned}$$

where we note that we can split the trace Tr into $\mathrm{Tr}_{\mathsf{A}} \mathrm{Tr}_{\mathsf{B}}$ or $\mathrm{Tr}_{\mathsf{B}} \mathrm{Tr}_{\mathsf{A}}$ in either order. Hence

$$S(\rho_{\mathsf{AB}}\|\rho_{\mathsf{A}} \otimes \rho_{\mathsf{B}}) = S(\rho_{\mathsf{A}}) + S(\rho_{\mathsf{B}}) - S(\rho_{\mathsf{AB}}) = I(\mathsf{A} : \mathsf{B}),$$

which proves the desired result. □

Note that the quantum mutual information is not a measure of entanglement, but rather a measure of *correlations*. The mutual information vanishes if and only if ρ is a product state with the form $\rho_{\mathsf{AB}} = \rho_{\mathsf{A}} \otimes \rho_{\mathsf{B}}$.

21.2.6 Other quantities

We compared the relative entropies of ρ_{AB} with ρ_{A} and ρ_{B}. It is also natural to consider I. If I is the identity operator on a d-dimensional space, then $\frac{1}{d}I$ is a normalized density operator. Let ρ be an arbitrary density operator. Now

$$0 \leq S(\rho \| \tfrac{1}{d} I) = \mathrm{Tr}[\rho \log \rho] - \mathrm{Tr}[\rho \log \tfrac{1}{d} I] = -S(\rho) + \log d.$$

This implies that $S(\rho) \leq \log d$.

We also consider the following:

$$\begin{aligned}
0 &\leq S(\rho_{\mathsf{AB}} \| \tfrac{1}{d_{\mathsf{A}}} I \otimes \rho_{\mathsf{B}}) \\
&= -S(\rho_{\mathsf{AB}}) - \mathrm{Tr}[\rho \log(\tfrac{1}{d_{\mathsf{A}}} I \otimes \rho_{\mathsf{B}})] \\
&= -S(\rho_{\mathsf{AB}}) - \mathrm{Tr}[\rho(\log(\tfrac{1}{d_{\mathsf{A}}} I_{\mathsf{A}}) \otimes I_{\mathsf{B}})] - \mathrm{Tr}[\rho(I_{\mathsf{A}} \otimes \log \rho_{\mathsf{B}})] \\
&= -S(\rho_{\mathsf{AB}}) + \log d_{\mathsf{A}} + S(\rho_{\mathsf{B}}) \\
&= \log d_{\mathsf{A}} + S(\mathsf{A}|\mathsf{B}).
\end{aligned}$$

Finally, we note that $S(\mathsf{A}|\mathsf{B}) =$ is concave as a function ρ_{AB}. (why?)

21.2.7 Proof of strong sub-additivity

Define a function T on tripartite states on ABC as follows:

$$\begin{aligned}
T(\rho_{\mathsf{ABC}}) &= S(\mathsf{A}) + S(\mathsf{B}) - S(\mathsf{A},\mathsf{C}) - S(\mathsf{B},\mathsf{C}) && (21.1)\\
&= -S(\mathsf{C}|\mathsf{A}) - S(\mathsf{C}|\mathsf{B}). && (21.2)
\end{aligned}$$

Since $S(\mathsf{C}|\mathsf{A})$ and $S(\mathsf{C}|\mathsf{B})$ are concave, it follows that T is convex in ρ_{ABC}. Now consider a state ρ_{ABC} in spectral decomposition

$$\rho_{\mathsf{ABC}} = \sum_i r_i |\psi_i\rangle\langle\psi_i|_{\mathsf{ABC}}.$$

Then, by convexity of T,

$$T(\rho_{\mathsf{ABC}}) \leq \sum_i r_i T(|\psi_i\rangle\langle\psi_i|_{\mathsf{ABC}}) \leq 0,$$

where $T(|\psi_i\rangle\langle\psi_i|_{\mathsf{ABC}}) = 0$ (i.e. T vanishes for pure states) because.....?

For the other part of strong sub-additivity, we want to prove that

$$S(\mathsf{A}) + S(\mathsf{B}) \leq S(\mathsf{A},\mathsf{B}) + S(\mathsf{B},\mathsf{C}).$$

Let ρ_{ABC} be an arbitrary density operator. Consider a purification $|\psi\rangle_{\mathsf{ABCD}}$ such that

$$\rho = \mathrm{Tr}_{\mathsf{D}} |\psi\rangle\langle\psi|_{\mathsf{ABCD}}.$$

Then it is clear that

$$S(\mathsf{B},\mathsf{C},\mathsf{D}) + S(\mathsf{B}) \leq S(\mathsf{C},\mathsf{D}) + S(\mathsf{B},\mathsf{C}).$$

But since $|\psi\rangle$ is pure, it holds that $S(\mathsf{B},\mathsf{C},\mathsf{D}) = S(\mathsf{A})$ and $S(\mathsf{C},\mathsf{D}) = S(\mathsf{A},\mathsf{B})$, which yields the desired result.

22 Lecture 22

(**7 April 2016**)

22.1 More entropic properties

Theorem 22.1. *(1)* $S(\mathsf{A}|\mathsf{B},\mathsf{C}) \leq S(\mathsf{A}|\mathsf{B})$

(2) $S(\mathsf{A}:\mathsf{B}) \leq S(\mathsf{A}:\mathsf{B},\mathsf{C})$

Theorem 22.2 (Data processing inequality)**.** *Let* $\mathcal{E}^{\mathsf{A}\to\mathsf{A}'}$ *and* $\mathcal{E}^{\mathsf{B}\to\mathsf{B}'}$ *be channels, and define* $\sigma_{\mathsf{A}'\mathsf{B}'} = \mathcal{E}^{\mathsf{A}\to\mathsf{A}'} \otimes \mathcal{E}^{\mathsf{B}\to\mathsf{B}'}(\rho_{\mathsf{AB}})$*. It holds that*

$$S(\mathsf{A}':\mathsf{B}')_\sigma \leq S(\mathsf{A}:\mathsf{B})_\rho.$$

Proof. Note that $S(\mathsf{A}':\mathsf{B}')_\sigma =$

□

Follows from strong subadditivity.

22.2 Wrap-up

All of quantum information theory can be viewed as a theory of interconversion between various resources.

The goals of quantum infromation theory are the following.

- Ifentify classes of resources in quantum mechanics
- Identify elemental dynamical ...

22.3 Accessible information and Holevo bound

Suppose Alice prepares a state ρ_x with $x \in \{1,2,\ldots,n\}$ corresponding to probabilities $p_1, p_2, \ldots, p_n$ of some random variable X. Alice sends ρ_x to Bob through an identity channel.

(image)

How can Bob determine x? He can perform a measurement with Kraus operators $\{K_y\} = \{K_1,\ldots,K_m\}$ with possible outcomes $y \in \{1,\ldots,m\}$ (random variable Y). The probability that y is obtained given that Alice sent x is $p_{y|x}$. Hence the joint distribution of X and Y is then $p_{x,y} = p_{y|x}p_x$, and

$$\mathrm{Tr}[E_y\rho_x] = p_{y|x}.$$

where $E_y = K_yK_y^*$.

The ***acessible information*** of the random variable X and Alice's encoding $\{\rho_x\}$ is defined by

$$A = \max_{\{K_y\}} H(X:Y),$$

where the maximization is taken over all possible measurements that Bob could perform. This concept doesn't exist classically, because it would always be equal to $H(X)$ (because we can set $Y = X$).

Theorem 22.3. *It holds that* $\max_{\{K_y\}} H(X:Y) \leq H(X)$*, with equality if and only if the optimal measurement that Bob can perform gives absolute certainty about what* X *is.*

Example 22.4. Alice prepares a pure state $|\psi\rangle$ with probability p and another pure state $|\varphi\rangle$ with probability $1-p$. However, two non-orthogonal states cannot be reliably distinguished....

Example 22.5 (No-cloning theorem)**.** Suppose there existed a 'cloning' machine
(Graphic)
Any quantum operation must be represented by a unitary operator.

$$|\psi\rangle \otimes |T\rangle \to^U U(|\psi\rangle \otimes ||T\rangle\rangle) = |\psi\rangle \otimes |\psi\rangle$$

$$\langle\psi|\phi\rangle = (\langle\psi|\phi\rangle)^2 \qquad \Rightarrow \qquad \langle\psi|\phi\rangle = 1 \text{ or } |\psi\rangle \perp |\phi\rangle$$

Example 22.6 (No-cloning theorem again)**.** Alice prepares a pure state $|\psi\rangle$ with probability p and another pure state $|\varphi\rangle$ with probability $1-p$. If Bob's accessible information is equal to $H(p)$ then Bob can clone the state!

Otherwise: If a cloning machine exists....

22.3.1 Holevo bound

The accessible information A is very difficult to compute, since it involves optimizing over all possible POVMs. The following theorem gives an upper bound to the accessible information.

Theorem 22.7. *Given a random variable X and en encoding $\{\}$, it holds that*

$$A = \max_{\{K_y\}} H(X:Y) \leq S(\rho) - \sum_{x=1}^{n} p_x S(\rho_x)$$

with $\rho = \sum_x p_x \rho_x$.

The quantity on the right.hand side is called the ***Holevo bound***. It is 'achievable asymptotically'. Most of the time it is a strict inequality, but in the asymptotic limit of many copies it can become an equality (in most cases). The Holevo bound is proven using strong subadditivity. Also note that this bound is *better* than the bound discussed earlier

$$A = \max_{\{K_y\}} H(X:Y) \leq S(\rho) - \sum_{x=1}^{n} p_x S(\rho_x) \leq H(X).$$

Proof. Define a quantum state

$$\rho_{\mathsf{XBY}} = \sum_x p_x |x\rangle\langle x|_{\mathsf{X}} \otimes \rho_{\mathsf{B},x} \otimes |0\rangle\langle 0|_{\mathsf{Y}}.$$

and define

$$\sigma_{\mathsf{X'B'Z'}} = \sum_{x,y} p_x |x\rangle\langle x|_{\mathsf{X}} \otimes (K_y \rho_{\mathsf{B},x} K_y^*) \otimes |y\rangle\langle y|_{\mathsf{Y}}.$$

This represents thee result after Bob measures using the Kraus operators $\{K_y\}$ and takes note of the result y that was obtained. Now

$$\begin{aligned} S(\mathsf{X}:\mathsf{B})_\rho &= S(\mathsf{X}:\mathsf{B},\mathsf{Y})_\rho \\ &\geq S(\mathsf{X'}:\mathsf{B'},\mathsf{Y'})_\sigma && \text{(by data processing inequality)} \\ &\geq S(\mathsf{X'}:\mathsf{Y'})_\sigma && \text{(from strong subadditivity).} \end{aligned}$$

However,

$$S(\mathsf{X}:\mathsf{B})_\rho = S(\mathsf{X}) + S(\rho) - S(\mathsf{X},\mathsf{B}) = H(\mathsf{X}) + S(\rho) - \sum_x p_x S(\rho_x) - H(X).$$

This gives the upper bound. Furthermore,

$$\begin{aligned}\sigma_{\mathsf{X}'\mathsf{Y}'} &= \mathrm{Tr}_{\mathsf{B}'}\, \sigma_{\mathsf{X}'\mathsf{B}'\mathsf{Y}'} \\ &= \sum_{x,y} p_x \underbrace{\mathrm{Tr}\Big[K_y^* K_y \rho_x\Big]}_{p_{y|x}} \otimes |y\rangle\langle y| \\ &= \sum_{x,y} p_x p_{y|x} |x\rangle\langle x| \otimes |y\rangle\langle y|,\end{aligned}$$

and $S(\mathsf{X}' : \mathsf{Y}')_\sigma = H(X') + H(Y') - H(X', Y')$. □

References

[1] Nielsen, Michael, and Chuang, Isaac. *Quantum Information and Quantum Computation.*

[2] Watrous, John. *Theory of Quantum Information.*

[3] Wilde, Mark M. *From Classical to Quantum Shannon Theory.*