QUANTUM INFORMATION THEORY

IAN LIM
LAST UPDATED MARCH 26, 2019

These notes were taken for the *Quantum Information Theory* course taught by Nilanjana Datta at the University of Cambridge as part of the Mathematical Tripos Part III in Lent Term 2019. I live-TEXed them using Overleaf, and as such there may be typos; please send questions, comments, complaints, and corrections to `itel2@cam.ac.uk`.
Many thanks to Arun Debray for the LATEX template for these lecture notes: as of the time of writing, you can find him at `https://web.ma.utexas.edu/users/a.debray/`.

Contents

Lecture 1.

Friday, January 18, 2019

Note. Here's the relevant admin content for the first day. The lecturer's email is `n.datta@damtp.cam.ac.uk`, and course notes can be found on the CQIF website under Part III Lectures.

Quantum information theory (QIT) was born out of classical information theory (CIT).

Definition 1.1. Classical information theory is the mathematical theory of information processing tasks, e.g. storage, transmission, processing of information.

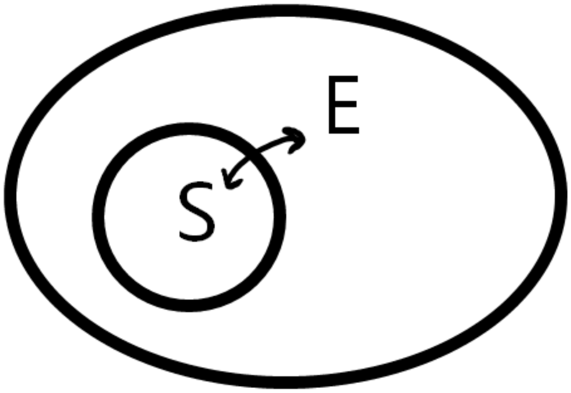


Figure 1. A sketch of the sort of systems we will be interested in in this class. We have an open system S which will naturally interact with its environment E.

In contrast, quantum information theory asks how these tasks can be performed if we harness quantum mechanical systems as information carriers. Such systems include electrons, photons, ions, etc.

QM has some novel features which are not present in our old Newtonian theories. We know that quantum systems obey the Heisenberg uncertainty principle, that energy is quantized in these systems, and QM systems cannot generically be copied (the famous no-cloning theorem). Quantum mechanically, one can describe the full state of a system without knowing the state of the subsystems– this is essentially the idea of entanglement.[1]

Here's a quick overview now of the structure of the course.

- Basic concepts of CIT
- Study of open quantum systems
- Mathematical tools for QIT
- Entanglement
- QIT itself

When we say open quantum systems, we mean quantum systems which interact with a broader environment. If we prepare a state and allow it to interact, what happens to the information stored in that state?

Classical information theory Historically, CIT was invented in 1948 with a pioneering paper by Claude Shannon. In this paper, he asked two critical questions.

Q1. What is the limit to which information can be *reliably* compressed?

Q2. What is the maximum rate at which information can be reliably sent through a communication channel?

That is, we may ask about how to encode information in such a way that it can still be recovered with a high probability of success. And we can ask how to send this information when our communication channels will naturally be noisy. The answers to these questions are known as *Shannon's Source Coding Theorem* and *Shannon's Noisy Channel Coding Theorem*, respectively.

What is information? We have an intuitive sense of what information means, but to formalize this takes a little work. In the loosest sense, information is associated to uncertainty and in particular information gain is related to a reduction in uncertainty.

Example 1.2. Suppose I have a system which takes some discrete values, e.g. I roll a fair die. The outcome is a variable x which takes values in some set, $J = \{1, 2, \ldots, 6\}$. We write that capital X is proportional to $p(x), x \in J$, where $P(X = x) = p(x) = 1/6 \, \forall x \in J$. That is, there is a probability mass function associated to the possible outcomes. The probability that we measure the system X in outcome x is $1/6$ for any outcome x in the set of outcomes.

[1] If you like, some composite states in a tensor product space cannot be decomposed into a direct product.

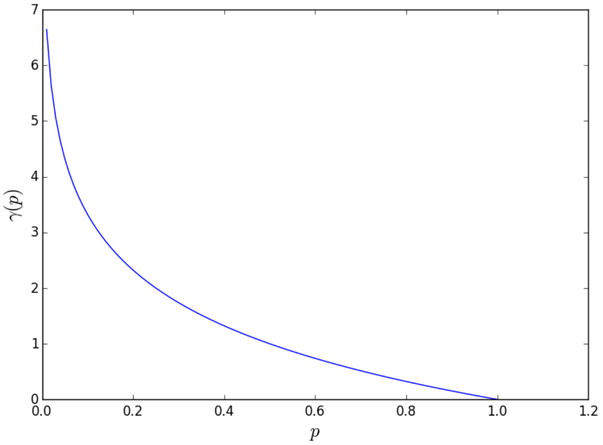


FIGURE 2. The surprisal $\gamma(p) \equiv -\log_2 p$ as a function of p, the probability of some event. Certainties ($p = 1$) are not very surprising, whereas very rare events ($p \ll 1$) are surprising, and so get $\gamma = 0$ and γ large respectively.

We also define the following quantity.

Definition 1.3. *Surprisal* is the quantity

$$\gamma(x) = -\log p(x). \tag{1.4}$$

When an event is very unlikely and it happens anyway... you are very surprised. For example, $p(x) = 1 \implies \gamma(x) = 0$ (certainties are not very surprising) while $p(x) \approx 0 \implies \gamma(x)$ large. See Fig. 2 for a plot of γ versus p.

This quantity has some features:

- It only depends on $p(x)$ and not on x.
- It is a continuous function of $p(x)$.
- It is additive for independent events.

This last property is easy to prove:

$$P(X = x, Y = y) = P_{XY}(x,y) = P_X(x)P_Y(y)$$

when X, Y are independent. Then

$$\gamma(x,y) = -\log P_{XY}(x,y) = \gamma(x) + \gamma(y).$$

Definition 1.5. We can now define the *Shannon entropy* of X to be

$$H(X) \equiv \mathbb{E}(\gamma(X)) = \sum_{x\in J}(-\log p(x))p(x), \tag{1.6}$$

the expected value of the surprisal. We see again that $H(X)$ does not depend on the actual outcomes themselves but only on the probability distribution $P(X)$.

As a matter of convention we will take logs to be $\log \equiv \log_2$, and for events which are impossible, $P(x) = 0$, we have $0 \log 0 = 0$ (which one can prove by taking the limit $\lim_{u\to 0} u \log u = 0$).

Binary entropy Consider an event which has two possible outcomes, $X \sim P(x), x \in J = \{0,1\}$ where $P(X = 0) = p$ and $P(X = 1) = 1 - p$. Then the Shannon entropy is

$$H(X) = -p \log p - (1-p)\log(1-p) \equiv h(p). \tag{1.7}$$

We see that if the probability is $p = 1/2$, then we have no information a priori about this systems– the entropy is maximized. $h(p)$ is a continuous function of p, and it is concave. See the illustration in Fig. 3.

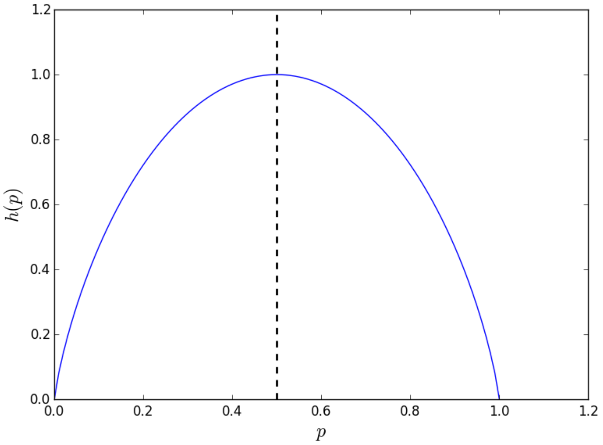


FIGURE 3. The Shannon entropy of a binary event where there are two possible outcomes, one of which happens with probability p and the other with $1 - p$. When $p = 0.5$, our ignorance is at a maximum– we know nothing a priori about what our generator will spit out.

Definition 1.8. We can also define a different entropy, the Rényi entropy, which is

$$H_\alpha(X) = \frac{1}{1-\alpha}\log\left(\sum_{x\in J} p(x)^\alpha\right), \tag{1.9}$$

with $\alpha \in (1, 2]$. As an exercise, we can verify that $\lim_{\alpha\to 1} H_\alpha(X) = H(X)$, i.e. the Renyi entropy reduces to the Shannon entropy.[2]

Why do we choose to work with the Shannon entropy? It has to do with the operational interpretation– the Shannon entropy represents an optimal rate of data compression, i.e. the data compression limit.

In CIT, a classical information source emits some messages/data/signals/information. For instance, J could output a binary output or perhaps telegraph English (26 letters and a space). Now, the simplest class of sources is *memoryless*– they are "independent identically distributed" sources (i.i.d.), which means that successive messages are independent of each other, and they are identically distributed.

Definition 1.10. Suppose we have some random variables $U_1, U_2, \ldots, U_n$ with $U_i \sim p(u), u \in J$. We say these are *identically distributed* if

$$p(u) = P(U_k = u), u \in J \quad \forall 1 \le k \le n.$$

We could study a signal emitted by n uses of the source to get some sequence $\underline{u}^{(n)} = (u_1, u_2, \ldots, u_n)$.

Definition 1.11. Moreover, if the probability mass function takes the form

$$\begin{aligned} p(\underline{u}^{(n)}) &= P(U_1, \ldots, U_n = u_n) \\ &= p(u_1) \ldots p(u_n). \end{aligned}$$

[2] The proof is fairly quick. First note that as $\alpha \to 1$, the denonimator $1 - \alpha$ goes to zero and the log becomes $\log(\sum_{x\in J} p(x)) = \log 1 = 0$, so we can apply L'Hôpital's rule and take some derivatives. Note also that $\frac{d}{dx}a^x = \frac{d}{dx}e^{\log a^x} = \frac{d}{dx}e^{x\log a} = \log a e^{x\log a} = a^x \log a$. Thus by L'Hôpital's rule,

$$\begin{aligned} \lim_{\alpha\to 1} H_\alpha(X) &= \lim_{\alpha\to 1}\frac{1}{1-\alpha}\log\left(\sum_{x\in J} p(x)^\alpha\right) \\ &= \lim_{\alpha\to 1}\frac{1}{(-1)}\frac{\sum_{x\in J}(p(x)^\alpha \log p(x))}{\sum_{x\in J} p(x)^\alpha} \\ &= -p(x)\log p(x) = H(X). \end{aligned}$$

Technically I have done this calculation with a natural log rather than a base 2 log, but the result is the same, since the numerical factor from taking the derivative of the log cancels with the factor from rewriting the derivative of $p(x)^a$ in terms of a base 2 log. ☒

If the source is indeed independent and identically distributed, then it makes sense to describe it by a sincle probability mass function, $U \sim p(u)$, so that the Shannon entropy of the source can be said to be

$$H(U) = -\sum_{u \in J} p(u) \log p(u). \tag{1.12}$$

Another guiding question. Why is data compression possible? Our information source has some *redundancy*. For instance, in the English language, certain letters are more common than others, so we can encode something that is more common in a shorter string in anticipation it will be used more often.

This sort of scheme is known as variable length coding, e.g. we might encode the letter "e" as the string 10 and the letter "z" as 11000. In contrast, we could also use a fixed length coding scheme where we have a "typical set", a subset of our total outcomes J^n (things we might like to encode). Our typical set then has a one-to-one mapping to the set of encoded messages, e.g. $\{0,1\}^m$, so we can always recover them precisely, while several outcomes outside the typical set might map to the same encoded message. There's some probability that we'll want to encode things outside the typical set, and in decoding we'll get the original message a little bit wrong. But if we choose the typical set well, this can be made to be a rare occurrence. We are usually interested in *asymptotic i.i.d.* settings, i.e. in the limit as the size of the set of possible messages to be encoded goes to ∞.

Example 1.13. Suppose we have a horse race with eight horses. They have labels $1, 2 \ldots, 8$, and the message we would like to encode is the label of the winning horse. A priori, we only need 3 bits to encode the label since 2^n different messages can be stored in n bits.

However, what if the horses are not all equally fast (i.e. likely to win)? Suppose that p_i is the probability of the ith horse winning, such that

$$p_i = 1/2, 1/4, 1/8, 1/16, 1/64, \ldots, 1/64.$$

Now we assign the following code words:

$$\begin{aligned} C(1) &= 0 \\ C(2) &= 10 \\ C(3) &= 110 \\ C(4) &= 1110 \\ C(5) &= 111100 \\ C(6) &= 111101 \\ C(7) &= 111110 \\ C(8) &= 111111. \end{aligned}$$

Let l_i be the length of the ith codeword, e.g. $l_5 = 6$. We can compute that the average length of a code is then $\sum p_i l_i = 2$, and we've chosen a "prefix-free code" so that a sequence like 10011001110 can be uniquely decoded to a sequence of winners from our code words. That is, no codeword is a prefix of any other code.[3]

Let's compute the expected length of the codeword– it is

$$\sum_i p_i l_i = 1 \times \frac{1}{2} + 2 \times \frac{1}{4} + 3 \times \frac{1}{8} + 4 \times \frac{1}{16} + 4 \times \frac{1}{64} \times 6 = 2, \tag{1.14}$$

and this is exactly the Shannon entropy of the system, as expected.

Lecture 2.

Monday, January 21, 2019

Last time, we introduced Shannon's Source Coding Theorem:

Theorem 2.1. *For an i.i.d (memoryless) source, the optimal rate of reliable data compression (i.e. the data compression limit) is precisely the Shannon entropy $H(X)$ of the source.*

[3] For the sequence 10011001110, we know that the first winner was the horse corresponding to 10, horse 2. The next winner was horse 1 with code 0. This sequence breaks up as 10|0|110|0|1110, so the winners were 2, 1, 3, 1, and 4 in that order.

We started by saying that if we have an iid source, we can model it by a collection of n sources $U_1, U_2 \ldots, U_n$ which outputs a length-n vector $\underline{u}^{(n)} = (u_1, \ldots, u_n) u_i \in J$. For an iid source, all the sources have the same probability mass function,

$$U_i \sim p(u), u \in J,$$

which means that we can equivalently model the source as a single source,

$$U \sim p(u), u \in J; p(\underline{u}^{(n)}) = \prod_{i=1}^{n} P(U_i = u_i) = p(u_1) \ldots p(u_n).$$

The Shannon entropy of the source is given as usal by

$$H(U) = -\sum_{u \in J} p(u) \log p(u). \tag{2.2}$$

Now let us define a compression map.

Definition 2.3. A *compression map* of *rate* R is a map $\mathcal{C}$ with

$$\mathcal{C}^n : \underline{u}^{(n)} = (u_1, \ldots, u_n) \mapsto \underline{x}^{m_n} = (x_1, \ldots, x_{m_n}) \in \{0,1\}^{m_n}. \tag{2.4}$$

That is, $\mathcal{C}$ maps our output string of length n to a compressed (encoded) string $\underline{x}$ of length m_n. We say that the *rate* of encoding is then

$$R = \frac{m_n}{n} = \frac{\text{number of bits in codeword}}{\text{number of uses of source}}. \tag{2.5}$$

If a compression scheme has rate R, then we assign unique codewords to $2^{\lceil nR \rceil}$ messages.

Question: when is such a map $\mathcal{C}^n$ a compression map? If our source outputs n values in the alphabet J, then we have total possibilities

$$|J|^n = 2^{n \log |J|}. \tag{2.6}$$

These can be stored in $n \log |J|$ bits. Thus c^n is a compression map if $m_n < n \log |J|$, i.e. if we encode the output in fewer bits than it would take to uniquely encode every single string in the naive binary way.

We can of course also define a decompression map:

Definition 2.7. A decompression map $\mathcal{D}^n$ is a map

$$\mathcal{D}^n : \underline{x}^{m_n} \in \{0,1\}^{m_n} \mapsto \underline{u}^{(n)} = (u_1, \ldots, u_n), \tag{2.8}$$

i.e. which takes us back to the original length-n strings of source outputs.

Now we can ask what the probability of a successful encoding and decoding is– namely,

$$\sum_{\underline{u}^{(n)} \in J^n} p(\underline{u}^{(n)}) P\Big(\mathcal{D}^n(\mathcal{C}^n(\underline{u}^{(n)})) \neq \underline{u}^{(n)}\Big) \tag{2.9}$$

is the average probability of error of the compression process. We write this as $P_{av}^{(n)}(C_n)$, where C_n denotes an encoding and decoding scheme.

Definition 2.10. C_n is a triple defined to be $C_n \equiv (\mathcal{C}^n, \mathcal{D}^n, R)$ which represents a choice of code. We say that a code is *reliable* if $P_{av}^{(n)} \to 0$ in the limit as $n \to \infty$. That is, $\forall \epsilon \in (0,1), \exists n$ such that $p_{av}^{(n)} \leq \epsilon$.

Then there is an optimal rate of data compression,

$$R_\infty = \inf\{R : \exists C_n(\mathcal{C}^n, \mathcal{D}^n, R) \text{ s.t. } p_{av}^{(n)}(C_n) \to 0 \text{ as } n \to \infty\}. \tag{2.11}$$

That is, R_∞ is effectively the minimum rate R of all reliable coding schemes. What Shannon's source coding theorem tells us is that $R_\infty = H(U)$. The lowest rate (highest density, if you like) we can reliably compress an iid source to is given by the Shannon entropy.

Definition 2.12. An ϵ-typical sequence is a sequence defined as follows. Fix $\epsilon \in (0,1)$ and take an iid source with $U \sim p(u), u \in J$ which gives us a length-n output $\underline{u}^{(n)} = (u_1, \ldots, u_n)$. Then if

$$2^{-n(H(U)+\epsilon)} \leq p(\underline{u}^{(n)}) \leq 2^{-n(H(U)-\epsilon)}, \tag{2.13}$$

we say that $\underline{u}^{(n)}$ is an ϵ-typical sequence.

Definition 2.14. An *ϵ-typical set* is then defined to be the set

$$T_\epsilon^{(n)} = \{\underline{u}^{(n)} \in J^n \text{ such that } 2.13 \text{ holds}\}. \tag{2.15}$$

In the asymptotic limit let us observe that

$$p(\underline{u}^{(n)}) \approx 2^{-nH(U)}, \tag{2.16}$$

so all ϵ-typical sequences are almost equiprobable since ϵ can be made arbitrarily small. Does this agree with our intuitive notion of a typical sequence? Yes– take a sequence $\underline{u}^{(n)} = (u_1, \dots, u_n), u_i \in J$. Note that for every $u \in J$, the number of times we expect to u to appear in a string $\underline{u}^{(n)}$ is simply $np(u)$.

Our intuition tells us that any typical sequence should therefore fit this expectation.[4] The probability of getting one specific typical sequence is

$$\begin{aligned} p(\underline{u}^{(n)}) &\simeq \prod_{u \in J} p(u)^{np(u)} \\ &= \prod_u 2^{np(u)\log p(u)} \\ &= 2^{n \sum p(u) \log p(u)} \\ &= 2^{-nH(U)}. \end{aligned}$$

So this agrees well with our formal definition of a typical sequence. Note that there is a difference between typical and high-probability– we'll investigate this distinction further on the example sheet.

Now, typical sequences have some nice properties.

Theorem 2.17 (Typical sequence theorem)**.** *$\forall \delta > 0$ and large n,*

- $H(U) - \epsilon \leq -\frac{1}{n} \log p(\underline{u}^{(n)}) \leq H(u) + \epsilon$[5]
- $P(T_\epsilon^{(n)}) := \sum_{\underline{u}^{(n)} \in T_\epsilon^{(n)}} p(\underline{u}^{(n)}) > 1 - \delta$*. That is, the probability of getting any typical sequence (as a subset of possible outputs) can be made arbitrarily close to* 1.
- $2^{n(H(U)-\epsilon)}(1-\delta) < |T_\epsilon^{(n)}| \leq 2^{n(H(U)+\epsilon)}$*, where* $|T_\epsilon^{(n)}|$ *is the number of typical (length n) sequences.*[6]

Since $\epsilon > 0$, we see that in the limit $\epsilon \to 0$,

$$|T_\epsilon^{(n)}| \to 2^{nH(U)}. \tag{2.18}$$

That is, we need $nH(U)$ bits to store all the typical sequences.

Now we can state Shannon's theorem formally.

[4] To make this more concrete, suppose we have a weighted coin. The weighted coin has outcomes h and t (heads and tails), and it produces h with probability $3/4$ and t with probability $1/4$. If we flip the coin n times, we expect to see about $n \times p(h) = n \times 3/4$ heads and $n \times p(t) = n \times 1/4$ tails since each flip is independent. If $n = 4$, for instance, then a "typical sequence" will have three heads and one tails.

Consider a specific example of a length-4 typical sequence, *hhht* in that order. The probability of getting this specific sequence is $p(h) \times p(h) \times p(h) \times p(t) = 27/256$. We could have written this as $(p(h))^{np(h)} \times (p(t))^{np(t)}$, or equivalently $2^{np(h) \times \log p(h)} \times 2^{np(t) \times \log p(t)}$. Combining terms, we see that this is just $2^{n(p(h)\log p(h) + p(t) \log p(t))} = 2^{-nH(U)}$.

This is not the probability of getting *any* sequence which fits the typical sequence condition! That probability would be something like $\binom{4}{3}$ times the probability we got, since we want exactly three heads. However, we will put a bound on this quantity shortly.

[5] This follows from taking the log of the definition of an ϵ-typical sequence and dividing by $-n$.

[6] Since the probability of any individual typical sequence is bounded from below by definition and there are $|T_\epsilon^{(n)}|$ such sequences, the probability of getting *any* typical sequence is bounded by

$$2^{-n(H(U)+\epsilon)} |T_\epsilon^{(n)}| \leq \sum_{\underline{u}^{(n)} \in T_\epsilon^{(n)}} p(\underline{u}^{(n)}) \leq 1.$$

This leads us to conclude that $|T_\epsilon^{(n)}| \leq 2^{n(H(U)+\epsilon)}$.

However, $|T_\epsilon^{(n)}|$ is also bounded from below. We know from the previous property and the definition of a typical sequence that

$$1 - \delta < \sum p(\underline{u}^{(n)}) \leq 2^{-n(H(U)-\epsilon)} |T_\epsilon^{(n)}|,$$

so $2^{n(H(U)-\epsilon)}(1-\delta) < |T_\epsilon^{(n)}|$.

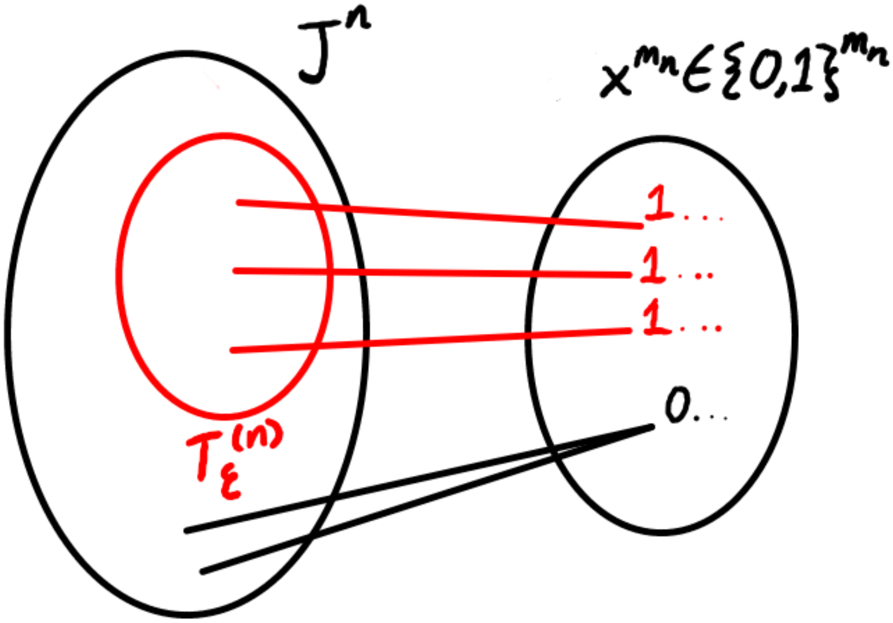


FIGURE 4. An illustration of the encoding procedure for the achievability part of the Shannon source coding theorem. Of our source's possible outputs J^n, we set up a one-to-one encoding of the typical set $T_\epsilon^{(n)}$, (red ellipse), and send all other elements of J^n to some random value in our set of codewords.

Theorem 2.19 (Shannon's Source Coding Theorem)**.** *Suppose we have an iid source U with Shannon entropy $H(U)$.*

(a) *(Achievability) Suppose $R > H(U)$. Then $\exists$ a reliable compression-decompression scheme of rate R.*
(b) *(Converse) For $R < H(U)$, any compression-decompression scheme is not reliable.*

Constructive proof of (a) Let us suppose that $R > H(U)$. We fix $\epsilon \in (0,1)$ such that $R > H(U) + \epsilon$ (for instance, $\epsilon = (R - H(U))/2$). Then we choose n large enough (i.e. the asymptotic limit) such that $T_\epsilon^{(n)}$ satisfies the conditions of the typical sequence theorem. Then we can write

$$|T_\epsilon^{(n)}| \leq 2^{n(H(U)+\epsilon)} < 2^{nR}. \tag{2.20}$$

Now we divide our set of sequences J^n into the typical set T_ϵ^n and its complement $A_\epsilon^n = J^n \setminus T_\epsilon^n$. Let us then order the elements of our typical set, i.e. we assign some labels/indices to all the elements. Since $|T_\epsilon^n| < 2^{nR}$, we need at most nR bits to store all the labels of the typical sequences (i.e. the ones we always want to recover reliably).[7]

With our encoding scheme for the typical set in hand, let us preface our encoding with a 1, i.e. a *flag bit*. So the typical set elements will be encoded as

$$\underline{u}^{(n)} \in T_\epsilon^n \mapsto 1\underbrace{011\ldots1}_{\lceil nR \rceil}. \tag{2.21}$$

Our codewords will be of length $\lceil nR \rceil + 1$, and we can assign the complement A_ϵ^n to some random codeword beginning with a 0 instead. This procedure is shown in Fig. 4. So our rate of success when we decode will not be exactly 1– we can perfectly decode typical set elements, but there is some loss when we encode elements outside the typical set. However, things are not so bad. Let us take the limit as $n \to \infty$

[7] As nR may not be an integer, we'll practically need at most $\lceil nR \rceil$ bits.

and look at the failure probability $p_{av}^{(n)}$.

$$\begin{aligned} p_{av}^{(n)} &:= \sum p(\underline{u}^{(n)}) P\Big(\mathcal{D}^n(\mathcal{C}^n(\underline{u}^{(n)})) \neq \underline{u}^{(n)}\Big) \\ &= \sum_{\underline{u}^{(n)} \in T_\epsilon^n} p(\underline{u}^{(n)}) P(\underline{u}'^{(n)} \neq \underline{u}^{(n)}) + \sum_{\underline{u}^{(n)} \in A_\epsilon^n} p(\underline{u}^{(n)}) P(\underline{u}'^{(n)} \neq \underline{u}^{(n)}). \end{aligned}$$

But the first term is zero since we can always decode typical set elements, and the second part can be made to be arbitrarily small ($< \delta$) by the typical sequence theorem. Therefore we conclude that our scheme is reliable.[8] ⊠

Lemma 2.22. *Suppose we have a set S^n which has size $|S^n| = 2^{nR}$, with $R < H(U)$. $\forall \delta > 0, S_n \subset J^n$ s.t. $|S^n| = 2^{nR}$ with $R < H(U)$, we have $P(S^n) < \delta$ for n large enough.*

This implies the converse, and is in the course notes (but is useful to think on by oneself).

Non-lectured aside: the converse I'll present here an argument for the above lemma. A similar exposition appears in the official course notes.

We have some set S^n with size $|S^n| = 2^{nR}$. That is, we can encode and decode at most 2^{nR} elements with perfect precision. What elements should we choose?

We know that the probability of our source producing any element in the atypical set $A_\epsilon^{(n)}$ becomes arbitrarily small by the typical sequence theorem, so in order to give our encoding scheme the best chance of success, we should not bother with encoding any elements in $A_\epsilon^{(n)}$. But note that

$$|S^n| = 2^{nR} < 2^{nH(U)} < |T_\epsilon^{(n)}|$$

for some $\epsilon > 0$, so we cannot encode the entire typical set. At best, we can encode a subset of $T_\epsilon^{(n)}$.

Let's do that, then. We take $S^n \subset T_\epsilon^{(n)}$, and note that the probability of any individual typical sequence is $2^{-nH(U)}$. Since we have 2^{nR} such sequences in S^n, the probability of our source producing any sequence in S^n is simply

$$P(S^n) = \sum_{\underline{u}^{(n)} \in S^n} p(\underline{u}^{(n)}) = 2^{nR} 2^{-nH(U)} = 2^{-n(H(U)-R)}. \tag{2.23}$$

Since $R < H(U)$ by assumption, $H(U) - R > 0 \implies P(S^n) = 2^{-n(H(U)-R)} \to 0$ as $n \to \infty$. Thus $\forall \delta > 0$, $\exists N$ such that $P(S^n) < \delta$ for $n \geq N$.

One interpretation of this is as follows– we tried to encode a subset of the typical set, hoping that any elements in $T_\epsilon^{(n)} \setminus S^n$ wouldn't totally ruin our encoding scheme. However, what we didn't account for was the limit $n \to \infty$. The number of typical sequences grows too fast for our encoding scheme to keep up, so that the probability of our source producing a typical sequence we didn't encode is

$$P(T_\epsilon^n) - P(S^n) > 1 - \delta - 2^{-n(H(U)-R)}, \tag{2.24}$$

which can be made arbitrarily close to 1. The moral of the story is that if we don't encode the entire typical set at a minimum, our scheme is doomed to fail.

Lecture 3.

Wednesday, January 23, 2019

Let's recall the statement of Shannon's source coding theorem. Shannon tells us that if we have an iid source $U \sim p(u); u \in J$ with Shannon entropy $H(U)$, then there is a fundamental limit on data compression given by $H(U)$ such that for any rate $R > H(U)$, there exists a reliable compression-decompression scheme of rate R, and conversely for any rate $R < H(U)$, any scheme of rate R will not be reliable.

[8] That is, since $P(T_\epsilon^n) > 1 - \delta$, it follows that $P(A_\epsilon^n) < \delta$ in the large-n limit. So the nonzero failure rate is washed out by the fact that

$$\sum_{\underline{u}^{(n)} \in A_\epsilon^n} p(\underline{u}^{(n)}) P(\underline{u}'^{(n)} \neq \underline{u}^{(n)}) \leq \sum_{\underline{u}^{(n)} \in A_\epsilon^n} p(\underline{u}^{(n)}) = P(A_\epsilon^n) < \delta$$

for δ arbitrarily small.

See my notes from last lecture for a heuristic argument of the converse. The formal argument can be made with ϵs and δs– for example, my statement that we need not consider elements in $A_\epsilon^{(n)}$ is equivalent to $\sum_{\underline{u}^{(n)} \in S^n \cap A_\epsilon^n} p(\underline{u}^{(n)}) \leq P(A_\epsilon^n \to 0$.

Entropies Consider a pair of random variables X, Y with *joint probability*

$$P(X = x, Y = y) = P_{XY}(x,y) = p(x,y). \tag{3.1}$$

Here, $x \in J_X$ some alphabet and similarly $y \in J_Y$. We can also define the conditional probability

$$P(Y = y|X = x) = p(y|x), \tag{3.2}$$

the probability of y given x.

Definition 3.3. Now we have the *joint entropy*, which is

$$H(X,Y) \equiv -\sum_{x \in J_X, y \in J_Y} p(x,y) \log p(x,y). \tag{3.4}$$

Definition 3.5. We also have the *conditional entropy*, which is

$$\begin{aligned} H(Y|X) &\equiv \sum_x p(x) H(Y|X = x) \\ &= -\sum_x p(x) \sum_y p(y|x) \log p(y|x). \end{aligned}$$

But we can simplify this to write

$$H(Y|X) = -\sum p(x,y) \log p(y|x), \tag{3.6}$$

which implies that

$$p(x,y) = p(x)p(y|x) = p(y)p(x|y). \tag{3.7}$$

This leads us to a chain rule,

$$H(X,Y) = H(Y|X) + H(X). \tag{3.8}$$

We also have the notion of a relative entropy, which measures a "distance" between two probability distributions. Suppose we have distributions $p = \{p(x)\}_{x \in J}$ and $q = \{q(x)\}_x \in J$, Let us assume that the supp$p \subseteq$ supp$q$, with supp$p = \{x \in J : p(x) > 0\}$. This implies that $q(x) = 0 \implies p(x) = 0$, which we denote $p \ll q$.

Definition 3.9. Thus we define the *relative entropy* to be

$$D(p||q) \equiv \sum_{x \in J} p(x) \log \frac{p(x)}{q(x)}. \tag{3.10}$$

If $p \ll q$, then this is well-defined (otherwise we might have $q \to 0$ with p nonzero). Taking $0 \log \frac{0}{q(x)} = 0$ we see that this represents a sort of distance,

$$D(p||q) \geq 0 \tag{3.11}$$

with equality iff $p = q$.

This is not quite a true metric, since it is not symmetric, $D(p||q) \neq D(q||p)$, and moreover it does not satisfy a triangle inequality, i.e. $D(p||r) \not\leq D(p||q) + D(q||r)$.

Using the relative entropy, we can now define a useful quantity known as the mutual information.

Definition 3.12. The mutual information between two sources X and Y is

$$\begin{aligned} I(X;Y) &= H(X) + H(Y) - H(X,Y) \\ &= H(X) - H(X|Y). \end{aligned}$$

The mutual information has some intuitive properties.

- $I(X : X) = H(X)$, since $I(X;X) = H(X) + H(X) - H(X,X) = H(X)$.
- $I(X;Y) = I(Y;X)$
- if X, Y independent, then $I(X;Y) = 0$.

Suppose now we have P, Q taking non-negative real values, with $Q(x) = 0 \implies P(X) = 0$. THus the relative entropy is

$$D(P||Q) = \sum P(x) \log \frac{P(x)}{Q(x)}.$$

What if $P(x) = p(x), x \in J$ and $Q(x) = 1 \forall x \in J$? Then

$$D(P||Q) = \sum_x p(x) \log p(x) = -H(X). \tag{3.13}$$

It's almost trivial to check that if $Q(x) = \frac{1}{|J|}$ instead, then we would get an additional factor of $-\log |J|$.

Exercise 3.14. Check that the mutual information satisfies

$$I(X;Y) = D(p(x,y)||p(x)p(y)). \tag{3.15}$$

Let's take a minute to prove the non-negativity of the relative entropy. That is, $D(p||q) \geq 0$.

Proof. By definition,

$$D(p||q) = \sum_{x \in J} p(x) \log \frac{p(x)}{q(x)}. \tag{3.16}$$

Let us define a set A such that

$$A = \{x \in J \text{ s.t. } p(x) >)\}.$$

Thus A is the support of J. We can compute

$$\begin{aligned} -D(p||q) &= \sum p(x) \log \frac{q(x)}{p(x)} &(3.17)\\ &= \mathbb{E}_p \left(\log \frac{q(X)}{p(X)} \right). &(3.18) \end{aligned}$$

Note that Xs denote random variables, while xs indicate the values they take.

Jensen's inequality from probability theory tells us that for convave functions f, $\mathbb{E}(f(X)) \leq f(\mathbb{E}(X))$.

We conclude that

$$\begin{aligned} -D(p||Q) &\leq \log(\mathbb{E}_p \frac{q(X)}{p(X)}) \\ &= \log \sum_{x \in A} p(x) \frac{q(x)}{p(x)} \\ &\leq \log \sum_{x \in J} q(x) \\ &= \log 1 = 0 \\ &\implies D(p||Q) \geq 0. \end{aligned}$$

⊠

Suppose we had a distribution $p = \{p(x)\}, q(x) \frac{1}{|J|} \forall x \in J$ as before. Then

$$\begin{aligned} 0 \leq D(p||q) &= \sum p(x) \log \frac{p(x)}{(1/|J|)} &(3.19)\\ &= -H(X) + \sum p(x) \log |J| &(3.20)\\ \implies H(X) &\leq \log |J|. &(3.21) \end{aligned}$$

Lecture 4.

Friday, January 25, 2019

Last time, we introduced many important classical concepts. We talked about the mutual (common) information $I(X:Y)$ between two sources, arguing that

$$\begin{aligned} I(X:Y) &= H(X) + H(Y) - H(X,Y) \\ &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X). \end{aligned}$$

In particular we find that $I(X:X) = H(X), I(X:Y) = I(Y:X)$, and $I(X:Y) = 0$ iff X, Y are independent.

We can also prove that the mutual information is non-negative,

$$I(X:Y) \geq 0, \tag{4.1}$$

which follows from writing in terms of the conditional entropy as $H(X) - H(X|Y) \geq 0$. Equivalently we should show that

$$H(X|Y) \leq H(X). \tag{4.2}$$

That is, *conditioning reduces entropy*.

We may describe the concavity of $H(X)$– that is, for two sources with $X, Y; J$ with $\lambda \in [0, 1]$

$$H(\lambda p_X + (1-\lambda)p_Y) \geq \lambda H(p_x) + (1-\lambda)H(p_Y), \tag{4.3}$$

which we will prove on the first examples sheet. This is analogous to what we showed in a few lines about the binary entropy.

The Shannon entropy of $H(X,Y)$ (where we simply replace $p(x)$s in the definition of $H(X)$ with $p(x,y)$) is constrained by the following inequality:

$$H(X,Y) \geq H(X) + H(Y). \tag{4.4}$$

This property is known as *subadditivity*.

We also have the property that the conditional entropy is non-negative–

$$H(X|Y) \geq 0. \tag{4.5}$$

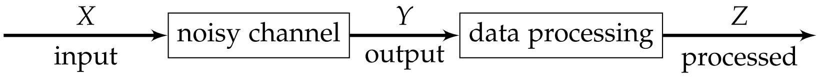
Equivalently, $H(X,Y) - H(Y) \geq 0$ We shall see that once we introduce quantum correlations, this will no longer be true.

Data processing inequality Suppose we have some variables $X_1, X_2, \ldots$. In a Markov chain, we say that the probability of some outcome $X_n = x_n$ in a chain is

$$P(X_n = x_n | X_1 = x_1 \ldots X_{n-1} = x_{n-1}) = P(X_n = x_n | X_{n-1} = x_{n-1}). \tag{4.6}$$

That is, the value of a Markov chain at a position n depends only on its value at $n-1$.

Consider a simple Markov chain with three variables, $X \to Y \to Z$.



Then by the definition of a Markov chain, $P(Z = z | X = x, Y = y) = P(Z = z | Y = y)$, and we can prove that

$$I(X:Z) \leq I(X:Y), \tag{4.7}$$

known as the *data processing inequality* (DPI). That is, there is no data processing that can increase the correlation between two random variables.

Chain rules Chain rules are relations between different entropy quantities, e.g. $H(X,Y) = H(X) + H(Y|X)$. Suppose we have three random variables X, Y, Z with a joint probability of $p(x,y,z)$.

Exercise 4.8. Prove that

$$H(X,Y,Z) = H(X) + H(Y|X) + H(Z|X,Y). \tag{4.9}$$

Definition 4.10. Now one can define the *conditional mutual information* as follows:

$$I(X:Y|Z) := H(X|Z) - H(X|Y,Z) \geq 0, \tag{4.11}$$

with equality when $X - Y - Z$ forms a Markov chain.

We have one more topic for classical information theory– it is *Shannon's Noisy Channel Coding Theorem.* As usual, let us work in the asymptotic iid limit.

Suppose we have some source X producing outputs in an alphabet J_X, and some received signals $Y \in J_Y$. We also have a noisy channel $\mathcal{N} : J_X \to J_Y$, and a stochastic map, defined to be a set of probabilities $\{p(y|x)\}$.

Here's the setup. Alice wants to send a message m to her friend Bob. To do this, she takes her message $m \in \mathcal{M}$ a set of messages and runs an encoding process $\mathcal{E}_n$ to produce a codeword $x_m^{(n)}$. She uses the (possibly noisy) channel $\mathcal{N}$ multiple times, say n times, to send a transmitted message $y_m^{(n)} \neq x_m^{(n)}$, which Bob then runs a decoding process $\mathcal{D}_n$ on to get a final decoded message m'.

If $m' \neq m$, we have gotten an error. In the $n \to \infty$ limit, we would like the probability of error $p_{err}^{(n)} = p(m' \neq m) \to 0$.

In some sense, encoding is like the dual process of compression. In encoding, we add redundancy in a controlled manner to account for the potential noise of the channel $\mathcal{N}$.

Definition 4.12. We define a *discrete channel* to be the following:

- An input alphabet J_X
- An output alphabet J_Y
- A set of conditional probabilities (dependent on the number of uses n) $\{p(\underline{y}^n|\underline{x}^n)\}$.

The input to n uses of the channel sends n uses of the source, $\underline{x}^{(n)} = (x_1, \ldots, x_n) \in J_X^n$ to an output $\underline{y}^{(n)} = (y_1 \ldots y_n) \in J_Y^n$ with probability $p(\underline{y}^n|\underline{x}^n)$.

We can consider memoryless channels, i.e. where the probability of n uses completely separates into n independent uses of the channel as

$$p(\underline{y}^n|\underline{x}^n) = \prod_{i=1}^n p(y_i|x_i). \tag{4.13}$$

For a memoryless channel, we may write the transition probabilities as a *channel matrix*,

$$\begin{bmatrix} p_{11} & \cdots & p_{1|J_Y|} \\ \vdots & & \vdots \\ p_{|J_X|1} & \cdots & p_{|J_X||J_Y|} \end{bmatrix}. \tag{4.14}$$

If the rows are permutations, then the channel matrix is symmetric.

Example 4.15. Consider a memoryless binary symmetric channel (m.b.s.c). Thus the set of inputs and the set of outputs are $J_X = J_Y = \{0,1\}$. If the channel sends $0 \mapsto 1$ with probability p and $0 \mapsto 0$ with probability $1-p$ (that is, $p(0|1) = p$), then the channel matrix takes the form

$$\begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}, \tag{4.16}$$

which we can represent in the following diagram (with initial states on the left and final states on the right).

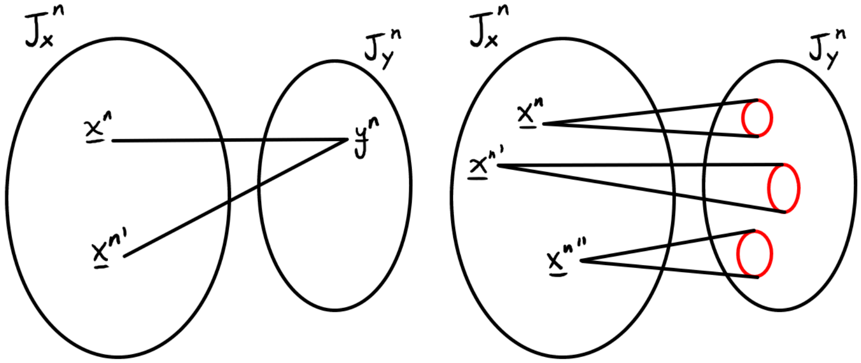
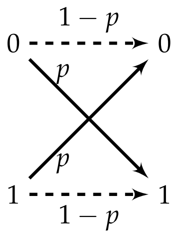


Figure 5. In a noisy channel, it might be the case that multiple inputs map to the same output, as in the left set of ovals. Both $\underline{x}^n$ and $\underline{x}^{n\prime}$ have been mapped to the same $\underline{y}^n$ with some probability. However, Shannon tells us that certain codewords will be transmitted as disjoint regions (red ovals) after being sent through the channel, so those codewords can be reliably decoded after transmission.



Now consider the following encoding scheme. We encode 0 → 000 and 1 → 111. Suppose we got 010 as the output. What do we think the input was?

Probably 0, since it could have come from 000 with the middle bit flipped. But it could have come from 111 with the first and last bits flipped, too.

Now, a simple exercise. For what p is this encoding-decoding scheme better than just sending the original message? Intuitively, we might guess $p = 1/2$, and this is correct. But we should prove it.

Moreover, what is the correspondence between the input and output of the channel? We see that it's certainly not one-to-one, from the last example. So we might have to guess what the original message was. However, what Shannon realized was that for certain elements of J_X^n, their images under the noisy channel map will be disjoint, so these elements will make very good codewords since we can always decode the output even after noise is introduced– see Fig. 5.

We won't do the full proof of the theorem today, but we can introduce the setup. Suppose Alice has a message $[M] = \{1, 2, \ldots, M\}$ she would like to send to Bob. She has a noisy channel $\mathcal{N} : J_X \to J_Y$ with some transition probabilities $p(\underline{y}^n | \underline{x}^n)$.

$$\begin{tikzcd} {} \arrow[r, "m \in M"'] & \boxed{\mathcal{E}_n} \arrow[r, "x_m"'] & \boxed{\mathcal{N}} \arrow[r, "y_m"'] & \boxed{\mathcal{D}_n} \arrow[r, "m'"'] & {} \end{tikzcd}$$

(a) First, Alice can choose an encoding scheme $\mathcal{E}_n : [M] \to J_X^n$ where $\forall m \in [M], \mathcal{E}_n(m) = \underline{x}^{(n)} \in J_X^n$.

(b) She then sends her message through the channel $\mathcal{N}^{(n)} : x^{(n)} \to y^{(n)}$, producing some transmitted messages $y^{(n)}$ with some given probabilities.

(c) Bob receives the message and performs the decoding with $\mathcal{D}_n$ to get some decoded message $\mathcal{D}_n(\mathcal{N}^{(n)}(\mathcal{E}_n(M))) = m'$.

Thus the *maximum probability of error* is

$$\max m \in [M] P(\mathcal{D}_n(\mathcal{N}^{(n)}(\mathcal{E}_n(M))) \neq m) = p(\mathcal{E}_n, \mathcal{D}_n). \tag{4.17}$$

We say that the *rate* is the number of the bits of the message transmitted per use of the channel. That is,

$$R = \frac{\log M}{n} \tag{4.18}$$

since $M \approx 2^{\lfloor nR \rfloor}$.

Definition 4.19. We say that a rate is R is *achievable* if there exists a sequence $(\mathcal{E}_n, \mathcal{D}_n)$ with $M = 2^{nR}$ such that

$$\lim_{n\to\infty} p(\mathcal{E}_n, \mathcal{D}_n) = 0, \tag{4.20}$$

i.e. the maximum probability of error tends to zero as n goes to ∞.

We make one final definition for today.

Definition 4.21. The *channel capacity* is defined to be

$$C(\mathcal{N}) = \sup\{R : R \text{ is an achievable rate}\}, \tag{4.22}$$

the maximum achievable rate for a channel.

Non-lectured: m.b.s.c encoding For Example 4.15, we were asked to consider a binary channel N with error probability p. That is, if we give it an input $x \in \{0,1\}$, we get an output $N(x) = y \in \{0,1\}$ such that $p(N(x) \neq x) = p$.

We came up with the following encoding scheme: send $0 \mapsto 000$ and $1 \mapsto 111$. To decode, we simply take a majority vote, e.g. 010 was "probably" 000, so the original message was 0. Now how much better can we do with this redundancy? Let's consider the possible inputs, how they would be encoded, and how often they would be correct.

Suppose we want to send $0 \mapsto 000$.

- With probability $(1-p)^3$, none of the three bits are flipped and we get 000 as the output. The process succeeds.
- With probability $3 \times p(1-p)^2$, exactly one of the three bits is flipped. (The factor of 3 comes from the fact that we could have flipped any of the three.) We still succeed.
- If two or three bits are flipped, we definitely fail.

By the symmetry of the problem, the success and failure probabilities are the same for $1 \mapsto 111$.

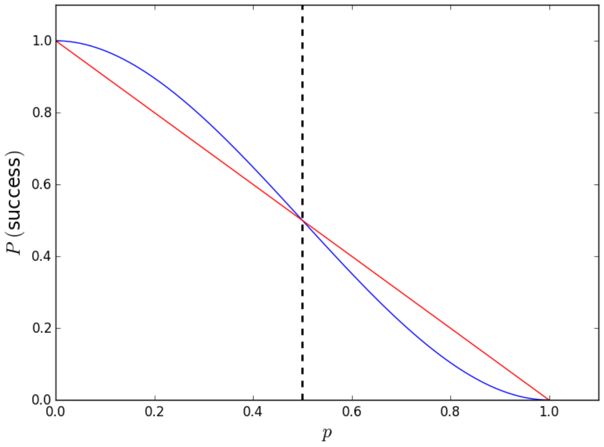
Let's add this up to get the total success probability:

$$(1-p)^3 + 3p(1-p)^2 = (1+2p)(1-p)^2. \tag{4.23}$$

When $p = 1/2$, the success probability of our scheme is

$$(1+2p)(1-p)^2 = (2)(1/2)^2 = 1/2. \tag{4.24}$$

We can nicely visualize this with the following graphic:



Here, the curved blue line is our three-bit scheme and the red line is the single-bit success probability $1-p$. For completeness, we can explicitly show that the crossover points occur when $P(\text{three bits}) - P(\text{one bit}) = (1+2p)(1-p)^2 - (1-p) = 0$. Rewriting, we have $(1-p)(1-2p)p = 0$, which clearly has zeroes at $p = 0, 1/2, 1$. If we now take a derivative, we see that $\frac{d}{dp}(P(\text{three bits}) - P(\text{one bit}))|_{p=1/2} = 1 - 6(1/2) + 6(1/2)^2 = -1/2$, so $P(\text{three bits}) > P(\text{one bit})$ for $p < 1/2$.

Lecture 5.

Monday, January 28, 2019

Last time, we introduced the setup of Shannon's second key theorem, the noisy channel theorem.

Recall our problem– Alice has a message she wants to send to Bob, but she only has access to a noisy channel (defined by a stochastic map) which has some probability of corrupting her message when she sends it. Therefore Alice selects a codeword, translating her message $m \in [M]$ to a codeword $x^{(n)}$ which she then sends through the noisy channel $\mathcal{N}$.

The channel then outputs a transmitted (still encoded) message $y^{(n)}$ with probability

$$p(y^{(n)}|x^{(n)}) \equiv \prod_{i=1}^{n} p(y_i|x_i), \tag{5.1}$$

and Bob then decodes this transmission to get a decoded message m'.

We say that a code $(\mathcal{E}_n, \mathcal{D}_n)$ (i.e. an encoding-decoding scheme) has rate R if $\lceil M \rceil \approx 2^{nR}$. Thus $R = \frac{\log|M|}{n}$. A rate is *achievable* if there exists a code with that rate such that the probability of error after decryption goes to zero in the limit as $n \to \infty$.

Shannon's theorem tells us that the capacity $C(\mathcal{N})$ (i.e. the supremum of all achievable rates) is precisely related to the mutual information between the inputs and outputs of the noisy channel:

$$C(\mathcal{N}) = \max_{\{p(x)\}_{x \in J_X}} I(X:Y). \tag{5.2}$$

Example 5.3. Consider our m.b.s.c. from last time. Recall the mutual information is defined

$$I(X:Y) = H(Y) - H(Y|X), \tag{5.4}$$

where $H(Y|X) = \sum_{x \in J_X} p(x) H(Y|X=x)$, with $H(Y|X=x) = -\sum_{y \in J_Y} p(y|x) \log p(y|x)$. But of course we can explicitly compute these entropies[9] and we can check that

$$H(Y|X=x) = h(p) \forall x \in \{0,1\}. \tag{5.5}$$

Thus

$$\begin{aligned} I(X:Y) &= H(Y) - \sum p(x) h(p) \\ &= H(Y) - h(p) \leq \log|J_Y| - h(p), \end{aligned}$$

so

$$\begin{aligned} C(\mathcal{N}) &= \max_{\{p(x)\}} I(X:Y) \\ &= H(Y) - h(p) \\ &\leq \log|J_Y| - h(p). \end{aligned}$$

[9] $p(y|x)$ is given to us in the channel matrix, so for example

$$\begin{aligned} H(Y|X=0) &= -\sum_{y \in J_Y} p(y|0) \log p(y|0) \\ &= -[(1-p)\log(1-p) + p \log p] \\ &= h(p), \end{aligned}$$

and $H(Y|X=1)$ is the same by the symmetry of the channel matrix.

Could we have equality? That is, $\{p(y)\}$ such that $H(Y) = \log |J_Y|$. This happens if we have outcomes y in a uniform distribution. What are the initial probabilities $\{p(x)\}$? Well,

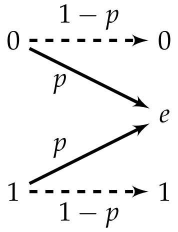
$$p(y) = \sum_x p(y|x) \cdot p(x), \tag{5.6}$$

and we find that $p(x) = 1/2$ for $x = 0$ and $1/2$ for $x = 1$, with $p(y) = 1/2$ for $y = 0$ and $1/2$ for $y = 1$.

We therefore find that the capacity of an m.b.s.c is

$$C(\mathcal{N}_{mbsc}) = \log |J_Y| - h(p) = 1 - h(p). \tag{5.7}$$

As a quick note, the input and output alphabets need not be of equal size. Consider the *binary erasure channel*, which transmits the input with probability $1 - p$ and erases the input with probability p. Thus $J_X = \{0, 1\}$ and $J_Y = \{0, 1, e\}$.



Recall the intuitive picture of the theorem. Shannon realized that for certain codewords, their images after applying the channel map $\mathcal{N}^n$ will represent disjoint subsets in J_Y^n in the asymptotic limit. The maximal rate is then the number of codewords with this property we can choose divided by n the codeword length, or equivalently the number of disjoint subsets we can pack into J_Y^n.

Now for each input sequence $\underline{x}^{(n)}$ of length n, how many typical Y sequences will we get? Recall that there are $|T_n| \approx 2^{nH(X)}$ typical sequences for our variable $X \sim p(x)$. So translating this formula through our channel, we expect to get

$$|T_n| \approx 2^{nH(Y|X)} \tag{5.8}$$

typical sequences in J_Y^n. The total number of possible typical Y sequences is $2^{nH(Y)}$ using the induced distribution $\{p(y)\}$. Therefore we expect to be able to partition the set of typical Y sequences into a number of disjoint typical sets given by

$$\frac{2^{nH(Y)}}{2^{nH(Y|X)}} \approx 2^{n(H(Y)-H(Y|X))} \tag{5.9}$$

$$= 2^{nI(X:Y)}, \tag{5.10}$$

so heuristically, $C(\mathcal{N}) = \max_{\{p(x)\}} I(X:Y)$.

Note that this theorem *does not* translate directly to the quantum case. The classical proof relies on a notion of joint probability of two typical sequences, which has no analogue in QI.[10]

QIT preliminaries Consider a quantum system A. Its states are described by a Hilbert space $\mathcal{H}_A$, where we will take $\dim H$ to be finite. That is, a finite-dimensional Hilbert space is a complex inner product space, i.e. a set with a vector space structure over the field $\mathbb{C}$ equipped with an inner product $(\cdot, \cdot) : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$.

Definition 5.11. An *inner product* is a bilinear function obeying the following properties:

- $(v, v') = (v', v)^* \, \forall v, v' \in \mathcal{H}$
- $(v, av') = a(v, v')$ and $(v, v_1 + v_2) = (v, v_1) + (v, v_2)$.
- $(v, v) \geq 0$ (positive semidefinite), with equality when $v = 0$.

The inner product induces a norm on $\mathcal{H}$, defined

$$||v|| = \sqrt{(v, v)}, \mathcal{H} \to \mathbb{R}. \tag{5.12}$$

The norm defines a distance between two vectors,

$$d(v, v') = ||v - v'||, \tag{5.13}$$

[10] I suspect this is due to entanglement.

which has the properties of being symmetric, with $d(v,v')=0$ iff $v=v'$, and obeying the triangle inequality,

$$d(u,v)\leq d(u,v')+d(v,v'). \tag{5.14}$$

The Cauchy-Schwarz inequality also holds, i.e.

$$\forall v,v'\in\mathcal{H}, |(v,v')|\leq\sqrt{(v,v)(v',v')}. \tag{5.15}$$

Linear maps/operators on $\mathcal{H}$

- We call a map $A:\mathcal{H}\to\mathcal{H}'$ a homomorphism, with the set $A\in B(\mathcal{H},\mathcal{H}')=\text{Homo}(\mathcal{H},\mathcal{H}')$.
- When $\mathcal{H}=\mathcal{H}$, we call such a map an endomorphism and denote the set of such maps $\text{End}(\mathcal{H})$.
- The simplest operator we can define is the identity map, $1\in B(\mathcal{H})$ such that $1v=v\forall v\in\mathcal{H}$.
- We may also define the adjoint of a homomorphism, $A^\dagger:\mathcal{H}'\to\mathcal{H}$. Thus if $A\in B(\mathcal{H},\mathcal{H}')$, then $A^\dagger\in B(\mathcal{H}',\mathcal{H})$. Thus $A^\dagger$ is defined to be the unique operator satisfying

  $$(v',AAv)=(A^\dagger v',v) \tag{5.16}$$

  with $(A^\dagger)^\dagger=A$, where $v\in\mathcal{H},v'\in\mathcal{H}'$. Note that the set of homomorphisms and endomorphisms can be promoted to Hilbert spaces by defining an inner product, the *Hilbert-Schmidt inner product*, defined as

  $$(A,B)_{HS}=\text{Tr}(A^\dagger B). \tag{5.17}$$

Matrix representation Since $\mathcal{H}$ is finite-dimensional by assumption, it can be given a basis $\{v_i\}_{i=1}^d$ where $d=\dim\mathcal{H}$ Thus an element $A\in B(\mathcal{H})$ can be represented by a matrix A with elements

$$A_{ij}=(v_i,Av_j). \tag{5.18}$$

If $\mathcal{H}=\mathbb{C}^d$, for instance, then $B(\mathcal{H})=B(\mathbb{C}^d)\equiv\mathcal{M}_d$, the set of $d\times d$ complex matrices. In $d=2$, we would have

$$A=\begin{pmatrix}a&b\\c&d\end{pmatrix}, A^\dagger=\begin{pmatrix}a^*&c^*\\b^*&d^*\end{pmatrix}. \tag{5.19}$$

Now maps A have the property that if $A=A^\dagger$, then $A\geq0$, i.e. A is positive semidefinite, so that $\forall v\in\mathcal{H},(v,Av)\geq0$. If $A\geq0, A^2=A$.

Lecture 6.

Wednesday, January 30, 2019

Today, we shall begin our discussion of quantum information theory. First, a quick review of Dirac's bra-ket notation– we denote a vector in Hilbert space $\mathcal{H}=\mathbb{C}^d$ by

$$|v\rangle=\begin{pmatrix}v_1\\\vdots\\v_d\end{pmatrix}, \tag{6.1}$$

and call this a *ket*. We also have the dual vectors (row vectors, if you like), called *bras*. such that

$$\langle v|=(v_1^*,\ldots,v_d^*). \tag{6.2}$$

The braket notation provides us with a natural inner product:

$$(u,v)=\langle u|v\rangle=\sum_{i=1}^d u_i^*v_i. \tag{6.3}$$

This space also comes equipped with an outer product, $|u\rangle\langle v|$, which is the matrix

$$|u\rangle\langle v|=\begin{pmatrix}u_1v_1^*&\ldots&\\\vdots&&\\u_dv_1^*&\ldots&u_dv_d^*\end{pmatrix}. \tag{6.4}$$

We can then take an orthonormal basis (onb) for $\mathcal{H}$, which we denote by $\{|e_i\rangle\}$ with $\langle e_i|e_j\rangle = \delta_{ij}$. Note that for any basis of $\mathcal{H}$, we can write the identity matrix as

$$I = \sum_{i=1}^{d} |e_i\rangle\langle e_i|. \tag{6.5}$$

There is a nice basis $|e_1\rangle = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ we can write down, so that for a general basis $\{|f_i\rangle\}_{i=1}^{d}$ related to the original by a unitary U, we find that

$$\sum_{i=1}^{d} |f_i\rangle\langle f_i| = \sum_{i=1}^{d} U|e_i\rangle\langle e_i|U^\dagger = UIU^\dagger = UU^\dagger = I. \tag{6.6}$$

Now in classical information, our simplest system was a binary bit, a system taking values 0 and 1. For quantum information theory, we have a *qubit*, a two-level system represented by a Hilbert space with $\mathcal{H} = \mathbb{C}^2$ and basis vectors $\{|0\rangle, |1\rangle\}$ or equivalently $\{|\uparrow\rangle, |\downarrow\rangle\}$. Physically, these could be the spin states of an electron or perhaps the polarizations of a photon.

Now, it is obvious that any state in Hilbert space can be decomposed in the basis of our choice, i.e.

$$|\psi\rangle = a|0\rangle + b|1\rangle, \tag{6.7}$$

with $a, b \in \mathbb{C}$. We shall require that our states are normalized under this inner product, so that

$$1 = \langle\psi|\psi\rangle = |a|^2 + |b|^2, \tag{6.8}$$

which means that $|a|^2$ and $|b|^2$ have the interpretation of probabilities.

We also have some important operators on Hilbert space. These are the Pauli matrices$\sigma_0, \sigma_x, \sigma_y, \sigma_z$. As it turns out, these operators form a basis. Note that we have a set of self-adjoint 2×2 complex matrices

$$B_{SA}(\mathbb{C}^2) = \{A \in B(\mathcal{H}) : A = A^\dagger\}, \tag{6.9}$$

and we can write a general matrix $M \in M_2/M_{sa}$ in terms of the Pauli matrices,

$$M = \frac{1}{2}(x_0\sigma_0 + \mathbf{x} \cdot \boldsymbol{\sigma}), \tag{6.10}$$

where $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$.

Spectral decomposition The spectral decomposition says that we can write a matrix in terms of its eigenvalues,

$$A = \sum_{i=1}^{d} \lambda_i |e_I\rangle\langle e_i|, \tag{6.11}$$

such that $A|e_i\rangle = \lambda_i|e_i\rangle$. Sometimes, we say that the eigenvalue decomposition is written in terms of projectors instead,

$$A = \sum_{i=1}^{m} \lambda_i \Pi_i \tag{6.12}$$

where Π_i projects onto some basis.

Given a self-adjoint operator $A = A^\dagger$ and a nice function f, what is the value $f(A)$? Note that A, being self-adjoint, can be diagonalized by a unitary. Thus

$$A_d = UAU^\dagger \implies A = U^\dagger A_d U, \tag{6.13}$$

so that

$$f(A) = U^\dagger \begin{pmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_d) \end{pmatrix}. \tag{6.14}$$

Thus for example

$$f(A) = e^{iA} = I + iA + \frac{i^2}{2!} + \dots.$$

QM postulates We consider the following postulates of quantum mechanics, which will in fact be qualified by the fact we are working in an open system.

(a) The state of a (closed) system is given by a ray in $\mathcal{H}$, i.e. a vector defined up to a global phase. Thus we cannot distinguish a state $|\psi\rangle$ and $e^{i\phi}|\psi\rangle$ by any physical measurement. We traditionally take a representative of this equivalence class, $|\psi\rangle$.

For an open system A, consider a system which is in states $|\psi_i\rangle$ with some coefficients $p_i, i = 1, \dots, m$. The state is characterized by an ensemble

$$\{p_i, |\psi_i\rangle\}_{i=1}^m. \tag{6.15}$$

Note that these $|\psi_i\rangle$s need not be mutually orthogonal,

$$\langle\psi_i|\psi_j\rangle \neq \delta_{ij}, \tag{6.16}$$

and moreover this is *not* a superposition but a statistical mixture. A superposition is a pure state where the state is normalized and can be written as

$$|\Psi\rangle = \sum_{i=1}^d a_i |\phi_i\rangle. \tag{6.17}$$

So a statistical mixture is instead described by a *density matrix* (or density operator). We could write our ensemble as

$$\rho \equiv \sum_{i=1}^m p_i |\psi_i\rangle\langle\psi_i|, \tag{6.18}$$

noting that the $|\psi_i\rangle$s in general *need not be orthogonal*.

Definition 6.19. A *density matrix* on $\mathcal{H}$ ($\dim\mathcal{H} = d$) is an operator ρ with the following properties:

- $\rho \geq 0$, i.e. ρ is positive semi-definite, $\langle\phi|\rho|\phi\rangle \geq 0$, which implies that $\rho = \rho^\dagger$.
- $\operatorname{Tr}\rho = 1$ (which gives it a probabilistic interpretation).

Let us remark that ρ is hermitian and therefore admits a spectral decomposition, i.e.

$$\rho = \sum_{j=1}^d \lambda_j |e_j\rangle\langle e_j| \tag{6.20}$$

in terms of an orthonormal basis. Thus

$$\rho = \sum_{i=1}^m p_i |\psi_i\rangle\langle\psi_i| = \sum_{j=1}^d \lambda_j |e_j\rangle\langle e_j|. \tag{6.21}$$

We will prove on Examples Sheet 2 that the set $\mathcal{D}(\mathcal{H})$ of density matrices is a convex set.

Pure and mixed states Consider a density matrix

$$\rho = \sum p_i |\psi_i\rangle\langle\psi_i|, \tag{6.22}$$

and suppose for example that $p_2 = 1, p_i = 0 \forall i \neq 2$. Then

$$\rho = |\psi_2\rangle\langle\psi_2|. \tag{6.23}$$

This is very nice, because we know precisely the state of the system (or equivalently the outcome of applying the operator ρ). We call this a *pure state*, referring either to the vector $|\psi_2\rangle$ or the operator $|\psi_2\rangle\langle\psi_2|$. Otherwise, ρ is a *mixed state*.

A pure state will have $\rho^2 = \rho$, so we can define the *purity* of a state by $\operatorname{Tr}\rho^2$. Conversely, we can define a completely mixed state by

$$\rho = I/d = \frac{1}{d}\sum_{i=1}^d |e_i\rangle\langle e_i|, \tag{6.24}$$

such that a completely mixed state has purity $1/d$ (where we get a factor of d from taking the trace of I).[11]

In classical probability, we remark that the convex set of probability distributions forms a *simplex*.

Now let's briefly discuss the expectation value of an observable (self-adjoint operator) in $B(\mathcal{H})$. For a state described by a density matrix ρ, we define the expectation value to be

$$\phi(A) \equiv \langle A\rangle_\rho = \operatorname{Tr}(A\rho). \tag{6.25}$$

This is a linear normalized functional–

- $\phi(aA + bB) = a\phi(A) + b\phi(B)$
- $\phi(A) \geq 0$ with equality when $A = I$.

Lecture 7.

Friday, February 1, 2019

Last time, we introduced the density matrix formulation of a statistical ensemble of states. For some arbitrary set of states $\{|i\rangle\}$, we describe a statistical mixture by

$$\{p_i, |\phi_i\rangle\}_{i=1}^m \leftrightarrow \rho = \sum_{i=1}^m p_i|\phi_i\rangle\langle\phi_i|. \tag{7.1}$$

These $|\phi_i\rangle$s need not be mutually orthogonal, though the p_is must form a probability distribution. In particular, if none of the p_is are equal to 1, then the state is called a mixed state. Conversely, if one of the p_is are equal to 1, then we call it a pure state.

We introduced the density matrix because we were interested in open (interacting) quantum systems. Let's take a minute to discuss the structure of composite systems. Suppose we have systems A, B with corresponding Hilbert spaces $\mathcal{H}_A, \mathcal{H}_B$. Then the composite system is the tensor product space

$$\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B. \tag{7.2}$$

For instance, if $\mathcal{H}_A, \mathcal{H}_b \simeq \mathbb{C}^2$, then for vectors

$$|v_A\rangle = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, |v_B\rangle = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

then their tensor product is

$$|v_A\rangle \otimes |v_B\rangle \begin{pmatrix} a_1b_1 \\ a_1b_2 \\ a_2b_1 \\ a_2b_2 \end{pmatrix}.$$

More generally if $\dim \mathcal{H}_A = m, \dim \mathcal{H}_B = n$, then the tensor product of a matrix A with $m \times m$ entries a_{ij} and a matrix B is a new $mn \times mn$ matrix where each of the entries a_{ij} in A are replaced by an $m \times m$ matrix, $a_{ij}B$.

In particular, an orthonormal basis can be constructed by simply taking tensor products of the basis elements for each of the individual Hilbert spaces.

States and the density matrix Suppose we have the density matrix for a state in a composite system,

$$\rho_{AB} = \sum_{i,j,\alpha,\beta} a_{i\alpha,j\beta}(|i_A\rangle \otimes |\alpha_B\rangle)(\langle j_A| \otimes \langle\beta_B|). \tag{7.3}$$

Then the state of system A is described by the *partial trace* over the subsystem B:

$$\begin{aligned}\rho_A &= \operatorname{Tr}_B \rho_{AB} & (7.4)\\ &= \operatorname{Tr}_B \sum_{i,j,\alpha,\beta} a_{i\alpha,j\beta}(|i_A\rangle \otimes |\alpha_B\rangle)(\langle j_A| \otimes \langle\beta_B|) & (7.5)\\ &= \sum a_{i\alpha,j\beta}|i_A\rangle\langle j_A|(\operatorname{Tr}|\alpha_B\rangle\langle\beta_B|). & (7.6)\end{aligned}$$

Note that $\operatorname{Tr}|\alpha_B\rangle\langle\beta_B| = \sum_{\gamma_B}\langle\gamma_B|\alpha_B\rangle\langle\beta_B|\gamma_B\rangle = \delta_{\alpha\beta}$, and similarly, $\operatorname{Tr}(|i\rangle\langle j|) = \langle i|j\rangle = \delta_{ij}$.

[11] Explicitly, the trace is $\operatorname{Tr}\rho^2 = \sum_{i=1}^d\sum_{j=1}^d \frac{1}{d^2}\langle e_j|e_i\rangle\langle e_i|e_j\rangle = \frac{1}{d^2}\sum_{i=1}^d\sum_{j=1}^d \delta_{ij}\delta_{ij} = \frac{1}{d^2}\sum_{i=1}^d 1 = 1/d$.

We conclude that the density matrix after taking the partial trace is

$$\rho_A = \operatorname{Tr}_B \rho_{AB} = \sum_{i\alpha j\beta} a_{i\alpha,j\beta} |i_A\rangle\langle j_A| \delta_{\alpha\beta} \tag{7.7}$$

$$= \sum_{ij\alpha} a_{i\alpha,j\alpha} |i_A\rangle\langle j_A| \in B(\mathcal{H}_A). \tag{7.8}$$

One can then show that $\rho_A \geq 0$ (is positive semi-definite) and $\operatorname{Tr} \rho_A = 0$, so ρ_A is in fact a density matrix. We call ρ_A the *reduced density matrix*, or a reduced state.

Recall that the ordinary trace is cyclic, $\operatorname{Tr}(ABC) = \operatorname{Tr}(CAB)$. However, the partial trace Tr_A is *not* in general cyclic. It may be an interesting exercise to try to figure out when the partial trace is cyclic. It's also easy to prove that the complete trace is given by taking the partial traces,

$$\operatorname{Tr}(\cdot) = \operatorname{Tr}_A \operatorname{Tr}_B(\cdot) = \operatorname{Tr}_B \operatorname{Tr}_A(\cdot). \tag{7.9}$$

Now let us consider an observable $M_{AB} \in B(\mathcal{H}_A \otimes \mathcal{H}_B)$. In particular, let

$$M_{AB} = M_A \otimes I_B. \tag{7.10}$$

The expectation value of this observable is given by

$$\begin{aligned} \langle M_{AB} \rangle_{\rho_{AB}} &= \operatorname{Tr}(M_{AB}\rho_{AB}) \\ &= \operatorname{Tr}((M_A \otimes I_B)\rho_{AB})) \\ &= \operatorname{Tr}_A \operatorname{Tr}_B((M_A \otimes I_B)\rho_{AB}) \\ &= \operatorname{Tr}(M_A\rho_A). \end{aligned}$$

For this reason, the partial trace is often defined such that for any M_{AB} of this form,

$$\operatorname{Tr}_B(M_{AB}\rho_{AB}) \equiv \operatorname{Tr}(M_A\rho_A). \tag{7.11}$$

Example 7.12. Consider a system with two qubits, so $\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2$. The full density matrix is

$$\rho_{AB} = \rho_1 \otimes \rho_2, \tag{7.13}$$

where

$$\rho_A = \operatorname{Tr}_B \rho_{AB} = \rho_1, \quad \rho_B = \operatorname{Tr}_A \rho_{AB} = \rho_2. \tag{7.14}$$

Example 7.15. Consider the same Hilbert space as before, but consider the system in a pure state,

$$|\phi^+_{AB}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle). \tag{7.16}$$

Here, we're using a fairly intuitive shorthand where $|00\rangle = |0_A\rangle \otimes |0_B\rangle$. Then the density matrix is

$$\rho = |\phi^+_{AB}\rangle\langle\phi^+_{AB}| = \frac{1}{2}(|0_A\rangle\langle 0_A| \otimes |0_B\rangle\langle 0_B| + \ldots). \tag{7.17}$$

Now we can check as an exercise[12] that ρ_A takes on a simple form–

$$\rho_A = \operatorname{Tr}_B \rho_{AB} = \frac{1}{2}[|0_A\rangle\langle 0_A| + |1_A\rangle\langle 1_A|] = \frac{I_A}{2}, \tag{7.18}$$

and similarly

$$\rho_B = \operatorname{Tr}_A \rho_{AB} = \frac{I_B}{2}. \tag{7.19}$$

[12] The full expansion of ρ_{AB} is

$$\rho_{AB} = \frac{1}{2}(|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|),$$

so taking the partial trace over B, we have

$$\begin{aligned} \operatorname{Tr}_B \rho_{AB} &= \frac{1}{2}(|0_A\rangle\langle 0_A|(\langle 0_B|0_B\rangle) + |0_A\rangle\langle 1_A|(\langle 1_B|0_B\rangle) + |1_A\rangle\langle 0_A|(\langle 0_B|1_B\rangle) + |1_A\rangle\langle 1_A|(\langle 1_B|1_B\rangle)) \\ &= \frac{1}{2}(|0_A\rangle\langle 0_A| + |1_A\rangle\langle 1_A|) = \frac{I_A}{2}. \end{aligned}$$

A similar calculation holds for the trace over A.

This should strike us as a bit strange– after taking the partial traces, we just get the identity matrix of each subsystem, i.e. a completely mixed state. In this way, we have information about the complete system but no information about the subsystems. This is the purely quantum phenomenon we call *entanglement*.

Definition 7.20. To state this more precisely, for a state $|\psi_{AB}\rangle$, if there exist $|\psi_A\rangle, |\psi_B\rangle$ such that

$$|\psi_{AB}\rangle = |\psi_A\rangle \otimes |\psi_B\rangle, \tag{7.21}$$

then we call $|\psi_{AB}\rangle$ a *product state*. Otherwise, it is *entangled*.

In fact, there are four entangled states which are special:

$$|\phi^{\pm}_{AB}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \tag{7.22}$$

$$|\psi^{\pm}_{AB}\rangle = \frac{1}{\sqrt{2}}(|01\rangle \pm |10\rangle). \tag{7.23}$$

These are the so-called *maximally entangled* states or "Bell states," i.e. bipartite pure states such that when we take the partial traces, their reduced states are completely mixed:

$$\rho_{AB} = |\psi_{AB}\rangle\langle\psi_{AB}| \text{ such that } \rho_B = I_B/2. \tag{7.24}$$

We then say that for a mixed state, if its density matrix can be written

$$\rho_{AB} = \sum p_i \omega_i^A \otimes \sigma_i^B, \tag{7.25}$$

we say it is *separable*. Otherwise, it is entangled.

Last time, we also referred to the Pauli matrices $\sigma_0, \sigma_x, \sigma_y, \sigma_z$, and remarked that their real span (i.e. linear combinations with real coefficients) is the set of 2×2 self-adjoint matrices,

$$A = x_0\sigma_0 + \mathbf{x} \cdot \boldsymbol{\sigma}$$

where $x_0, x_1, x_2, x_3 \in \mathbb{R}$. If $A = \rho$ is a density matrix, then $\operatorname{Tr}\rho = 1 \implies x_0 = 1/2$ since $\rho = I/2 + \mathbf{x} \cdot \boldsymbol{\sigma}/2$, and the σ_i are traceless.

Next lecture, we will talk about three concepts:

- Schmidt decomposition
- Purification
- No-cloning theorem

We'll briefly state the first of these: for any state $|\psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, then there exists an orthonormal basis

$$\{|i_A\rangle\}_{i=1}^{d_A}, \{|i_B\rangle\}_{i=1}^{d_B} \tag{7.26}$$

such that

$$|\psi_{AB}\rangle = \sum_{i=1}^{\min\{d_A, d_B\}} \lambda_i |i_A\rangle \otimes |i_B\rangle, \tag{7.27}$$

with $\lambda_i \geq 0, \sum \lambda_i^2 = 1$. Then the density matrix is

$$\rho_{AB} = |\psi_{AB}\rangle\langle\psi_{AB}| = \sum \lambda_i \lambda_j |i_A\rangle\langle j_A| \otimes |i_B\rangle\langle j_B|. \tag{7.28}$$

Taking the partial trace over B, we get a δ_{ij} and therefore find that

$$\rho_A = \sum_{i=1}^{\min(d_A, d_B)} \lambda_i^2 |i_A\rangle\langle i_A|. \tag{7.29}$$

Lecture 8.

Monday, February 4, 2019

Today we shall discuss the Schmidt decomposition, purification, and the no-cloning theorem.

Theorem 8.1 (Schmidt decomposition)**.** *For any pure state $|\psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, there exists an orthonormal basis*

$$\{|i_A\rangle\}_{i=1}^{d_A}, \{|i_B\rangle\}_{i=1}^{d_B} \tag{8.2}$$

such that

$$|\psi_{AB}\rangle = \sum_{i=1}^{\min\{d_A,d_B\}} \lambda_i |i_A\rangle \otimes |i_B\rangle, \tag{8.3}$$

with $\lambda_i \geq 0, \sum \lambda_i^2 = 1$.

Proof. Let $\{|r_a\rangle\}_{r=1}^{d_A}$ and $\{|\alpha_B\rangle\}_{\alpha=1}^{d_B}$ be orthonormal bases for $\mathcal{H}_A, \mathcal{H}_B$. Thus

$$\{|r_A\rangle \otimes |\alpha_B\rangle\}_{r,\alpha} \tag{8.4}$$

forms an onb of $\mathcal{H}_A \otimes \mathcal{H}_B$. A general state can be expressed in this basis as

$$|\psi_{AB}\rangle = \sum_{r,\alpha} a_{r\alpha} |r_A\rangle \otimes |\alpha_B\rangle. \tag{8.5}$$

Here, $a_{r\alpha}$ form elements of A, a $d_A \times d_B$ matrix.

We now apply the singular value decomposition to write A in terms of unitaries U (a $d_A \times d_A$ matrix) and V ($d_B \times d_B$) as

$$A = U \underbrace{D}_{d_A \times d_B} V, \tag{8.6}$$

with elements $d_{ij} = d_{ii}\delta_{ij}, d_{ii} \geq u$. That is, D is diagonal, though it is not square.

Then the coefficients may be written as

$$a_{r\alpha} = \sum_{i=1}^{d_A} \sum_{\beta=1}^{d_B} u_{ri} d_{i\beta} v_{\beta\alpha}. \tag{8.7}$$

Since $d_{i\beta} = \delta_{i\beta} d_{ii}$, we rewrite our state as

$$|\psi_{AB}\rangle = \sum_{i=1}^{\min(d_A,d_B)} \lambda_i \underbrace{\left(\sum_r^{d_A} u_{ri} |r_A\rangle\right)}_{|i_A\rangle} \underbrace{\left(\sum_\alpha^{d_B} v_{i\alpha} |\alpha_B\rangle\right)}_{|i_B\rangle}, \tag{8.8}$$

where we recognize $d_{ii} = \lambda_i$. Thus we have written the state as

$$|\psi_{AB}\rangle = \sum_{i=1}^{\min(d_A,d_B)} \lambda_i |i_A\rangle |i_B\rangle. \tag{8.9}$$

We can check that $\langle j_A | i_A \rangle = \delta_{ij}$ by using unitarity:[13]

$$\begin{aligned} \langle j_A | i_A \rangle &= \sum_{r,r'} (u^*_{r'j} \langle r'_A|)(u_{ri}|r_A\rangle) \\ &= \sum_r u^*_{rj} u_{ri} \\ &= \sum (U^\dagger)_{ir} (U)_{ri} = U^\dagger U. \end{aligned}$$

The proof for the second basis vector is equivalent.

To prove that the λ_is squared add to 1, we write the density matrix

$$\rho_{AB} = |\psi_{AB}\rangle\langle\psi_{AB}|, \tag{8.10}$$

[13] A slightly quicker way to do this is to recognize that we're just taking $\langle j_A|i_A\rangle = \langle Ur'_A|Ur_A\rangle = \langle r'_A|U^\dagger U r_A\rangle = \langle r'_A|r_A\rangle$.

so that for instance

$$\begin{aligned}\rho_A &= \mathrm{Tr}_B |\psi_{AB}\rangle\langle\psi_{AB}| \\ &= \mathrm{Tr}_B \sum_j \lambda_j \lambda_i (|i_A\rangle|i_B\rangle)(\langle j_A|\langle j_B|) \\ &= \sum_i \lambda_i \lambda_j |i_A\rangle\langle j_A| \\ &= \sum_{i=1}^{d_m} \lambda_i^2 |i_A\rangle\langle i_A|\end{aligned}$$

since $\mathrm{Tr}(|i_B\rangle\langle j_B|) = \delta_{ij}$.

What we observe is that while the dimensions of ρ_A and ρ_B are different, they have the same number of nonzero eigenvalues, λ_1 through λ_k where k is the rank of ρ_A.

Let $d_m = \min(d_A, d_B)$. It follows that we can write

$$\rho_A = \sum_{i=1}^{d_m} \lambda_i^2 |i_A\rangle\langle i_A| = \sum_{i=1}^{\mathrm{rk}(\rho_A)} \lambda_i^2 |i_a\rangle\langle i_A|. \tag{8.11}$$

The state itself can therefore be written as

$$|\psi_{AB}\rangle = \sum_{i=1}^{\min(\mathrm{rk}\rho_A, \mathrm{rk}\rho_B)} \lambda_i |i_A\rangle|i_B\rangle, \tag{8.12}$$

which is exactly the Schmidt decomposition as claimed. ☒

Note that the Schmidt decomposition is unique if all the eigenvalues of ρ_A and ρ_B are nondegenerate. If so, we can construct the Schmidt decomposition by pairing eigenvectors of ρ_A and ρ_B which share the same eigenvalue.

Definition 8.13. We say that the *Schmidt rank* of $|\psi_{AB}\rangle$ is then $n(\psi_{AB}) =$ the number of positive Schmidt coefficients, where the λ_is are the Schmidt coefficients.

Theorem 8.14. *A state $|\psi_{AB}\rangle$ is entangled iff $n(\psi_{AB}) > 1$, where $n(\psi_{AB})$ is the Schmidt rank of $|\psi_{AB}\rangle$.*

n.b. if $n(\psi_{AB}) = 1$, then $|\psi_{AB}\rangle = |i_A\rangle \otimes |i_B\rangle$.

Purification Generally, it is nicer to work with pure states than mixed states. We would therefore like to be able to associate a pure state (perhaps in a larger Hilbert space) with any mixed state.

That is, given a density matrix $\rho_A \in \mathcal{H}_A$, we would like to define a purifying reference system R with Hilbert space $\mathcal{H}_R$ and a new state $|\psi_{AR}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_R$ such that

$$\rho_A = \mathrm{Tr}_R |\psi_{AR}\rangle\langle\psi_{AR}|. \tag{8.15}$$

We claim that this is always possible, and will explicitly construct the purified state.

Proof. Let us take $\mathcal{H}_R \simeq \mathcal{H}_A$. Look at the spectral decomposition of our state,

$$\rho_A = \sum_{i=1}^{d_A} p_i |i_A\rangle\langle i_A| \tag{8.16}$$

where $\{|i_A\rangle\}$ is an onb for $\mathcal{H}_A$. We can equivalently take a set of elements $\{|i_R\rangle\}$ to be an onb for $\mathcal{H}_R$. Since $\mathcal{H}_R$ is a copy of $\mathcal{H}_A$, we can define a bigger state $|\psi_{AR}\rangle$ as

$$|\psi_{AR}\rangle \equiv \sum_{i=1}^{d} \sqrt{p_i} |i_A\rangle|i_R\rangle, \tag{8.17}$$

where $d = \dim \mathcal{H}_A = \dim \mathcal{H}_R$. However, note that this is none other than the Schmidt decomposition we just defined, with $\lambda_i = \sqrt{p_i}$.

We now claim that

$$\rho_{AB} = |\psi_{AR}\rangle\langle\psi_{AR}| \tag{8.18}$$

is a pure state, since the Schmidt coefficients λ_is of this state satisfy $\sum \lambda_i^2 = \sum p_i = 1$.[14] A quick computation[15] confirms that

$$\mathrm{Tr}_R |\psi_{AR}\rangle\langle\psi_{AR}| = \rho_A, \tag{8.19}$$

Thus ρ_{AB} is a purification of ρ_A. ☒

Let's also observe that if we have a system AB in a state Ω_{AB} such that $\mathrm{Tr}_B\, \Omega_{AB} = \psi_A$ is a pure state, then Ω_{AB} must itself be a product state, $\Omega_{AB} = \psi_A \otimes \omega_B$, where $\psi_A = |\psi_A\rangle\langle\psi_A|$.

This also tells us that correlations contains in a pure state are *monogamous*, i.e. for a bipartite state $A = A_1A_2$, with $|\psi\rangle = |\psi_{A_1A_2}\rangle$, then the bigger system $AB = A_1A_2B$ will have a state of the form

$$\Omega_{A_1A_2B} = \psi_{A_1A_2} \otimes \omega_B \tag{8.20}$$

No-cloning theorem In popular language, the no-cloning theorem says that there does not exist a quantum copier. More formally, $\nexists$ a unitary operator which can perfectly copy an unknown $|\psi\rangle$.

Proof. Let $|\psi\rangle \in \mathcal{H}$ with $\mathcal{H}$ some Hilbert space, and suppose there exists such a unitary $U \in \mathcal{B}(\mathcal{H} \otimes \mathcal{H})$. That is, we can take an arbitrary reference state $|\psi\rangle$ and a "blank" state $|s\rangle$ and get out two copies of $|\psi\rangle$. Thus

$$U(|\phi\rangle \otimes |s\rangle) = |\phi\rangle \otimes |\phi\rangle \tag{8.21}$$
$$U(|\psi\rangle \otimes |s\rangle) = |\psi\rangle \otimes |\psi\rangle \tag{8.22}$$

for two distinct but otherwise arbitrary reference states $|\psi\rangle, |\phi\rangle$. Let us now take the inner products of the LHS and RHS of 8.21 and 8.22. We get

$$((\langle\phi| \otimes \langle s|)U^\dagger U(|\psi\rangle \otimes |s\rangle) = ((\langle\phi| \otimes \langle\phi|)(|\psi\rangle \otimes |\psi\rangle). \tag{8.23}$$

Now we see that since U is a unitary, we get

$$\langle\phi|\psi\rangle\langle s|s\rangle = \langle\phi|\psi\rangle^2. \tag{8.24}$$

WLOG, we can choose our blank state to be normalized, $\langle s|s\rangle = 1$. But so $\langle\phi|\psi\rangle = \langle\phi|\psi\rangle^2 \implies \langle\phi|\psi\rangle = 0$ or 1. That is, either the states are orthogonal or they are identical. Therefore our copier does not work on arbitrary reference states, and we have reached a contradiction. ☒

Example 8.25. Let's see a concrete example of this: suppose we take $|\psi\rangle \in \mathbb{C}^2$, where $|\psi\rangle = a|0\rangle + b|1\rangle$, and take the "blank" state $|s\rangle = |0\rangle$. Then we would like a unitary operator U such that

$$U(|\psi\rangle \otimes |0\rangle) = |\psi\rangle \otimes |\psi\rangle. \tag{8.26}$$

Under linearity, our state must be

$$aU|00\rangle + bU|10\rangle. \tag{8.27}$$

We can certainly prepare a unitary such that $U|0\rangle \otimes |0\rangle = |00\rangle$ and $U|1\rangle \otimes |0\rangle = |11\rangle$. However, when we now operate on an arbitrary state $|\psi\rangle \otimes |0\rangle$, we see that

$$U(|\psi\rangle \otimes |0\rangle) = aU|00\rangle + bU|10\rangle = a|00\rangle + b|11\rangle. \tag{8.28}$$

[14] The state $|\psi_{AR}\rangle$ is normalized, since

$$\langle\psi_{AR}|\psi_{AR}\rangle = \sum_{i,j} \sqrt{p_i p_j}\langle j_A|i_A\rangle\langle j_R|i_R\rangle = \sum_i p_i = 1,$$

so it follows that

$$\rho_{AB}^2 = |\psi_{AR}\rangle\langle\psi_{AR}|\psi_{AR}\rangle\langle\psi_{AR}| = |\psi_{AR}\rangle\langle\psi_{AR}| = \rho_{AB},$$

i.e. ρ_{AB} is a pure state.

[15] Explicitly,

$$|\psi_{AR}\rangle\langle\psi_{AR}| = \sum_{i,j} \sqrt{p_i p_j}|i_A\rangle|i_R\rangle\langle j_A|\langle j_R|,$$

so

$$\begin{aligned}\mathrm{Tr}_R\, |\psi_{AR}\rangle\langle\psi_{AR}| &= \sum_{i,j} \sqrt{p_i p_j}|i_A\rangle\langle j_A| \underbrace{\langle j_R|i_R\rangle}_{\delta_{ij}} \\ &= \sum_i p_i\, |i_A\rangle\langle i_A| = \rho_A.\end{aligned}$$

But this is an *entangled state*, and in particular it is certainly not $|\psi\rangle \otimes |\psi\rangle$, since

$$|\psi\rangle \otimes |\psi\rangle = a^2|00\rangle + b^2|11\rangle + ab|01\rangle + ba|10\rangle. \tag{8.29}$$

We see that it's because of linearity and the tensor product structure of composite quantum systems that our unitary operator cannot copy a generic unknown state.

Some concluding remarks: observe that if we have a single unknown state, we cannot make copies by the no-cloning theorem, but if we already have many copies, we could measure those copies in some bases and then prepare new copies of the state. In addition, the process of quantum teleportation (which we haven't discussed) does *not* contradict no-cloning because the original state becomes inaccessible to us– its information is all in the teleported state after the measurement procedure.

We've now shown that the first postulate of QM in a closed system (states as rays in Hilbert space) is replaced by the density matrix formalism, with some important consequences. Soon we'll consider the second postulate, that the dynamics of a quantum system are determined by a unitary operator.

Maximally entangled states Consider a state

$$|\psi_{AB}\rangle = \sum_{i=1}^{d_m} \lambda_i |i_A\rangle|i_B\rangle, \tag{8.30}$$

with $m = \min(\dim A, \dim B)$ as before. If $\lambda_i = 1/\sqrt{d_m}$, we call this a maximally entangled state. A maximally entangled state is a state such that its partial trace yields a completely mixed state– cf. the Bell state $|\phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$.

Lecture 9.

Wednesday, February 6, 2019

Last time, we introduced maximally entangled states. As it turns out, these states have a few interesting properties. Recall that such states are defined on composite Hilbert spaces such that for

$$\mathcal{H}_A \otimes \mathcal{H}_B \simeq \mathbb{C}^d \otimes \mathbb{C}^d \tag{9.1}$$

equipped with a (fixed) onb for $\mathbb{C}^d$ given by $\{|i\rangle\}_{i=1}^d$, a maximally entangled state is then a state which is written

$$|\Omega\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |i\rangle|i\rangle. \tag{9.2}$$

- Every MES $|\Phi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$ can be written in the form

  $$|\Phi\rangle = (I_d \otimes U)|\Omega\rangle \tag{9.3}$$

  for some unitary U. One should check explicitly[16] that

  $$\mathrm{Tr}_2|\Phi\rangle\langle\Phi| = \frac{I}{d} \text{ and } \mathrm{Tr}_1|\Phi\rangle\langle\Phi| = \frac{I}{d}. \tag{9.4}$$

[16] The proof is quick.

$$\begin{aligned}
\mathrm{Tr}_2(|\Phi\rangle\langle\Phi|) &= \mathrm{Tr}_2\left((I\otimes U)\frac{1}{\sqrt{d}}\sum_i |i\rangle|i\rangle\right)\left(\frac{1}{\sqrt{d}}\sum_j \langle j|\langle j|(I\otimes U^\dagger)\right)\\
&= \frac{1}{d}\sum_{i,j} |i\rangle\langle j|\,\mathrm{Tr}(U\,|i\rangle\langle j|\,U^\dagger)\\
&= \frac{1}{d}\sum_{i,j} |i\rangle\langle j|\,\mathrm{Tr}(|i\rangle\langle j|\,U^\dagger U)\\
&= \frac{1}{d}\sum_{i,j} |i\rangle\langle j|\,\delta_{ij}\\
&= \frac{I}{d}.
\end{aligned}$$

The proof for tracing over the first subsystem is almost the same. Strictly, what this shows is that every state of this form is maximally entangled. We haven't shown that every maximally entangled state admits this form.

- Lemma: for any $A, B \in B(\mathbb{C}^d)$,
  - $\langle\Omega|A \otimes B|\Omega\rangle = \frac{1}{d}\operatorname{Tr}(A^T B)$, where transposition is done in the basis $\{|i\rangle\}_{i=1}^{d}$.
  - $(A \otimes I)|\Omega\rangle = (I \otimes A^T)|\Omega\rangle$, a property we shall call "ricochet." The proofs of these lemmas are an exercise, and are done at the end of this lecture's notes.
- We can write down a purification of a state ρ in terms of $|\Omega\rangle$: we claim it is

  $$|\psi\rangle = \sqrt{d}(\sqrt{\rho} \otimes I)|\Omega\rangle. \tag{9.5}$$

  Let us check:

  $$\begin{aligned}|\psi\rangle\langle\psi| &= d(\sqrt{\rho} \otimes I)|\Omega\rangle\langle\Omega|(\sqrt{\rho} \otimes I)\\ &= \sum_{i,j}\sqrt{\rho}|i\rangle\langle j|\sqrt{\rho} \otimes |i\rangle\langle j|.\end{aligned}$$

  Tracing over the second system, $|i\rangle\langle j| = \delta_{ij}$, so the partial trace is then

  $$\operatorname{Tr}_2|\psi\rangle\langle\psi| = \sqrt{\rho}\sum_i|i\rangle\langle i|\sqrt{\rho} = \rho. \tag{9.6}$$

- Every bipartite pure state $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ can be written in the form

  $$|\psi\rangle = (I \otimes R)|\Omega\rangle \tag{9.7}$$

  for some operator R.

  Proof. Let $|\psi\rangle = |\psi_{AB}\rangle = \sum\lambda_i|i_A\rangle|i_B\rangle$, by the Schmidt decomposition. Let V, W be isometries such that

  $$V|i\rangle = |i_A\rangle\forall i; \quad V : \mathbb{C}^d \to \mathcal{H}_A \tag{9.8}$$

  $$W|i\rangle = |i_B\rangle\forall i; \quad W : \mathbb{C}^d \to \mathcal{H}_B. \tag{9.9}$$

  The proof is constructive. Choose $R \equiv WQV^T$, where Q is defined in terms of the Schmidt coefficients,

  $$Q = \sum\sqrt{d}\lambda_j|j\rangle\langle j|. \tag{9.10}$$

  Let us look at the RHS of 9.7. For this choice of R, it is

  $$\begin{aligned}&= (I \otimes WQV^T)|\Omega\rangle\\ &= (I \otimes W)(I \otimes Q)(I \otimes V^T)|\Omega\rangle\\ &= (I \otimes W)(I \otimes Q)(V \otimes I)|\Omega\rangle\\ &= (V \otimes W)(I \otimes Q)|\Omega\rangle\\ &= (V \otimes W)\frac{1}{\sqrt{d}}\sum_i|i\rangle \otimes Q|i\rangle\sum_j\sqrt{d}\lambda_j|j\rangle\langle j|\langle i|\\ &= (V \otimes W)\sum_i\lambda_i|i\rangle \otimes |i\rangle\\ &= \sum\lambda_i|i_A\rangle|i_B\rangle.\end{aligned}$$

  ⊠

  Here, we have used the "ricochet" property to interchange $I \otimes V^T$ to $V \otimes I$, and moved $V \otimes I$ through $I \otimes Q$ since they act on independent parts of the composite system.

Time evolution of open systems Question: what is the most general description of the dynamics of an *open* quantum system? Answer: it is given by a linear *completely positive trace-preserving* (CPTP) map. The advantage of such a map is that it gives us a description of the effect of *any* allowed physical process on your system, including operations like measurement. In particular, it also allows us to describe discrete state changes.

As all reasonable evolution operators should be linear, we will usually omit this from the description and just speak of a CPTP map. We can also reasonably call this a *quantum operator* or a *quantum channel*. That is, we have a map $\Lambda : \mathcal{D}(\mathcal{H}) \to \mathcal{D}(\mathcal{H})$, e.g. it takes the density matrix ρ from $\rho \mapsto \Lambda(\rho) = \rho'$. We call this a *superoperator* because it is a map from operators to operators.

Example 9.11. We've constructed a general description of open quantum systems, but it should include our previous description of closed quantum systems as a special case. Take Λ to be a unitary transformation, such that

$$\rho' = \Lambda(\rho) = U\rho U^\dagger. \tag{9.12}$$

Let us now unpack some of the properties of CPTP maps.

- This map satisfies linearity:

  $$\Lambda(a\rho_1 + b\rho_2) = a\Lambda(\rho_1) + b\Lambda(\rho_2).$$

  We want our CPTP maps to be linear so that we can interpret mixed state density matrices in a probabilistic way. That is, if we have some distribution of density matrices ρ_i given with some probabilities p_i (i.e. a set $\{p_i, \rho_i\}_{i=1}^m$, then we can describe the system as a new density matrix

  $$\sigma = \sum_{i=1}^m p_i \rho_i,$$

  and thus the map Λ should also represent a valid map on the full system σ:

  $$\Lambda(\sigma) = \sum_{i=1}^m p_i \Lambda(\rho_i).$$

- Positivity: for $\rho \geq 0, \rho' = \Lambda(\rho) \geq 0$. We say Λ is a positive (or positivity-preserving) map if

  $$\Lambda(A) \geq 0 \forall A \geq 0. \tag{9.13}$$

- Λ must be trace-preserving, i.e. for ρ with $\operatorname{Tr}\rho = 1$, we want

  $$\operatorname{Tr}(\Lambda\rho) = \operatorname{Tr}\rho' = 1. \tag{9.14}$$

These conditions are necessary, but not sufficient. In fact we, require Λ to be *completely positive*, as we'll define now.

Definition 9.15. Let $\Lambda : \mathcal{D}(\mathcal{H}_A) \to \mathcal{D}(\mathcal{H}'_A)$, where $\mathcal{H}_A$ is the Hilbert space of this system. Consider an extension of $\mathcal{H}_A$ to the bigger space $\mathcal{H}_A \otimes \mathcal{H}_B$. That is, we add another system B, called the ancilla or (for obvious reasons) the environment.

Note that I_B is the identity operator on B, i.e. $I_B \in \mathcal{B}(\mathcal{H}_B)$, whereas id_B is the superoperator $\mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_B)$ such that $id_B Q = Q \, \forall Q \in B(\mathcal{H}_B)$. We then say that Λ is *completely positive* if $\Lambda \otimes id_B$ is positive for all such extensions.

For instance, suppose the composite system AB is initially in a state $\rho_A \otimes \omega_B$. Thus a completely positive map yields a state

$$(\Lambda \otimes id_B)(\rho_A \otimes \omega_B) = \sigma_{AB}, \tag{9.16}$$

where σ_{AB} is guaranteed to be a legitimate state of the composite system AB.

Example 9.17. Let Λ be the transposition map. This is certainly positive:

$$\Lambda \equiv T : \rho \to \rho^T. \tag{9.18}$$

That is, if ρ had no negative eigenvalues, then transposition will preserve the eigenvalues and therefore preserve positivity.

We will now show that there exists a composite state which is positive, but not positive after the application of $\Lambda \otimes id_B$. Let the composite system $\mathcal{H}_A \otimes \mathcal{H}_B$ with $\mathcal{H}_A, \mathcal{H}_B \simeq \mathbb{C}^d$ be described by the density matrix

$$\rho_{AB} = |\Omega\rangle\langle\Omega| \tag{9.19}$$

where

$$|\Omega\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^d |i\rangle|i\rangle \tag{9.20}$$

is a MES. Now we hit the first part with the transpose:

$$\begin{aligned}(\Lambda \otimes id_B)|\Omega\rangle\langle\Omega| &= \frac{1}{d}\sum T(|i\rangle\langle j| \otimes |i\rangle\langle j| \\ &= \frac{1}{d}\sum |j\rangle\langle i| \otimes |i\rangle\langle j| \equiv \tilde{\rho}.\end{aligned}$$

Now we ask whether $\tilde{\rho} \geq 0$. The factor d certainly doesn't change the positivity of the state, so take $Q \equiv d\tilde{\rho}$ and consider its action on some states $|\phi\rangle = \sum_k a_k|k\rangle$, $|\psi\rangle = \sum_l b_l|l\rangle$. Then

$$\begin{aligned}Q(|\phi\rangle \otimes |\psi\rangle) &= \left(\sum |j\rangle\langle i| \otimes |i\rangle\langle j|\right)\left(\sum a_k|k\rangle \otimes \sum b_l|l\rangle\right) \\ &= \sum_{i,j} a_i|j\rangle \otimes b_j|i\rangle \\ &= \sum_j b_j|j\rangle \otimes \sum_i a_i|i\rangle = |\psi\rangle \otimes |\phi\rangle.\end{aligned}$$

What we see is that Q has swapped the states between the Hilbert spaces,

$$Q(|\phi\rangle \otimes |\psi\rangle) = |\psi\rangle \otimes |\phi\rangle \implies Q^2 = I. \tag{9.21}$$

This tells us that the eigenvalues of Q are ± 1, which means that we have constructed an operator which is positive but not completely positive.

Non-lectured aside: extra proofs These proofs were originally footnotes, but I thought it might be useful to collect them here at the end of the lecture to avoid clutter.

Proof. Trace and $|\Omega\rangle$: we show that $\langle\Omega|A \otimes B|\Omega\rangle = \frac{1}{d}\operatorname{Tr}(A^T B)$.

Note that by the usual laws of matrix multiplication, if $A = a_{ij}|i\rangle\langle j|$ and similarly $B = b_{ij}|i\rangle\langle j|$, then $A^T B = a_{ji}B_{jl}|i\rangle\langle l|$ and so

$$\operatorname{Tr}(A^T B) = a_{ji}b_{jl}\langle l|i\rangle = A_{ji}b_{ji}. \tag{9.22}$$

Now by explicit computation, we see that

$$\begin{aligned}\langle\Omega|A \otimes B|\Omega\rangle &= \frac{1}{\sqrt{d}}\langle\Omega|(a_{ij}|i\rangle\langle j|k\rangle \otimes b_{lm}|l\rangle\langle m|k\rangle) \\ &= \frac{1}{\sqrt{d}}\langle\Omega|(a_{ik}|i\rangle \otimes b_{lk}|l\rangle) \\ &= \frac{1}{d}a_{ik}\langle n|i\rangle b_{lk}\langle n|l\rangle \\ &= \frac{1}{d}a_{nk}b_{nk} \\ &= \frac{1}{d}\operatorname{Tr}(A^T B),\end{aligned}$$

where we have swapped $|\Omega\rangle$s freely for their expressions in terms of an orthonormal basis and evaluated the Kronecker deltas implicitly rather than writing them out. ☒

Proof. Ricochet property: we wish to prove that

$$(A \otimes I)|\Omega\rangle = (I \otimes A^T)(|\Omega\rangle).$$

For brevity, I'm suppressing the sums in the following expressions. All sums are taken over 1 to d. Let $A = a_{ij}|i\rangle\langle j|$, and thus $A^T = a_{ji}|i\rangle\langle j|$. Then

$$\begin{aligned}(A \otimes I)|\Omega\rangle &= a_{ij}|i\rangle\langle j|k\rangle \otimes |k\rangle \\ &= a_{ij}|i\rangle\delta_{jk} \otimes |k\rangle \\ &= a_{ik}|i\rangle \otimes |k\rangle \\ &= a_{ki}|k\rangle \otimes |i\rangle \\ &= |k\rangle \otimes a_{ji}|i\rangle\delta_{jk} \\ &= |k\rangle \otimes a_{ji}|i\rangle\langle j|k\rangle \\ &= (I \otimes A^T)|\Omega\rangle,\end{aligned}$$

where we have simply relabeled i and k in the fourth line since both sums run from 1 to d. ⊠

Lecture 10.

Monday, February 11, 2019

Admin note: there was no lecture (and hence no notes) for Friday, February 8, as Prof. Datta sustained an injury which prevented her from giving the lecture.

Quantum operations and CPTP maps To recap from last time, any allowed physical process on a quantum system is given by a quantum operation. The map must be completely positive (CP) in order to allow us to properly couple an ancilla (environment) to our system, and it must be linear and trace-preserving in order to take density matrices to other density matrices.

Consider a map $\Lambda : B(\mathcal{H}) \to B(\mathcal{K})$, where $\mathcal{H} \simeq \mathbb{C}^m, \mathcal{K} \simeq \mathbb{C}^n$. Let $\mathcal{M}_m, \mathcal{M}_m^+$ bu $m \times m$ complex positive semi-definite matrices. The set of density matrices on $\mathbb{C}^n$ is given by

$$\mathcal{D}(\mathbb{C}^m) = \{\rho \in \mathcal{M}_m^+ ; \operatorname{Tr}\rho = 1\}. \tag{10.1}$$

Definition 10.2. A map $\Lambda : \mathcal{M}_m \to \mathcal{M}_n$ is positive if

$$\Lambda(A) \in \mathcal{M}_n^+ \forall A \in \mathcal{M}_m^+. \tag{10.3}$$

Definition 10.4. For a given positive integer k, Λ is k-positive if $(\Lambda \otimes \mathrm{id}_k)$ is positive, where id_k is the identity (super)operator, $\mathrm{id}_k : \mathcal{M}_k \to \mathcal{M}_k$ such that $\mathrm{id}_k(Q) = Q \forall Q \in \mathcal{M}_k$.

Definition 10.5. The map Λ is completely positive (CP) if it is k-positive $\forall k \in \mathbb{Z}^+$, positive integers.

Theorem 10.6 (Necessary and sufficient condition for CP)**.** *A linear map $\Lambda : \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^{d'})$ is completely positive $\iff (\Lambda \otimes id_d)(|\Omega\rangle\langle\Omega|) \geq 0$, where*

$$|\Omega\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |i\rangle|i\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d. \tag{10.7}$$

That is, it suffices to check positivity on the density matrix corresponding to the maximally entangled d-dimensional state.

Proof. Necessity follows immediately from the definition of CP. To show sufficiency, consider an arbitrary $k \geq 1$. For a state $\rho \in \mathcal{D}(\mathbb{C}^d \otimes \mathbb{C}^k)$, we have a spectral decomposition

$$\rho = \sum \lambda_i \, |\phi_i\rangle\langle\phi_i| \tag{10.8}$$

where $|\phi_i\rangle \in \mathbb{C}^d \otimes \mathbb{C}^k$. Now we have

$$\begin{aligned}(\Lambda \otimes \mathrm{id}_k)\rho \geq 0 &\implies \sum_i \lambda_i (\Lambda \otimes \mathrm{id}_k)(|\phi_i\rangle\langle\phi_i|) \geq 0 &(10.9)\\ &\implies \forall i, (\Lambda \otimes \mathrm{id}_k)\, |\phi_i\rangle\langle\phi_i| \geq 0. &(10.10)\end{aligned}$$

We saw that for each of the basis states $|\phi_i\rangle$, we could write it as

$$|\phi_i\rangle = (I \otimes R_i)|\Omega\rangle \tag{10.11}$$

for some $R_i \in \mathcal{B}(\mathbb{C}^d, \mathbb{C}^k)$. Thus we can rewrite the basis states in our inequality to get

$$(\Lambda \otimes \mathrm{id}_k)(I \otimes R_i) |\Omega\rangle\langle\Omega| (I \otimes R_i^\dagger) \geq 0. \tag{10.12}$$

Note that with the following definition

$$(\mathrm{id}_d \otimes f_i)(\omega) := (I \otimes R_i)(\omega(I \otimes R_i^\dagger)), \tag{10.13}$$

our inequality becomes

$$(\Lambda \otimes \mathrm{id}_k)(\mathrm{id}_d \otimes f_i)(|\Omega\rangle\langle\Omega|) \geq 0. \tag{10.14}$$

Rewriting, this expression becomes

$$(\mathrm{id}_{d'} \otimes f_i)(\Lambda \otimes \mathrm{id}_d) |\Omega\rangle\langle\Omega| = \underbrace{(I_{d'} \otimes R_i)}_{A} \underbrace{(\Lambda \otimes \mathrm{id}_d)(|\Omega\rangle\langle\Omega|)}_{B} \underbrace{(I_{d'} \otimes R_i^\dagger)}_{A^\dagger}. \tag{10.15}$$

This is equivalent to the condition on matrices that $ABA^\dagger \geq 0$, and it turns out that for $ABA^\dagger \geq 0$, it suffices to have $B \geq 0$.[17] Thus

$$(\Lambda \otimes \mathrm{id}_d) |\Omega\rangle\langle\Omega| \geq 0. \tag{10.16}$$

☒

This construction we have defined is known as the *Choi matrix* (a Choi state of Ω), i.e.

$$J \equiv J(\Lambda) = (\Lambda \otimes \mathrm{id}) |\Omega\rangle\langle\Omega|). \tag{10.17}$$

Theorem 10.18 (Stinespring's dilation theorem). *Let $\Lambda : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ be a quantum operator. Then there exists a Hilbert space $\mathcal{H}'$ and a unitary operator $U \in \mathcal{B}(\mathcal{H} \otimes \mathcal{H}')$ such that $\forall \rho \in \mathcal{D}(\mathcal{H})$,*

$$\Lambda(\rho) = \mathrm{Tr}_{\mathcal{H}'}(U(\rho \otimes \phi)U^\dagger) \tag{10.19}$$

where ϕ is some fixed (pure) state in $\mathcal{H}'$.

That is, to perform a quantum operation we can couple to an ancilla, perform the unitary operation, and trace over the degrees of freedom in the ancilla $\mathcal{H}'$.

Stinespring's dilation theorem is a result from operator theory, but we'll see shortly that there are two more equivalent and relevant formulations, known as the Kraus Representation Theorem and the C-J isomorphism. We'll discuss this first one today.

Theorem 10.20 (Kraus Rep'n Theorem). *A linear map $\Lambda : \mathcal{M}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ is CP if*

$$\Lambda(\rho) = \sum_{k=1}^{r} A_k \rho A_k^\dagger \tag{10.21}$$

where $\{A_k\}_{k=1}^r$ is a finite set of linear operators in $\mathcal{B}(\mathcal{H}, \mathcal{K})$. Moreover it is TP if

$$\sum_{k=1}^{r} A_k^\dagger A_k = I_{\mathcal{H}}. \tag{10.22}$$

[17] Basically, if $B \geq 0$ then $\langle v|B|v\rangle \geq 0 \forall v$. But then define $A^\dagger v' = v$, and we see that

$$\langle v|B|v\rangle = \langle A^\dagger v'|B|A^\dagger v'\rangle = \langle v'|ABA^\dagger|v'\rangle \geq 0 \, \forall v'.$$

Proof. We start by proving that the latter holds if the map is trace preserving and 10.21 holds. That is, trace preserving tells us that

$$\begin{aligned} 1 &= \operatorname{Tr}\Lambda(\rho) \\ &= \operatorname{Tr}\sum_k A_k\rho A_k^\dagger \\ &= \sum_k \operatorname{Tr}(A_k\rho A_k^\dagger) \\ &= \sum_k \operatorname{Tr}(A_k^\dagger A_k\rho) \\ &= \operatorname{Tr}\left((\sum_k A_k^\dagger A_k)\rho\right) \forall\rho \\ &\implies \sum_k A_k^\dagger A_k = I_{\mathcal{H}}. \end{aligned}$$

Here, we have done nothing other than use definitions and the linearity and cyclic property of the trace. ☒

Kraus Rep'n Thm $\equiv$ restatement of Stin. D. Thm. WLOG assume $\phi \equiv |\phi\rangle\langle\phi| \in \mathcal{D}(\mathcal{H}')$. Let $\{|e_k\rangle\}_k$ be an onb for $\mathcal{H}'$. By Kraus, we can write

$$\Lambda(\rho) = \sum_k \langle e_k|U(\rho\otimes\phi)U^\dagger|e_k\rangle = \sum_k A_k\rho A_k^\dagger. \tag{10.23}$$

with ϕ defined as above. That is, $\Lambda(\rho) = \operatorname{Tr}_{\mathcal{H}'}(U(\rho\otimes\phi)U^\dagger)$. We define

$$A_k := \langle e_k|U|\phi\rangle \tag{10.24}$$

where $U \in \mathcal{B}(\mathcal{H}\otimes\mathcal{H}')$ and it follows that

$$\begin{aligned} \sum_k A_k^\dagger A_k &= \sum_k \langle\phi|U^\dagger\, |e_k\rangle\langle e_k|\, U|\phi\rangle \\ &= \langle\phi|\phi\rangle I_{\mathcal{H}} = I_{\mathcal{H}}. \end{aligned}$$

We call the A_k Kraus operators. Some of the details are an exercise to fill in later.

Choi-Jamilkowski (C-J) isomorphism We saw that $\Lambda : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ where $\mathcal{H} \simeq \mathbb{C}^d, \mathcal{K} \simeq \mathbb{C}^{d'}$ is CP iff $J(\Lambda) = (\Lambda\otimes \mathrm{id}_d)\,|\Omega\rangle\langle\Omega| \geq 0$. In fact, it turns out that $\exists$ an isomorphism between linear maps and positive operators. This is a great result, since positive operators are much nicer to work with.

Theorem 10.25. *The following equation provides a bijection between linear maps $\Lambda : \mathcal{M}_d \to \mathcal{M}_{d'}$ and operators $J \in \mathcal{B}(\mathbb{C}^{d'}\otimes\mathbb{C}^d)$, with J defined as follows:*

$$J \equiv (\Lambda\otimes\mathrm{id}_d)\,|\Omega\rangle\langle\Omega| \tag{10.26}$$

and

$$\operatorname{Tr}(A\Lambda(B)) = d\operatorname{Tr}(J(A\otimes B^T)) \tag{10.27}$$

$\forall A \in \mathcal{M}_{d'}, B \in \mathcal{M}_d$. *The maps $\Lambda \to J \to \Lambda$ defined by 10.26 and 10.27 are mutual inverses and lead to the following correspondence:*

(a) *Λ is CP* $\iff$ $J \geq 0$.
(b) *Λ is TP* $\iff$ $\operatorname{Tr}_A J = I_d/d$.

Proof. We'll first prove that 10.26→10.27. The RHS of 10.27 is

$$\begin{aligned} \text{RHS} &= d\operatorname{Tr}(J(A\otimes B^T)) \\ &= d\operatorname{Tr}((\Lambda\otimes\mathrm{id})(\Omega)(A\otimes B^T)). \end{aligned}$$

Note we will need the concept of the *adjoint* Λ^* of a map Λ w.r.t. the Hilbert-Schmidt inner product. That is, if $\Lambda : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$, then $\Lambda^* : \mathcal{B}(\mathcal{K}) \to \mathcal{B}(\mathcal{H})$ where

$$\operatorname{Tr}(A\Lambda(B)) = \operatorname{Tr}(\Lambda^*(A)B). \tag{10.28}$$

Thus writing in terms of the adjoint, we have

$$\begin{aligned}\text{RHS} &= d\operatorname{Tr}((\Lambda\otimes \mathrm{id}_d)(\Omega)(A\otimes B^T))\\ &= d\operatorname{Tr}((A\otimes B^T)(\Lambda\otimes \mathrm{id}_d)(\Omega))\\ &= d\operatorname{Tr}((\Lambda^*(A)\otimes B^T)(|\Omega\rangle\langle\Omega|)).\end{aligned}$$

Note this is slightly different from how it was presented in lecture. Here, I've used the cyclic property of the trace to switch the order of J and $A\otimes B^T$, where I'm considering both as elements of $M_{d'}\otimes M_d$, and then I used the definition of the adjoint to change the Λ into a Λ^*.[18] Of course, we can split up the tensor product as

$$\begin{aligned}(\Lambda^*(A)\otimes B^T)\,|\Omega\rangle\langle\Omega| &= (\Lambda^*(A)\otimes I)(I\otimes B^T)\,|\Omega\rangle\langle\Omega|\\ &= (\Lambda^*(A)\otimes I)(B\otimes I)\,|\Omega\rangle\langle\Omega|\\ &= (\Lambda^*(A)B\otimes I)\,|\Omega\rangle\langle\Omega|\\ &= (A\Lambda(B)\otimes I)\,|\Omega\rangle\langle\Omega|,\end{aligned}$$

where we have used the ricochet property in the second line to change a B^T into a B and turned Λ^* back into a Λ. Finally, observe that this object (which after all is just $J(A\otimes B^T)$) lives in $M_{d'}\otimes M_d$. Let us denote a partial trace over the $M_{d'}$ subsystem by $\operatorname{Tr}_{d'}$ and over M_d, by Tr_d. In this notation, we see that

$$\begin{aligned}d\operatorname{Tr}((A\Lambda(B)\otimes I)\,|\Omega\rangle\langle\Omega|) &= \operatorname{Tr}\left[(A\Lambda(B)\otimes I)\sum_{i,j}|i\rangle\langle j|\otimes|i\rangle\langle j|\right]\\ &= \operatorname{Tr}_{d'}\operatorname{Tr}_d\left[(A\Lambda(B)\otimes I)\sum_{i,j}|i\rangle\langle j|\otimes|i\rangle\langle j|\right]\\ &= \operatorname{Tr}_{d'}\left[(A\Lambda(B)\otimes I)\sum_{i,j}|i\rangle\langle j|\,\delta_{ij}\right]\\ &= \operatorname{Tr}_{d'}\left[\sum_i(A\Lambda(B))\,|i\rangle\langle i|\right]\\ &= \operatorname{Tr}(A\Lambda(B)),\end{aligned}$$

where we recognize $\sum_i |i\rangle\langle i|$ as just the identity. We conclude that

$$\operatorname{Tr}(A\Lambda(B)) = d\operatorname{Tr}(J(A\otimes B^T)). \qquad \boxtimes$$

Lecture 11.

Wednesday, February 13, 2019

Last time, we continued our discussion of quantum operations as linear CPTP maps. We proved that a map Λ is CP $\iff$ $J(\Lambda) = (\Lambda\otimes \mathrm{id})\,|\Omega\rangle\langle\Omega| \geq 0$, so it suffices to check positivity on the maximally entangled state. We mentioned the Stinespring Dilation Theorem from operator theory, and showed that from Stinespring we can get the Kraus Rep. Theorem. Finally, we started setting up the C-J isomorphism, which establishes an isomorphism between linear maps and positive operators.

The C-J isomorphism says that for

$$J \equiv (\Lambda\otimes \mathrm{id})\,|\Omega\rangle\langle\Omega|, \tag{11.1}$$

we have

$$\operatorname{Tr}(A\Lambda(B)) = d\operatorname{Tr}(J(A\otimes B^T)). \tag{11.2}$$

We proved last time that Λ is CP $\iff$ $J\geq 0$. Next, we will show that Λ is TP $\iff$ $\operatorname{Tr}_A J = I_d/d$.

[18] It is also fairly clear that the adjoint of id is another identity operator on the appropriate space of matrices. Notice that $\operatorname{Tr}(A\,\mathrm{id}(B)) = \operatorname{Tr}(AB)$ and $\operatorname{Tr}(A\,\mathrm{id}\,B) = \operatorname{Tr}(\mathrm{id}^*(A)B)$. For this to be true for all A, B, it must be that $\mathrm{id}^* = \mathrm{id}$, so the identity operator is self-adjoint. Thus we've sort of skipped a line here– $(A\otimes B^T)(\Lambda\otimes \mathrm{id}_d) = \Lambda^*(A)\otimes \mathrm{id}^*(B^T) = \Lambda^*(A)\otimes B^T$. The result then follows.

Proof. Suppose that Λ is trace-preserving. Then $\operatorname{Tr}\Lambda(B) = \operatorname{Tr}(I_{d'}\Lambda(B)) = \operatorname{Tr}(\Lambda^*(I_{d'})B) = \operatorname{Tr} B \, \forall B$, so

$$\Lambda^*(I_{d'}) = I_d. \tag{11.3}$$

Now the trace of J is

$$\begin{aligned}\operatorname{Tr} J &= \operatorname{Tr}((\Lambda \otimes \mathrm{id}_d)\Omega) \\ &= \operatorname{Tr}((I_A \otimes I_B)(\Lambda \otimes \mathrm{id}_d)\Omega) \\ &= \operatorname{Tr}((\Lambda^*(I_{d'}) \otimes I_d)\Omega) \\ &= \operatorname{Tr}((I_d \otimes I_d)\Omega) = \operatorname{Tr}(\Omega).\end{aligned}$$

We can break the trace up into the partial traces:

$$\begin{aligned}\operatorname{Tr}_B(\operatorname{Tr}_A J) &= \operatorname{Tr}_B \operatorname{Tr}_A(\Omega) \\ &= \operatorname{Tr}_B(I_d/d) \implies \operatorname{Tr}_A J = I_d/d.\end{aligned}$$

⊠

We now claim that 11.1 and 11.2 define an isomorphism, i.e. a map that is both injective and surjective.

CJ→ Kraus Suppose we have Λ a linear CP map. Thus CJ tells us that

$$J(\Lambda) = (\Lambda \otimes \mathrm{id}) \, |\Omega\rangle\langle\Omega| \geq 0.$$

We know that $\operatorname{Tr} J(\Lambda) = 1$, and we also know that we can decompose a state $|\psi_i\rangle$ as

$$|\psi_i\rangle = (R_i \otimes I)|\Omega\rangle \tag{11.4}$$

for some R_i.[19] These operators are $R_i \in \mathcal{B}(\mathbb{C}^d, \mathbb{C}^{d'})$, and thus we get a decomposition

$$J = \sum p_i \, |\psi_i\rangle\langle\psi_i| = \sum_i p_i (R_i \otimes I) \, |\Omega\rangle\langle\Omega| \, (R_i^\dagger \otimes I). \tag{11.5}$$

Thus with $A_i := \sqrt{p_i} R_{i'}$, we get

$$J(\Lambda) = \sum_{i'} (A_i \otimes I) \, |\Omega\rangle\langle\Omega| \, (A_i^\dagger \otimes I). \tag{11.6}$$

Comparing to the original definition of $J(\Lambda)$ in terms of Λ, we see that

$$\Lambda(\rho) = \sum_{I=1}^{r} A_i \rho A_i^\dagger. \quad \boxtimes \tag{11.7}$$

Kraus → Stinespring We want to show that we can get Stinespring (a linear map Λ written in terms of unitaries U, a reference state ϕ, and a partial trace over the ancilla) from Kraus. One possible isometry is

$$|\Psi\rangle \equiv U(|\psi\rangle \otimes |\phi\rangle) = \sum_{k=1}^{r} A_k |\psi\rangle \otimes |k\rangle \tag{11.8}$$

where $\{|k\rangle\}$ is an onb in $\mathcal{H}'$. One may check that U is indeed an isometry, i.e.

$$\langle\Psi|\Psi\rangle = \langle\psi|\psi\rangle \tag{11.9}$$

using $\{|k\rangle\}$ an onb and $\sum A_k^\dagger A_k = I$.

We see that

$$U(\rho \otimes |\phi\rangle\langle\phi|)U^\dagger = \sum p_i (U|\psi_i\rangle \otimes |\phi\rangle)(\ldots)^\dagger. \tag{11.10}$$

Taking the partial trace over H' we see that

$$\sum A_k \rho A_k^\dagger = \Lambda(\rho) = \operatorname{Tr}_{\mathcal{H}'}(U(\rho \otimes \phi)U^\dagger). \tag{11.11}$$

[19] Previously, this was $|\psi\rangle = (I \otimes R)|\Omega\rangle$. But by ricochet, we can just move this over to some $(R^T \otimes I)|\Omega\rangle$ and relabel $R^T = R_i$.

Measurement Here is the third postulate of quantum mechanics, the "von Neumann/projective" measurement formalism. In a closed system, we have:

- A system in state $|\psi\rangle$
- Measure an observable A
- The outcome is an eigenvalue of A, some $\{a\}$.
- The probability of outcome a is given by a projection,
$$p(a) = \langle\psi|P_a|\psi\rangle \tag{11.12}$$
where $A = \sum aP_a = \sum a\,|e_a\rangle\langle e_a|$.
- The post-measurement state if the outcome was a is then
$$|\psi\rangle \to |\psi'\rangle = \frac{P_a|\psi\rangle}{\sqrt{\langle\psi|P_a|\psi\rangle}}. \tag{11.13}$$

Example 11.14. Suppose your friend goes the the lab and prepares an electron in the spin state $|\psi\rangle$, where
$$\boldsymbol{\sigma}\cdot\hat{n}|\psi\rangle = |\psi\rangle \tag{11.15}$$
where $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ and $\hat{n}$ is a unit vector. For instance, if $\hat{n} = (0, 0, 1)$, then $|\psi\rangle = |0\rangle$ the up-spin state.

We can ask the reasonable question: "What is the direction of $\hat{n}$?" This is a perfectly legitimate question, but $\hat{n}$ does not represent an observable (i.e. a Hermitian operator), so we cannot answer this question with the existing measurement formalism.

Note that these projection operators P_a had better be positive semidefinite in order for our outcomes to have a probabilistic interpretation, and
$$\sum p(a) = 1 \implies 1 = \sum_a \langle\psi|P_a|\psi\rangle \implies \sum_a P_a = I. \tag{11.16}$$
Since we measure with some self-adjoint operator A, it must be that
$$P_a P_b = \delta_{ab} P_a. \tag{11.17}$$
That is, our projections are orthogonal. It is this postulate we will drop.

Generalized measurement postulate In our broader formalism, measurements are described by some operators $\{M_a\}$. We assume nothing about these M_a. The as label possible outcomes, such that
$$\sum_a M_a M_a^\dagger = I, \tag{11.18}$$
a completeness relation. Now if the system is in a state $|\psi\rangle$, then we say the probability of getting a is
$$p(a) = \langle\psi|M_a^\dagger M_a|\psi\rangle. \tag{11.19}$$
The post-measurement state is then
$$|\psi\rangle \to |\psi'\rangle = \frac{M_a|\psi\rangle}{\sqrt{\langle\psi|M_a^\dagger M_a|\psi\rangle}}. \tag{11.20}$$
We see that in the special case where $M_a = P_a$, since $M_a^\dagger M_a = P_a^\dagger P_a = P_a^2 = P_a$, we get back the old projective measurement postulate,
$$|\psi'\rangle = \frac{P_a|\psi\rangle}{\sqrt{\langle\psi|P_a^\dagger P_a|\psi\rangle}} = \frac{P_a|\psi\rangle}{\sqrt{\langle\psi|P_a|\psi\rangle}}.$$

POVMs We now introduce *positive operator-valued measures*, or POVMs. We had $p(a) = \langle\psi|M_a^\dagger M_a|\psi\rangle$, so let us define $E_a := M_a^\dagger M_a \geq 0$. $\sum_a E_a = \sum_a M_a^\dagger M_a = I$, and clearly $E_a^\dagger = E_a$.

Of course, it follows that $E_a \geq 0 \implies p(a) \geq 0$. One may define that $p(a) = \mathrm{Tr}(E_a\rho)$. In addition, since $\sum_a E_a = I \implies \sum p(a) = 1$. We call these E_a POVM elements.

Definition 11.21. A POVM is defined by any partition of the identity I into a finite set of positive semidefinite operators $\{E_a\}$ acting on the Hilbert space $\mathcal{H}$ of the system to be measured, i.e.
$$E_a \geq 0, \quad \sum_a E_a = I.$$

Lecture 12.

Friday, February 15, 2019

Last time, we introduced the measurement formalism with our generalized measurement postulate. Thus for a set of operators $\{M_a\}$ acting on a state $|\psi\rangle$ or equivalently a density matrix ρ, we can define a probability of an outcome a by

$$p(a) = \langle\psi|M_a^\dagger M_a|\psi\rangle, \quad p(a) = \mathrm{Tr}(M_a^\dagger M_a \rho). \tag{12.1}$$

Unlike in the previous formalism, M_a need not be self-adjoint. Thus the post-measurement state is given by

$$|\psi\rangle \to |\psi'\rangle = \frac{M_a|\psi\rangle}{\langle\psi|M_a^\dagger M_a|\psi\rangle}, \quad \rho \to \rho' = \frac{M_a \rho M_a^\dagger}{\mathrm{Tr}(M_a^\dagger M_a \rho}. \tag{12.2}$$

Naimark's Theorem We shall discuss the implementation of a general measurement, following Stinespring. Consider a system $\mathcal{H}_A$ with initial state $|\psi\rangle$, and some measurement operators $\{M_a\}$.

(a) Add an ancilla B with Hilbert space $\mathcal{H}_B$ such that $\dim \mathcal{H}_B = |\{M_a\}| = \#$ of posssible outcomes. Equip B with an onb $\{|e_a\rangle\}$.

(b) Consider B to be in some state $|\phi\rangle$ so that the initial combined state is

$$|\psi\rangle \otimes |\phi\rangle, \tag{12.3}$$

where the states of A, B are initially uncorrelated.

(c) Stinespring tells us we will need a unitary U acting on $\mathcal{H}_A \otimes \mathcal{H}_B$ to implement our measurement. Let us define

$$|\Psi_{AB}\rangle = U(|\psi\rangle \otimes |\phi\rangle) := \sum_a M_a|\psi\rangle \otimes |e_a\rangle. \tag{12.4}$$

One may check that U preserves inner products on states of AB of the form $|\psi\rangle \otimes |\phi\rangle$, i.e. for

$$|\Phi\rangle = U(|\varphi\rangle \otimes |\phi\rangle = \sum_a M_a|\varphi\rangle \otimes |e_a\rangle, \tag{12.5}$$

we have

$$\langle\Phi|\Psi\rangle = \langle\varphi|\psi\rangle \tag{12.6}$$

using only the properties that $\{|e_a\rangle\}$ form an onb and $\sum M_a^\dagger M_a = I$. Vecotrs of the form $|\chi\rangle \otimes |\phi\rangle$ for a fixed $|\phi\rangle$ span a subspace $\mathcal{H}_S$ of $\mathcal{H}_A \otimes \mathcal{H}_B$. Thus

$$U : \mathcal{H}_S \to \mathcal{H}_A \otimes \mathcal{H}_B. \tag{12.7}$$

Note that such an operator U can be extended to a unitary on the full Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$, i.e. $\exists$ some U' unitary with

$$U' : \mathcal{H}_A \otimes \mathcal{H}_B \to \mathcal{H}_A \otimes \mathcal{H}_B \quad \text{s.t. } U'(|\chi\rangle \otimes |\phi\rangle) \equiv U(|\chi\rangle \otimes |\phi\rangle. \tag{12.8}$$

That is, U' agrees with U on all the states in $\mathcal{H}_S$.

(d) To finish the theorem, we make a projective measurement on the state $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ to get back to the system A. A projective measurement consistes of a set of projection operators $\{P_a\}$ where

$$P_a = I_A \otimes |e_a\rangle\langle e_a| \,. \tag{12.9}$$

Note that a is an index and not summed over! One may check these are indeed projective, i.e. $P_a P_{a'} = \delta_{aa'} P_a$. Now the probability of an ouctome a is given by

$$p(a) = \langle\Psi|P_a|\Psi\rangle. \tag{12.10}$$

Substituting directly, we see that

$$\begin{aligned} p(a) &= \langle\psi| \otimes \langle\phi|U^\dagger P_a \underbrace{U|\psi\rangle \otimes |\phi\rangle}_{\sum_{a'} M_{a'}|\psi\rangle \otimes |e_{a'}\rangle} \\ &= \langle\psi|M_a^\dagger M_a|\psi\rangle. \end{aligned}$$

Moreover, the post-measurement state will be

$$\begin{aligned}|\Psi\rangle \to |\Psi'\rangle &\propto P_a|\Psi\rangle \\ &\sim M_a|\psi\rangle \otimes |e_a\rangle\end{aligned}$$

up to a normalization factor. Once we trace over the ancilla, we get

$$\mathrm{Tr}_B|\Psi'\rangle\langle\Psi'| \propto M_a|\psi\rangle\langle\psi|M_a^\dagger, \tag{12.11}$$

which is exactly the correct post-measurement state we expected from applying M_a directly.

Thus our procedure can be summed up as follows. Add an ancilla B. Define unitary dynamics (depending on $\{M_a\}$). Perform the projective measurement in AB. Finally, take a partial trace over the ancilla B to get the post-measurement state.

Example 12.12. Let's return to our previous example of trying to find the direction of the spin of an electron. Someone prepares a spin in a state

$$\boldsymbol{\sigma} \cdot \hat{n}|\psi\rangle = \psi \tag{12.13}$$

where $\hat{n} \in \{\hat{n}_a\}$ is some finite set such that $\exists\{\lambda_a\}$ with $\sum \lambda_a \hat{n}_a = 0, \lambda_a \geq 0, \sum_a \lambda_a = 1$.

Recall we defined POVMs, which were measurements $\{E_a\}$ where we don't care about the post-measurement state. They obeyed $E_a \geq 0$ and $\sum_a E_a = I$, such that for a density matrix ρ, $p(a) = \mathrm{Tr}(E_a\rho)$.

In this case, we see that this measurement admits a POVM:

$$E_a := \lambda_a(I + \hat{n}_a \cdot \boldsymbol{\sigma}). \tag{12.14}$$

We now claim that

$$E_a = 2\lambda_a P_{\hat{n}_a}, \tag{12.15}$$

where $P_{\hat{n}_a} = |\uparrow_{\hat{n}_a}\rangle\langle\uparrow_{\hat{n}_a}|$ is a projective operator. Thus

$$\hat{n}_a \cdot \boldsymbol{\sigma}|\uparrow_{\hat{n}_a}\rangle = |\uparrow_{\hat{n}_a}\rangle \tag{12.16}$$

$$\hat{n}_a \cdot \boldsymbol{\sigma}|\downarrow_{\hat{n}_a}\rangle = |\downarrow_{\hat{n}_a}\rangle. \tag{12.17}$$

It follows that

$$\begin{aligned}E_a|\uparrow_{\hat{n}_a}\rangle &= \lambda_a(I + \hat{n}_{a'}\boldsymbol{\sigma})|\uparrow_{\hat{n}_a}\rangle \\ &= 2\lambda_a|\uparrow_{\hat{n}_a}\rangle.\end{aligned}$$

We have $P_{\hat{n}_a}|\uparrow_{\hat{n}_a}\rangle = |\uparrow_{\hat{n}_a}\rangle$ and $P_{\hat{n}_a}|\downarrow_{\hat{n}_a}\rangle = 0$ with our choice of P as above.

Thus $E_a \geq 0$, and

$$\begin{aligned}\sum E_a &= \sum \lambda_a I + \sum_a \lambda_a \hat{n}_a \cdot \boldsymbol{\sigma} \\ &= I,\end{aligned}$$

where the second term is zero. Thus the $\{E_a\}$ form a POVM.

Consider the case where $\hat{n} \in \{\hat{n}_1, \hat{n}_2\}$, with $\lambda_1 = \lambda_2 = 1/2$. Then $\hat{n}_1 + \hat{n}_2 = 0$. It follows that

$$E_1 = 2\lambda P_{\hat{n}_1} = P_{\hat{n}_1} \tag{12.18}$$

$$E_2 = I - P_{\hat{n}_1}. \tag{12.19}$$

Thus our POVM is really a projective measurement. One should check that given an initial state $|\psi\rangle$ such that $\boldsymbol{\sigma} \cdot \hat{n}_1|\psi\rangle = |\psi\rangle$,

$$p(\hat{n}_1) = \langle\psi|E_1|\psi\rangle, \quad p(\hat{n}_2) = 0. \tag{12.20}$$

In the example sheet, we will consider the case of three spin states and $E_a = \frac{2}{3}\hat{P}_{n_a}$.

In a similar vein, on the examples sheet we will consider the case where Alice prepares a state $|\psi\rangle$ which is either $|0\rangle$ or $|+\rangle = \frac{|0\rangle+|1\rangle}{\sqrt{2}}$. For this setup, we can actually prepare a POVM such that we never make an error of misidentification– our POVM may tell us that the state is $|0\rangle$, and it is definitely correct, or $|+\rangle$, and it is definitely correct. But sometimes it will conclude that we can't decide what the state is. A pure projective measurement could not have told us this.

We may also define a *pure POVM*, which is some E_a of the form

$$E_a = |\psi_a\rangle\langle\psi_a|\,.$$

Bipartite entanglement Consider a pure state $|\psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$. We call a pure state a *product state* if $\exists|\chi_A\rangle \in \mathcal{H}_A, |\phi_B\rangle \in \mathcal{H}_B$ such that

$$|\psi_{AB}\rangle = |\chi_A\rangle \otimes |\phi_B\rangle. \tag{12.21}$$

Otherwise, the state is *entangled.*

Similarly, consider a mixed state $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$. If

$$\exists \omega_A \in \mathcal{D}(\mathcal{H}_A), \sigma_B \in \mathcal{D}(\mathcal{H}_B) \text{ s.t. } \rho_{AB} = \omega_A \otimes \sigma_B, \tag{12.22}$$

we call the state a (mixed) *product state*. On the other hand, if

$$\exists \{p_i\}, \rho_i^A \in \mathcal{D}(\mathcal{H}_A), \rho_i^B \in \mathcal{D}(\mathcal{H}_B) \text{ s.t. } \rho_{AB} = \sum_i p_i \rho_i^A \otimes \rho_i^B, \tag{12.23}$$

we call this state *separable*. Clearly, product states are a subset of separable states where just one of the p_is is nonzero. Otherwise, the state is *entangled*.

- Product states have no correlation between the two systems. Alice and Bob prepare their systems separately and never coordinate.
- Separable states have *classical* correlations. Alice and Bob use a classical communication channel, e.g. A and B share a random number generator that produces outcome i with probability p_i. They decide to construct by local operations (LO) the state $\rho_i^A \otimes \rho_i^B$.
- Otherwise, the state is entangled and exhibits purely quantum correlations.

Lecture 13.

Monday, February 18, 2019

Entanglement We defined the notion of entanglement last time. Note that entanglement cannot be created or increased via LOCC (local operation classical channels). However, it will turn out to be a valuable resource (e.g. for use in algorithms).

Some of the simplest entangled states we can write down are the Bell states in $\mathcal{H}_A \otimes \mathcal{H}_B \simeq \mathbb{C}^2 \otimes \mathbb{C}^2$. They are

$$|\phi^{\pm}_{AB}\rangle = \frac{1}{\sqrt{2}}(|00\rangle \pm |11\rangle) \tag{13.1}$$

$$|\psi^{\pm}_{AB}\rangle = \frac{1}{\sqrt{2}}(|01\rangle \pm |10\rangle). \tag{13.2}$$

These four states can be characterized by two bits– a parity bit (are the two bits parallel, e.g. $|00\rangle$, or antiparallel, $|01\rangle$) and a phase bit (is the sign of the phase $+$ or $-$). For instance, in this notation, 01 (with parity the first bit, phase the second) indicates $|\phi^-\rangle$.

Two bits can therefore be encoded in a Bell state. This information can be recovered/decoded by a *joint* measurement on the 2 qubits. Suppose we want to send a message to a friend, but we only have a quantum channel, i.e. we can only send qubits. What is the measurement we will make? It is a *Bell measurement*, a projective measurement with the following four operators:

$$P_{00} = |\phi^+\rangle\langle\phi^+| \tag{13.3}$$

$$P_{01} = |\phi^-\rangle\langle\phi^-| \tag{13.4}$$

$$P_{10} = |\psi^+\rangle\langle\psi^+| \tag{13.5}$$

$$P_{11} = |\psi^-\rangle\langle\psi^-|. \tag{13.6}$$

Say the state received was $|\phi^-\rangle$. Making this projective measurement, we get $p(11) = 0$ and indeed $p(10) = p(00) = 0$. Only $p(01) = 1$. Moreover, our post-measurement state when we get 1 is undisturbed. We got 1 and we didn't destroy the state in the process since

$$|\phi^{-\prime}\rangle \propto |\phi^-\rangle\langle\phi^-|\phi^-\rangle = |\phi^-\rangle. \tag{13.7}$$

"Distant labs" From now on, we shall look at the "distant labs" paradigm. That is, Alice and Bob each have one qubit, say one qubit of a Bell state, e.g. $|\phi^-_{AB}\rangle$. Suppose now Alice makes a measurement with a local unitary operator (i.e. she can only affect her qubit), e.g.

$$(\sigma_z^A \otimes I_B). \tag{13.8}$$

It's straightforward to see that since $\sigma_z|0\rangle = 0, \sigma_z|1\rangle = -|1\rangle$,

$$|\phi^+\rangle \leftrightarrow |\phi^-\rangle \tag{13.9}$$

$$\frac{|00\rangle + |11\rangle}{\sqrt{2}} \leftrightarrow \frac{|00\rangle - |11\rangle}{\sqrt{2}}. \tag{13.10}$$

Similarly,

$$|\psi^+\rangle \leftrightarrow |\psi^-\rangle.$$

Under $\sigma_x^A \otimes I_B$, we see that the Bell states will be exchanged as follows:

$$|\phi^+\rangle \leftrightarrow |\psi^+\rangle \tag{13.11}$$

$$|\phi^-\rangle \leftrightarrow |\psi^-\rangle. \tag{13.12}$$

Now suppose that Alice and Bob have a classical channel (e.g. a telephone), so they can coordinate their measurements. For instance, Alice and Bob agree to both perform σ_z on their respective qubits. The outcome is ±1 for each of them. They can communicate the outcomes and infer *either* the phase bit or the parity bit, but not both.

Example 13.13. Say the initial state (unknown to A and B) is $|\phi^-\rangle$. Suppose they measure $\sigma_z^A \otimes \sigma_z^B$, and they get the outcomes $+1, +1$. The post-measurement state is then given by acting with the projective operator $P_{1,1} = |0\rangle\langle 0| \otimes |0\rangle\langle 0|$.[20] Then the post-measurement state is

$$\propto P|\phi^-\rangle = (|0\rangle\langle 0| \otimes |0\rangle\langle 0|)\left(\frac{|00\rangle - |11\rangle}{\sqrt{2}}\right) \tag{13.14}$$

$$= |00\rangle. \tag{13.15}$$

Thus they have determined the parity bit to be zero, but in doing so they've destroyed the entanglement in the original state and cannot recover the phase bit.

Generalized measurement of Bell states How does the story change if we do a generalized measurement? Suppose A and B share

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

Alice does a generalized measurement with

$$M_1 = \begin{pmatrix} \cos\theta & 0 \\ 0 & \sin\theta \end{pmatrix}, \quad M_2 = \begin{pmatrix} \sin\theta & 0 \\ 0 & \cos\theta \end{pmatrix}. \tag{13.16}$$

The possible outcomes are 1 and 2. If the outcome is 1, then the post-measurement state is proportional to

$$(M_1 \otimes I_B)|\phi^+\rangle = \cos\theta|00\rangle + \sin\theta|11\rangle$$

and if the outcome is 2,

$$(M_2 \otimes I_B)|\phi^+\rangle = \cos\theta|11\rangle + \sin\theta|00\rangle,$$

where it's a simple exercise to check that these are the final states.

Based on her measurement, Alice makes a decision. If she got outcome 1, she does nothing ($I \otimes I$), and if she gets 2, she performs σ_x^A on her qubit (the NOT operation). Thus the new states are

$$\cos\theta|00\rangle + \sin\theta|11\rangle,$$

$$\cos\theta|01\rangle + \sin\theta|10\rangle.$$

[20] That is, the post-measurement state is given by the projection operator made from the eigenvector corresponding to the eigenvalue we measured.

Finally, Alice tells Bob what she measured, whereupon if the measurement was 1, Bob does nothing, and if the measurement was 2, Bob uses σ_x^B on his qubit so that either way, the final state shared between A and B is

$$|\phi_{AB}^+\rangle \to \cos\theta|00\rangle + \sin\theta|11\rangle \equiv |\chi\rangle. \tag{13.17}$$

One can readily check[21] that in general,

$$\rho_A = \mathrm{Tr}_B|\chi\rangle\langle\chi| \neq I/2, \tag{13.18}$$

so the Schmidt rank of this state is 2. By LOCCs, we have gone from a maximally entangled state to a non-maximally entangled state.

Suppose now Alice and Bob share a general state $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$. Can they change it to a desired state $|\phi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ via LOCC? The answer to this question is captured in *Nielsen's majorization theorem.*

What is majorization? To understand the theorem, we'll have to know what majorization is. Let $\mathbf{x} = (x_1, \ldots, x_n), \mathbf{y} = (y_1, \ldots, y_n)$ with $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. We say that $\mathbf{x}$ is *majorized* by $\mathbf{y}$, denoted $\mathbf{x} \prec \mathbf{y}$ if

$$\sum_{i=1}^k x_i^\downarrow \leq \sum_{i=1}^k y_i^\downarrow \quad \forall 1 \leq k \leq n-1 \tag{13.19}$$

and

$$\sum_{i=1}^n x_i^\downarrow = \sum_{i=1}^n y_i^\downarrow, \tag{13.20}$$

where $x_1^\downarrow \geq x_2^\downarrow \geq \ldots \geq x_n^\downarrow$ orders the elements of $\mathbf{x}$ in non-increasing order. For

$$\mathbf{x} = (1/n, \ldots, 1/n), \quad \mathbf{y} = (1, 0, 0, \ldots, 0), \tag{13.21}$$

we see that $\mathbf{x}$ is majorized by $\mathbf{y}$ since $1/n \leq 1, 2/n \leq 1, \ldots$ and $k/n \leq 1$.[22]

Theorem 13.22. $\mathbf{x} \prec \mathbf{y}$ *iff* $\exists\{p_i\}, \{P_i\}$ *with* P_i *some permutation matrices such that*

$$\mathbf{x} = \sum_i p_i P_i \mathbf{y}. \tag{13.23}$$

We also have Birkhoff's theorem:

Theorem 13.24. *For a matrix* $D = \sum_i p_i P_i$, *we say that if* $\sum_i D_{ij} = 1$ *and* $\sum_j D_{ij} = 1$, *then* D *is doubly stochastic.* $\mathbf{x} \prec \mathbf{y}$ *iff* $\exists D$ *(doubly stochastic) such that* $\mathbf{x} = D\mathbf{y}$.

In the quantum case, we say that for density matrices ρ, σ, ρ is majorized by σ if

$$\lambda(\rho) \prec \lambda(\sigma), \tag{13.25}$$

where $\lambda(\rho) = (r_1, \ldots, r_n)$ is the vector of the eigenvalues of ρ. Thus if $\rho \prec \sigma$, then $\exists\{p_i\}, \{U_i\}$ s.t.

$$\rho = \sum_i p_i U_i \sigma U_i^\dagger. \tag{13.26}$$

Nielsen's majorization theorem gives us the condition for the construction of an arbitrary state $|\phi\rangle$ from a given state $|\psi\rangle$.

Theorem 13.27 (Nielsen's majorization thm)**.** $|\psi\rangle \to |\phi\rangle$ *by LOCC iff* $\lambda_\psi \prec \lambda_\phi$ *where* λ_ψ *is the vector of eigenvalues* $\lambda(\rho_\psi)$ *with* $\rho_\psi = \mathrm{Tr}_B\, |\psi\rangle\langle\psi|$ *and* $\lambda_\phi = \lambda(\rho_\phi)$ *where* $\rho_\phi = \mathrm{Tr}_B\, |\phi\rangle\langle\phi|$.

Note it doesn't matter whether we trace over A or B by the Schmidt decomposition since for a pure bipartite state the nonzero eigenvalues after doing a partial trace are the same.

[21] The density matrix is

$$|\chi\rangle\langle\chi| = \cos^2\theta\, |00\rangle\langle 00| + \cos\theta \sin\theta (|00\rangle\langle 11| + |11\rangle\langle 00|) + \sin^2\theta\, |11\rangle\langle 11|\,,$$

so tracing over B (for instance) gives

$$\rho_A = \mathrm{Tr}_B(|\chi\rangle\langle\chi|) = \cos^2\theta\, |0\rangle\langle 0| + \sin^2\theta\, |1\rangle\langle 1| \neq I/2$$

except for in special cases like where $\theta = \pm\pi/4$.

[22] In words, order the elements of the vectors $\mathbf{x}, \mathbf{y}$ from largest to smallest. Take the partial sums of the first k elements in the ordered vectors. If every partial sum of the ordered $\mathbf{y}$ is greater than the corresponding partial sum of $\mathbf{x}$, with the full sums being equal, then $\mathbf{y}$ majorizes $\mathbf{x}$.

Lecture 14.

Wednesday, February 20, 2019

We started asking about what states can be constructed in a composite space $\mathcal{H}_A \otimes \mathcal{H}_B$ by LOCC.

There is a connection between majorization and the transfer of entanglement, established in Nielsen's Majorization Theorem.

Consider the reduced states

$$\rho_\psi = \mathrm{Tr}_B \,|\psi\rangle\langle\psi|\,; \quad \rho_\phi = \mathrm{Tr}_B \,|\phi\rangle\langle\phi|\,, \tag{14.1}$$

with $\lambda_\psi = \lambda(\rho_\psi)$ and $\lambda_\phi = \lambda(\rho_\phi)$ with λ the vector of eigenvalues.

Nielsen's Majorization theorem tells us that

$$|\psi\rangle \to |\phi\rangle \iff \lambda_\psi \prec \lambda_\phi. \tag{14.2}$$

We denote $\rho \prec \sigma$ if $\lambda(\rho)\mathrm{C}^{\mathrm{pre}}\lambda(\sigma)$. In fact, Uhlmann's theorem says that $\rho \prec \sigma \iff \exists$ a set of unitaries $\{U_i\}$ such that

$$\rho = \sum_i p_i U_i \sigma U_i^\dagger. \tag{14.3}$$

Recall that

$$\mathbf{x} \prec \mathbf{y} \iff \mathbf{x} = \sum_i p_i P_i \mathbf{y} \tag{14.4}$$

$$\iff \mathbf{x} = D\mathbf{y} \tag{14.5}$$

where D is doubly stochastic.

If $|\psi\rangle \to_{LOCC} |\phi\rangle$, then the operation can be implemented as follows. This generalizes the process we came up with last time.

(a) Alice does a single measurement $\{M_a\}$, getting an outcome a, and based on that outcome she performs a unitary W_a (may be the identity).
(b) By the classical channel (CC), Alice tells Bob that she measured the outcome a.
(c) Bob does his own local unitary (LU) U_a.

Proof. We prove this in the forward direction. If $|\psi\rangle \to_{LOCC} |\phi\rangle$, then $\lambda_\psi \prec \lambda_\phi$.

Alice makes her single measurement $\{M_a\}$, measures a, and performs a unitary W_a. Her initial state is $\rho_\psi = \mathrm{Tr}_B \,|\psi\rangle\langle\psi|$, and her final state is ρ_ϕ since she's successfully constructed her half of $|\phi\rangle$.

If we got the outcome a, then the post-measurement state of Alice is

$$\frac{M_a \rho_\psi M_a^\dagger}{p(a)}, \tag{14.6}$$

and after Alice performs the unitary W_a, she has

$$W_a \frac{M_a \rho_\psi M_a^\dagger}{p(a)} W_a^\dagger = \rho_\phi, \tag{14.7}$$

since Bob's unitary doesn't affect the half that Alice has. One may check that

$$\mathrm{Tr}_B (I \otimes U_a) \sigma_{AB} (I \otimes U_a^\dagger) = \mathrm{Tr}_B \,\sigma_{AB}. \tag{14.8}$$

Rearranging, we have

$$W_a M_a \rho_\psi M_a^\dagger W_a^\dagger = p(a)\rho_\phi. \tag{14.9}$$

We now apply the polar decomposition, which says that we can write an operator as

$$A = \sqrt{AA^\dagger} V. \tag{14.10}$$

Therefore it follows that

$$W_a M_a \sqrt{\rho_\psi} = \sqrt{W_a M_a \rho_\psi M_a^\dagger W_a^\dagger} V_a, \tag{14.11}$$

where we recognize the quantity in the square root as none other than $p(a)\rho_\phi$. Therefore

$$W_a M_a \sqrt{\rho_\psi} = \sqrt{p(a)} \sqrt{\rho_\phi} V_a. \tag{14.12}$$

Now

$$\sum_a \sqrt{\rho_\psi} M_a^\dagger W_a^\dagger W_a M_a \sqrt{\rho_\psi} = \sum_a p(a) V_a^\dagger \rho_\phi V_a. \tag{14.13}$$

Since $\sum M_a^\dagger M_a = I$, we find that

$$\rho_\psi = \sum p(a) V_a^\dagger \rho_\phi V_a \implies \lambda_\psi \prec \lambda\phi \tag{14.14}$$

by Uhlmann's theorem. ☒

Now, Nielsen's theorem has the following implications.

- LOCC cannot increase the Schmidt number of a state. That is, with $|\psi\rangle; n_\psi$ and $|\phi\rangle; n_\phi$, if $|\psi\rangle \to_{LOCC} |\phi\rangle$, then $n_\psi \geq n_\phi$.
- This implies that LOCC cannot increase the entanglement of a pure state.

Proof. Let $\lambda_\psi = (\nu_1, \ldots, \nu_d)$ and $\lambda_\phi = (\mu_1, \ldots, \mu_d)$ be the vectors of eigenvalues of ρ_ψ, ρ_ϕ respectively, where $d = \dim \mathcal{H}_A$. WLOG they are already ordered, $\nu_1 \geq \nu_2 \geq \ldots; \mu_1 \geq \mu_2 \geq \ldots$.

The proof is by contradiction. Assume $|\psi\rangle \to_{LOCC} |\phi\rangle$, with $n_\psi < n_\phi$. Thus

$$\lambda_\psi = (\nu_1, \ldots, \nu_j, 0, 0, \ldots, 0)$$
$$\lambda_\phi = (\mu_1, \ldots, \mu_j, \ldots, \mu_m, 0, \ldots, 0).$$

Thus $\exists$ some integer m such that $\mu_m \neq 0$ but $\nu_m = 0$. It follows that since all the other ν_i are zero,

$$\sum_{i=1}^{m-1} \nu_1 = 1 \text{ but } \sum_{=1}^{m-1} \mu_1 < 1. \tag{14.15}$$

By Nielsen's theorem, $|\psi\rangle \to_{LOCC} |\phi\rangle$ iff $\lambda_\psi \prec \lambda_\phi$, i.e.

$$\sum_{i=1}^{k} \nu_i \leq \sum_{i=1}^{k} \mu_1 \quad \forall 1 \leq k \leq d. \tag{14.16}$$

But we've just seen that if we take $k = m - 1 \leq d$, we have the LHS $= 1$ and the RHS < 1. Thus $n_\psi < n_\phi \implies \lambda_\psi \not\prec \lambda_\phi$, so $n_\psi \geq n_\phi$. ☒

We now define a measure of entanglement for a pure bipartite state, the *entanglement entropy*.

Definition 14.17. For a state $|\psi_{AB}\rangle$ with reduced density matrices ρ_A, ρ_B, the *entanglement entropy* is denoted $S(\rho_A) = S(\rho_B)$, where

$$S(\rho) = -\operatorname{Tr}(\rho \log \rho). \tag{14.18}$$

Theorem 14.19. *Let $|\psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B; d_A = \dim \mathcal{H}_A$.*

(a) $S(\rho_A) = 0 \iff |\psi_{AB}\rangle$ *is a product state (separable).* $S(\rho_A) > 0 \iff |\psi_{AB}\rangle$ *is entangled.*

(b) $S(\rho_A) = \log d_A$ *for the maximal mixed state* $\iff |\psi_{AB}\rangle$ *is a MES.*

Note that ρ admits a spectral decomposition,

$$\rho = \sum \lambda_i |e_i\rangle\langle e_i|, \tag{14.20}$$

so then $\log \rho = \sum (\log \lambda_i) |e_i\rangle\langle e_i| \implies S(\rho) = -\sum \lambda_i \log \lambda_i \equiv H(\{\lambda_i\})$, the Shannon entropy of the vector of eigenvalues. In particular, we see that a pure state has $S(\rho_A) = 0$ and $S(\rho_A) = \log d$ when $\{\lambda_1\} = (1/d, \ldots, 1/d)$.

Proof. $S(\rho_A) = 0 \iff \rho_A$ is pure $\iff$ the Schmidt number of $|\psi_{AB}\rangle = 0$. But then our state is

$$|\psi_{AB}\rangle = |\chi_A\rangle \otimes |\omega_B\rangle \tag{14.21}$$

is a separable (product) state. ☒

We can also see that if $|\psi\rangle \to_{LOCC} |\phi\rangle$ then $n_\psi \geq n_\phi$. In addition, the entropy is non-increasing: $S(\rho_\psi) \geq S(\rho_\phi)$.

There is a property known as Schur concavity: for $\rho \prec \sigma$, we have

$$\lambda(\rho) \prec \lambda(\sigma) \implies S(\rho) \geq S(\sigma). \tag{14.22}$$

This is a special case of the property for vectors that a function f is Schur concave if $\mathbf{x} \prec \mathbf{y} \implies f(\mathbf{x}) \geq f(\mathbf{y})$. It will turn out that any function which is both concave and symmetric is Schur concave.

Applications of entanglement We will now illustrate why entanglement is such a useful, fungible resource.

Superdense coding: Suppose Alice has 2 bits she wants to send to Bob, but her telephone line has been cut. She has no classical channel, and is only allowed to send 1 qubit via a noiseless quantum channel.

Can she send her two bits? Yes, *if* Alice and Bob already share a Bell state, e.g.

$$|\psi_{AB}^{+}\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}. \tag{14.23}$$

Alice has two bits she wants to send, and depending on what her message is, she acts locally on her qubit A as follows:

- $00 \to \sigma_0 : |\phi^+\rangle \mapsto |\phi^+\rangle$
- $01 \to \sigma_z : |\phi^+\rangle \mapsto |\phi^-\rangle$
- $10 \to \sigma_x : |\phi^+\rangle \mapsto |\psi^+\rangle$
- $11 \to i\sigma_y \equiv \sigma_z\sigma_x : |\phi^+\rangle \mapsto |\psi^-\rangle$.

She then sends her qubit to Bob, who now possesses the full state AB. But in fact, Bob can now make a Bell measurement, a projective measurement with the projectors

$$|\psi^{\pm}\rangle\langle\psi^{\pm}|, |\psi^{\pm}\rangle\langle\psi^{\pm}|. \tag{14.24}$$

But there's something even better about this– if there is a malicious eavesdropper (Eve) who intercepts Alice's qubit, she cannot recover the message because Alice's qubit alone is in a completely mixed state thanks to the magic of entanglement.

In contrast to superdense coding (send 2 classical bits using a qubit), we also have quantum teleportation (send a quantum state using a classical channel). These are some nice applications and we'll go over teleportation next time.

Lecture 15.

Friday, February 22, 2019

Last time, we discussed Nielsen's majorization theorem, which stated that a state $|\phi\rangle$ can be constructed from $|\psi\rangle$ by LOCC iff $\lambda_\psi \prec \lambda_\phi$. This theorem has two implications– first, $n_\psi \geq n_\phi$, so we cannot create/increase entanglement by local operations. Second, the entanglement entropy cannot increase under LOCC,

$$S(\rho_\psi) \geq S(\rho_\phi).$$

However, note that $n_\psi \geq n_\phi \not\Longrightarrow \lambda_\psi \prec \lambda_\phi$: to see this, consider

$$\lambda_\psi = (1-\epsilon, \epsilon/2, \epsilon/2, 0, \ldots, 0), \quad n_\psi = 3$$
$$\lambda_\phi = (1/2, 1/2, 0, \ldots, 0), \quad n_\phi = 2,$$

but $\lambda_\psi \not\prec \lambda_\phi$. For now, we'll set aside teleportation to discuss something a bit different.

Separability problem Consider a state

$$\rho \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B).$$

Is it separable or entangled?

Theorem 15.1. *Let $\rho \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$. ρ is separable iff $\forall$ positive (P) maps $\Lambda : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$,*

$$(\Lambda \otimes \mathrm{id}_B)(\rho) \geq 0. \tag{15.2}$$

Proof. $\Longrightarrow$: Suppose ρ is separable. Then ρ can be written

$$\rho = \sum p_i \omega_i^A \otimes \sigma_i^B, \tag{15.3}$$

so

$$(\Lambda \otimes \mathrm{id})\rho = \sum p_i \Lambda(\omega_i^A) \otimes \sigma_i^B \geq 0, \tag{15.4}$$

where this is a convex combination of positive semidefinite operators and is therefore overall positive.

Converse (Horodecki): If ρ is entangled, then $\exists$ some P map such that

$$(\Lambda \otimes \mathrm{id})\rho \not\geq 0. \tag{15.5}$$

In particular this map is not completely positive. For instance, take $\Lambda = T$ the transposition map. We consider the Peres-Horodecki PPT criterion (Positive Partial Transpose). That is, let

$$\rho = \sum p_{i\alpha,j\beta} |i\rangle\langle j| \otimes |\alpha\rangle\langle\beta| . \tag{15.6}$$

Thus the partial transpose is

$$\rho^{T_A} = (T \otimes \mathrm{id})\rho = \sum p_{i\alpha,j\beta} |j\rangle\langle i| \otimes |\alpha\rangle\langle\beta| . \tag{15.7}$$

We say that a density matrix ρ is PPT is $\rho^{T_A} \geq 0$.

The PPT criterion states that

(a) Separable $\implies$ PPT
(b) Separable $\iff$ PPT for $2 \times 2, 2 \times 3$ systems.

We'll show the first part, sep $\implies$ PPT: suppose ρ is separable. Then

$$\rho = \sum p_i \omega_i^A \otimes \sigma_i^B ,$$

so

$$\rho^{T_A} \equiv (T \otimes \mathrm{id}_B)\rho = \sum p_i (\omega_i^A)^T \otimes \sigma_i^B \geq 0 \tag{15.8}$$

since the transpose of a positve operator is still positive.

Now the second part, PPT $\implies$ separable for $2 \times 2, 2 \times 3, 3 \times 2$.

Lemma 15.9. *(Stormer, Wo) Any P map $\Lambda : \mathcal{M}_d \to \mathcal{M}_{d'}$ s.t. $dd' \leq 6$ has the form*

$$\Lambda = \Lambda_1 + \Lambda_2 T \quad \textit{where } \Lambda_1, \Lambda_2 \equiv CP. \tag{15.10}$$

Using this lemma, we see that

$$\begin{aligned}(\Lambda \otimes \mathrm{id})\rho &= ((\Lambda_1 + \Lambda_2 T) \otimes \mathrm{id})\rho \\ &= (\Lambda_1 \otimes \mathrm{id})\rho + (\Lambda_2 \otimes \mathrm{id}) \underbrace{(T \otimes \mathrm{id})\rho}_{\rho^{T_A} \geq 0} .\end{aligned}$$

But since we assume PPT, this partial transpose is positive, and Λ_1, Λ_2 are completely positive, so

$$(\Lambda \otimes \mathrm{id})\rho \geq 0. \tag{15.11}$$

⊠

Reduction map For $X \in \mathcal{B}(\mathcal{H}_A)$ the space of operators on A, we define the reduction map

$$\Lambda_R(X) := (\operatorname{Tr} X)I - X. \tag{15.12}$$

One may then check that

$$(\mathrm{id} \otimes \Lambda_R)(\rho) = \rho_A \otimes I - \rho \tag{15.13}$$

$$(\Lambda_R \otimes \mathrm{id})\rho = I \otimes \rho_B - \rho. \tag{15.14}$$

The *reduction criterion* is as follows, using Thm. 15.1:

$$\rho \text{ is sep} \iff \rho_A \otimes I - \rho \geq 0, I \otimes \rho_B - \rho \geq 0. \tag{15.15}$$

Separability criterion involving observables

Definition 15.16. An observable $W \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ (i.e. self-adjoint operator) is an *entanglement witness* (EW) if

$$\operatorname{Tr}(W\sigma) \geq 0 \quad \forall \sigma \text{ sep} \tag{15.17}$$

and $\exists$ at least one ρ entangled such that

$$\operatorname{Tr}(W\rho) < 0. \tag{15.18}$$

Therefore if W is an EW and ρ is a state s.t. $\operatorname{Tr}(W\rho) < 0$, then we infer that ρ is entangled. W is an EW which "detects" the entangled state ρ.

Theorem 15.19. *$\forall$ entangled state $\rho \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B), \exists$ an EW W.*

This follows from the Hahn-Banach theorem, a result in functional analysis.

Observe that

$$\mathrm{Tr}(W\rho) = (W, \rho)_{HS} = \mathrm{Tr}\, W^\dagger \rho \equiv \mathrm{Tr}\, W\rho, \tag{15.20}$$

since W is self-adjoint and HS indicates the Hilbert-Schmidt inner product.

Consider $\mathbf{w}, \mathbf{r}$ unit vectors in $\mathbb{R}^3$. The inner product is then $(\mathbf{w}, \mathbf{r}) = \cos\theta$, so there is in principle some plane P such that with $\mathbf{r} \in P, (\mathbf{w}, \mathbf{r}) = 0$, and this plane divides $\mathbb{R}^3$ into two regions where $(\mathbf{w}, \mathbf{r}) > 0$ on one side and < 0 on the other.

Theorem 15.21 (Hahn-Banach)**.** *If S_1, S_2 are two convex, closed sets in a real Banach space and one of them is compact, then $\exists$ a continuous linear functional ϕ and an $\alpha \in \mathbb{R}$ such that $\forall$ pairs $\omega_1 \in S_1, \omega_2 \in S_2$,*

$$\phi(\omega_1) < \alpha \leq \phi(\omega_2). \tag{15.22}$$

We claim that Thm. 15.19 follows directly from the Hahn-Banach theorem.

Consider the space

$$\mathcal{A}_{AB} \equiv \mathcal{B}_{sa}(\mathcal{H}_A \otimes \mathcal{H}_B) \tag{15.23}$$

equipped with the H-S inner product ("sa" indicates self-adjoint). Choose as S_1 the set $S_1 = \{\rho_0\}$ for ρ_0 entangled, and $S_2 = \{\sigma \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B) : \sigma \text{ sep}\}$. The first set is a singleton and therefore compact, while this second is closed and convex. Thus Hahn-Banach guarantees that $\exists \phi, \alpha$ such that

$$\phi(\rho) < \alpha \leq \phi(\sigma) \forall \sigma \in S_2. \tag{15.24}$$

Note that the Riesz Rep'n Thm states that we can write a general functional ϕ as $\phi(\rho) = \mathrm{Tr}(A\rho)$ for some $A \in \mathcal{B}_{sa}(\mathcal{H}_A \otimes \mathcal{H}_B)$. Thus WLOG we can write this as

$$\mathrm{Tr}(A\rho) < \alpha \leq \mathrm{Tr}(A\sigma), \tag{15.25}$$

and thus we take

$$W = A - \alpha I \tag{15.26}$$

so that

$$\mathrm{Tr}(W\rho) = \mathrm{Tr}(A\rho) - \alpha < 0, \tag{15.27}$$

$$\mathrm{Tr}(W\sigma) = \mathrm{Tr}(A\sigma) - \alpha \geq 0 \forall \sigma \text{ sep.} \quad \boxtimes \tag{15.28}$$

We also have notions of a "finer" entanglement witness, e.g. for two EWs, the entangled states detected by one EW are a strict subset of the EWs detected by the other.

Teleportation in 3 minutes Suppose Alice and Bob share $|\phi^+\rangle$, and Alice also has a qubit in state $|\psi_C\rangle$. Thus the whole system is in state

$$|\psi_C\rangle \otimes |\psi^+_{AB}\rangle. \tag{15.29}$$

WLOG we can write $|\psi_C\rangle = a|0\rangle + b|1\rangle$. Expanding out the tensor product state, we have

$$|\psi_C\rangle \otimes |\phi^+_{AB}\rangle = \frac{1}{2}\left[|\phi^+_{CA}\rangle(a|0_B\rangle + b|1_B\rangle) + |\phi^-\rangle(a|0\rangle - b|1\rangle) + |\psi^+\rangle(a|1\rangle + b|0\rangle) + |\psi^-\rangle(a|1\rangle - b|0\rangle)\right] \tag{15.30}$$

$$= \frac{1}{2}\left[|\phi^+_{CA}\rangle(|\psi_B\rangle) + |\phi^-\rangle(\sigma_z|\psi\rangle) + |\psi^+\rangle(\sigma_x|\psi\rangle) + |\psi^-\rangle(-i\sigma_y|\psi\rangle)\right]. \tag{15.31}$$

Alice makes a Bell measurement on her two qubits, measuring $00, 01, 10, 11$ and communicates the result to Bob, who can recover the state by a local unitary.

Lecture 16.

Monday, February 25, 2019

Distance measures Given two states ρ, σ, how well can we distinguish between them? We have a few measures to test this.

Trace distance We define the *trace distance* $D(\rho, \sigma)$ between two density matrices ρ, σ as follows:

$$D(\rho,\sigma) = \frac{1}{2}||\rho - \sigma||_1 \tag{16.1}$$

$$||A||_1 = \mathrm{Tr}|A| = \mathrm{Tr}\sqrt{A^\dagger A} \tag{16.2}$$

Thus define $A := \rho - \sigma$. We take A to be self-adjoint, $A^\dagger = A$. Then we can decompose A into its eigenvalues,

$$\begin{aligned} A &= \sum a_i \, |\phi_i\rangle\langle\phi_i| \\ &= \sum_{a_i \geq 0} a_i \phi_i + \sum_{a_i < 0} a_i \phi_i \\ &= Q - R \end{aligned}$$

where Q, R are now positive semi-definite. Thus by the linearity of the trace,

$$\begin{aligned} D(\rho,\sigma) &= \frac{1}{2}\,\mathrm{Tr}|A| \\ &= \frac{1}{2}\left[\sum_{a_i \geq 0} - \sum_{a_i \geq 0} a_i - \sum_{a_i < 0} a_i\right] \\ &= \frac{1}{2}(\mathrm{Tr}\, Q + \mathrm{Tr}\, R), . \end{aligned}$$

However, note that $A = \rho - \sigma = Q - R$. Since A is traceless,[23] it follows that $\mathrm{Tr}\, Q = \mathrm{Tr}\, R$, which implies that

$$D(\rho,\sigma) = \mathrm{Tr}\, Q = \mathrm{Tr}\, R. \tag{16.3}$$

Lemma 16.4.

$$D(\rho,\sigma) = \max \mathrm{Tr}(P(\rho - \sigma)), \quad 0 \leq P \leq I. \tag{16.5}$$

Here, we use $P \leq I$ to indicate that $X \equiv (I - P) \geq 0$ is positive semi-definite as an operator.

Proof. Since $D(\rho, \sigma) = \mathrm{Tr}\, Q$, let P be the projector onto the support of Q. Then

$$\begin{aligned} \mathrm{Tr}(P(\rho - \sigma)) &= \mathrm{Tr}(P(Q - R)) \\ &= \mathrm{Tr}\, PQ - \mathrm{Tr}\, PR \\ &= \mathrm{Tr}\, Q \\ &= D(\rho,\sigma), \end{aligned}$$

since the supports of Q and R are orthogonal. Conversely, $\forall 0 \leq P \leq I$, we have

$$\begin{aligned} \mathrm{Tr}\, P(\rho - \sigma) &= \mathrm{Tr}(P(Q - R)) \\ &\leq \mathrm{Tr}\, PQ \\ &\leq \mathrm{Tr}\, Q \\ &= D(\rho,\sigma) \end{aligned}$$

since $P \leq I$.[24] Combining these, we see that

$$D(\rho,\sigma) = \max_{0 \leq P \leq I} \mathrm{Tr}\, P(\rho - \sigma).$$

⊠

This trace distance has some nice properties. It forms a metric on $\mathcal{D}(\mathcal{H})$, as it is

- Symmetric, $D(\rho,\sigma) = D(\sigma,\rho)$[25]

[23] This follows since ρ and σ are density matrices and have trace 1, so by the linearity of the trace, $\mathrm{Tr}(A) = \mathrm{Tr}(\rho) - \mathrm{Tr}(\sigma) = 1 - 1 = 0$.

[24] The second line follows since R is positive semi-definite, and PR is also positive semi-definite. The third line follows because $\mathrm{Tr}\, PQ = \mathrm{Tr}\, PQ^{1/2}Q^{1/2} = \mathrm{Tr}\, Q^{1/2}PQ^{1/2} \leq \mathrm{Tr}\, Q$.

[25] Follows since $D(\rho,\sigma) = \frac{1}{2}||A||_1$, and $||A||_1 = \mathrm{Tr}\sqrt{A^\dagger A} = \mathrm{Tr}\sqrt{(-A)^\dagger(-A)} = || - A||_1$, as it should be. Thus $D(\rho,\sigma) = \frac{1}{2}||A||_1 = \frac{1}{2}|| - A||_1 = D(\sigma,\rho)$.

- $D(\rho,\sigma) = 0$ iff $\rho = \sigma$
- Triangle inequality, $D(\rho,\omega) \leq D9\rho,\sigma) + D(\sigma,\omega)$.

The second property follows from $D(\rho,\sigma) = \operatorname{Tr} Q = \operatorname{Tr} R = 0 \implies a_i = 0 \forall i \implies A = 0 \implies \rho = \sigma$. The final follows from using Lemma 16.4 and noting that

$$\begin{aligned} D(\rho,\sigma) &= \operatorname{Tr}(P(\rho-\sigma)) \\ &= \operatorname{Tr} P(\rho-\omega) + \operatorname{Tr} P(\omega-\sigma) \\ &\leq D(\rho,\omega) + D(\omega,\sigma). \end{aligned}$$

Lemma 16.6. *Monotonicity under quantum operations Λ:*

$$D(\Lambda(\rho),\Lambda(\sigma)) \leq D(\rho,\sigma). \tag{16.7}$$

Proof. We know that

$$D(\Lambda(\rho),\Lambda(\sigma)) = \operatorname{Tr}(P(\lambda(\rho) - \Lambda(\sigma)),$$

where equality is reached for some P. We saw that $\operatorname{Tr} Q = \operatorname{Tr} R \implies \operatorname{Tr}(\Lambda(Q)) = \operatorname{Tr}(\Lambda(R))$ since Λ is CPTP. Now

$$\begin{aligned} D(\rho,\sigma) = \operatorname{Tr} Q = \operatorname{Tr} \lambda(Q) &\geq \operatorname{Tr}(P\Lambda(Q)) \\ &\geq \operatorname{Tr}(P(\Lambda(Q) - \Lambda(R))) \\ &= D(\Lambda(\rho,\Lambda(\sigma)), \end{aligned}$$

using the fact that $0 \leq P \leq I$. ☒

Fidelity Let us define

$$F(\rho,\sigma) = \operatorname{Tr}\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}} \tag{16.8}$$

$$= ||\sqrt{\sigma}\sqrt{\rho}||_1. \tag{16.9}$$

This quantity $F(\rho,\sigma)$, called the fidelity, has some nice properties.

- $F(\rho,\sigma) = F(\sigma,\rho)$.
- $0 \leq F(\rho,\sigma) \leq 1$ with equality on the right if $\rho = \sigma$.

Case 1: Note that if $[\rho,\sigma] = 0$, then ρ and σ admit a simultaneous eigenbasis,

$$\rho = \sum_i \lambda_i |e_i\rangle\langle e_i|, \quad \sigma = \sum_i \mu_i |e_i\rangle\langle e_i|.$$

Thus since $\sqrt{|\psi\rangle\langle\psi|} = |\psi\rangle\langle\psi|$,

$$F(\rho,\sigma) = \operatorname{Tr}(\sum_i \sqrt{\lambda_i\mu_i}\,|e_i\rangle\langle e_i| \tag{16.10}$$

$$= \sum_i \sqrt{\lambda_i\mu_i} \equiv F_{\text{cl}}(\lambda,\mu), \tag{16.11}$$

where F_{cl} is now the classical fidelity and we've moved the trace inside the sum by linearity.

Case 2: $\rho = |\phi\rangle\langle\phi|, \sigma$. Exercise: show that

$$F(|\phi\rangle\langle\phi|,\sigma) = \sqrt{\langle\phi|\sigma|\phi\rangle}. \tag{16.12}$$

Case 3: $\rho = |\phi\rangle\langle\phi|, \sigma = |\psi\rangle\langle\psi|$.

$$F(\rho,\sigma) = \sqrt{|\langle\phi|\psi\rangle|^2} = |\langle\phi|\psi\rangle|. \tag{16.13}$$

For instance, this tells us that $F(U\rho U^\dagger, U\sigma U^\dagger) = F(\rho,\sigma)$.

Theorem 16.14 (Uhlmann)**.** *Let $\rho_A, \sigma_A \in \mathcal{D}(\mathcal{H}_A)$. Then the fidelity $F(\rho_A,\sigma_A)$ is*

$$F(\rho_A,\sigma_A) = \max_{|\psi_\rho^{AB}\rangle, |\psi_\sigma^{AB}\rangle} |\langle\psi_\rho^{AB}|\psi_\sigma^{AB}\rangle| \tag{16.15}$$

where these $|\psi_{\rho/\sigma}^{AB}\rangle$ are purifications of the original density matrices.

Proof. We will need the following lemma:

Lemma 16.16. $||A||_1 = \max_{U\ unitary} |\mathrm{Tr}(UA)| \quad \forall A \in \mathcal{B}(\mathcal{H})$.

Proof. We use the polar decomposition, so $\exists V$ a unitary such that $A = |A|V$. If we now choose $U = V^\dagger$, then the inequality will be saturated:

$$|\mathrm{Tr}(UA)| = \mathrm{Tr}(V^\dagger |A| V) = \mathrm{Tr}(|A|).$$

Hence $\forall U$ unitary,

$$\begin{aligned} |\,\mathrm{Tr}(AU)| &= |\,\mathrm{Tr}(|A|VU)| \\ &\leq \sqrt{(\mathrm{Tr}\,|A|)\,\mathrm{Tr}(U^\dagger V^\dagger |A| VU)} \\ &= \mathrm{Tr}\,|A| \equiv ||A||_1 \end{aligned}$$

where we recognize this trace in the first line as a Hilbert-Schmidt inner product $(x,y)_{HS}$ with $x = |A|^{1/2}, y = |A|^{1/2}VU$ and apply Cauchy-Schwartz. In the last line, we use the cyclic property of the trace to get $\sqrt{(\mathrm{Tr}\,|A|)^2} = \mathrm{Tr}\,|A|$.

This proves the lemma. ☒

Now we can use the lemma to prove the theorem. We want to prove that

$$F(\rho_A, \sigma_A) = \max_{|\psi_\rho^{RA}\rangle, |\psi_\sigma^{RA}\rangle} |\langle \psi_\rho^{RA} | \psi_\sigma^{RA} \rangle|,$$

where RA indicates the purified system.

Recall that there is a unitary freedom of purifications– any two purifications of the same ρ are related by a unitary transformation acting only on the reference system. Equivalently, we can maximize over unitaries. If we fix $|\psi_\sigma\rangle_{RA}$, then

$$F(\rho_A, \sigma_A) = \max_{U^\rho, U^\sigma} |\langle \psi_\rho | (U_\rho^{R\dagger} \otimes I_A)(U_\sigma^{R\dagger} \otimes I_A) | \psi_\sigma \rangle.| \tag{16.17}$$

Since the set of unitaries is closed, take $U \equiv U_\rho^{R\dagger} U_\sigma^R$, and we maximize over U:

$$\max_U \langle \psi_\rho | U \otimes I | \psi_\sigma \rangle.$$

Now we know that we can generically write purified states as

$$|\psi_\sigma^{RA}\rangle = \sqrt{d}(I_R \otimes \sqrt{\sigma_A} |\Omega\rangle, \tag{16.18}$$

and so

$$\begin{aligned} \max_U \langle \psi_\rho | U \otimes I | \psi_\sigma \rangle &= d \max_U |\langle \Omega | (I_R \otimes \sqrt{\rho_A})(U \otimes I)(I_R \otimes \sqrt{\sigma_A}) |\Omega\rangle| \\ &= d \max_U |\langle \Omega | (I_R \otimes \sqrt{\rho_A}\sqrt{\sigma_A})(U \otimes I) |\Omega\rangle| \\ &= d \max_U |\langle \Omega | (I_R \otimes \sqrt{\rho_A}\sqrt{\sigma_A} U^T |\Omega\rangle|. \end{aligned}$$

Hence this becomes

$$\begin{aligned} \max_U \langle \psi_\rho | U \otimes I | \psi_\sigma \rangle &= \max_U |\sum_i \langle i | \sqrt{\rho_A}\sqrt{\sigma_A} U^T | i \rangle| \\ &= \max_U |\mathrm{Tr}(\sqrt{\rho_A}\sqrt{\sigma_A} U^T)| \\ &= \max_V |\mathrm{Tr}(\sqrt{\rho_A}\sqrt{\sigma_A} V| = ||\sqrt{\rho_A}\sqrt{\sigma_A}||_1 = F(\rho, \sigma). \end{aligned}$$

☒

Lecture 17.

Wednesday, February 27, 2019

Last time, we discussed the trace distance and the fidelity, two measures of distance between states. We also stated Uhlmann's theorem, which says that the fidelity of two states ρ, σ is related to the maximal overlap between two purifications $|\psi_\rho^{AB}\rangle, |\sigma_\rho^{AB}\rangle$.

Some properties follow from the theorem.

(a) $0 \leq F(\rho, \sigma \leq 1$ and $F(\rho, \sigma) = 1$ iff $\rho = \sigma$. The first follows since

$$0 \leq |\langle \psi_\rho^{AR} | \psi_\sigma^{AR} \rangle| \leq 1,$$

and the latter since $\rho = \sigma \iff F(\rho, \sigma) = 1$

(b) $F(\rho, \sigma) = F(\sigma, \rho)$, which is clear from the inner product definition.

Lemma 17.1 (Monotonicity under partial trace). *For $\rho_{AB}, \sigma_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$,*

$$F(\rho_{AB}, \sigma_{AB}) \leq F(\rho_A, \sigma_A). \tag{17.2}$$

Proof. We use Uhlmannn's theorem. Thus $\exists |\psi_\rho^{ABC}\rangle, |\psi_\sigma^{ABC}\rangle$ which are purifications of ρ_{AB}, σ_{AB} s.t.

$$F(\rho_{AB}, \sigma_{AB}) = \langle \psi_\rho^{ABC} | \psi_\sigma^{ABC} \rangle. \tag{17.3}$$

But $|\psi_\rho^{ABC}\rangle, |\psi_\sigma^{ABC}\rangle$ are also purifications of ρ_A, σ_A, and so for this particular purification, the fidelity must be $\geq \langle \psi_\rho^{ABC} | \psi_\sigma^{ABC} \rangle$. Thus

$$F(\rho_{AB}, \sigma_{AB}) \leq F(\rho_A, \sigma_A). \tag{17.4}$$

☒

We now state the Fuchs-van de Graaf inequality:

$$1 - F(\rho, \sigma) \leq D(\rho, \sigma) \leq \sqrt{1 - F^2(\rho, \sigma)} \tag{17.5}$$

Operational interpretation of $D(\rho, \sigma)$ We now discuss (binary) quantum hypothesis testing. Suppose Alice prepares either state ρ_0 or ρ_1 with probability $1/2$. She then sends her state to Bob, who wants to distinguish the states. Bob makes a measurement– he constructs a (binary) POVM $\{E_0, E_1 = I - E_0\}$.

What is Bob's probability of error? It is

$$\begin{aligned} p_{\text{err}}(\{E_0, E_1\}) &= \frac{1}{2}[\text{Tr}(E_0 \rho_1) + \text{Tr}(E_1 \rho_0)] && (17.6) \\ &= \frac{1}{2}[1 - \text{Tr}(E_0(\rho_0 - \rho_1))], && (17.7) \end{aligned}$$

where we've taken the probability of ρ being a given state (1/2) and multiplied by the likelihood of measuring 0 when the state was ρ_1 ($\text{Tr}(E_0 \rho_1)$) and the same for measuring 1 when the state was ρ_0. In the second line, we just rewrote $E_1 = I - E_0$. Now the minimum error probability is

$$\begin{aligned} p_{\text{err}}^* &= \min_{0 \leq E_0 \leq I} p_{\text{err}}(E_0 && (17.8) \\ &= \frac{1}{2}\left[1 - \max_{0 \leq E_0 \leq I} \text{Tr}(E_0(\rho_0 - \rho_1))\right] && (17.9) \\ &= \frac{1}{2}[1 - D(\rho_0, \rho_1)]. && (17.10) \end{aligned}$$

Thus the maximum success probability is

$$p_{\text{suc}}^* = 1 - p_{\text{err}}^* = \frac{1}{2}[1 + D(\rho_0, \rho_1)]. \tag{17.11}$$

One can show that p_{suc}^* is achieved via a projective measurement, $E_0 = P_0, E_1 = P_1$. Recall that we could write $\rho_0 - \rho_1 = Q - R$ in terms of positive and negative eigenvalues. Here, P_0, P_1 are the projections onto the supports of Q and R, respectively. The proof of this is due to Holevo and Helstrom.

Quantum entropy The notion of entropy for quantum systems is also called the von Neumann entropy.

Definition 17.12. For a state $\rho \in \mathcal{D}(\mathcal{H})$, the *von Neumann entropy* is

$$S(\rho) = -\operatorname{Tr}(\rho \log \rho). \tag{17.13}$$

Here, the log is base 2 (as is typical in information theory).

Since ρ admits a spectral decomposition, $\rho = \sum \lambda_i |e_i\rangle\langle e_i|$, the von Neumann entropy reduces to

$$S(\rho) = -\sum \lambda_i \log \lambda_i = H(\{\lambda_i\}), \tag{17.14}$$

the Shannon entropy of the set of eigenvalues.

The von Neumann entropy has the following properties:

(a) $S(\rho) \geq 0$ with $S(\rho) = 0 \iff \rho$ is pure (i.e. ρ has one non-zero eigenvalue).
(b) $S(U^\dagger \rho U) = S(\rho) \forall U$ unitary (since the eigenvalues are not changed under a unitary).
(c) $S(\rho) \leq \log d$ with equality when $\rho = I/d$.

To prove this last property, let us define a parent quantity, the quantum relative entropy $D(\rho||\sigma)$. Let $\rho \in \mathcal{D}(\mathcal{H}), \sigma \geq 0$.

Definition 17.15. The *quantum relative entropy* is the quantity

$$D(\rho||\sigma) := \operatorname{Tr}(\rho \log \rho - \rho \log \sigma), \tag{17.16}$$

which is well-defined if $\operatorname{supp}\rho \subseteq \operatorname{supp}\sigma$.

This is the quantum analogue of the KL divergence (classical relative entropy), defined $D(p||q) = \sum_i p_i \log \frac{p_i}{q_i}$. Moreover, we see that the quantum relative entropy is indeed a parent quantity for the von Neumann entropy:

$$S(\rho) = -D(\rho||I) \tag{17.17}$$

since $\rho \log \sigma|_{\sigma=I} = 0$.

A useful property is the *Klein's inequality*,

$$D(\rho||\sigma) \geq 0. \tag{17.18}$$

Proof. If $\rho = \sum \lambda_i |i\rangle\langle i|$, $\sigma = \sum a_\alpha |\alpha\rangle\langle\alpha|$, then

$$\begin{aligned} D(\rho||\sigma) &= \sum \lambda_i \log \lambda_i - \operatorname{Tr}\left((\sum_i \lambda_i |i\rangle\langle i|)(\sum_\alpha \log(a_\alpha) |\alpha\rangle\langle\alpha| \right) \\ &= \sum_i \lambda_i \log \lambda_i - \sum_{i,\alpha} \lambda_i \log a_\alpha |\langle i|\alpha\rangle|^2. \end{aligned}$$

But observe that for $p_{i\alpha} = |\langle i|\alpha\rangle|^2$, we have $\sum_i p_{i\alpha} = \sum_i |\langle i|\alpha\rangle|^2 = \sum_i \langle\alpha|i\rangle\langle i|\alpha\rangle = 1$, and the same is true if we sum over α. This tells us that $p_{i\alpha}$ are the elements of a doubly stochastic matrix. Thus

$$D(\rho||\sigma) = \sum \lambda_i \log \lambda_i - \sum_{i,\alpha} \lambda_i \log a_\alpha p_{i\alpha}. \tag{17.19}$$

Note that $f(x) = \log x$ is a concave function, so treating the $p_{i\alpha}$s as a probability, we have

$$\sum_\alpha p_{i\alpha} \log a_\alpha \leq \log(\sum_\alpha p_{i\alpha} q_\alpha), \tag{17.20}$$

which tellls us that

$$D(\rho||\sigma) \geq \sum \lambda_i \log \lambda_i - \sum_i \lambda_i \log(\sum_\alpha p_{i\alpha} a_\alpha). \tag{17.21}$$

Defining $r_i := \sum_\alpha p_{i\alpha} a_\alpha$ where $r_i \geq 0$ and $\sum_i r_i = 1$, we find that

$$\begin{aligned} D(\rho||\sigma) &\geq \sum \lambda_i (\log \lambda_i - \log r_i) \\ &= D_{KL}(\lambda||r) \\ &\geq 0, \end{aligned}$$

where we recognize the classical relative entropy on $\lambda = \{\lambda_i\}, r = \{r_i\}$. ☒

We can also prove the upper bound on $S(\rho) \leq \log d$. Take $\sigma = I/d$, and then

$$\begin{aligned} 0 \leq D(\rho||\sigma) &= \operatorname{Tr}(\rho \log \rho - \rho \log I/d) \\ &= -S(\rho) - \rho \operatorname{diag}(\log 1/d, \ldots, \log 1/d) \\ &= -S(\rho) - \log 1/d \operatorname{Tr} \rho \\ &= -S(\rho) + \log d \operatorname{Tr} \rho. \end{aligned}$$

Thus $S(\rho) \leq \log d$.

Note that $S(\rho)$ is concave,

$$S(\sum p_i \rho_i) \geq \sum p_i S(\rho_i). \tag{17.22}$$

One can prove this by taking $f(x) = -x \log x$ and considering $\rho_i, \bar{\rho} = \sum p_i \rho_i$.

Composite systems There are some notions of entropy for composite quantum systems, $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$.

(a) The *joint entropy* is

$$S(\rho_{AB}) = -\operatorname{Tr} \rho_{AB} \log \rho_{AB} \tag{17.23}$$

.

(b) The *quantum conditional entropy* is

$$S(A|B)_\rho = S(AB) - S(B) = S(\rho_{AB}) - S(\rho_B) \tag{17.24}$$

(c) The *quantum mutual information* is

$$\begin{aligned} I(A:B) &= S(A) + S(B) - S(AB) \\ &= S(A) - S(A|B) \\ &= S(B) - S(B|A) \\ &= I(B:A). \end{aligned}$$

Note that this last quantity does not satisfy some properties of the classical mutual information. Classically, for X, Y random variables,

$$H(X) \leq H(XY) \implies H(XY) - H(X) \geq 0, \tag{17.25}$$

and this latter expression is just a conditional entropy. Thus $H(Y|X) \geq 0$. But in the quantum case, $S(A|B)$ need not be ≥ 0. To see this, let us take $\rho_{AB} = |\phi^+\rangle\langle\phi^+|$. Thus

$$\begin{aligned} S(A|B) &= S(AB) - S(A) \\ &= -\log d < 0 \end{aligned}$$

since $S(AB) = 0$ and A is a completely mixed state.

Naturally, the von Neumann entropy has a nice additive structure under tensor products

$$S(\rho \otimes \sigma) = S(\rho) + S(\sigma), \tag{17.26}$$

and additionally,

$$S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B), \tag{17.27}$$

which can be proved using $D(\rho||\sigma) \geq 0$.

Lecture 18.

Friday, March 1, 2019

Last time, we introduced some ideas of entropy in quantum systems. We stated Klein's inequality,

$$D(\rho||\sigma) \geq 0 \forall \rho, \sigma \in \mathcal{D}(\mathcal{H}). \tag{18.1}$$

We also defined the von Neumann entropy, $S(\rho)$. This quantitiy has some nice properties:

(a) Composite states $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ also obey the property of subadditivity,

$$S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B), \tag{18.2}$$

which follows from the positivity of the mutual information,

$$I(A:B) = S(\rho_A) + S(\rho_B) - S(\rho_{AB}) \geq 0. \tag{18.3}$$

Thus $I(A:B) = D(\rho_{AB}||\rho_A \otimes \rho_B) \geq 0$.

(b) Note that if $\rho_{AB} = |\psi_{AB}\rangle\langle\psi_{AB}|$ is a pure state, then $S(\rho_A) = S(\rho_B)$ by the Schmidt decomposition (the reduced states share the same non-zero eigenvalues, and the von Neumann entropy depends only on the eigenvalues).

(c) Triangle inequality/Araki-Lieb inequality: $S(\rho_{AB}) \geq |S(\rho_A) - S(\rho_B)|$.

Suppose we have a purification $\rho_{AB} \to |\psi_{ABR}\rangle$. By (a), we have $S(A,R) \leq S(A) + S(R)$. But by (b) we also have $S(A,R) = S(B)$ and $S(A,B) = S(R)$. Substituting in, we have

$$S(B) \leq S(A) + S(A,B) \implies S(A,B) \geq S(B) - S(A) \tag{18.4}$$

and similarly

$$S(A,B) \geq S(A) - S(B) \implies S(A,B) \geq |S(A) - S(B)|. \tag{18.5}$$

(d) If $\rho = \sum p_I \rho_i$ where the ρ_i have mutually orthogonal supports, then (proof in example sheet 3)

$$S(\sum_i p_i \rho_i) = H(\{p_i\}) + \sum p_i S(\rho_i) \tag{18.6}$$

The von Neumann entropy also obeys the property of *strong subadditivity (SSA)*. The original proof is due to Lieb and Ruskai (1973). Suppose we have a tripartite system ρ_{ABC}. Then

$$S(\rho_{ABC}) + S(\rho_B) \leq S(\rho_{AB}) + S(\rho_{BC}). \tag{18.7}$$

This property has some interesting consequences.

(a) Conditioning reduces entropy,

$$S(A|BC) \leq S(A|B). \tag{18.8}$$

This is immediate– since $S(A|BC) = S(ABC) - S(BC)$, just move the terms around to get $S(A|BC) = S(ABC) - S(BC) \leq S(AB) - S(B) = S(A|B)$.

(b) Discarding quantum systems never increases mutual information, i.e.

$$I(A:B) \leq I(A:BC). \tag{18.9}$$

Proof: just add $S(A)$ to both sides and rearrange to get

$$I(A:B) = S(A) + S(B) - S(AB) \leq S(BC) + S(A) - S(ABC) = I(A:BC). \tag{18.10}$$

(c) Quantum operations never increase mutual information. That is, we have a bipartite system AB and we perform a CPTP map Λ on the B part. Thus with $\rho_{A'B'} = (\mathrm{id}_A \otimes \Lambda)\rho_{AB}$,

$$I(A':B') \leq I(A:B). \tag{18.11}$$

Proof. We use Stinespring. That is, to implement the operation Λ we introduce the ancilla $\mathcal{H}_C$ with some reference state $\phi \in \mathcal{H}_C$ and a unitary U_{BC} such that

$$\mathrm{Tr}_C(U_{BC}(\rho_B \otimes \phi)U_{BC}^\dagger) = \Lambda(\rho_B) \equiv \rho_{B'}. \tag{18.12}$$

Then we can prove that

$$I(A:B) = \underbrace{I(A:BC)}_{\rho_{AB}\otimes\phi_C}, \tag{18.13}$$

since C is uncorrelated with A, and we can rewrite this mutual information as

$$D(\rho_{AB}||\rho_A \otimes \rho_{BC}) = D(\rho_{AB} \otimes \phi_C||\rho_A \otimes \rho_B \otimes \phi_C) \tag{18.14}$$

One may check the following properties:

- ◦ $D(\rho \otimes \omega||\sigma \otimes \nu) = D(\rho||\sigma) + D(\omega||\nu)$
- ◦ $\forall U$ unitary, $D(U\rho U^\dagger||U\sigma U^\dagger) = D(\rho||\sigma)$.
- ◦ Joint convexity:

$$D(\sum p_i \rho_i|| \sum p_i \sigma_i) \leq \sum p_i D(\rho_i||\sigma_i). \tag{18.15}$$

With the first property, we can rewrite this as

$$D(\rho_{AB}||\rho_A \otimes \rho_B) + D(\phi_C||\phi_C) = I(A:B). \tag{18.16}$$

Now using the second properties since ρ_{ABC} and $\rho_{A'B'C'}$ are related by a unitary transformation,

$$\rho_{A'B'C'} = (\mathrm{id}_A \otimes U_{BC})(\rho_{ABC})(\mathrm{id}_A \otimes U^\dagger_{BC}), \tag{18.17}$$

We can just trace over C to complete the proof. ☒

Let us now consider the quantum relative entropy and the data processing inequality: for Λ a quantum operation,

$$D(\rho||\sigma) \geq D(\Lambda(\rho)||\Lambda(\sigma)). \tag{18.18}$$

Consider a qudit, $\mathcal{H} \simeq \mathbb{C}^d$ with some basis $\{|j\rangle\}_{j=0}^{d-1}$. There are generalizations of the Pauli matrices σ_x, σ_Z– call them X, Z such that

$$X^k|j\rangle = |j \oplus k\rangle, \tag{18.19}$$

$$Z^m|j\rangle = e^{2\pi i m j/d}|j\rangle \tag{18.20}$$

where $\oplus$ indicates addition $\mod d$, with $k, m \in \{0, 1, \ldots, d-1\}$. So for instance with $k = 1, d = 2$ we have

$$X|0\rangle = |1\rangle; \quad X|1\rangle = 0$$

$$Z|0\rangle = |0\rangle; \quad Z|1\rangle = -|1\rangle.$$

Thus on qubits these operators reduce to the old σ_x, σ_z.

Let us introduce some combination unitary operators

$$W_{k,m} = X^k Z^m \in \mathcal{B}(\mathbb{C}^d). \tag{18.21}$$

There are d^2 such operators, called *Heisenberg-Weyl operators*. For $A \in \mathcal{B}(\mathbb{C}^d)$, we have as an exercise the following proof:

$$\frac{1}{d^2}\sum W_{k,m} A W^\dagger_{k,m} = (\mathrm{Tr}\, A)\tau \tag{18.22}$$

where $\tau \equiv I/d$ is the completely mixed state.

Now to prove the DPI, note first that

$$D(\Lambda(\rho)||\Lambda(\sigma)) = D(\Lambda(\rho) \otimes \tau||\Lambda(\sigma) \otimes \tau) \tag{18.23}$$

where τ is as above. We can certainly couple an unrelated system. But now using Stinespring, we can implement Λ as

$$\Lambda(\rho) = \mathrm{Tr}_2\, U(\rho \otimes \phi)U^\dagger, \tag{18.24}$$

so the LHS of 18.22 can be rewritten

$$\begin{aligned}\frac{1}{d^2}\sum(I \otimes W_{k,m})U(\rho \otimes \phi)U^\dagger(I \otimes W^\dagger_{k,m}) &= (\mathrm{Tr}_2\, U(\rho \otimes \phi)U^\dagger) \otimes \tau \\ &= \Lambda(\rho) \otimes \tau.\end{aligned}$$

Defining $I \otimes W_{km} \equiv \tilde{W}_{km}$, we have an expression for $\Lambda(\rho) \otimes \tau$. Thus

$$D(\Lambda(\rho)||\Lambda(\sigma)) = D(\Lambda(\rho) \otimes \tau||\Lambda(\sigma) \otimes \tau) \tag{18.25}$$

$$= D(\frac{1}{d^2}\sum_{k,m}\tilde{W}_{km}U(\rho \otimes \phi)U^\dagger\tilde{W}^\dagger_{km}||\frac{1}{d^2}\sum -k, m\tilde{W}_{km}U(\sigma \otimes \phi)U^\dagger\tilde{W}^\dagger_{km}) \tag{18.26}$$

$$= D(\frac{1}{d^2}\sum_{km}V_{km}(\rho \otimes \phi)V^\dagger_{km}||\frac{1}{d^2}\sum_{km}V_{km}(\sigma \otimes \phi)V^\dagger_{km} \tag{18.27}$$

where we've defined $\tilde{W}_{km}U \equiv V_{km}$ a unitary. We use joint convexity to turn this into

$$\begin{aligned} D(\Lambda(\rho)||\Lambda(\sigma)) &\leq \frac{1}{d^2}\sum_{km} D(V_{km}(\rho\otimes\phi)V_{km}^\dagger || V_{km}(\sigma\otimes\phi)V_{km}^\dagger) \\ &= \frac{1}{d^2}\sum_{k,m} D(\rho\otimes\phi||\sigma\otimes\phi) \\ &= \frac{1}{d^2}\sum_{k,m} D(\rho||\sigma) \\ &\implies D(\Lambda(\rho)||\Lambda(\sigma)) \leq D(\rho||\sigma) \quad \boxtimes \end{aligned}$$

We haven't proved the joint convexity, but it is implied by Lieb's concavity theorem– let X be a matrix and $0 \leq t \leq 1$ such that

$$f(A,B) := \mathrm{Tr}(X^\dagger A^t X B^{1-t}) \tag{18.28}$$

is jointly concave in A, B. Then

$$f(\sum p_i A_i, \sum_i p_i B_i) \geq \sum p_i f(A_i, B_i) \tag{18.29}$$

Lecture 19.

Monday, March 4, 2019

Today we will discuss quantum data compression. Let Q be a quantum information source, e.g. a highly attenuated laser emitting single monochromatic photons. Hence the source produces some signals $|\Psi_k\rangle$ with probability p_k, and these signals lie in a Hilbert space $\mathcal{H}$ with dimension $d = \dim\mathcal{H}$. Then we assign a density matrix

$$\rho = \sum p_k |\Psi_k\rangle\langle\Psi_k| \tag{19.1}$$

to our source. Note that the outputs need not be orthogonal: $\langle\Psi_j|\Psi_k\rangle \neq \delta_{ij}$.

Just like in the classical case, we will be interested in the asymptotic limit, i.e. for n copies/uses of the source, we can produce a string of outputs

$$|\Psi_k^{(n)}\rangle \in \mathcal{H}^{\otimes n} \text{ with probability } p_k^{(n)} \tag{19.2}$$

where

$$\rho^{(n)} = \sum p_k^{(n)} \left|\Psi_k^{(n)}\right\rangle\left\langle\Psi_k^{(n)}\right|. \tag{19.3}$$

Thus our source is described by $\{\rho^{(n)}, \mathcal{H}^{\otimes n}\}$ (or more formally, $\{|\Psi_k^{(n)}\rangle, p_k^{(n)}, \mathcal{H}^{\otimes n}\}$.

Now let us define a data compression map

$$\mathcal{C}^{(n)} : \left|\Psi_k^{(n)}\right\rangle\left\langle\Psi_k^{(n)}\right| \mapsto \tilde{\rho}_k^{(n)} \in \mathcal{D}(\tilde{\mathcal{H}}_n), \tag{19.4}$$

where $\mathcal{H}_n$ is some new Hilbert space. For this to be a compression map, we must have the dimension of the target space be smaller than the dimension of the source,

$$d_c^{(n)} = \dim\tilde{\mathcal{H}}_n < d^n. \tag{19.5}$$

Let us also suppose there exists a decompression map

$$\mathcal{D}^{(n)} : \mathcal{D}(\tilde{\mathcal{H}}_n) \to \mathcal{D}(\mathcal{H}^{\otimes n}). \tag{19.6}$$

In particular, these maps must be quantum operations and hence linear CPTP maps. The *rate* of this compression-decompression scheme is

$$R = \frac{\log d_c^{(n)}}{n}. \tag{19.7}$$

This is simply the number of qubits in the compressed version divided by the number of uses of the source. We can also invert the relationship and write

$$\dim\tilde{H}_n \equiv d_c^{(n)} = 2^{nR}. \tag{19.8}$$

With our source and compression/decompression maps in hand, what does it mean to say that such a scheme is reliable? Quantum signals are not completely distinguishable, unlike the classical case. However, we can use the fidelity instead. We have the following criterion, the *ensemble average fidelity*, and say that the scheme is reliable if

$$\lim_{n\to\infty} \bar{F}_n = 1, \tag{19.9}$$

where this ensemble average fidelity is defined by

$$\bar{F}_n := \sum_k p_k^{(n)} \langle \Psi_k^{(n)} | \mathcal{D}^{(n)}(\tilde{\rho}_k^{(n)} | \Psi^{(}_{k}n)\rangle, \tag{19.10}$$

such that $0 \leq \bar{F}_n \leq 1$ with $\bar{F}_n = 1$ iff $\mathcal{D}^{(n)}(\tilde{\rho}_k^{(n)} = \left|\Psi_k^{(n)}\right\rangle\!\left\langle\Psi_k^{(n)}\right|$.

The original proof about quantum data compression was due to Schumacher, who worked with a memoryless (iid) source. this proof relies on the notion of a *typical subspace*. Where in classical information theory, a sequence was either in the typical set or not, in a quantum system this is weakened slightly. What we can say instead is that an output sequence $|\Psi_k^{(n)}\rangle$ has a large component in the typical subspace $J_\epsilon^{(n)} \subset \mathcal{H}^{\otimes n}$.

Consider a source output $\rho^{(n)} \in \mathcal{D}(\mathcal{H}^{\otimes n}$ described by an iid source. Thus

$$\rho^{(n)} = \pi^{\otimes n}, \quad \pi \in \mathcal{D}(\mathcal{H}) \tag{19.11}$$

where

$$\pi = \sum q_i \left|\phi_i\right\rangle\!\left\langle\phi_i\right|. \tag{19.12}$$

In particular, we can then write

$$\rho^{(n)} = \sum_{\mathbf{i}} \lambda_{\mathbf{i}}^{(n)} \left|\chi_{\mathbf{i}}^{(n)}\right\rangle\!\left\langle\chi_{\mathbf{i}}^{(n)}\right| \tag{19.13}$$

where

$$\lambda_{\mathbf{i}}^{(n)} = q_{i_1,\ldots,i_n}, \quad |\chi_{\mathbf{i}}^{(n)}\rangle = |\phi_{i_i}\rangle \otimes \ldots \otimes |\phi_{i_n}\rangle. \tag{19.14}$$

That is, we label the eigenvalues and eigenvectors of $\rho^{(n)}$ in terms of sequences $\mathbf{i} = (i_1, \ldots, i_n)$.

In the classical case, the data compression limit was given by the Shannon entropy. In the quantum case, we will need the von Neumann entropy:

$$S(\rho^{(n)}) = S(\pi^{\otimes n}) = nS(\pi), \tag{19.15}$$

since the von Neumann entropy adds under tensor products. We now define the set of ϵ-typical sequences $T_\epsilon^{(n)}$ as sequences $\mathbf{i}$ with probability $\lambda_{\mathbf{i}}^{(n)} = q_{i_1} \ldots q_{i_n}$ such that

$$2^{-n(S(\pi)+\epsilon)} \leq \lambda_{\mathbf{i}}^{(n)} \leq 2^{-n(S(\pi)-\epsilon)}, \tag{19.16}$$

where S is now the von Neumann entropy.

We can now define the typical subspace $\mathcal{T}_\epsilon^{(n)} \subset \mathcal{H}^{\otimes n}$ as

$$\mathcal{T}_\epsilon^{(n)} := \operatorname{span}\{|\chi_{\mathbf{i}}^{(n)}\rangle : \mathbf{i} \in T_\epsilon^{(n)}\}. \tag{19.17}$$

Theorem 19.18 (Typical subspace theorem)**.** *Fix $\epsilon > 0$. Then $\forall \delta > 0, \exists n_0(\delta) > 0$ such that $\forall n \geq n_0(\delta)$ and $\rho^{(n)} = \pi^{\otimes n}$. Then*

$$\operatorname{Tr}(P_\epsilon^{(n)} \rho^{(n)} > 1 - \delta \tag{19.19}$$

where $P_\epsilon^{(n)}$ is the projection onto the typical subspace $\mathcal{T}_\epsilon^{(n)}$ and

$$(1-\delta)2^{n(S(\pi)-\epsilon)} \leq |\dim \mathcal{T}_\epsilon^{(n)}| \leq 2^{n(S(\pi)+\epsilon)}. \tag{19.20}$$

Proof.

$$\operatorname{Tr}(P_\epsilon^{(n)} \rho^{(n)}) = \sum -i \in T_\epsilon^{(n)} \lambda_{\mathbf{i}}^{(n)} = \sum_{\mathbf{i} \in T_\epsilon^{(n)}} p(\mathbf{i}) = Pr(T\epsilon^{(n)}) > 1 - \delta. \tag{19.21}$$

That is, we use $P_\epsilon^{(n)} = \sum_{\mathbf{i} \in T_\epsilon^{(n)}} \left|\chi_{\mathbf{i}}^{(n)}\right\rangle\!\left\langle\chi_{\mathbf{i}}^{(n)}\right|$.

The second part of the theorem is proved with analogy to the classical typical sequence theorem, since $\dim \mathcal{T}_\epsilon^{(n)} = |T_\epsilon^{(n)}|$. ☒

Theorem 19.22 (Schumacher)**.** *For an iid memoryless source $\{\pi, \mathcal{H}\}$,*

(1) *If $R > S(\pi)$ then $\exists$ a reliable compression-decompression scheme of rate R.*
(2) *If $R < S(\pi)$ then no compression-decompression scheme of rate R is reliable.*

This sounds a lot like Shannon's theorem. We'll try to prove at least the first part today.

Proof. Let $R > S(\pi)$. Our proof is constructive. Choose $\epsilon > 0$ such that $R > S(\pi) + \epsilon$. By the typical subspace theorem, for any $\delta > 0$ and n large enough we have

$$\dim \mathcal{T}_\epsilon^{(n)} \leq 2^{n(S(\pi)+\epsilon)} < 2^{nR}. \tag{19.23}$$

What is the compression map? We have

$$\mathcal{C}^{(n)} : \left|\Psi_k^{(n)}\right\rangle\!\left\langle\Psi_k^{(n)}\right| \mapsto \tilde{\rho}_k^{(n)} \tag{19.24}$$

where

$$\tilde{\rho}_k^{(n)} = \alpha_k^2 \left|\tilde{\psi}_k^{(n)}\right\rangle\!\left\langle\tilde{\psi}_k^{(n)}\right| + \beta_k^2 \left|\Phi_0\right\rangle\!\left\langle\Phi_0\right|, \tag{19.25}$$

with

$$|\tilde{\Psi}_k^{(n)}\rangle = \frac{P_\epsilon^{(n)}|\Psi_k^{(n)}\rangle}{\sqrt{\langle\Psi_k^{(n)}|P_\epsilon^{(n)}|\Psi_k^{(n)}\rangle}}, \tag{19.26}$$

the projection of $|\Psi_k^{(n)}\rangle$ onto the typical subspace with a normlization. $|\Phi_0\rangle$ is some fixed state in the typical subspace, and we fix $\alpha_k^2 = \langle\Psi_k^{(n)}|P_\epsilon^{(n)}|\Psi_k^{(n)}\rangle$ with $\alpha_k^2 + \beta_k^2 = 1$.

Now our decompression map is simply

$$\mathcal{D}^{(n)}(\tilde{\rho}_k^{(n)}) = \tilde{\rho}_k^{(n)} \oplus \mathbf{0}, \tag{19.27}$$

where $\mathbf{0}$ is there to pad the decompressed matrix with zeroes. Now what is the average ensemble fidelity?

$$\begin{aligned}
\bar{F}_n &= \sum_k p_k^{(n)} \langle\Psi_k^{(n)}|\tilde{\rho}_k^{(n)}|\Psi_k^{(n)}\rangle \\
&= \sum p_k^{(n)} \left[\alpha_k^2 \underbrace{|\langle\Psi_k^{(n)}|\tilde{\Psi}_k^{(n)}\rangle|^2}_{\alpha_k^2} + \beta_k^2 \underbrace{|\langle\Psi_k^{(n)}|\Phi_0\rangle|}_{\geq 0} \right] \\
&\geq \sum p_k^{(h)} \alpha_k^4 \\
&\geq \sum p_k^{(n)} (2\alpha_k^2 - 1)
\end{aligned}$$

since $(1-x)^2 \geq 0 \implies x^2 \geq 2x - 1$. We conclude that

$$\begin{aligned}
\bar{F}_n &\geq 2 \sum_k p_k^{(n)} \langle\Psi_k^{(n)}|P_\epsilon^{(n)}|\Psi_k^{(n)}\rangle - 1 \\
&= 2\,\mathrm{Tr}(P_\epsilon^{(n)}\rho^{(n)}) - 1.
\end{aligned}$$

But by the typical subspace theorem, we have $\mathrm{Tr}(P_\epsilon^{(n)}\rho^{(n)}) > 1 - \delta$, so

$$\bar{F}_n \geq 1 - 2\delta, \tag{19.28}$$

where δ can be made arbitrarily small in the limit as $n \to \infty$. ☒

Lecture 20.

Wednesday, March 6, 2019

Last time, we proved the first half of Schumacher's theorem. That is, for an iid memoryless ource $\{\pi, \mathcal{H}\}$, our signals are $\rho^{\otimes n} = \pi^{\otimes n}$, where the signals take the form $|\Psi_k^{(n)}\rangle$ with probability $p_k^{(n)}$. A single use of the source produces

$$\pi = \sum q_i |\phi_i\rangle, \tag{20.1}$$

while the full output is

$$\rho^{\otimes n} = \sum \lambda_{\mathbf{i}}^{(n)} \left|\chi_{\mathbf{i}}^{(n)}\right\rangle\!\left\langle\chi_{\mathbf{i}}^{(n)}\right| \tag{20.2}$$

with $|\chi_{\mathbf{i}}^{(n)}\rangle = |\phi_{i_1}\rangle \otimes \ldots \otimes |\phi_{i_n}\rangle$.

Last time, we proved by construction that for any rate $R > S(\pi), \exists$ a reliable compression-decompression scheme of rate R. Today, we will show that if $R < S(\pi)$, no compression-decompression scheme is reliable.

Proof. Let $R < S(\pi)$, and choose $\epsilon > 0$ such that $R = S(\pi) - 2\epsilon$. We have compression and decompression maps $\mathcal{C}^{(n)}$ and $\mathcal{D}^{(n)}$ such that

$$\mathcal{C}^{(n)} : \mathcal{D}(\mathcal{H}^{\otimes n}) \to \mathcal{D}(\tilde{\mathcal{H}}_n), \tag{20.3}$$

$$\mathcal{D}^{(n)} : \mathcal{D}(\tilde{H}_n) \to \mathcal{D}(\mathcal{H}^{\otimes n}). \tag{20.4}$$

We say that

$$\dim \tilde{\mathcal{H}}_n \approx 2^{nR}. \tag{20.5}$$

Let us denote a single compressed input as $\mathcal{C}^{(n)}(\left|\Psi_k^{(n)}\right\rangle\!\left\langle\Psi_k^{(n)}\right|) = \tilde{\rho}_k^{(n)}$ and the decompressed output as $\mathcal{D}^{(n)}(\tilde{\rho}_k^{(n)} \equiv \sigma_k^{(n)} \in \mathcal{D}(\mathcal{H}^{\otimes n})$. We also denote by $\tilde{P}_n$ the orthogonal projection operator onto $\tilde{\mathcal{H}}_n$.

Now the ensemble average fidelity is

$$\bar{F}_n = \sum_k p_k^{(n)} \langle\Psi_k^{(n)}|\sigma_k^{(n)}|\Psi_k^{(n)}\rangle. \tag{20.6}$$

We shall insert the identity $P_\epsilon^{(n)} + \bar{P}_\epsilon^{(n)}$, where $P_\epsilon^{(n)}$ is the projection onto the typical subspace $\mathcal{T}_\epsilon^{(n)}$ and $\bar{P}_\epsilon^{(n)} = I - P_\epsilon^{(n)}$. Sandwiching $\sigma_k^{(n)}$ with the identity gives us four terms when we expand out.

The first term looks like

$$\begin{aligned}
(I) &= \sum_k p_k^{(n)} \langle\Psi_k^{(n)}|P_\epsilon^{(n)} \sigma_k^{(n)} P_\epsilon^{(n)}|\Psi_k^{(n)}\rangle \\
&\leq \sum_k p_k^{(n)} \langle\Psi_k^{(n)}|P_\epsilon^{(n)} \sigma_k^{(n)} P_\epsilon^{(n)}|\Psi_k^{(n)}\rangle \\
&\leq \mathrm{Tr}(\rho^{(n)} P_\epsilon^{(n)} \mathcal{D}^n(\tilde{P}_n) P_\epsilon^{(n)}) \\
&= \sum_{\mathbf{i}\in\mathcal{T}_\epsilon^{(n)}} \lambda_i^{(n)} \langle\chi_{\mathbf{i}}^{(n)}|\mathcal{D}^n(\tilde{P}_n)|\chi_{\mathbf{i}}^{(n)}\rangle \\
&= 2^{-n(S(\pi-\epsilon)} \sum_i \langle\chi_{\mathbf{i}}^{(n)}|D^{(n)}(\tilde{P}_n)|\chi_{\mathbf{i}}^{(n)}\rangle \\
&= 2^{-n(S(\pi)-\epsilon)}\, \mathrm{Tr}(\mathcal{D}^n(\tilde{P}_n)).
\end{aligned}$$

where we've used the fact that $\rho_k^{(n)} \leq \tilde{P}_n$, so $\mathcal{D}^n(\rho_k^{(n)}) \leq \mathcal{D}^n(\tilde{P}_n)$ and rearranged terms to a trace by recognizing that $\rho^n = \sum_k p_k^{(n)} \left|\Psi_k^{(n)}\right\rangle\!\left\langle\Psi_k^{(n)}\right|$. We then used the projectors to turn the sum into a sum over only states in the typical subspace $\mathcal{T}_\epsilon^{(n)}$, and then we used the bound on $\lambda_{\mathbf{i}}^{(n)} \leq 2^{-n(S(\pi)-\epsilon}$ from the typical subspace theorem.

The second term is simpler: we have a term such that

$$\begin{aligned}(II) &= \sum p_k^{(n)} \langle \Psi_k^{(n)} | \bar{P}_\epsilon^{(n)} \underbrace{\mathcal{D}^{(n)}(\tilde{\rho}_k^{(n)})}_{\sigma_k^{(n)} \leq I} \bar{P}_\epsilon^{(n)} | \Psi_k^{(n)} \rangle \\ &\leq \sum_{\mathbf{i} \notin \mathcal{T}_\epsilon^{(n)}} \lambda_{\mathbf{i}}^{(n)} \\ &= \Pr(A_\epsilon^{(n)}) \to 0 \text{ as } n \to \infty.\end{aligned}$$

The cross terms $(III) + (IV)$ take the form

$$\begin{aligned}(III) + (IV) &= \sum p_k^{(n)} \langle \Psi_k^{(n)} | P_\epsilon^{(n)} \sigma_k^{(n)} \bar{P}_\epsilon^{(n)} + \bar{P}_\epsilon^{(n)} \sigma_k^{(n)} P_\epsilon^{(n)} | \Psi_k^{(n)} \rangle \\ &= \mathrm{Tr}(A^\dagger B + B^\dagger A)\end{aligned}$$

where $A = \sqrt{\sigma_k^{(n)}} P_\epsilon^{(n)} \sqrt{\rho^{(n)}}; B = \sqrt{\sigma_k^{(n)}} \bar{P}_\epsilon^{(n)} \sqrt{\rho^{(n)}}$. From here, we can observe that

$$[\mathrm{Tr}(A^\dagger B + B^\dagger A)]^2 = (2\mathrm{Re}(\mathrm{Tr}\, A^\dagger B))^2 \tag{20.7}$$

$$\leq 4|Tr(A^\dagger B)|^2 \tag{20.8}$$

$$\leq 4|(A, B)_{HS}|^2 \tag{20.9}$$

$$\leq (A, A)(B, B) \tag{20.10}$$

by Cauchy-Schwartz (and using the fact that $\overline{\mathrm{Tr}\, \mathrm{X}} = \mathrm{Tr}\, \mathrm{X}^\dagger$). Hence this is bounded by

$$4\,\mathrm{Tr}(\rho^{(n)} P_\epsilon^{(n)} \sigma_k^{(n)} P_\epsilon^{(n)})\,\mathrm{Tr}(\rho^{(n)} \bar{P}_\epsilon^{(n)} \sigma_k^{(n)} \bar{P}_\epsilon^{(n)}), \tag{20.11}$$

where recognizing that $P_\epsilon^{(n)} \leq I, \sigma_k^{(n)} \leq I$, we have the final bound

$$[\mathrm{Tr}(A^\dagger B + B^\dagger A)]^2 \leq 4\,\mathrm{Tr}(\bar{P}_\epsilon^{(n)} \rho^{(n)} \bar{P}_\epsilon^{(n)}) = \Pr(A_\epsilon^{(n)}) \to 0. \tag{20.12}$$

Hence this completes the proof of Schumacher's theorem– we have shown that all the terms are bounded and vanish in the $n \to \infty$ limit. ☒

Quantum channels Let us consider sending information via qubit. That is, Alice prepares a state ρ and sends it to Bob through a *quantum channel* represented by a map Λ, and what Bob receives is $\Lambda(\rho) \neq \rho$, where we anticipate there is noise in the channel.

Take a qubit, $\rho \in \mathbb{C}^2$, and recall that we can write the qubit state on the Bloch sphere as

$$\rho = \frac{1}{2}(I_2 + \mathbf{s} \cdot \boldsymbol{\text{œ}}), \tag{20.13}$$

with $\mathbf{s} = (s_x, s_y, s_z)$.

The first channel we'll consider is the bit flip channel:

$$\Lambda(\rho) = p\sigma_x \rho \sigma_x + (1-p)\rho = \sum A_k \rho A_k^\dagger, \tag{20.14}$$

where this channel admits a Kraus representation with

$$A_1 = \sqrt{1-p} I, \quad A_2 = \sqrt{p}\sigma_x. \tag{20.15}$$

One can easily check that $\sum_{k=1}^2 A_k^\dagger A_k = I$. Now we can put the Bloch sphere decomposition into our expression for $\Lambda(\rho)$. Recalling that $\sigma_i \sigma_j = \delta_{ij} + i\epsilon_{ijk}\sigma_k$, we can show that the final state can also be written in a Bloch representation as

$$\mathbf{s} = (s_x, s_y, s_z) \tag{20.16}$$

$$\to \mathbf{s}' = (\mathbf{s}_x, (1-2p)s_y, (1-2p)s_z). \tag{20.17}$$

The next channel is the *phase flip* channel, which is

$$\Lambda(\rho) = p\sigma_z \rho \sigma_z + (1-p)\rho, \tag{20.18}$$

with output

$$\mathbf{s}' = ((1-2p)s_x, (1-2p)s_y, s_z). \tag{20.19}$$

In general, we may consider random unitary (mixing-enhancing) channels, i.e. convex combinations of unitaries which generally produce CPTP maps. That is,

$$\sigma \equiv \Lambda(\rho) = \sum_i p_i U_i \rho U_i^\dagger. \tag{20.20}$$

This should remind us of Uhlmann's theorem– recall the idea of majorization. Uhlmann told us that

$$\mathbf{x} \prec \mathbf{y} \iff \mathbf{x} = \sum p_j P_j \mathbf{y}, \tag{20.21}$$

and in the quantum case, we have

$$\omega \prec \nu \iff \omega = \sum p_i U_i \nu U_i^\dagger. \tag{20.22}$$

Thus we see that for a general unitary channel, the output is majorized by the input,

$$\sigma \prec \rho \implies S(\sigma) \geq S(\rho) \tag{20.23}$$

by Schur concavity.

Here's another channel– the depolarizing channel, with

$$\Lambda(\rho) = (1-p)\rho + \frac{p}{3}(\sigma_x \rho \sigma_x + \sigma_y \rho \sigma_y + \sigma_z \rho \sigma_z), \tag{20.24}$$

with four natural Kraus operators $A_1 = \sqrt{1-p}I, A_2 = \sqrt{p/3}\sigma_x$, and A_3, A_4 the same with σ_y, σ_z. If we compute this, we find that

$$\mathbf{s}' = (f(p)s_x, f(p)s - y, f(p)s_z) \tag{20.25}$$

with $f(p) = 1 - \frac{4p}{3}$. Thus the depolarization channel scales down vectors from the Bloch sphere.

Example 20.26. Prove that

$$\Lambda(\rho) = (1-q)\rho + q\frac{I}{2} \tag{20.27}$$

is an alternate form for the depolarizing channel. Find the relation between p and q, using the identity

$$\frac{I}{2} = \frac{1}{4}\left[\rho + \sigma_x \rho \sigma_x + \sigma_y \rho \sigma_y + \sigma_z \rho \sigma_z\right]. \tag{20.28}$$

Lecture 21.

Friday, March 8, 2019

Last time, we discussed the qubit depolarizing channel. On e can show as an exercise that it has two equivalent forms, where $q = 4p/3, q \leq 1 \implies p \leq 3/4$.

Amplitude-damping channel Suppose we have a two-level atom where there is a ground state $|0_A\rangle$ and an excited state $|1_A\rangle$. By releasing a photon γ, for instance, this system can transition from the excited state to the ground state. This photon goes into the environment, i.e. the electromagnetic field, which is initially in a vacuum state $|0_E\rangle$

Hence the evolution of this system is described by both the atom (A) and the environment (E) and hence either

$$|1_A\rangle \to_p |0_A\rangle + \gamma \tag{21.1}$$

or

$$|1_A\rangle \to_{1-p} |1_A\rangle. \tag{21.2}$$

By Stinespring, we can implement this as some unitaries:

$$\Lambda(\rho) = \mathrm{Tr}_E\left[U(\rho \otimes |0\rangle\langle 0|_E)U^\dagger\right], \tag{21.3}$$

where

$$\rho = \sum_{i=0}^{1} \rho_{ij} |i\rangle\langle j| \tag{21.4}$$

where

$$U|0_A\rangle \otimes |0_E\rangle = |0_A\rangle \otimes |0_E\rangle \tag{21.5}$$

$$U|1_A\rangle \otimes |0_E\rangle = \sqrt{1-p}|1_A\rangle \otimes |0_E\rangle + \sqrt{p}|0_A\rangle \otimes |1_E\rangle. \tag{21.6}$$

Now our CPTP map Λ can be written as

$$\Lambda(\rho) = \begin{pmatrix} \rho_{00} + p\rho_{11} & \sqrt{1-p}\rho_{01} \\ \sqrt{1-p}\rho_{10} & (1-p)\rho_{11} \end{pmatrix}. \tag{21.7}$$

We can check that this U does actually implement Λ. For instance, look at $i = 0, j = 1$:

$$\begin{aligned} \rho_{10} \operatorname{Tr}_E\left[U\, |0\rangle\langle 1|_A \otimes |0\rangle\langle 0|_E\, U^\dagger\right] &= \rho_{01} \operatorname{Tr}_E(U|0_A\rangle \otimes |0_E\rangle)(\langle 1_A| \otimes \langle 0_E|U^\dagger) \\ &= \rho_{01} \operatorname{Tr}_E(|0_A\rangle \otimes |0_E\rangle)(\sqrt{1-p}\langle 1_A| \otimes \langle 0_E| + \sqrt{p}\langle 0_A| \otimes \langle 1_E|) \\ &= \sqrt{1-p}\rho_{01}\, |0_A\rangle\langle 1_A|\,. \end{aligned}$$

The others can be checked easily.

We can also write a Kraus representation for this operation. With

$$\Lambda(\rho) = \sum_{k=1}^{2} A_k \rho A_k^\dagger, \tag{21.8}$$

we have

$$A_1 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & \sqrt{p} \\ 0 & 0 \end{pmatrix}. \tag{21.9}$$

One can see explciilty that

$$A_1|0\rangle = |0\rangle, A_1|1\rangle = \sqrt{1-p}|1\rangle \tag{21.10}$$

$$A_1|0\rangle = 0, A_2|1\rangle = \sqrt{p}|0\rangle. \tag{21.11}$$

So A_1 represents the state staying as it is and A_2 represents the decay process.

Notice that we could have N successive uses of Λ. Hence

$$\begin{pmatrix} \rho_0 0 & \rho_{01} \\ \rho_{10} & \rho_{11} \end{pmatrix} \to_\Lambda \begin{pmatrix} \rho_0 0 + p\rho_{11} & \sqrt{1-p}\rho_{01} \\ \sqrt{1-p}\rho_{10} & (1-p)\rho_{11} \end{pmatrix} \to_\Lambda \begin{pmatrix} \rho'_{00} + p\rho'_{11} & \sqrt{1-p}\rho'_{01} \\ \sqrt{1-p}\rho'_{10} & (1-p)\rho'_{11} \end{pmatrix}. \tag{21.12}$$

Letting $q = 1 - p$, we find that after n uses, we end up in the state

$$\Lambda(\ldots\Lambda(\Lambda(\rho))\ldots) = \begin{pmatrix} \rho_{00} + \rho_{11} p[1 + q + q^2 + \ldots + q^{n-1}] & q^{n/2}\rho_{01} \\ q^{n/2}\rho_{10} & q^n\rho_{11}, \end{pmatrix} \tag{21.13}$$

and in the limit as $n \to \infty$ we get

$$\begin{pmatrix} \rho_{00} + \rho_{11} & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = |0\rangle\langle 0|_A\,. \tag{21.14}$$

Notice that $S(\rho) > 0$, but $\lim_{n\to\infty} S(\Lambda^n(\rho)) = S(|0\rangle\langle 0|) = 0$. Note that this *does not* mean that the information in the state has increased. Really, what's happened is that we've thrown away the original state with many uses of the channel. Hence we see that $S(\Lambda(\rho)) \not\geq S(\rho)$. We will shortly try to describe a measure of information that does decrease monotonically, $\chi(\Lambda(\rho)) \leq \chi(\rho)$.

Suppose we have a decay rate δ. This is probability that A decays per unit time, and hence in time Δt, the probability of decay is $p = \delta\Delta t = \delta\frac{t}{n}$. Equivalently $t = n\Delta t$. We can see from our matrix that

$$\begin{aligned} \rho_{11} \mapsto q^n\rho_{11} &= (1-p)^n\rho_{11} \\ &= \left(1 - \frac{\delta t}{n}\right)^n \rho_{11} \\ &= e^{-\delta t}\rho_{11} \end{aligned}$$

in the $n \to \infty$ limit. Thus our final state in the continuous limit is

$$\begin{pmatrix} \rho_{00} + \rho_{11}(1 - e^{-\delta t}) & e^{-\delta t/2}\rho_{01} \\ e^{-\delta t/2}\rho_{10} & e^{-\delta t}\rho_{11} \end{pmatrix} \to \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \tag{21.15}$$

Accessible information and the Holevo bound We shall define a quantity χ known as the Holevo χ quantity. It arises as an upper bound in a quantum information theoretic task: suppose we have a message-sender Alice who has a classical source producing signals $X \sim p(x), x \in J$. But she only has a quantum channel, so she has instead an encoder $x \mapsto \rho_x$. For now, suppose Alice has a noiseless quantum channel. She therefore sends ρ_x through the channel, where Bob receives ρ_x and decodes with a measurement (WLOG a POVM) $\{E_y\}_{y\in J}$ to get an inference Y.

Q: How much information can Bob gain about X through his measurement? It is simply the mutual information $I(X:Y)$. To get the maximum possible information, we ought to maximize

$$I_{\text{acc}}(\underbrace{\{p_x, \rho_x\}}_{\epsilon}) = \max I(X:Y), \tag{21.16}$$

where we maximize over the encoded states and the probability of the inputs. Holevo proved that

$$I_{\text{acc}}(\epsilon) \leq \chi(\epsilon), \tag{21.17}$$

where

$$\chi(\epsilon) := S(\underbrace{\sum_x p_x\rho_x}_{\rho}) - \sum_x p_x S(\rho_x) \tag{21.18}$$

Thus if $\{p_x, \rho_x\}$ is an ensemble of pure states, $\chi(\epsilon) = S(\rho)$, so the mutual information is total. Remarkably, this is completely independent of what measurement we do.

Theorem 21.19 (Holevo bound)**.** *The mutual information is bounded by the Holevo χ quantity:*

$$I(X:Y) \leq \chi(\{p_x, \rho_x\}) \tag{21.20}$$

so that

$$I_{acc}(\{p_x, \rho_x\}) \leq \chi(\{p_x, \rho_x\}), \tag{21.21}$$

where equality holds if all the ρ_xs commute, and the measurment is a projection onto the simultaneous eigenbasis.

Proof. The proof is by strong subadditivity. Recall that we need three systems to make SSA work. Call them A, Q, B– we'll define them as follows.

I We start by embedding the random variable X into a quantum system A with Hilbert space $\mathcal{H}_A$; $\dim \mathcal{H} = |J|$. For $x \in J$, we can assign the orthonormal basis $\{|x\rangle\}$.

II The second system Q is then Alice's encoded state $\rho_x \in \mathcal{D}(\mathcal{H}_Q)$.

III The final system B is Bob's measuring device, with a Hilbert space $\mathcal{H}_B$. We assume it is in a fixed pure state $|\phi_B\rangle$.

We can describe the overall process with a state $AQB \in \mathcal{H}_A \otimes \mathcal{H}_Q \otimes \mathcal{H}_B$.

The initial state is

$$\rho_{AQB} = \left(\sum p_x \underbrace{|x\rangle\langle x|}_{A} \otimes \underbrace{\rho_x}_{Q} \right) \otimes \underbrace{|\phi_B\rangle\langle\phi_B|}_{B}. \tag{21.22}$$

Now let us describe the measurement– associate a quantum operation Λ acting on Q, B such that

$$\Lambda(\sigma_Q \otimes |\phi_B\rangle\langle\phi_B|) = \sum_y \sqrt{E_y}\sigma_Q\sqrt{E_y} \otimes |y_B\rangle\langle y_B|, \tag{21.23}$$

where we now sum over the measurement outcomes y. Recall that we went from projective measurements $\{M_y\}$ to POVMs $E_a = M_a^\dagger M_a$ such that $M_a = \sqrt{E_a}$. Notice this is legitimate since the trace is preserved,

$$\begin{aligned} \operatorname{Tr}\Lambda(\sigma_Q \otimes |\phi_B\rangle\langle\phi_B|) &= \operatorname{Tr}\sum_y \sqrt{E_y}\sigma_Q\sqrt{E_y} \\ &= \sum_y \operatorname{Tr} E_y\sigma_Q = \operatorname{Tr}\sigma_Q = 1. \end{aligned}$$

Now for a more general state of the system QB, we have

$$\Lambda(\omega_{QB}) = \sum_y A_y \omega_{QB} A_y^\dagger, \tag{21.24}$$

where we can check as an exercise that

$$A_y = \sqrt{E_y} \otimes U_y, \quad U_y|\phi_B\rangle = |y_b\rangle \tag{21.25}$$

such that $\sum_y A_y^\dagger A_y = I, \sum_y E_y = I$.

Now our system has gone from an initial state $\rho_{AQB} \to \rho_{A'Q'B'}$ where

$$\rho_{A'Q'B'} = \sum_{x,y} p_x \, |x\rangle\langle x| \otimes \sqrt{E_y}\rho_x\sqrt{E_y} \otimes |y_B\rangle\langle y_B| \, . \tag{21.26}$$

We'll complete the proof next time.

Also, a quick note that the converse of the data compression theorem is a bit algebra-heavy, and will not be examinable.

Lecture 22.

Monday, March 11, 2019

Today we will complete the proof of the Holevo bound. We said that Alice's message, the encoded state, and Bob's measurement device were written as a tripartite initial state

$$\rho_{AQB} = \sum_x p_x \, |x\rangle\langle x| \otimes \rho_x \otimes |\phi_B\rangle\langle\phi_B| \, , \tag{22.1}$$

which became

$$\rho_{A'Q'B'} = \sum_{x,y} p_x \, |x\rangle\langle x| \otimes \sqrt{E_y}\rho_x\sqrt{E_y} \otimes |y_B\rangle\langle y_B| \, . \tag{22.2}$$

We know that SSA has the following consequences:

(1) $I(A:B) \leq I(A:BC)$
(2) $I(A':B') \leq I(A:B)$

where $\rho_{A'B'} = (\mathrm{id}_A \otimes \Lambda)\rho_{AB}$. Thus notice that

$$I(A:Q) = I(A:QB) \tag{22.3}$$

since B is uncorrelated initially.

$$I(A:QB) \geq I(A':Q'B') \tag{22.4}$$

by (2) above, and

$$I(A':Q'B'') \geq I(A':B') \tag{22.5}$$

by (1). Hence

$$I(A':B') \leq I(A:Q). \tag{22.6}$$

In fact, this is the Holevo bound. For notice that the RHS is $I(A:Q)$, such that

$$\begin{aligned} \rho_{AQ} &= \sum_x p_x \, |x\rangle\langle x| \otimes \rho_x \\ \rho_A &= \sum p_x \, |x\rangle\langle x| \\ \rho_Q &= \sum p_x \rho_x = \rho. \end{aligned}$$

The entropies of these states are

$$S(\rho_A) = H(\{p_x\}), \quad S(\rho_Q) = S(\sum p_x\rho_x). \tag{22.7}$$

Recall that on Examples Sheet 3, we showed that for ω_x with mutually orthogonal supports,

$$S(\sum p_x\omega_x) = H(\{p_x\}) + \sum p_x S(\omega_x). \tag{22.8}$$

Thus for the state $\rho_{AQ} = \sum p_x \omega_x$ wth $\omega_x = |x\rangle\langle x| \otimes \rho_x$, we have

$$S(\rho_{AQ}) = H(\{p_x\}) + \sum p_x S(\rho_x) \tag{22.9}$$

since $S(|x\rangle\langle x| \, x) = 0$.

Thus the mutual information $I(A:Q)$ is

$$\begin{aligned} I(A:Q) &= S(A) + S(Q) - S(AQ) \\ &= H(\{p_x\}) + S(\sum p_x\rho_x) - H(\{p_x\}) - \sum p_x S(\rho_x) \\ &= \chi(\{p_x, \rho_x\}). \end{aligned}$$

We're nearly done. All that remains to to evaluate the LHS,

$$I(A' : B') = S(A') + S(B') - S(A'B'). \tag{22.10}$$

Returning to our expression for the statement after measurement, we have

$$\rho_{A'B'} = \sum p_x \left|x\right\rangle\!\left\langle x\right| \operatorname{Tr}(E_y\rho_x) \otimes \left|y_B\right\rangle\!\left\langle y_B\right| \tag{22.11}$$

by the cyclicity of the trace. Notice that $\operatorname{Tr}(E_y\rho_x) = p(y|x)$, the probability of measuring outcome y given that the state was ρ_x, so we have

$$\rho_{A'B'} = \sum_{x,y} p(x)p(y|x) \left|xy\right\rangle\!\left\langle xy\right|_{A'B'} = \sum_{x,y} p(x,y) \left|xy\right\rangle\!\left\langle xy\right|_{A'B'}. \tag{22.12}$$

Since these are mutually orthogonal states, we find that

$$S(A'B') = H(\{p(x,y)\}) = H(XY), \tag{22.13}$$

the joint Shannon entropy. Now

$$\rho_{A'} = \sum p(x) \left|x\right\rangle\!\left\langle x\right| \implies S(A') = H(\{p_x\}) = H(X) \tag{22.14}$$

and similarly

$$\rho_{B'} = \sum_{x.y} p(x,y) \left|y\right\rangle\!\left\langle y\right| = \sum p(y) \left|y\right\rangle\!\left\langle y\right| \implies S(B') = H(Y). \tag{22.15}$$

Thus we find that

$$I(A' : B') = H(X) + H(Y) - H(XY) \equiv I(X : Y), \tag{22.16}$$

and therefore we conclude that

$$I(X;Y) \leq \chi(\{p_x, \rho_x\}). \tag{22.17}$$

This is the Holevo bound. ☒

Properties of the Holevo χ quantity Let us denote the average state $\rho = \sum p_x\rho_x$.

(a) $\chi(\{p_x, \rho_x\}) \geq 0$ (follos from concavity of $S(\rho)$).

(b) $\chi(\mathcal{E}) \to S(\rho)$ when the ρ_x are pure.

(c) $\chi(\mathcal{E}) = \sum p_x D(\rho_x||\rho)$. This is true because

$$\begin{aligned} &\sum p_x D(\rho_x||\rho) \\ &\qquad = \sum p_x[-S(\rho_x) - \operatorname{Tr}\rho_x \log\rho] \\ &\qquad = -\sum p_x S(\rho_x) + S(\sum p_x\rho_x). \end{aligned}$$

Notice that $D(\rho_x||\rho) \geq 0$.

(d) By the data processing inequality, the relative entropy is non-increasing under quantum operations,

$$D(\Lambda(\rho_x)||\Lambda(\rho)) \leq D(\rho_x||\rho). \tag{22.18}$$

Taking $\mathcal{E} = \{p_x, \rho_x\}, \mathcal{E}' = \{p_x, \Lambda(\rho_x)\}$, we find that

$$\chi(\mathcal{E}') \leq \chi(\mathcal{E}). \tag{22.19}$$

(e) For an ensemble $\{p_x, \rho_x\}$, we can embed the state in terms of the classical labels as

$$\rho_{XA} = \sum p_i \left|x\right\rangle\!\left\langle x\right| \otimes \rho_x, \tag{22.20}$$

known as a classical-quantum (c-q) state.[26] Then the Holevo quantity of such an ensemble is

$$\chi(\{p_x, \rho_x\}) = I(X;A)_\rho. \tag{22.21}$$

[26] We saw this on the examples sheet as well.

Moreover, We could run the quantum part through a quantum channel $\Lambda : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ to get a new state

$$\tilde{\rho}_{XB} = \sum p_i \, |i\rangle\langle i| \otimes \Lambda(\rho_x), \tag{22.22}$$

and then it follows that

$$I(X:B)_{\tilde{\rho}} \leq I(X:A)_{\rho}. \tag{22.23}$$

Noisy quantum channels What happens if our noiseless quantum channel is replaced by a noisy quantum channel Λ? Clearly, Bob no longer receives ρ_x but $\Lambda(\rho_x)$, and must decode this new state. The maximum information Bob can retrieve for a *single use* of the channel is then bounded by

$$\chi(\{p_x, \Lambda(\rho_x)\}), \tag{22.24}$$

the Holevo χ quantity. In fact, we can do better by using the channel multiple times. Preskill has an argument that more uses of the channel always improves the outcome.

Now what is the classical capacity of the channel Λ? Suppose we have a memoryless quantum channel, i.e.

$$\Lambda^{(n)} \equiv \Lambda^{\otimes n}. \tag{22.25}$$

That is, the noise is uncorrelated between uses.

Definition 22.26. The *classical capacity* is the maximum rate of reliable transmission of classical information evaluated in the asymptotic limit, i.e. in the limit of channel uses $n \to \infty$.

Here's something strange about quantum channels as opposed to classical channels. For a classical memoryless channel $\mathcal{N}$, we model this channel by a set of conditional probabilities $\{p(y|x)\}$, where the capacity was

$$C(\mathcal{N}) = \max_{p(x)} I(X;Y). \tag{22.27}$$

This was a unique value. On the other hand, consider a quantum channel Λ. We have some options which will affect the capacity.

(a) Information sent– classical or quantum

(b) The encoded state $\rho^{(n)}$ (input to n uses of the channel)– entangled or product state. Notice that if $\rho^{(n)} \equiv \rho_1 \otimes \ldots \otimes \rho_n$ is a product state, then the outcome state $\Lambda(\rho_1) \otimes \ldots \otimes \Lambda(\rho_n)$ is also a product state.

(c) The measurement/decoding protocol could act on the n uses individually or collectively.

(d) Presence of auxiliary resources, e.g. Alice and Bob share an entangled state, and Alice uses her half in the encoding of her classical message.

We'll restrict ourselves to the following scenario. Alice has a classical information source and she wants to send a message through a (noisy) memoryless quantum channel Λ. Thus Alice has some messages

$$\mathcal{M} = \{1, 2, \ldots, |\mathcal{M}|\} \tag{22.28}$$

and these messages get encoded as

$$M \in \mathcal{M} \mapsto \rho_M^{(n)}, \tag{22.29}$$

a quantum state which is then transmitted through the channel as

$$\sigma_M^{(n)} = \Lambda^{\otimes n}(\rho_M^{(n)}). \tag{22.30}$$

Bob performs a POVM $\{E_M^{(n)}\}_{M \in \mathcal{M}}$. Call our encoding scheme $\mathcal{E}^{(n)}$ and the decoding scheme $\mathcal{D}^{(n)}$.

If the message M was sent, the probability of error is then

$$p_{\text{err}} = 1 - \operatorname{Tr}(E_M^{(n)} \sigma_M^{(n)}). \tag{22.31}$$

The maximum probability of error of $\mathcal{C}^{(n)} = (\mathcal{E}^{(n)}, \mathcal{D}^{(n)})$ is then

$$p_{\max}(\mathcal{C}^{(n)}) = \max_{m \in \mathcal{M}} \Big[1 - \operatorname{Tr}(E_M^{(n)} \sigma_M^{(n)}))\Big]. \tag{22.32}$$

We say the information transmssion is reliable if

$$\lim_{n \to \infty} p_{\max}^{(n)} = 0. \tag{22.33}$$

That is, the maximum probability of error tends to zero in the asymptotic limit. We then say that the rate of the channel is

$$R = \frac{\log|\mathcal{M}|}{n}. \tag{22.34}$$

Lecture 23.

Wednesday, March 13, 2019

Last time, we set up the problem of sending classical information through a quantum channel. For our case, we shall be interested in Alice preparing a product state input,

$$\rho_M^{(n)} = \rho_M^1 \otimes \ldots \otimes \rho_M^n, \tag{23.1}$$

which is transmitted through the memoryless channel $\Lambda^{\otimes n} : \rho_M^{(n)} \to \sigma_M^{(n)} = \sigma_M^1 \otimes \ldots \otimes \sigma_M^n$ to Bob as another (generally different) product state. Bob then applies his decoding protocol $\mathcal{D}$ on the entire σ_M^n once all the messages have come in.

The product state capacitiy is given by $C^{(1)}(\Lambda) = \sup\{R : R \text{ achievable}\}$ where R is the rate of a code.

Theorem 23.2 (Holevo-Schumacher-Westmoreland (HSW))**.** *The product state capacity is given by*

$$C^{(1)}(\Lambda) = \max_{\{p_x,\rho_x\}_{x=1}^{d_A^2}} \chi(\{p_x, \Lambda(\rho_x)\}) \equiv \chi^*(\Lambda) \tag{23.3}$$

That is, for any rate $R < \chi^*(\Lambda)$, there exists a code $\mathcal{C}^{(n)}$ with that rate which is reliable. Conversely, for all codes $\mathcal{C}^{(n)}$ with rate $R > \chi^*(\Lambda)$, no such code is reliable.

In the examples sheet, we will show that by transmitting n qubits, Alice can send at most n bits of classical message to B. Moreover, for today we will use the generalized Fano's inequality, i.e. for two random variables X, Y with values $x_i, y_i, i = 1, \ldots, m$, for $\epsilon \in (0, 1)$,

$$\sum_{i=1}^{m} P(X = x_i, Y = y_i) = 1 - \epsilon. \tag{23.4}$$

Equivalently

$$H(X|Y) \le h(\epsilon) + \epsilon \log(m-1), \tag{23.5}$$

or

$$P(X = Y) = 1 - \epsilon. \tag{23.6}$$

Proof. Assume WLOG that the messages are uniformly distributed. We can do this since we are only interested in the maximum probability of error, which should not depend on the distribution of the message.

If the message M is sent, the probability of error is then

$$P(\hat{M} \neq M) = 1 - \mathrm{Tr}(E_M^{(n)} \sigma_M^{(n)}, \tag{23.7}$$

i.e. $1 -$ the probability of a successful decoding. Now nR is the number of bits of classical information sent on n uses. The maximum number of qubits Alice can send is $\log d_B^n = n \log d_B$ where $\mathrm{d}_B^n = \dim \mathcal{H}_B^{\otimes n}$. That is, in each use Alice sends at most $\log d_B$ qubits corresponding to a $\sigma_M^i \in \mathcal{H}_B$. Hence

$$nR \le n \log d_B \tag{23.8}$$

Using our second fact, the generalized Fano's inequality, let $q = P(\hat{M} \neq M)$. We want to prove that $\forall \mathcal{C}^{(n)}$ of rate $R > \chi^*(\Lambda)$, $p_{\max}^{(n)} \to 0$ as $n \to \infty$.

The generalized Fano's inequality then says that

$$\begin{aligned} H(M|\hat{M}) &\le h(q) + q \log(|\mathcal{M}| - 1) \\ &\le h(q) + q \log|M| \\ &= h(q) + qnR \le h(q) + qn \log d_B m \end{aligned}$$

using the definition $R = \frac{\log|\mathcal{M}|}{n}$ and the inequality $R \le \log d_B$.

Hence

$$\begin{aligned} q_n \log d_B \geq (M|\hat{M}) - h(q) \\ &= H(M\hat{M}) - H(\hat{M}) + (-H(M) + H(M)) - h(q) \\ &= H(M) - I(M:\hat{M}) - h(q). \end{aligned}$$

The Holevo bound now tells us that

$$I(M:\hat{M}) \leq \chi\left(\left\{\frac{1}{|\mathcal{M}|}, \sigma_M^{(n)}\right\}\right), \tag{23.9}$$

where we have taken a uniform distribution so $p_x = 1/|\mathcal{M}|\, \forall x$ and substituted the definition $\Lambda(\rho_M^{(n)}) = \sigma_m^{(n)}$ for n uses of the channel.

Recalling the definition of the Holevo χ quantity, we can write

$$I(M:\hat{M}) \leq S\left(\underbrace{\frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} \sigma_M^{(n)}}_{\bar{\sigma}_j}\right) - \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} S(\sigma_M^{(n)}). \tag{23.10}$$

Notice that $\sigma_M^{(n)} \in \mathcal{H}_B^{\otimes n}$. If we write the state

$$\omega_{B_1\ldots B_n} := \frac{1}{|\mathcal{M}|}\sum \sigma_M^{(n)} = \frac{1}{|\mathcal{M}|\sum_M \in \mathcal{M}}(\sigma_M^1 \otimes \ldots \otimes \sigma_M^n), \tag{23.11}$$

bu subadditivity we know that

$$S(\omega_{B_1\ldots B_n}) \leq \sum_{j=1}^{n} S(\omega_{B_j}), \tag{23.12}$$

where $\omega_{B_j} = \mathrm{Tr}_{\not{j}}\,\omega_{B_1\ldots B_n} = \frac{1}{|\mathcal{M}|}\sum_{M\in\mathcal{M}} \sigma_M^j = \bar{\sigma}^j$, the average output state.

For a product state, notice that

$$S(\sigma_M^{(n)} = \sum_{j=1}^{n} S(\sigma_M^J). \tag{23.13}$$

Hence our mutual information is bounded by

$$\begin{aligned} I(M:\hat{M}) &\leq \sum_{j=1}^{n} S(\bar{\sigma}_j) - \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}}\sum_{j=1}^{n} S(\sigma_M) \\ &= \sum_{j=1}^{n}\left[S(\frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} \sigma_M^j) - \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} S(\sigma_M^j)\right], \end{aligned}$$

so we conclude that

$$I(M:\hat{M}) \leq \sum_{j=1}^{n} \chi\left(\left\{\frac{1}{|\mathcal{M}|}, \sigma_M^j\right\}\right). \tag{23.14}$$

Notice that the Holevo capacity bounds the individual Holevo χ quantity

$$\begin{aligned} \chi^*(\Lambda) &= \max_{\{p_x,\rho_x\}} S(\sum p_x \Lambda(\rho_x)) - \sum p_x S(\Lambda(\rho_x)) \\ &\geq S\left(\frac{1}{|\mathcal{M}|}\sum \sigma_M^j\right) - \frac{1}{|\mathcal{M}|}\sum S(\sigma_M^j) \\ &= \chi\left(\{\frac{1}{|\mathcal{M}|}, \sigma_M^j\}\right), \end{aligned}$$

and therefore

$$I(M:\hat{M}) \leq \sum_{j=1}^{n} \chi^*(\Lambda) = n\chi^*(\Lambda). \tag{23.15}$$

Using this bound on the mutual information we now write

$$\begin{aligned} qn\log d_B &\geq H(M) - I(M:\hat{M}) - h(q) \\ &\geq H(M) - n\chi^*(\Lambda - h(q) \\ &= \log|\mathcal{M}| - n\chi^*(\Lambda - h(q) \\ &= nR - n\chi^*(\Lambda) - h(q). \end{aligned}$$

Dividing through by $n\log d_B$, we have

$$q \geq \frac{n(R-\chi^*(\Lambda))}{n\log d_B} - \frac{h(q)}{n\log d_B}. \tag{23.16}$$

We see that in the $n\to\infty$ limit, this second term goes to zero and $p^{(n)}_{\max} \geq q$, so in the $n\to\infty$ lmit we have

$$\frac{R-\chi^*(\Lambda)}{\log d_B} > 0 \implies p^{(n)}_{\max} \not\to 0. \tag{23.17}$$

☒

Lecture 24.

Friday, March 15, 2019

Last time, we discussed the HSW theorem, stating that

$$C^{(1)}(\Lambda) = \chi^*(\Lambda) = \max_{\{p_x,\rho_x\}} \chi(\{p_x, \Lambda(\rho_x)\}). \tag{24.1}$$

Lemma 24.2. *Any quantum channel can transmit classical information as long as it is not a constant channel, i.e. $\Lambda(\rho)$ not identical for all ρ.*

Proof. If Λ is not a constant channel, $\exists|\psi\rangle, |\phi\rangle$ s.t.

$$\Lambda(|\psi\rangle\langle\psi|) \neq \Lambda(|\phi\rangle\langle\phi|). \tag{24.3}$$

Hence for the ensemble $\mathcal{E} = \{p_1 = p_2 = 1/2, \rho_1 = |\psi\rangle\langle\psi|, \rho_2 = |\phi\rangle\langle\phi|\}$, we have a Holevo χ quantity

$$\chi(\mathcal{E}) = S(\frac{1}{2}\Lambda(\psi) + \frac{1}{2}\Lambda(\phi)) - \left[\frac{1}{2}S(\Lambda(\psi)) + \frac{1}{2}S(\Lambda(\phi))\right] > 0 \tag{24.4}$$

since equality is achieved in concavity of S only for $\Lambda(\psi) = \Lambda(\phi)$, which we required was false. Hence $\chi^*(\Lambda) > 0$ since it is the maximum over all ensembles, so we can send classical information using product states. ☒

χ^* has the following property: *superadditivity of $\chi^*(\Lambda)$.*

$$\chi^*(\Lambda_1 \otimes \Lambda_2) \geq \chi^*(\Lambda_1 + \chi^*(\Lambda_2). \tag{24.5}$$

The proof is an exercise on Examples Sheet 4. It follows by iteration that

$$\begin{aligned} \chi^*(\Lambda^{\otimes n}) &= \chi^*(\Lambda \otimes \Lambda^{\otimes n-1}) \\ &\geq \chi^*(\Lambda) + \chi*(\Lambda^{\otimes n-1}) \\ &\geq n\chi^*(\Lambda) \end{aligned}$$

Question: can one increase the classical capacity of a quantum channel by using entangled input states? To address this question, we have the *additivity conjecture of χ^** (since resolved):

$$\chi^*(\Lambda_1 \otimes \Lambda_2) = \chi(\Lambda_1) + \chi^*(\Lambda_2). \tag{24.6}$$

We have a formal expression for the classical capacity:

$$C_{cl}(\Lambda) = \lim_{n\to\infty} \frac{1}{n}\chi^*(\Lambda^{\otimes n}). \tag{24.7}$$

Notice that applying superadditivity, we have

$$C_{cl}(\Lambda) \geq \lim_{n\to\infty} \frac{n}{n}\chi^*(\Lambda), \tag{24.8}$$

so that if *additivity* holds,

$$C_{cl}(\Lambda) = \chi^*(\Lambda). \tag{24.9}$$

Therefore if additivity is true, then entangled inputs do not improve the capacity of the channel. However, an insightful proof by Matt Hastings established a counterexample to the additivity conjecture. Finding a counterexample is very difficult– it's more like an existence proof. But many examples of additivity do exist, e.g.

$$\chi^*(\Lambda) \text{ for } \Lambda(\rho) = (1-p)\rho + p\frac{I}{2} \tag{24.10}$$

the depolarizing channel. Hastings used a powerful theorem by Peter Shor linking four different quantities, $\chi^*, S_{\min}(\Lambda) = \min_\rho S(\Lambda(\rho))$, and two others– which says that if one is additive, then all are, and conversely if one is not additive, then all are not. What Hastings found was a counterexample to the additivity of $S_{\min}$. (This is non-examinable.)

Quantum capacity In transmitting a quantum state, we're not interested in the probability of error but rather the fidelity between the initial and final states, $F(\rho, \sigma)$.

Theorem 24.11 (LSD). *Define*

$$Q(\Lambda) = \lim_{n\to\infty} \frac{1}{n} \max_{\rho^{(n)} \in \mathcal{D}(\mathcal{H}^{\otimes n})} I_C(\Lambda^{\otimes n}, \rho^{(n)}) \tag{24.12}$$

where $I_C(\Lambda, \rho)$ is the coherent information with respect to input ρ.

Let $\Lambda : Q \to Q'$. It has a Stinespring implementation $U_\Lambda : QE \to Q'E$, where E is an ancilla we can take WLOG to be in a pure state $|0_E\rangle$. We can also purify Q by introducing a reference system R. Hence the initial state is $|\psi_{RQE}\rangle = |\psi^\rho_{RQ}\rangle \otimes |0_E\rangle$. After the operation,

$$\begin{aligned} I_C(\Lambda, \rho) &:= -S(R, Q')_{\rho'} \\ &= -\left[S(\rho'_{RQ'}) - S(\rho'_{Q'}\right] \\ &= -\left[S(\rho'_E) - S(\rho'_{RE}\right] \\ &= S(R|E)_{\rho'} \end{aligned}$$

But now

$$\begin{aligned} I_C(|Lambda, \rho) &= S(R|E)_{\rho'} \\ &= S(\rho'_{RE}) - S(\rho'_E) \\ &\leq S(\rho'_R) + S(\rho'_E) - S(\rho'_E) \\ &= S(\rho_R) = S(\rho_Q) \end{aligned}$$

since RQ is a pure state, and we have used subadditivity to rewrite $S(\rho'_{RE})$. We conclude that

$$I_C(\Lambda, \rho) \leq S(\rho). \tag{24.13}$$

DPI for quantum systems Classically, we talked about the data processing inequality, by which a system $X \to Y \to Z$ obeyed $I(X:Y) \geq I(X:Z)$. That is, data processing cannot improve the mutual information between two systems.

Here, the quantum analogue is

$$S(\rho) \geq I_C(\Lambda_1, \rho) \geq I_C(\Lambda_2 \circ \Lambda_1, \rho). \tag{24.14}$$

The proof is by strong subadditivity.

Proof. We can describe applying the two maps Λ_1, Λ_2 with their Stinespring implementations. BY strong subadditivity, we know that for a tripartite system ABC, $S(A|BC) \leq S(A|B)$. Hence with the final state $\rho''_{RE_1E_2}$, we have

$$S(R|E_1E_2)_{\rho''} \leq S(R|E_1)_{\rho''}, \tag{24.15}$$

where the LHS is just $I_C(\Lambda_2 \circ \Lambda_1, \rho)$. What is the right side? IT is

$$\begin{aligned} S(R|E_1)_{\rho''} &= S(\rho''_{RE_1}) - S(\rho''_{E_1}) \\ &= S(\rho'_{RE_1}) - S(\rho'_{E_1}) \\ &= S(R|E_1)_{\rho'} \\ &= I_C(\Lambda_1, \rho). \end{aligned}$$

Hence the DPI holds:

$$S(\rho) \geq I_C(\Lambda_1, \rho) \geq I_C(\Lambda_2 \circ \Lambda_1, \rho). \tag{24.16}$$

⊠

The entanglement fidelity is non-examinable. The converse of the Schumacher compression theorem is non-examinable (the one with four inequalities). Justify manipulations like SSA