Lecture 2

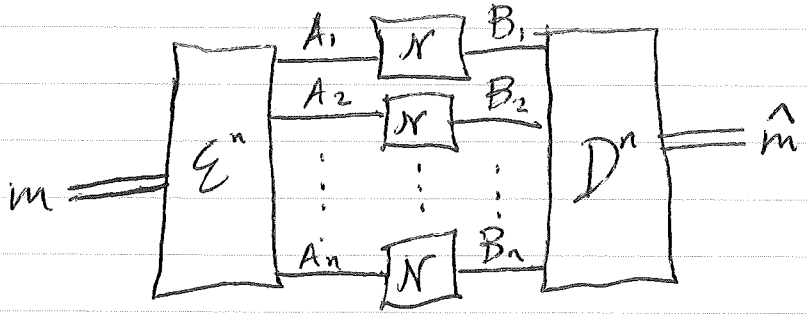
Classical Communication over Quantum Channels

Fundamental question: How much data can we send reliably over a communication channel? If we take advantage of quantum effects, can we send more data than if we did not?

To address this question, we need to define codes and capacity.

Given is a quantum channel $\mathcal{N}_{A \to B}$
We allow a communication scheme to use the channel n times.
Formally, an (n, R, ε) code consists of an encoding $\mathcal{E}^n$ & decoding $\mathcal{D}^n$ w/ the following structure

A message m is selected from a set $\mathcal{M} = \{1, \ldots |\mathcal{M}|\}$



(classical to quantum channel)

1) encoding $\mathcal{E}^n$ is a preparation of a quantum state $\rho_{A^n}^m = \rho_{A_1 \cdots A_n}^m$

2) $\rho_{A^n}^m$ gets transmitted through channel, leading to $\left(\mathcal{N}_{A_1 \to B_1} \otimes \mathcal{N}_{A_2 \to B_2} \otimes \cdots \otimes \mathcal{N}_{A_n \to B_n}\right)\left(\rho_{A^n}^m\right)$

$$= \mathcal{N}^{\otimes n}\left(\rho_{A^n}^m\right)$$

3) Decoding is a measurement (quantum to classical channel)

POVM $\left\{\Lambda_{B^n}^m\right\}$ $\quad \Lambda_{B^n}^m \geq 0$ & $\sum_m \Lambda_{B^n}^m = I_{B^n}$

Probability of successful decoding is

$$\Pr\left\{\hat{M} = m \,\middle|\, M = m\right\} = \mathrm{Tr}\left\{\Lambda_{B^n}^m \mathcal{N}^{\otimes n}\left(\rho_{A^n}^m\right)\right\}$$

An (n, R, ε) code has

$$\Pr\{\hat{M} = m \mid M = m\} \geq 1 - \varepsilon \qquad \forall m \in \mathcal{M}$$

the rate is equal to $\frac{\#\text{ bits}}{\text{channel use}}$

i.e., $R = \frac{\log_2 |\mathcal{M}|}{n}$

these three parameters should trade off against each other. Clearly, we cannot have n small, ε small, & R high

Shannon's idea was to allow for n to be large so that the others can go in the directions we wish

Asymptotic info. theory focuses on rate as quantity of primary interest

Define a rate R to be <u>achievable</u> if $\forall \varepsilon > 0$ & sufficiently large n, there exists an (n, R, ε) code as defined above.

The capacity $C(N)$ = supremum of all achievable rates

To prove a capacity theorem, one must do it in two parts:

1) achievability - demonstrate the existence of a sequence of (n, R, ε) protocols
   i.e., a lower bound for $C(N)$
2) converse - prove that rates exceeding capacity are not achievable
   i.e., an upper bound for $C(N)$

if lower bound & upper bound coincide, then theorem is established.

It is desirable for the capacity to be efficiently computable for a given channel.

Let's prove the converse first.

our goal is to relate the rate, n, & error to an information quantity.

Suppose that we have an (n, R, ε) w/ encoding and decoding.

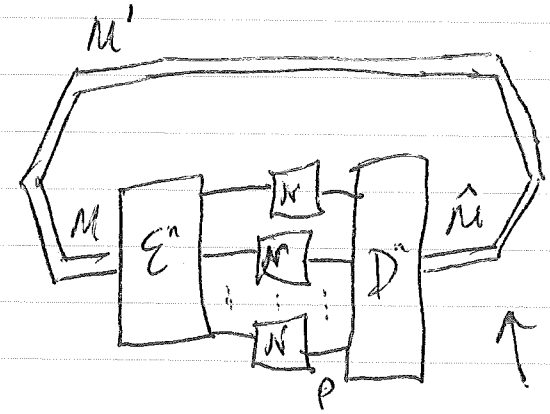
Then we could always prepare at the input a maximally correlated state of the form $\Phi_{M'M} = \frac{1}{|M|} \sum_{m,m'} |m\rangle\langle m|_{M'} \otimes |m\rangle\langle m|_M$

and then get a state at the output ε-close to $\Phi_{M'M}$.

How to quantify? Use trace distance

$\| \rho - \sigma \|_1$ where $\|A\|_1 = \mathrm{Tr}\{\sqrt{A^\dagger A}\}$

So this protocol looks like



↑ let $\omega_{M'\hat{M}}$ denote final state

Exercise: Show that an (n, R, ε) code satisfies $\| \Phi_{M'M} - \omega_{M'\hat{M}} \|_1 \leq 2\varepsilon$

First step: $\log |\mathcal{M}| = I(\mathcal{M}'; \hat{\mathcal{M}})_{\Phi}$

2nd step: continuity of entropy

for $\|\rho - \sigma\|_1 \leq \varepsilon$, $|H(\rho) - H(\sigma)| \leq c_1 \varepsilon \log d + c_2 h_2(\varepsilon)$

w/ $h_2(\varepsilon) = -\varepsilon \log_2 \varepsilon - (1-\varepsilon)\log_2(1-\varepsilon)$

+ satisfying $\lim_{\varepsilon \to 0} h_2(\varepsilon) = 0$

$\Rightarrow$ $I(\mathcal{M}'; \hat{\mathcal{M}})_{\Phi} \leq I(\mathcal{M}'; \hat{\mathcal{M}})_{\omega} + c_1' \varepsilon \log |\mathcal{M}| + c_2' h_2(\varepsilon)$

Now use an inequality called quantum data processing which follows from monotonicity of relative entropy

we get $\hat{\mathcal{M}}$ from the decoder acting on B^n

So $I(\mathcal{M}'; \hat{\mathcal{M}})_{\omega} \leq I(\mathcal{M}'; B^n)_{\rho}$

where $\rho_{\mathcal{M}B^n} = \sum_m \frac{1}{|\mathcal{M}|} |m\rangle\langle m|_{\mathcal{M}'} \otimes \mathcal{N}^{\otimes n}(\rho_A^m)$

define Holevo information of a channel $\mathcal{R}$ as $\chi(\mathcal{R}) = \max_{\{p(x), \rho^x\}} I(X; B)_{\sigma}$

where $\sigma_{XB} = \sum_x p(x) |x\rangle\langle x|_X \otimes \mathcal{R}(\rho^x)$

so $I(M'; B^n)_\rho \leq \chi(N^{\otimes n})$

and we finally get

$$\log |M| \leq \chi(N^{\otimes n}) + c_1' \varepsilon \log |M| + c_2' h_2(\varepsilon)$$

divide by n to get

$$R \leq \frac{1}{n}\chi(N^{\otimes n}) + c_1' \varepsilon R + c_2' \frac{h_2(\varepsilon)}{n}$$

This bound applies for all (n, R, ε) codes

Now take the limit as $n \to \infty$ & then $\varepsilon \to 0$ to get an upper bound on capacity (i.e., on any achievable rate)

$$\Rightarrow R \leq \lim_{n\to\infty} \frac{1}{n}\chi(N^{\otimes n})$$

So we've now reduced the upper bound to the study of the mathematical properties of this function

There is a straightforward upper bound on this quantity called the maximum output entropy. I.e., one can show that

$$\chi(\mathcal{N}^{\otimes n}) \leq n \max_{\rho} H(\mathcal{N}(\rho))$$

$$\Rightarrow C(\mathcal{N}) \leq \max_{\rho} H(\mathcal{N}(\rho))$$

We will show that this upper bound is tight for a special class of channels called pure-state channels.

Suppose that the channel is of the form $x \to |\psi_x\rangle\langle\psi_x|$.

That is, we input a classical letter and get out a pure quantum state.

We will show that the upper bound given above is achievable for this channel.

Non-commutative union bound

Given $\sigma : \sigma \geq 0 \quad Tr\{\sigma\} \leq 1$

& projectors $P_1, \ldots, P_L$

$$Tr\{\sigma\} - Tr\{P_L \cdots P_1 \sigma P_1 \cdots P_L\}$$

$$\leq 2\sqrt{\sum_{i=1}^{L} Tr\{(I-P_i)\sigma\}}$$

Classical union bound.

For events $A_1, \ldots, A_L$

$$Pr\{(A_1 \cap \cdots \cap A_L)^c\}$$

$$= Pr\{A_1^c \cup \cdots \cup A_L^c\}$$

$$\leq \sum_{i=1}^{L} Pr\{A_i^c\}$$

First we need the notion of the typical subspace.

Let ρ be a density operator &

let $\rho = \sum_z p(z) |z\rangle\langle z|$

be a spectral decomposition for it.

Consider that

$$\rho^{\otimes n} = \sum_{z_1} p(z_1) |z_1\rangle\langle z_1| \otimes \cdots \otimes \sum_{z_n} p(z_n) |z_n\rangle\langle z_n|$$

$$= \sum_{z_1,\ldots,z_n} p(z_1)\cdots p(z_n) \left(|z_1\rangle \cdots |z_n\rangle\right)\left(\langle z_1| \cdots \langle z_n|\right)$$

$$\equiv \sum_{z^n} p(z^n) \; |z^n\rangle\langle z^n|$$

Let $\tau^n_{\rho,\delta}$ be the typical subspace defined as

$$\text{span}\left\{ |z^n\rangle : \left| -\frac{\log p(z^n)}{n} - H(\rho) \right| \leq \delta \right\}$$

($-\frac{\log p(z^n)}{n}$: sample entropy; $H(\rho)$: true entropy)

let $\tau^n_{p(z),\delta}$ denote the set of classical sequences z^n satisfying this property

Consider that

$$\frac{-\log p(z^n)}{n} = \frac{1}{n}\sum_{i=1}^{n}\left[-\log p(z_i)\right]$$

↙ sample average

& $H(\rho) = -\mathrm{Tr}\{\rho \log \rho\}$

$$= \sum_z p(z)\left[-\log p(z)\right]$$

↑ true average

define a typical projector as

$$\Pi^n_{\rho,\delta} = \sum_{z^n \in \Upsilon^n_{p(z),\delta}} |z^n\rangle\langle z^n|$$

then one can apply the law of large numbers to say that

$\forall \varepsilon > 0$ & sufficiently large n,

$$\mathrm{Tr}\{\Pi^n_{\rho,\delta}\, \rho^{\otimes n}\} \geq 1 - \varepsilon$$

that is, we have an arbitrarily large probability to project into this subspace.

can prove some other properties

$$2^{-n[H(\rho)+\delta]}\Pi_{\rho,\delta}^{n} \leq \Pi_{\rho,\delta}^{n}\,\rho^{\otimes n}\,\Pi_{\rho,\delta}^{n} \leq 2^{-n[H(\rho)-\delta]}\Pi_{\rho,\delta}^{n}$$

& $\mathrm{Tr}\{\Pi_{\rho,\delta}^{n}\} \leq 2^{n[H(\rho)+\delta]}$

↑ typical subspace

↑ eigenvalue cutoff

is exponentially smaller than all of Hilbert space.

Back to designing a code:

channel is $x \to |\psi_x\rangle$

Shannon's idea was simply to pick a code completely @ random

That is, pick each symbol of each codeword independently @ random according to $p(x)$

$x_1(1)\ x_2(1)\ x_3(1)\ \cdots\ x_n(1)$ ← 1st codeword

$x_1(2)\ x_2(2)\ x_3(2)\ \cdots\ x_n(2)$ ← 2nd codeword

$x_1(|\mathcal{M}|)\ x_2(|\mathcal{M}|)\ \cdots\ x_n(|\mathcal{M}|)$ ← $|\mathcal{M}|$th codeword

abbreviate as

$$x^n(m) = x_1(m)\, x_2(m) \cdots x_n(m) \quad \leftarrow m\text{th codeword}$$

sending $x^n(m)$ through channel leads to

$$\psi_{x^n(m)} = |\psi_{x_1(m)}\rangle\langle\psi_{x_1(m)}| \otimes |\psi_{x_2(m)}\rangle\langle\psi_{x_2(m)}| \otimes \cdots \otimes |\psi_{x_n(m)}\rangle\langle\psi_{x_n(m)}|$$

How to decode? Use a sequential decoder which asks: Was it the first codeword?

Was it the second?

⋮

Was it the mth?

until getting a "hit"

that is, if mth codeword is transmitted, a correct sequence of events would be

Was it 1st? no. Was it 2nd? no.

... Was it $(m-1)$th? no.

Was it mth? yes.

How to ask?

Perform measurement:

$$\left\{ \psi_{x^n(m)} \,,\; I - \psi_{x^n(m)} \right\}$$

$$\psi_{x^n(m)} = \Pi_m \qquad I - \psi_{x^n(m)} = \hat{\Pi}_m$$

Success probability is then

$$\mathrm{Tr}\left\{ \Pi_m \hat{\Pi}_{m-1} \cdots \hat{\Pi}_1 \, \psi_{x^n(m)} \hat{\Pi}_1 \cdots \hat{\Pi}_{m-1} \Pi_m \right\}$$

$\Rightarrow$ error probability is $1 -$ (the above)

$$= \mathrm{Tr}\left\{ \psi_{x^n(m)} \right\} - \mathrm{Tr}\left\{ \Pi_m \hat{\Pi}_{m-1} \cdots \hat{\Pi}_1 \, \psi_{x^n(m)} \hat{\Pi}_1 \cdots \hat{\Pi}_m \Pi_m \right\}$$

Idea: [redacted] code was chosen according to dist. $p(x)$

[redacted] average output density operator is $\sum_x p(x) |\psi_x\rangle\langle\psi_x| = \rho$

So, b/c of way code is chosen, we would expect that on average, codewords are in typical subspace of $\rho^{\otimes n}$ b/c

$$\sum_x p(x^n) \mathrm{Tr}\left\{ \Pi^{\delta}_{\rho^{\otimes n}} |\psi_{x^n}\rangle\langle\psi_{x^n}| \right\} = \mathrm{Tr}\left\{ \Pi^{\delta}_{\rho^{\otimes n}} \rho^{\otimes n} \right\} \geq 1 - \epsilon$$

Instead analyze expectation of average error probability

$$\mathbb{E}_{X^n(1),\ldots,X^n(|\mathcal{M}|)} \left\{ \frac{1}{|\mathcal{M}|} \sum_m \left[\mathrm{Tr}\{\Lambda_{X^n(m)}\} - \mathrm{Tr}\{\Pi_m \hat{\Pi}_{m-1} \cdots \hat{\Pi}_1 \Lambda_{X^n(m)} \hat{\Pi}_1 \cdots \hat{\Pi}_{m-1} \Pi_m\} \right] \right\}$$

$$\approx \mathbb{E}_{X^n(1),\ldots,X^n(|\mathcal{M}|)} \left\{ \frac{1}{|\mathcal{M}|} \sum_m \left[\mathrm{Tr}\{\Pi^n_{\rho,\delta} \Lambda_{X^n(m)} \Pi^n_{\rho,\delta}\} - \mathrm{Tr}\{\Pi_m \hat{\Pi}_{m-1} \cdots \hat{\Pi}_1 \Pi^n_{\rho,\delta} \Lambda_{X^n(m)} \Pi^n_{\rho,\delta} \hat{\Pi}_1 \cdots \hat{\Pi}_{m-1} \Pi_m\} \right] \right\}$$

Apply tool: non-commutative union bound

$$\leq \mathbb{E}_{X^n(1),\ldots,X^n(|\mathcal{M}|)} \left\{ \frac{1}{|\mathcal{M}|} \sum_m 2 \left[\mathrm{Tr}\{(I-\Pi_m)\Pi^n_{\rho,\delta} \Lambda_{X^n(m)} \Pi^n_{\rho,\delta}\} + \sum_{i=1}^{m-1} \mathrm{Tr}\{\Pi_i \Pi^n_{\rho,\delta} \Lambda_{X^n(m)} \Pi^n_{\rho,\delta}\} \right]^{1/2} \right\}$$

use concavity to bring expectations inside square root of square root

$$\leq 2 \left[\mathbb{E}_{X^n(1),\ldots,X^n(|\mathcal{M}|)} \left\{ \frac{1}{|\mathcal{M}|} \sum_m \left[\cdots \right] \right\} \right]^{1/2}$$

Consider that $\mathrm{Tr}\{(I-\Pi_m)\Lambda_{X^n(m)}\} = 0$

can then use typicality to argue that

$$\mathbb{E}\left\{ \mathrm{Tr}\{(I-\Pi_m)\Pi^n_{\rho,\delta} \Lambda_{X^n(m)} \Pi^n_{\rho,\delta}\} \right\} \approx \varepsilon$$

For the other term

$$\sum_{i=1}^{m-1} \mathrm{Tr}\{\Pi_i \cdots\} \leq \sum_{i \neq m} \mathrm{Tr}\{\Pi_i \cdots\}$$

↑ focus on this

$$\mathrm{Tr}\{\Pi_i \Pi_{\rho,\delta}^n \psi_{x^n(m)} \Pi_{\rho,\delta}^n\} = \mathrm{Tr}\{\psi_{x^n(i)} \Pi_{\rho,\delta}^n \psi_{x^n(m)} \Pi_{\rho,\delta}^n\}$$

From the way that code is chosen, consider that random variables $X^n(i)$ & $X^n(m)$ are independent when $i \neq m$

$$\Rightarrow \mathbb{E}\{\cdot\} = \mathrm{Tr}\{\rho^{\otimes n} \Pi_{\rho,\delta}^n \rho^{\otimes n} \Pi_{\rho,\delta}^n\}$$

$$\leq 2^{-n[H(\rho)-\delta]} \mathrm{Tr}\{\rho^{\otimes n} \Pi_{\rho,\delta}^n\}$$

$$\leq 2^{-n[H(\rho)-\delta]}$$

$$\Rightarrow \text{second term is} \leq |\mathcal{M}| 2^{-n[H(\rho)-\delta]}$$

& overall bound is

$$2\left[\varepsilon + |\mathcal{M}| 2^{-n[H(\rho)-\delta]}\right]^{1/2}$$

Pick $|\mathcal{M}| = 2^{n[H(\rho)-2\delta]} \Rightarrow 2\left[\varepsilon + 2^{-n\delta}\right]^{1/2}$

$\to 0$ as $n \to \infty$

We have a bound on expectation of average error probability

$\Rightarrow \exists$ some code w/ average error probability $\leq 2\left[\varepsilon + 2^{-n\delta}\right]^{1/2}$

& rate is $\frac{\log |\mathcal{M}|}{n} = H(\rho) - 2\delta$

now remove half the codewords and we get a code w/ maximal error prob.

$\leq 4\left[\varepsilon + 2^{-n\delta}\right]^{1/2}$

& rate $\frac{\log\left[|\mathcal{M}|/2\right]}{n} = H(\rho) - 2\delta - \frac{1}{n}$

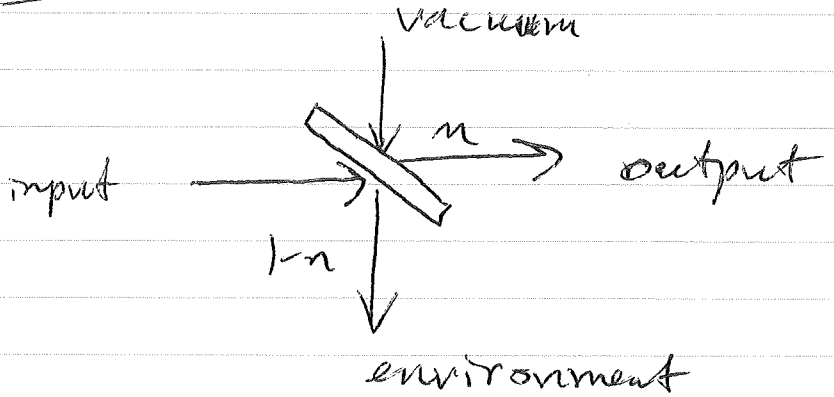
↑ becomes arbitrarily small as $n \to \infty$.

So we have a sequence of codes w/ $\varepsilon \to 0$ as $n \to \infty$ and rate approaches $H(\rho)$. Can take maximum over input dist's & rate is $\max_{p(x)} H\left(\sum_x p(x) f_x(\rho)\right)$

this matches the upper bound.

So we now know capacity for pure-state channels

Important example: optical beamsplitter channel



$n \in [0, 1]$

↑ fraction of input photons that make it to output on average.

simple model of photon loss for free space communication

"quantum Gaussian channel"

special state is coherent state

$$|\alpha\rangle = e^{-\frac{1}{2}|\alpha|^2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle$$

↑ photon number states

these states are good for classical communication $\alpha \in \mathbb{C}$ b/c beamsplitter preserves their purity:

$$|\alpha\rangle|0\rangle \rightarrow |\sqrt{n}\alpha\rangle|\sqrt{1-n}\alpha\rangle$$

We can induce a pure-state channel in this way:

$$\alpha \in \mathbb{C} \rightarrow |\sqrt{n}\,\alpha\rangle$$

Pick codewords according to a Gaussian distribution

$$P_{N_S}(\alpha) = \frac{1}{\pi N_S} \exp\left\{ \frac{-|\alpha|^2}{N_S} \right\}$$

where $N_S \in (0, \infty)$

if there is a mean photon number constraint, then pick N_S to match this.

the average density operator for code is then

$$\int P_{N_S}(\alpha)\, |\sqrt{n}\,\alpha\rangle\langle\sqrt{n}\,\alpha|\, d^2\alpha$$

$$= \underbrace{\frac{1}{N_S+1} \sum_{n=0}^{\infty} \left(\frac{N_S}{N_S+1} \right)^n |n\rangle\langle n|}_{\text{thermal state.}} = \Theta(N_S)$$

then

$H(\Theta(N_S))$ is achievable rate $\forall$

$$= (N_S+1) \log (N_S+1) - N_S \log N_S$$

How to implement sequential decoder?

Consider that they are tests of the following form

$$\left\{ |\alpha^n\rangle\langle\alpha^n| \,,\; I^{\otimes n} - |\alpha^n\rangle\langle\alpha^n| \right\}$$

where $|\alpha^n\rangle = |\alpha_1\rangle \otimes |\alpha_2\rangle \otimes \cdots \otimes |\alpha_n\rangle$

Define displacement operator

$$D(\alpha) = \exp\left(\alpha \hat{a}^\dagger - \alpha^* \hat{a}\right)$$

$\hat{a}^\dagger$ - creation
$\hat{a}$ - annihilation

can show that $|\alpha\rangle = D(\alpha)|0\rangle$

$$\Rightarrow |\alpha^n\rangle = \left[D(\alpha_1) \otimes \cdots \otimes D(\alpha_n)\right] |0\rangle^{\otimes n}$$

↑ vacuum

can implement displacement operators w/ linear optics & using that

$$|\alpha^n\rangle\langle\alpha^n| = D(\alpha^n)\left(|0\rangle\langle 0|\right)^{\otimes n} D(\alpha^n)^\dagger$$

we see that we've reduced to having to implement a vacuum or not measurement.

$$\left\{ \left(|0\rangle\langle 0|\right)^{\otimes n} \,,\; I^{\otimes n} - \left(|0\rangle\langle 0|\right)^{\otimes n} \right\}$$