AQIT

2025-26

[Axioms of QM?] **Lecture 1**

Def. Quantum Channel.

A completely positive, trace preserving (CPTP) map $N: L(A) \to L(B)$.

Positive: $N(\rho) \geq 0$, if $\rho \geq 0$.

Note that $A \geq 0$ means $\langle\psi|A|\psi\rangle \geq 0 \ \forall |\psi\rangle$,

and $A \geq B$ means $A - B \geq 0$.

Completely Positive: $N_A \otimes I_B(\rho_{AB}) \geq 0$ if $\rho_{AB} \geq 0$; for all B systems.

Trace Preserving: $\mathrm{Tr}\, N(\rho) = \mathrm{Tr}\, \rho \ \ \forall \rho$.

Elementary Elements.

- The identity map

$$I(X) = X.$$

[AKA $id(x) = X$.]

Identity Matrix: $\mathbb{1}$. Identity map I or id.

$$I(\rho) = \mathbb{1}\rho\mathbb{1} = \rho.$$

- Isometry $V: A \to B$.

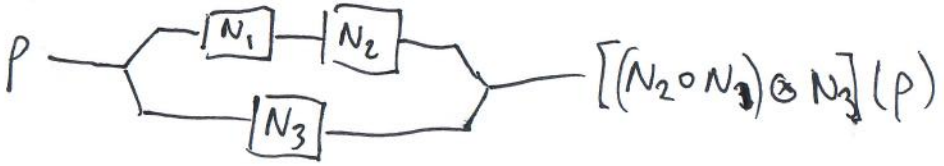
$$V^\dagger V = \mathbb{1}_A, \quad VV^\dagger = \Pi_{B(A)}.$$

- Trace and Partial Trace.

$$\mathrm{Tr}_A \rho_{AB} = \blacksquare\, \rho_B$$

- Compositions & Tensor Products

$N \circ M$ $\qquad$ $N \otimes M$



- Convex Combos.

$$N = \sum_i p_i N_i.$$

Thm. The following are equivalent.

1. N is CPTP

2. N is TP and $I_{A'} \otimes N$ is positive where $N: A \to B$ and $A' \equiv A$.

3. (Choi). ~~$J(N) = (I_{AB} N_{A \to B}) \Phi_{A'A}$~~ is positive and $\mathrm{tr}_B J(N) = \frac{1}{|A|} \mathbb{1}_A$.

3. The Choi operator $J(N)^{AB}$ is positive $J(N) \geq 0$ and $\mathrm{tr}_B J(N) = \frac{\mathbb{1}_A}{|A|}$.

$$J(N) = I_A \otimes N_{A' \to B} \Phi_{AA'}$$

$\Phi = |\Phi\rangle\langle\Phi|$ where $|\Phi\rangle = \frac{1}{\sqrt{N}} \sum_i |i\rangle_A \otimes |i\rangle_{A'}$

$\Phi = \frac{1}{N} \sum_{ij} |i\rangle\langle j| \otimes |i\rangle\langle j|$.

4. (Stinespring)

$N(\rho) = \mathrm{Tr}_E(V \rho V^\dagger)$ for some isometry V.

5. (Kraus).

$$N(\rho) = \sum_\alpha K_\alpha \rho K_\alpha^\dagger \,, \quad \sum_\alpha K_\alpha^\dagger K_\alpha = \mathbb{1}.$$

Pf.

(Do proofs on board.)

[Cyclic proof from Andreas's lecture notes, see main lecture notes PDF.]

Examples.

- Identity ✓
- Constant/Preparation Channel

$$N_\sigma(\rho) = \mathrm{tr}(\rho)\sigma$$

- Measure & Prepare (POVM $M = \{M_x\}_x$).

$$N(\rho) = \sum_x \mathrm{tr}(M_x \rho)\,\sigma_x$$

- Depolarizing (Mixing w/ Noise)

$$N(\rho) = (1-\lambda)\rho + \lambda \frac{\mathbb{1}}{d}$$

- Dephasing

$$N(\rho) = (1-\lambda)\rho + \lambda \sum_k |k\rangle\langle k|\rho|k\rangle\langle k|$$

- Qubit Pauli Error ✓

$$N(\rho) = p_0\rho + p_x X\rho X + p_y Y\rho Y + p_z Z\rho Z.$$

- Erasure channel.

$$N(\rho) = (1-\lambda)\rho + \lambda |\perp\rangle\langle\perp|$$

Classical channels.

A classical channel is a map

$$\Lambda(p) = \sum_i \Lambda_{j|i} p_i,$$

where $\Lambda_{j|i} \geq 0$ & $\sum_j \Lambda_{j|i} = 1$.

- "Stochastic Matrix"
- Conditional probability
- Maps prob to prob.
- Can embed classical prob theory as diagonal density matrices.

CQ channel: "Classical-Quantum"

$$N(\rho) = \sum_x \mathrm{tr}(\rho |x\rangle\langle x|)\,\sigma_x$$

Maps classical label x ($\equiv |x\rangle\langle x|$) to state ρ_x

QC Channel:

$$N(\rho) = \sum_x \mathrm{tr}(\rho M_x)\,|x\rangle\langle x|$$

After break: Adjourds.

IP extra time

- Classical Comm
- EA Comm
- Quantum Comm
- 1-Shot vs. Asymptotic.

Lecture 2

~~Advanced~~ ~~Quantum~~ Information Theory

- I promise we'll get very advanced and quantum. But not today. Today: Info Theory.
- Quantum theory is actually much older than Information Theory (1948). Even though "Quantum" sounds fancier.

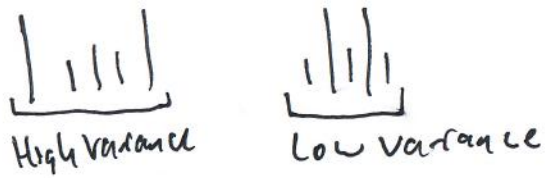
Info theory was an amazing breakthrough! Before Shannon it was not clear that
- Info could be unambiguously quantified.
- Channels have a definite capacity for transmission.

Shannon (1948) showed these are fundamental, not just practical, bounds.

So the topic of today's lecture:

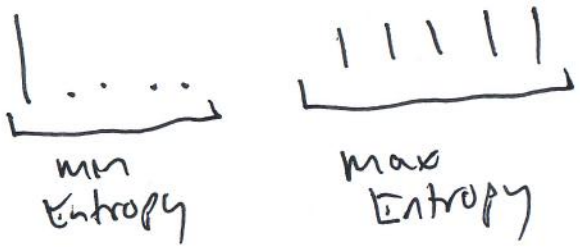
Shannon Entropy & Information.

① Shannon Entropy as uncertainty in a probability distribution.



Same Entropy

Entropy doesn't care about outcome space only probabilities.



As a first try to capture uncertainty, consider

$$\langle p_i \rangle = \sum_i p_i^2.$$

If $p_i = \delta_{ij}$ then $\langle p_i \rangle = 1$. [Large expected prob.]

If $p_i = \frac{1}{N}$ then $\langle p_i \rangle = \frac{1}{N}$. [Small expected prob.]

This is already a good uncertainty measure.

Why $H(p) = -\sum_i p_i \log p_i$?

We will see 3 reasons...

↳ Axiomatic Derivation.

↳ How many bits needed to represent outcomes?

↳ Asymptotic message rate.

By the way, $S(\rho) = -\mathrm{tr}\rho \log \rho = H(\text{e-vals})$.

Ⓐ Axiomatic.

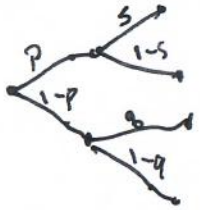
Shannon derives $H(p)$ from 3 axioms.

1. Monotonicity

$H(\frac{1}{n}, \dots, \frac{1}{n}) = A(n)$ monotone of n.

2. Continuity in p.

3. Chain Rule



$= H(\{p, 1-p\}) + p\, H(\{s, 1-s\}) + (1-p)\, H(\{q, 1-q\})$.

Thm. (Shannon) These axioms enforce

$$H(p) = -k \sum_i p_i \log p_i .$$

Pf. Shannon's Paper.

(See also Renyi — grown up chain rule.)

This is nice, but the practical reasons are more powerful.

How Many Bits?

- Shannon Entropy $H(p) = -\sum_i p_i \log p_i$

"# bits to encode outcomes drawn from p with optimal encoding"

- Cross Entropy $H(p;q) = -\sum_i p_i \log q_i$

"# of bits to encode from p when optimized for q"

- Relative Entropy $D(p||q) = H(p;q) - H(p)$

"The cost of non-optimal encoding"
# of extra bits.

Goal: Write down English text efficiently as binary strings.

26 Letters a, b, c, ..., x, y, z

Simplest Strategy — label using binary numbers $0 \to 26$ with $\lceil \log_2 26 \rceil = 5$ bit strings.

a → 00000
b → 00001
⋮
y → 11001
z → 11010
[space] → 11011
. → 11100
⋱

a → 00000

letters

codewords

But this is inefficient!

- Uses: 5 bits per letter.

Why?

- Some letters (a, e, t) are common, others (q, z) are rare.

Strategy... use short strings for common letters.

a → 001
q → 11101011

Let's be more precise.

Consider a league of 5 teams. You will write down the winner each year, in binary, to store records.

A	B	C	D	E
.1	.4	.3	.1	.1

← Win probabilities

$H(p) \approx 2.05$ bits

The naive strategy is to use $\lceil \log_2 5 \rceil = 3$ bits per letter.

Here's a better strategy

$A \to 00$
$B \to 01$
$C \to 10$
$D \to 110$
$E \to 111$

This uses, on average,

$\langle L \rangle = \sum p_i L_i = .8 * 2 + .2 * 3$

$= 2.2$ bits/letter

That's nearly optimal.

Let's prove optimal length.

Lemma. (Kraft-McMillan). For any uniquely decodable code, with codeword lengths L_i, the lengths obey

$$\sum_i 2^{-L_i} \leq 1.$$

Pf. CLRS Thm 5.2.1, 5.5.1.

Thm. The expected codeword length

$$L = \langle L_i \rangle = \sum_i p_i L_i$$

is $\geq H(p)$, the Shannon Entropy.

Pf.
$$L = \sum_i p_i L_i = \sum_i p_i (-\log 2^{-L_i})$$

$$= -\sum_i p_i \log \frac{2^{-L_i}}{\sum_i 2^{-L_i}} - \log \underbrace{\sum_i 2^{-L_i}}_{\leq 1}$$

$$\geq -\sum_i p_i \log r_i \quad \leftarrow \text{Reference (Bad) Distribution}$$

$$= H(p) + D(p||r) \quad \leftarrow \text{Penalty}$$

$$\geq H(p).$$

Thm. The optimal codeword length is between

$$H(p) \le L \le H(p)+1$$

Pf. C&T.

Now what if we look at long strings? The source is

$$p^n(x^n) = \prod_k p(x_k),$$

with entropy

$$H(p^n) = nH(p).$$

If we do "block coding", the # of bits per letter is

$$\frac{1}{n}H(p^n) \le L \le \frac{1}{n}(H(p)+1)$$

$$H(p) \le L \le H(p) + \frac{1}{n},$$

with optimal scheme.

Coding long blocks, expected length is

$$L \approx H(p)$$

Bits per letter.

(C) Asymptotic # of messages.

- Consider strings of n messages drawn from $p(x)$.

$$\hat{p}(x^n) = p(x_1)p(x_2)\cdots p(x_n).$$

This is "IID".

- Law of large numbers says

$$\frac{1}{n}\sum_k x_k \to \langle x \rangle \text{ as } n \to \infty,$$

for any IID x_k.

(Prob(DIFF $\ge \epsilon$) $\to 0$).

- What is probability of a given string?

$$-\frac{1}{n}\log p^n(x^n) = \frac{1}{n}\sum_k (-\log p(x_k))$$

$$\xrightarrow[n\to\infty]{} \langle -\log p(x) \rangle = H(p).$$

The prob that a string obeys

$$-\log p^n(x^n) \underset{\epsilon}{\approx} nH(p)$$

goes to one.

Summary. Repeatedly drawing letters from $p(x)$ produces

$2^{nH(p)}$

equally likely "typical" messages, each with probability

$2^{-nH(p)}$

Any other message has "negligible" probability.

Pf. CLT (Asymptotic Equipartition). Add ϵ's, δ's to make precise.

The Typical Set.

- Typical messages have to have the "right number" of each letter.
- In english "A" is likely and "Z" unlikely but AAAA & ZZZZ are both atypical (highly unlikely).
- Typical: ALHENHTTPA OOBTTVA

So...

Having n copies of a source $p(x)$ is like having a uniform source of $2^{nH(p)}$ different messages.

An aside...

Back to writing down numbers term in binary.

- Suppose you initially know nothing. You simply have to guess all equally likely. $\log N$ bits per outcome.
- Now you're given distribution $p(x)$ (the new probabilities). This saves you

$$D(p\|1/N) = \log N - H(p)$$

bits per letter.

- "Destruction is worth $D(p\|1/N)$ bits per letter."

- Now you get the actual outcome of the league, already knowing $p(x)$. This gives you

$$H(p)$$

more bits.

"Outcome worth $H(p)$ bits".

- What if you had wrong guess at p, but got to update it?

- (4) How Much Information is in a book?

  - Depends on decoding mechanism.
  - Go to projector? Show Shannon papers.

(5) If time remaining, Review either:

- Classical Channels
- Relative Entropy Props

In prep for next lecture.

Lecture 3

① Intro

Goal: Create Mutual Information between Sender and receiver.

---

② Prob. Theory Notation

$P_X(x)$ ← Particular Value
↖ Distribution

X = Random Variable
= Function of outcome of $P(x)$.

Joint Distribution

$$P_{XY}(x,y)$$

Marginal Distribution

$$P_X(x) = \sum_y P_{XY}(x,y)$$

Conditional

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

I will sometimes just write

P_{XY}, $P_{X|Y}$, P_X, etc, for shorthand.

$H(X) = H(P_X)$

$H(X|y) = H(X|Y=y) = H(P_{X|y}(x|y))$ ← fixed

$H(X|Y) = \sum P_Y(y)\, H(X|y)$

$H(X,Y) = H(P_{X,Y}(x,y))$

---

③ Conditional Probability.

A conditional prob dist

$$P_{X|Y}(x|y)$$

is a P.D. over x for each y

"Prob of x, given y."

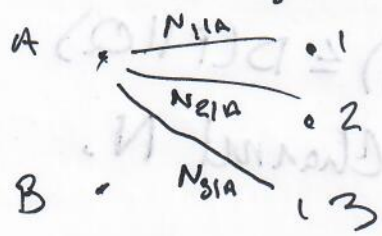
Thus $p(x|y) \geq 0 \quad \forall x, y.$

$$\sum_x P(x|y) = 1, \quad \forall y.$$

Also known as a "Stochastic Matrix" or "Stochastic Channel".

(4) A Classical Channel is defined by a conditional probability distribution from inputs to outputs.

$$N_{y|x}$$



A map from P.D.'s to P.D.'s:

$$q = N(p)$$

means

$$q_y = \sum_x N_{y|x} p_x .$$

It is a positive ~~trace-preserving~~ sum linear map.

(5) Markov Chain

$$X \to Y \to Z$$

is a markov chain if

$$p(x,y,z) = p(x)\, p(y|x)\, p(z|y)$$

Fact. This implies also $Z \to Y \to X$.

Pf. C&T.

(6) RE & Monotonicity

Relative Entropy:

$$D(P\|Q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

"How different ARE P, Q?"

Thm. (Monotonicity)

$$D(N(P)\|N(Q)) \leq D(P\|Q)$$

For Any classical Channel N.

(7) Mutual Information.

Joint distribution $P_{XY}(x,y)$.

$$I(X:Y) \equiv D(P_{XY}\|P_X P_Y)$$

Fact.

$$I(X:Y) = H(X) + H(Y) - H(X,Y)$$

$$= H(X) - H(X|Y)$$

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

"How different is P_{XY} from its marginals"

"How much do X, Y tell about each other?"

(8) Data Processing Inequality

Thm. If $X \to Y \to Z$ is a Markov chain, then $I(X:Z) \le I(X:Y)$.

Cor. Also, $I(X:Z) \le I(Y:Z)$.

~~Pf. Consider the channel ... $N(P) = \sum_y N$~~

Pf. Consider channel $N_{Y\to Z}$ defined by $N_{Z|Y} = P_{Z|Y}$ from the Chain.

$$P_{XYZ} = P_X P_{Y|X} P_{Z|Y}$$

~~$P_{XZ} = \sum_y P_{XYZ} = \sum_y P_{Z|Y}(P_X P_{Y|X})$~~ * Cor. b/c $X \to Y \to Z$.

$$P_{XY} = \sum_z P_{XYZ} = P_X P_{Y|X}$$

$$N(P_{XY}) = \sum_y P_{Z|Y} P_X P_{Y|X} = \sum_y P_{XYZ} = P_{XZ}$$

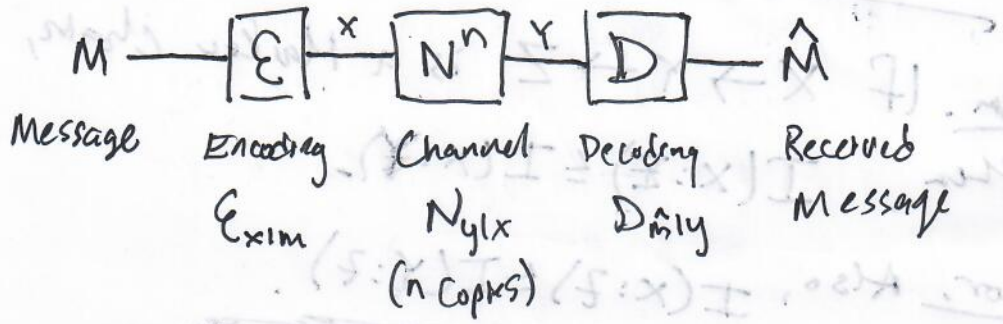
Similarly, $N(P_X P_Y) = P_X P_Z$.

* FACT. Monotonicity!

Thus $I(X:Z) = D(N(P_{XY}) \| N(P_X P_Y))$

$\le D(P_{XY} \| P_X P_Y) = I(X:Y)$

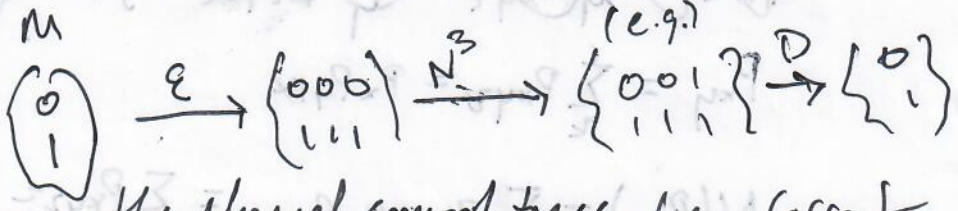
(1) The Communication Problem.



Question. At what Rate can we send messages Reliably using the channel N.

Example. Consider Bit Channel ~~Perfect~~ with error:

No single bit can be sent reliably in a single shot! But...



Using the channel several times, I can correct errors. This is a rate of

$$\frac{\log |M|}{n} = \frac{\log 2}{3} = \frac{1}{3} \text{ bits/signal}$$

Best possible Rate with negligible errors?

② Channel Capacity

Def. The capacity of channel N is the maximum amount of mutual information between the channel input & channel output,

$$C_N = \sup_{P_X} I(X : N(X)).$$

(2) Channel Capacity.

- Consider the mutual information between sent message and received message

$$I(M:\hat{M}) = \underbrace{H(M)}_{\text{Input Rate}} - \underbrace{H(M|\hat{M})}_{\text{Equivocation}}.$$

- The equivocation relates to Probability of error via Fano's Inequality:

$$H(M|\hat{M}) \leq 1 + P_e \log|M|,$$

where $P_e = \sum_m P_m P_e(m) = \sum_m P_m (1 - P(m|m))$.

- By D.P.I., since $M \to X^n \to Y^n \to \hat{M}$, clearly

$$\underbrace{I(M:\hat{M})}_{\text{Depends on encoding \& Decoding.}} \leq \underbrace{I(X^n:Y^n)}.$$

- But for <u>any</u> encoding/decoding,

$$I(X^n:Y^n) \leq n C_N,$$

where $C_N = \sup_{P_X} I(X:N(X))$ across a single use of the channel. This is the <u>channel capacity</u>.

- So for any $Q(D)$,

$$nC_N \geq I(M:\hat{M}) = H(M) - H(M|\hat{M}).$$

- Suppose $p_M(m)$ is uniform over $|M| = 2^{nR}$ messages [Rate of R message bits per use of channel].

The worst case probability of error

$$P_e^* = \max_m P_e(m) \geq P_e^{\text{uniform}} = \sum_m \frac{1}{2^{nR}} P_e(m).$$

We want an encoding ensuring <u>all</u> messages get through successfully, so P_e^* is small.

- Combining above with Fano, using uniform input

$$nC_N \geq H(m) - H(M|\hat{M})$$

$$\geq \log|M| - (1 + P_e^{\text{uniform}} \log|M|)$$

$$\geq nR - 1 - P_e^{\text{uniform}} nR$$

$$\geq nR - 1 - P_e^* nR$$

or...

$$nR P_e^* \geq nR - nC - 1$$

$$\boxed{P_e^* \geq 1 - \frac{C}{R} - \frac{1}{n}}$$

*For $R > C$, small P_e^* becomes impossible!

Thm. (Shannon Noisy Channel Coding.)

- For any rate $R < C$, it is possible to transmit across the channel with negligible probability of error. (Achievability)

- For any rate $R > C$, ~~[illegible]~~ the probability of errors is at least $P_e^{(n)} \geq 1 - \frac{C}{R} - \epsilon$, $\forall \epsilon > 0$. (Converse).

- (Strong Converse). In fact, if $R > C$ the probability that the whole message is correct $\to 0$.

Pf. We already proved converse, we won't prove strong converse. But suppose you transmit $R_0 = C$ successfully and throw out $R_1 = R - C$. Then

$$P_{succ} = \frac{2^{nC}}{2^{nR}} = \frac{1}{2^{n(R-C)}}$$

which $\to 0$ exponentially with n.

The hard part is achievability.

(3) Achievability

- The capacity is defined as

$$C_N = \sup_{P_X} I(X:N(X))$$

which is really

$$\sup_{P_X} D(P_X N_{y|x} \| P_X \sum_x N_{y|x} P_x)$$

$$= \sup_{P_X} D(P_X N_{y|x} \| P_X Q_Y), \quad Q = N(P).$$

(Joint dist over input & output; input; output.)

~~• Need to find a code that achieves capacity.~~

~~• Strategy: Random Encoding~~
~~Jointly Typical Decoding.~~

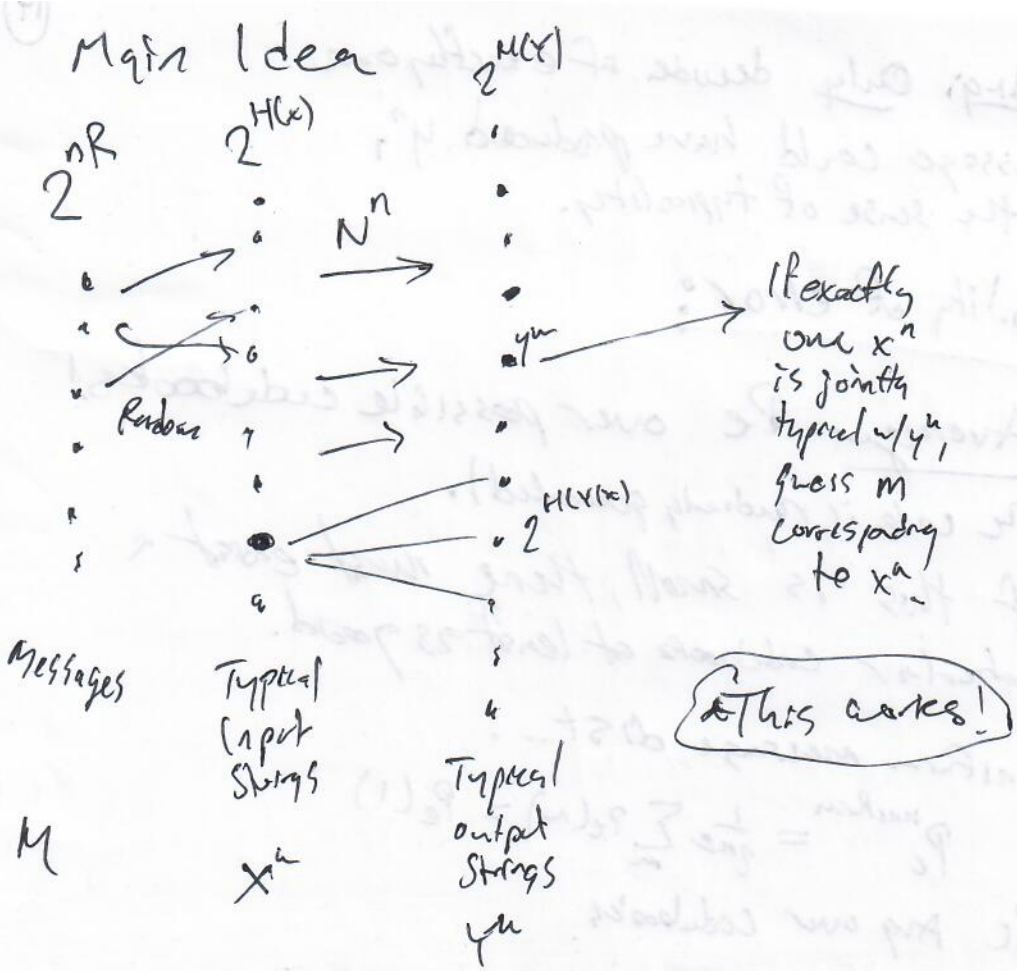
- Need to find a code that achieves the capacity.
- Strategy: • Random Coding
  • Jointly Typical Decoding

- Joint Typicality
  - x^n, y^n are jointly typical if x^n typical of $p(x)$, y^n typical of $p(y)$, and x^n, y^n typical of $p(x,y)$.
  - Drawing IID from $p(x,y)$, almost always gives jointly typical seq.
  - The probability that two independent marginally typical $\tilde{x}^n, \tilde{y}^n$ are jointly typical is $2^{-nI(x;y)}$.

- We need to find a code that achieves the capacity.

~~Strategy~~ : Random Coding
: Jointly typical Decoding.

Main Idea



Encoding: For each m, generate codeword $c(m)$ from $P(x)$, IID, where $P(x)$ is capacity-achieving input rate.

Transmission: Codeword x^n transmits to y^n with probability $P(y^n|x^n) = P(y_1|x_1)\ldots P(y_n|x_n)$.
So together x^n, y^n drawn from
$P(x^n)P(y^n|x^n) = P(x,y)^n$.

- Decoding: Only decode if exactly one message could have produced y^n, in the sense of typicality.

Probability of error:

- Average P_e over possible codebooks! (The code is randomly generated).
- If this is small, there must exist a particular codebook at least as good.
- Uniform message dist...

$$P_e^{\text{uniform}} = \frac{1}{2^{nR}} \sum_m P_e(m) = P_e(1)$$

b/c Avg over codebooks.

- Calculate $P_e(1)$.

$$P_{\text{decode}} = P_{x^n(m) \text{ jointly typical w/ } y^n} - P_{\text{Another } x^n \text{ is too}}$$

$$\geq 1 - 2^{n(R-I)} \cdot 2^{-nI} - \epsilon$$

$$\geq 1 - 2^{-n(I-R)}$$

$$P_e^{\text{uniform}} \leq 2^{-n(I-R)}.$$

which $\xrightarrow{n\to\infty} 0$ if $R < I$. ∎

- Expurgate half! $P_e^* \leq 2 P_e^{\text{uniform}}$, $R' = R - \frac{1}{n}$. ∎

(99) Relative Entropy Review

- Classical RE. (KL Divergence).

$$D(p\|q) = \sum_i p_i \log \frac{p_i}{q_i}.$$

Key Props.

① Non-Negativity

$$D(p\|q) \geq 0,$$

with equality iff $p = q$.

② Joint Convexity

$$D\left(\sum_k \lambda_k p_k \middle\| \sum_k \lambda_k q_k\right) \leq \sum_k \lambda_k D(p_k\|q_k)$$

③ Monotonicity

$$D(N(p)\|N(q)) \leq D(p\|q)$$

for any classical channel N.

④ Chain Rule

$$D(p_{xy}\|q_{xy}) = D(p_x\|q_x) + \sum_x p_x D(p_{y|x}\|q_{y|x}).$$

- Quantum RE

$$D(\rho\|\sigma) = \operatorname{tr}(\rho \log \rho - \rho \log \sigma)$$

Key Props.

① Non-Negativity

$$D(\rho\|\sigma) \geq 0$$

Equality IFF $\rho = \sigma$.

② Joint Convexity

$$D\left(\sum_k d_k \rho_k \,\middle\|\, \sum_n d_n \sigma_n\right) \leq \sum_n d_n D(\rho_k\|\sigma_k)$$

③ Monotonicity

$$D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) \leq D(\rho\|\sigma)$$

For any CPTP map $\mathcal{N}$.

No cloning Rule.

These Props are your new best friend.

• Other divergences: Renyi entropy.

Notice that

$$D(p\|q) = \left\langle -\log\left(\frac{p}{q}\right)\right\rangle_p$$

You could similarly try

$$-\log\left\langle \frac{p}{q}\right\rangle_p .$$

More generally what about

$$-\log\left\langle \left(\frac{p}{q}\right)^s\right\rangle^{1/s} ?$$

This is the Renyi α-Divergence

$$D_\alpha(p\|q) = -\log\left\langle \left(\frac{p}{q}\right)^s\right\rangle_p^{1/s}$$

where $\alpha = s-1$.

• Everything is Relative entropies.

$$H(\rho) = -D(\rho \| \mathbb{1})$$

$$I(A:B)_\rho = D(\rho_{AB} \| \rho_A \otimes \rho_B)$$

$$H(A|B)_\rho = -D(\rho_{AB} \| \rho_A \otimes \mathbb{1}_B)$$

And so on...

• Later we'll also see Generalized Divergences... which obey similar properties.

Lecture 4

Goals; Understand:

- Definition of one-shot comm. theory (plain and E.A.).
- ~~understand~~ Rate, Error Probability, codes.
- Basic Properties & Examples of oneshotrates.
- Superdense Coding
- Propagating Info through the diagram.

---

① One Shot Communication

- You are sending messages using the channel N only once! (Not Asymptotic $N \to \infty$).
- You want to send one from a set of K messages.

Rate: $R = \log K$

(K messages $\leftrightarrow$ $\log K$ $\frac{\text{message}}{\text{bits}}$).

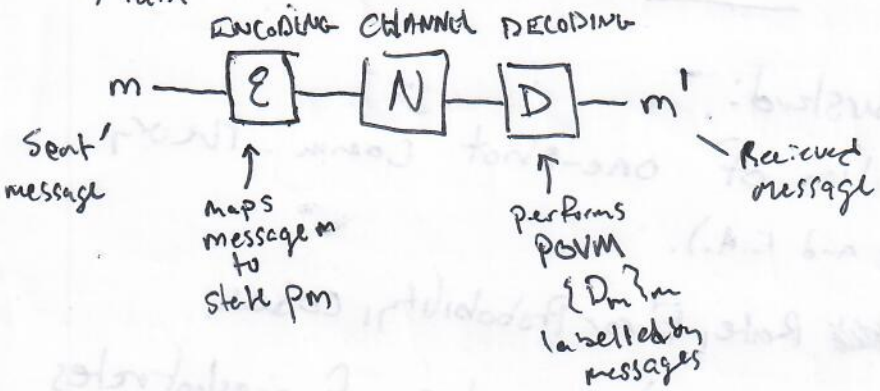
- Is it possible?

No! At least, not without error. (unlike $N \to \infty$).

- The question: What Rate R can you send messages at error level $P_e = ?$

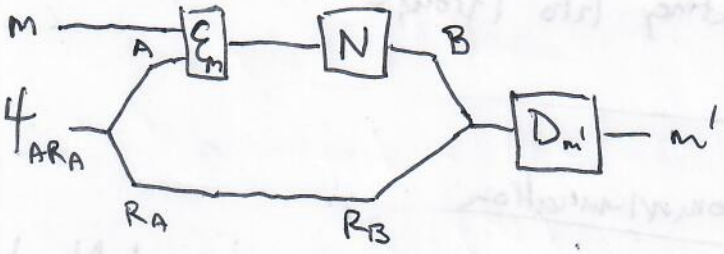
- Two Protocols: (like Asymptotic case)

Plain



Entanglement Assisted

Now $\mathcal{E}_m$ is modulation.



- Different ways to define $\mathcal{E}, D$ are around, but you can always just treat both as channels ($\mathcal{E}$ is $CQ \to Q$, D is QC).

We'll treat both in detail.

② Plain Classical Comm.

~~The encoder E maps the message~~

- The encoder $\mathcal{E}$ maps messages to quantum states,

$$m \xrightarrow{\mathcal{E}} \rho_m = \mathcal{E}(m).$$

You can think of this as a quantum channel, where the initial classical message is

$$m \longrightarrow |m\rangle\langle m|.$$

But often defined as map from alphabet to states.

- The decoder is a POVM but you can think of it as a qc channel

$$\mathcal{D}(\rho) = \sum_{m'} \mathrm{tr}(D_{m'}\rho)\,|m'\rangle\langle m'|.$$

- The probability to recieve m' given m is

$$\boxed{P(m'|m) = \mathrm{tr}\left(D_{m'} N(\rho_m)\right)}.$$

~~• Given uniform input distribution $P(m) = \frac{1}{K}$, the success/failure probability is~~

- Given uniform input dist, $p_m = \frac{1}{K}$, the Probability of error (unknown, average) is

$$\boxed{P_e = 1 - \frac{1}{K}\sum_m \mathrm{tr}\left(D_m N(\rho_m)\right)}$$

- For a given input size K (rate $R = \log K$), the optimal error probability is

$$P_e(R) = \inf_{\substack{(\mathcal{E},\mathcal{D}) \\ \text{with message} \\ \text{size } K = 2^R}} P_e .$$

More messages, higher error.

- We're interested in the complement of this,

$$C_\epsilon(N).$$

Highest rate possible given $P_e \leq \epsilon$.

This is the ϵ-one-shot capacity.

- Now formal definitions.

Def. A code for N is a pair $(\mathcal{E}, \mathcal{D})$. The rate of the code is $R = \log K$, with K size of message space. The (average, uniform) error probability is:

$$P_e = 1 - \frac{1}{K} \sum_m \mathrm{tr}\left(\mathcal{D}_m N(\mathcal{E}_m)\right).$$

Def. The ϵ-one-shot capacity of N is:

$$C_\epsilon(N) = \sup_{\substack{(\mathcal{E},\mathcal{D}) \\ \text{s.t.} \\ P_e \leq \epsilon}} R(\mathcal{E},\mathcal{D}).$$

Largest Rate For fixed error probability.

- What is $C_0(N)$ for B.S.C.?

Examples.

- Identity channel I_A.

  Let $K = d_A$. Let $\mathcal{E}(m) = |m\rangle\langle m| \equiv \rho_m$.

  Let $\mathcal{D}_m = |m\rangle\langle m|$. $\longrightarrow P_e = 0$. Thus

  $$C_\epsilon(I_A) \geq C_0(I_A) \geq \log d_A.$$

On the other hand, suppose K messages and error $P_e \le \epsilon$. Then for any such $(\mathcal{E}, \mathcal{D})$

$$\frac{1}{K}\sum_m tr(D_m \rho_m) \le \frac{1}{K}\sum_m tr(D_m) = \frac{1}{K} tr\, \mathbb{1}_A = \frac{d_A}{K}$$

$$\vee\!\mid$$

$$1-\epsilon.$$

Thus $\frac{d_A}{K} \ge 1-\epsilon$, so $R \le \log d_A - \log(1-\epsilon)$

So

$$\boxed{\log d_A \le C_\epsilon(I_A) \le \log d_A - \log(1-\epsilon)}.$$

- Constant Channel P_σ.

Intuitively, transmit nothing — in/out are independent

But even broken clock is right 2x/day.

For any $(\mathcal{E}, \mathcal{D})$,

$$P_e = 1 - \frac{1}{K}\sum_m tr(D_m \sigma) = 1 - \frac{1}{K}.$$

Thus ~~$C_\epsilon(P_\sigma) \dots \log \dots$~~

$$R \le \log \frac{1}{1-\epsilon}$$

A code exists, ~~such that~~ whenever $R \le \log \frac{1}{1-\epsilon}$.

$$\boxed{C_\epsilon(P_\sigma) = \log \left\lfloor \frac{1}{1-\epsilon} \right\rfloor}.$$

These examples are outliers... usually difficult to evaluate.

- Now an important "Data Processing" type result.

Prop. One shot capacity is monotonic under pre- and post- processing, and also under increasing ϵ. Namely,

$$C_\epsilon(B \circ N \circ A) \le C_\epsilon(N) \le C_{\epsilon' \ge \epsilon}(N).$$

Proof. Convert code for $B \circ N \circ A$ into a code for N with same Rate/P_e.

Let $(\mathcal{E}, \mathcal{D})$ be code for $B \circ N \circ A$.

$$\mathcal{E}(m) = \rho_m, \quad \mathcal{D} = \{D_m\}.$$

Then let $(\mathcal{E}', \mathcal{D}')$ be a code for N,

$$\mathcal{E}'(m) = A(\rho_m) \leftarrow \text{Encoding map} \checkmark$$

$$D'_m = B^\dagger(D_m) \leftarrow \text{POVM} \checkmark.$$

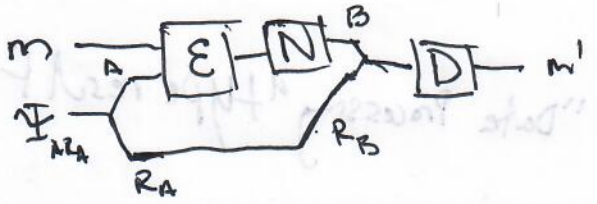
But

$$\operatorname{tr}[B^\dagger(D_m) N(A(\rho_m))] = \operatorname{tr}[D_m (B \circ N \circ A)(\rho_m)]$$

so $P'_e = P_e$. Thus $(\mathcal{E}', \mathcal{D}')$ is an (R, ϵ) code for N if $(\mathcal{E}, \mathcal{D})$ is an (R, ϵ) code for $B \circ N \circ A$.

Absorb into sup.

(3) Entanglement Assisted.



- Alice can prepare arbitrary Ψ_{AR_A} and pass the reservoir part to Bob via $I_{R_A \to R_B}$. In other words Alice and Bob share state Ψ_{AR_B}.
- The encoder is now a channel $\mathcal{E}_m$ that depends on message m. That is $\mathcal{E}_m^{A \to B}$ is a channel for each m. Aka "modulation".
  - ↳ Alternately, $\mathcal{E}(|m\rangle\langle m| \otimes \rho_m) \to \rho_B$ is a channel $\mathcal{E}^{M_A \to B}$.
- This is clearly at least as good as plain — it contains it! Suppose $\Psi = \omega_A \otimes \omega_{R_A}$ is a product state, and $\mathcal{E}_m = P_{\rho_m}$, and $D_m = D_m^B \otimes \mathbb{1}_{R_B}$. Then
$$P_{m'|m} = \mathrm{tr}(D_{m'}^B N(\rho_m))$$
equivalent to an arbitrary plain code.
- Is it better? Super dense coding proves yes.

- Super Dense Coding

N = qubit Identity channel I_2

$\Psi_{AR_A} = \Phi^+$ = Bell Pair $\left(|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)\right)$.

$K = 4$ messages $(R = 2)$

$\mathcal{E}_m = \{I, X, Y, Z\}$

D_m = Bell Basis Measurement

$P_e = 0$.

Thus, $C_\epsilon^{EA}(I_2) \geq C_0^{EA}(I_2) \geq 2$.

Def. An EA code for N is $(\mathcal{E}, D, \Psi)$. The rate is still $R = \log K$. Now

$$p_{m'|m} = \mathrm{tr}\left[D_{m'}\left[(N \circ \mathcal{E}_m) \otimes I\right](\Psi)\right].$$

$$P_e = 1 - \frac{1}{K}\sum_m p_{m|m}.$$

Def. The ϵ-one-shot EA capacity is

$$C_\epsilon^{EA}(N) = \sup_{(R,\epsilon)\ EA\ \text{codes}} R.$$

Prop. ϵ-one-shot EA capacity is monotonic under Pre-/post-processing, and under increasing ϵ.

PF. Same as plain case.

---

(4) Aside...

Any questions about HW?

---

(5) Overview of coming results.

Asymptotic

Plain Holevo $\chi(N) = \sup_{P_{XA} \in CQ} I(X:B)_\omega$ [Regularize]

EA Mutual $I(N) = \sup_{P_{RA}} I(R:B)_\omega$

$\omega = N_{A\to B}(\rho)$

One-Shot [Not equal but near]

Plain $X_h^\epsilon(N)$

EA I_h^ϵ

→ Change to ϵ-Hypothesis Testing Relative Entropy.

•Converses: Use DPI to remove D_1 the largest M.I.

(6) If necessary...

- Square Root Measurement
- Hayashi/Nagaoka
- $D_n^{\epsilon}(\rho \| \sigma)$

Lecture 5

Today:

Goal: Prove Achievability bounds for plain ; EA class. comm.

⟶ Define a code, check its (R, ϵ).

① Preliminaries
- Hyp Test RE
- SQRT Measurement
- H-N Lemma

② Plain: Random Coding [Wang/Renner 2012]

③ EA: Position-based Coding [Anshu et al 2019]

① Preliminaries

ⓐ Hypothesis Testing Relative Entropy

- Hypothesis Testing Problem:
  Decide whether unknown state is σ or ρ.
- To decide, do a povm $\{M_\sigma, \mathbb{1} - M_\rho\}$.
  If result is 0, guess ρ. If 1, guess σ.

$$P_{\substack{\text{guess} \\ \rho | \rho}} = \mathrm{tr}(M_\rho \rho) \qquad P_{\sigma|\sigma} = \mathrm{tr}[(\mathbb{1} - M_\rho)\sigma]$$

$$P_{\sigma|\rho} = \mathrm{tr}[(\mathbb{1} - M_\rho)\rho] \qquad P_{\rho|\sigma} = \mathrm{tr}[M_\rho \sigma]$$

$P_{\sigma|\rho}$: Type I Error; $P_{\rho|\sigma}$: Type II Error

- Given fixed bound on type I error,

$$\mathrm{tr}(M_\rho) \geq 1 - \epsilon$$

Minimize the type II error:

$$\beta_\epsilon = \min_{\substack{M \\ \mathrm{tr} M_\rho \geq 1-\epsilon}} \mathrm{tr}(M\sigma).$$

If small, very distinguishable.
If big, not distinguishable.

log of this is like a relative entropy

Def. ϵ-Hypothesis Testing Relative entropy

For two states ρ, σ on same system, $0 \le \epsilon \le 1$,

$$D_h^\epsilon(\rho\|\sigma) = -\log\left(\min_{0 \le M \le 0,\ \mathrm{tr} M\rho \ge 1-\epsilon} \mathrm{tr}(M\sigma)\right).$$

- Big = Distinguishable. Small = Indistinguishable. (Just like RE). [Recall $D(\rho\|\sigma) = \mathrm{tr}(\rho\log\rho - \rho\log\sigma)$]

Prop.

1. For any ρ, σ,
$$D_h^\epsilon(\rho\|\sigma) \ge D_h^\epsilon(\rho\|\rho) \le -\log(1-\epsilon) \ge 0.$$

2. For any ρ, σ, and any channel N,
$$D_h^\epsilon(\rho\|\sigma) \ge D_h^\epsilon(N(\rho)\|N(\sigma)).$$

3. For $\tau = \frac{\mathbb{1}}{d}$, $\psi = |\psi\rangle\langle\psi|$,
$$D_h^\epsilon(\psi\|\tau) = \log d - \log(1-\epsilon).$$

Note: compare to
$$D(\rho\|\sigma) \ge 0$$
$$D(\rho\|\sigma) \ge D(N(\rho)\|N(\sigma))$$
$$D(\psi\|\tau) = \log d.$$

Proofs.

(2). Let M be optimal for $D_h^\epsilon(N(\rho)\|N(\sigma))$.
Then $\mathrm{tr}(MN(\rho)) \ge 1-\epsilon$, ~~$D_h^\epsilon = -\log$~~
~~Both~~ $N^\dagger(M)$ is a POVM element such that
$$\mathrm{tr}(N^\dagger(M)\rho) = \mathrm{tr}(MN(\rho)) \ge 1-\epsilon,$$
Thus
$$D_h^\epsilon(\rho\|\sigma) = \sup_{M' : \mathrm{tr}M'\rho \ge 1-\epsilon} \left(-\log \mathrm{tr}(M'\sigma)\right)$$
$$\ge -\log \mathrm{tr}[N^\dagger(M)\sigma]$$
$$= -\log \mathrm{tr}[MN(\sigma)]$$
$$= D_h^\epsilon(N(\rho)\|N(\sigma)).$$

(1) ~~First~~, $D_h^\epsilon(\rho\|\sigma) \ge D_h^\epsilon(P_\rho(\rho)\|P_\rho(\sigma)) = D_h^\epsilon(\rho\|\rho)$.
Then, $D_h^\epsilon(\rho\|\rho) =$ ~~$-\log \mathrm{tr}M\rho$, $\mathrm{tr}M\rho \ge 1-\epsilon$~~
$$-\log \inf_{\mathrm{tr}M\rho \ge 1-\epsilon} \mathrm{tr}M\rho = -\log(1-\epsilon).$$

(3) $\mathrm{tr}(M\tau) = \frac{\mathrm{tr}M}{d} \ge \frac{\mathrm{tr}(M\psi)}{d} \ge \frac{1-\epsilon}{d}$. Equal for $M = (1-\epsilon)\psi$.

Relations to $D(\rho\|\sigma)$

Thm. "Quantum Stein's Lemma".

$$\lim_{n\to\infty} \frac{1}{n} D_n^{\epsilon}(\rho^{\otimes n}\|\sigma^{\otimes n}) = D(\rho\|\sigma)$$

for all $0<\epsilon<1$.

Prop. upper bound

$$D_n^{\epsilon}(\rho\|\sigma) \le \frac{D(\rho\|\sigma) + H_2(\epsilon)}{1-\epsilon}$$

with H_2 the binary entropy.

Pf. We'll prove these in later sections.

(b) SQRT Measurement.

Def. For any set $\{S_x\}$ of positive operators,

$$D_x = \left(\sum_x S_x\right)^{-1/2} S_x \left(\sum_x S_x\right)^{-1/2}$$

are the elements of a POVM called the square root measurement.

- Used to convert unnormalized set of POVM elements to a POVM.

(c) Hayashi-Nagaoka Lemma

- This is important because it relates the probabilities for D_x to S_x in the $\sqrt{\text{meas}}$.

Lemma. For $0 \le S \le \mathbb{1}$, $T \ge 0$, and any $c>0$,

$$\boxed{\mathbb{1} - (S+T)^{-1/2} S (S+T)^{-1/2} \le (1+c)(\mathbb{1}-S) + (2+c+c^{-1})T}$$

Outline of proof.

- For any $c>0$, $(A-cB)^{\dagger}(A-cB) \ge 0$.
- Let $A = \sqrt{T}$ and $B = \sqrt{T}\left((S+T)^{-1/2} - \mathbb{1}\right)$.
- Use $T \le (S+T)$...
- Use that $(\,)^{1/2}$ is operator Monotone, $(S+T)^{1/2} \ge S^{1/2}$.
- Use $S^{1/2} \ge S$ for $0 \le S \le \mathbb{1}$.
- Bunch of algebra.

OK, now ready for achievability proofs i.

(2) Classical Communication: Random Coding.

- We want to find lower bounds on the capacity

$$C_\epsilon(N) = \sup_{\substack{(\mathcal{E},\mathcal{D}) \\ P_e \le \epsilon}} R$$

Do this by defining a code and calculating (R, F).

Encoding:

- Choose any $P(x)$ and $\{\rho_x\}$.
- Set of messages is $m = 1 \dots K$.
- Random Codebook: For each m, choose an x_m randomly w/ prob $P(x)$.
  - ↳ Some $x_m = x_{m'}$ might be the same.

[redacted]

- Let $\rho_{XA} \equiv \sum_x P(x)|x\rangle\langle x| \otimes \rho_x$ and

$$\omega_{XB} \equiv (I_X \otimes \mathcal{N}_{A\to B})(\rho_{XA}) = \sum_x P(x)|x\rangle\langle x| \otimes \mathcal{N}(\rho_x).$$

Let $\omega_x \equiv \mathcal{N}(\rho_x)$.

Decoding:

- Let S be a POVM for the hypothesis test

$$\omega_{XB} \quad \text{vs} \quad \omega_X \otimes \omega_B$$

Suppose that $\mathrm{tr}(S\omega_{XB}) = 1 - \epsilon'$.

- Let $S_x = \mathrm{tr}_A\left((|x\rangle\langle x| \otimes \mathbb{1})S\right)$.

↳ $\mathrm{tr}_B(\omega_x S_x) = \mathrm{tr}\left((|x\rangle\langle x| \otimes \omega_x)S\right)$.

[redacted]

Pf. $\mathrm{tr}\left[(\mathbb{1} \otimes \omega_x)(|x\rangle\langle x| \otimes \mathbb{1})S\right] = \mathrm{tr}(\omega_x \mathrm{tr}_A(\dots))$.

- Let

$$D_m = \left(\sum_{m=1}^K S_{x_m}\right)^{-1/2} S_{x_m} \left(\sum_{m=1}^K S_{x_m}\right)^{-1/2}$$

be decoding POVM.

Now we calculate error for this scheme.

By definition,

$$P_e = 1 - \frac{1}{K}\sum_m \mathrm{tr}\left(D_m \mathcal{N}(\rho_{x_m})\right) = \frac{1}{K}\sum_m \mathrm{tr}\left(\omega_{x_m}(\mathbb{1} - D_m)\right)$$

But by H-N Lemma,

$$\mathbb{1} - D_m \le (1+c)(\mathbb{1} - S_{x_m}) + (2 + c + c^{-1})\sum_{m' \ne m} S_{x_{m'}}.$$

So

$$P_e \le \frac{1+c}{K} \sum_m \mathrm{tr}(\omega_{x_m} (\mathbb{1} - S_{x_m})) + \frac{2+c+c^{-1}}{K} \sum_m \sum_{m' \neq m} \mathrm{tr}(\omega_{x_m} S_{x_{m'}}).$$

Now <u>average over codes</u>. Imagine we generate the random codebook many indepent times.

On average,

$$P_e \le \frac{1+c}{K} \sum_m \sum_x P(x) \mathrm{tr}(\omega_x (\mathbb{1} - S_x)) + \frac{2+c+c^{-1}}{K} \sum_m \sum_{m' \neq m} \sum_{x x'} P(x) P(x') \mathrm{tr}(\omega_x S_{x'})$$

After averaging, all terms m-independent,

so $\sum_m \to K$, $\sum_m \sum_{m' \neq m} \to (K(K-1))$.

$$P_e \le (1+c) \sum_x P(x) \mathrm{tr}(\omega_x (\mathbb{1} - S_x)) + (2+c+c^{-1})(K-1) \sum_{x x'} P(x) P(x') \mathrm{tr}(\omega_x S_{x'})$$

$$= (1+c)\left[1 - \sum_x P(x) \mathrm{tr}(\omega_x S_x)\right] + \cdots$$

Now observe that

$$\sum_x P(x) \mathrm{tr}(\omega_x S_x) = \sum_x P(x) \mathrm{tr}(|x\rangle\langle x| \otimes \omega_x S) = \mathrm{tr}(\omega S).$$

And,

$$\mathrm{tr}(\omega_X \otimes \omega_B S) = \mathrm{tr}\left[\sum_x P(x) |x\rangle\langle x| \otimes \sum_{x'} P(x') \omega_{x'} S\right]$$

$$= \sum_{x x'} P(x) P(x') \mathrm{tr}\left[(\mathbb{1} \otimes \omega_{x'})(|x\rangle\langle x| \otimes \mathbb{1}) S\right]$$

$$= \sum_{x x'} P(x) P(x') \mathrm{tr}\left[\omega_{x'} S_x\right].$$

Plugging back in,

$$P_e \le (1+c)\underbrace{\left[1 - \mathrm{tr}(\omega S)\right]}_{1-\epsilon'} + (2+c+c^{-1})(K-1)\underbrace{\left[\mathrm{tr}(\omega_X \otimes \omega_B S)\right]}_{\text{Choose optimal } S \text{ to minimize}}$$

If we take the optimal S for the Hyp. Test.

$$P_e \le (1+c)\epsilon' + (2+c+c^{-1}) 2^R 2^{-D_h^{\epsilon'}(\omega_{XB} \| \omega_X \otimes \omega_B)}$$

Used $K-1 \le K = 2^R$. Let $P_e = \epsilon$ (relabel). Rearranging

$$2^{R - D_h^{\epsilon'}} \ge \frac{\epsilon - (1+c)\epsilon'}{2+c+c^{-1}}$$

$$R \ge D_h^{\epsilon'} - \log \frac{2+c+c^{-1}}{\epsilon - (1+c)\epsilon'}.$$

Optimizing over c gives $c = \frac{\epsilon - \epsilon'}{1+\epsilon'}$, so

$$R \ge D_h^{\epsilon'} - \log\left(\frac{4\epsilon}{(\epsilon - \epsilon')^2}\right)$$

- This is <u>average over many codes</u>. It follows at least one code in the class has P_e below the average.

Suppose we want $P_e \le \epsilon$ for fixed ϵ.

$$P_e \le (1+c)\epsilon' + (2+c+c^{-1})2^R 2^{-D_h^{\epsilon'}} \le \epsilon.$$

Possible only if $\epsilon' < \epsilon$. If so, we need

$$2^{R - D_h^{\epsilon'}} \le \frac{\epsilon - (1+c)\epsilon'}{2+c+c^{-1}}$$

for some $c > 0$. Optimizing gives $c = \frac{\epsilon - \epsilon'}{\epsilon + \epsilon'}$,

$$R \le D_h^{\epsilon'} - \log\left(\frac{4\epsilon}{(\epsilon-\epsilon')^2}\right).$$

This is <u>average</u> rate over many random codes. At least one code exists better than average.

We have proved:

Thm.
Fix $\epsilon > 0$, $\epsilon' < \epsilon$, $P(x)$, $\{P_x\}$. If

$$R \le D_h^{\epsilon'}(W_{XB} \| W_X \otimes W_B) - \log\left(\frac{4\epsilon}{(\epsilon-\epsilon')^2}\right)$$

then $\exists$ code with

$$P_e \le \epsilon.$$

Thus $\forall \epsilon' < \epsilon$,

$$C_\epsilon(N) \ge \sup_{P_{XA}} D_h^{\epsilon'}(W_{XB} \| W_X \otimes W_B) - \log\left(\frac{4\epsilon}{(\epsilon-\epsilon')^2}\right).$$

- What is this $D_h^{\epsilon}(\omega \| \omega \otimes \omega)$?

$$\chi(N) = \sup_{P_{XA}} I(X:B)_{\omega = N(\rho)}$$

$$\chi_\epsilon^\epsilon(N) = \sup_{P_{XA}} I_h^\epsilon(X:B)_{\omega = N(\rho)}$$

$D(\omega \| \omega \otimes \omega)$ vs $D_h^\epsilon(\omega \| \omega \otimes \omega)$

Generalized Mutual Information.

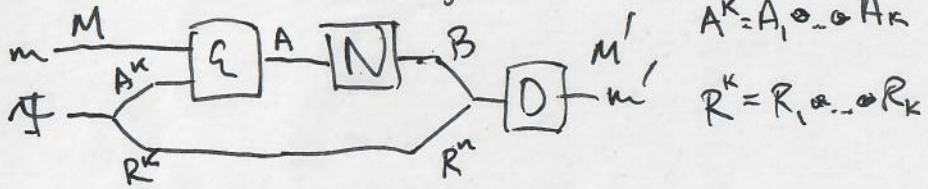
Lecture 6

Today:

(... Finish Previous Lecture: Random Coding)

- EA: Position Based Coding
- Converse Bounds
- Asymptotic Limit?

① Position Based Coding



Setup:

- Alice and Bob share K copies of ρ_{AR}:
$$\Psi_{A^K R^K} = (\rho_{AR})^{\otimes K}.$$
- Messages $m = 1 \dots K$.
- Let (one-copy)
$$\omega_{BR} = \mathcal{N}_{A\to B} \otimes I_R(\rho_{AR})$$
so
$$\omega_B = \mathrm{tr}_R \omega_{BR}, \quad \omega_R = \rho_R.$$

Encoding:

- Alice sends mth system to Bob through $\mathcal{N}$,
$$\mathcal{N} \circ \mathcal{E}_m = \mathcal{N}_{A_m \to B}.$$
Equivalently, $\mathcal{E}_m = \mathrm{tr}_{A^n/A_m}$ (trace out all n but m)

Decoding:

- Again let S be hypothesis tester for ω_{BR} vs $\omega_B \otimes \omega_R$.
- Let
$$S_m = S^{BR_m}.$$
- Again let
$$D_m = \left(\sum_m S_m\right)^{-1/2} S_m \left(\sum_m S_m\right)^{-1/2}.$$

Analysis:

- What does Bob see? Bob's state is
$$\mathcal{N}_{A_m \to B}(\rho^{A_1 R_1} \otimes \dots \otimes \rho^{A_n R_n})$$
But can only access BR^K, so really
$$\mathrm{tr}_{A^n/A_m}(\omega_m).$$

Thus,

$$\sigma_m^{BR^n} = \rho^{R_1} \otimes \dots \otimes N_{A_m \to B}(\rho^{AR_m}) \otimes \dots \otimes \rho^{R_n}$$

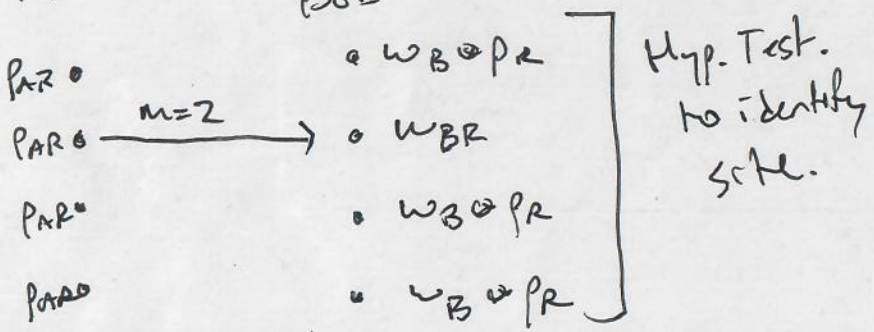
$$= \rho^{R_1} \otimes \dots \otimes \omega^{R_m B} \otimes \dots \otimes \rho^{R_n}.$$

The 2-party marginals of this are

$$\sigma_m^{BR_m} = \omega^{R_m B}, \quad m = m$$

$$\sigma_m^{BR_{m'}} = \omega^B \otimes \rho^{R_{m'}}$$

$$= \omega^B \otimes \omega^{R_{m'}}, \quad m' \neq m.$$



Error Prob:

- Like before

$$P_e = \frac{1}{K} \sum_m \mathrm{tr}\left(\sigma_m (\mathbb{1} - D_m)\right)$$

$$\mathbb{1} - D_m \leq (1+c)(\mathbb{1} - S_m) + (2 + c + c^{-1}) \sum_{i \neq m} S_m$$

- Plugging this in with optimal S_1

$$\mathrm{tr}(S_m \sigma) = \mathrm{tr}(S^{BR_m} \sigma)$$

$$= \mathrm{tr}(S \sigma^{BR_m})$$

$$\mathrm{tr}(S \omega^{RB}) = 1 - \epsilon' \quad (m = m)$$

$$\mathrm{tr}(S \omega^R \otimes \omega^B) = 2^{-D_h^{\epsilon'}} \quad (m' \neq m)$$

- This is the same for every term in sums, so again $\sum_m \to K$, $\sum_m \sum_{i \neq m} \to K(K-1)$.

So yet again

$$P_e \leq (1+c)\epsilon' + (2 + c + c^{-1}) 2^{R - D_h^{\epsilon'}} \leq \epsilon.$$

Solving we have then

Lemma. Fix $\epsilon > 0$, $\epsilon' < \epsilon$, pair-

if

$$R \leq D_h^{\epsilon'} - \log\left(\frac{4c}{(\epsilon - \epsilon')^2}\right),$$

there exists an (R, ϵ) EA code.

Thm. For any $0 \le \varepsilon' < \varepsilon$,

$$C_\varepsilon^{EA}(N) \ge \sup_{\rho_{RA}} D_h^{\varepsilon'}(\omega_{RB} \| \omega_R \otimes \omega_B) - \log \frac{4\varepsilon}{(\varepsilon-\varepsilon')^2}$$

Additionally, the sup can be limited to pure states with $R \sim A$.

Proof. We just showed the existence of such a code for any ρ_{RA}. For the sup restriction:

- Purify ρ_{RA} to ψ_{RAE}. This changes nothing in the code, which acts only on RA. Thus for any ρ_{RA} code, there also exists a ψ_{RAE} code.
- Now consider $\sup_{\psi_{RA}}$ (w.l.g.) over pure states, any pure ψ_{RA}, in the Schmidt Basis, looks like a state on AA'. Can delete the rest of the space without changing anything.

- We can even use this lemma to prove the Plain capacity theorem!
  - Any Prod-Assisted code = A plain code.
  - Note that P_e is linear in Ψ,

$$P_e = 1 - \frac{1}{M} \sum_m \mathrm{tr}\left(D_m \left(N \circ \mathcal{E}_m(\Psi) \right) \right).$$

  Thus if $\Psi = \sum_i p_i \Psi_i$, at least one Ψ_i must do better than Ψ.

  - Thus, starting from any CQ state

$$\rho_{XA} = \sum_x p(x) |x\rangle\langle x| \otimes \rho_x,$$

  we can convert P.B.C. into Plain code.

Thm (Again). For any $0 \le \varepsilon' < \varepsilon$,

$$C_\varepsilon(N) \ge \sup_{\substack{\rho_{XA} \\ (CQ)}} D_h^{\varepsilon'} - \log(M).$$

To summarize so far; let us define

$$\chi_n^\epsilon(N) = \sup_{P_{XA}} I_\epsilon^\epsilon(X:B)_\omega$$

$$I_n^\epsilon(N) = \sup_{P_{RA}} I_n^\epsilon(R:B)_\omega,$$

with $\omega = N_{A\to B}(\rho)$.

- We've proved achievable rates, it turns out they're nearly maximal. Next we'll find, that

$$\chi_n^{\epsilon'} - \log\frac{4\epsilon'}{(\epsilon-\epsilon')^2} \le C_\epsilon \le \chi_n^\epsilon$$

$$I_n^{\epsilon'} - \log\frac{4\epsilon'}{(\epsilon-\epsilon')^2} \le C_\epsilon^{EA} \le I_n^\epsilon.$$

Achievability Bounds ✓

Converse Bounds (Next.)

(2) Converse Bounds.

Generalized Divergences

Def. A function $\mathbb{D}(\rho\|\sigma)$ on states ρ & semi-definite σ from the same system such that

1. $\mathbb{D}$ takes values in $\mathbb{R}\cup\{\infty\}$.
2. $\mathbb{D}(\rho\|\sigma) \ge 0$ $\forall$ states ρ,σ, with 0 iff $\rho=\sigma$.
3. $\mathbb{D}(\rho\|\sigma) \ge \mathbb{D}(\mathcal{N}(\rho)\|\mathcal{N}(\sigma))$ $\forall$ CPTP $\mathcal{N}$.

is a generalized divergence.

Examples.

- Umegaki: $D(\rho\|\sigma) = \mathrm{tr}(\rho\log\rho - \rho\log\sigma)$
- Hyp. Test: $D_\epsilon^h(\rho\|\sigma) = \inf_{\mathrm{tr}(M\rho)\ge 1-\epsilon} \mathrm{tr}(M\sigma)$
- Petz Renyi: $D_\alpha(\rho\|\sigma) = \frac{1}{\alpha-1}\log\mathrm{tr}\left(\rho^\alpha\sigma^{1-\alpha}\right)$, $\alpha\in(0,\infty)$ [is divergence for $\alpha\in(0,2)$]
- Sandwiched Renyi: $\tilde{D}_\alpha(\rho\|\sigma) = \frac{1}{\alpha-1}\log\mathrm{tr}\left(\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\,\sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha\right)$, [Divergence for $\alpha\in(\frac{1}{2},\infty)$.]

② Converse Bounds:

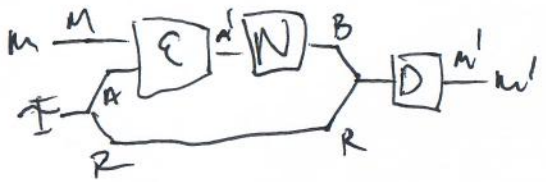
Thm.

$$C_\epsilon(N) \le \chi_n^\epsilon(N)$$

$$C_\epsilon^{EA}(N) \le I_n^\epsilon(N).$$

Proof.

EA first. Consider EA setup



Alice starts with state

$$\gamma_{MRA} = \frac{1}{K}\sum_m |m\rangle\langle m| \otimes \Psi_{AR}$$

This turns into

$$\rho_{mm'} = \mathcal{D}_{RB\to m'} \circ \mathcal{N}_{A'\to B} \circ \mathcal{E}_{MA\to A'm}(\gamma_{MRA})$$

$$= \sum_{mm'} \frac{1}{K} P_{m'|m} |m\rangle\langle m| \otimes |m'\rangle\langle m'|$$

(Alice "keeps a copy".)

By defn,

$$I_n^\epsilon(N) = \sup_{\rho_{RA}} D_h^\epsilon(\omega_{RB} \| \omega_R \otimes \omega_B)_{\omega = \mathcal{N}_{A\to B}(\rho_{RA})}$$

Absorb Alice's message system into $\bar{R}$, letting

$$\rho_{\bar{R}A} = \mathcal{E}_{MA\to MA'}(\gamma_{MRA})$$

$$= \frac{1}{K}\sum_m |m\rangle\langle m| \otimes \mathcal{E}_m^{A\to A'}(\Psi_{RA}).$$

Then $\omega_{\bar{R}B} = \mathcal{N} \circ \mathcal{E}(\gamma)$

$$= \frac{1}{K}\sum_m |m\rangle\langle m| \otimes \mathcal{N}_{A'\to B} \circ \mathcal{E}_m^{A\to A'}(\Psi_{RA})$$

And also

$$\rho_{mm'} = \mathcal{D}_{RB\to m'}(\omega_{\bar{R}B}).$$

B/c of sup, in particular

$$I_h^\epsilon(N) \ge D_h^\epsilon(\omega_{\bar{R}B} \| \omega_{\bar{R}} \otimes \omega_B).$$

$$\ge \inf_{\sigma_B} D_h^\epsilon(\omega_{\bar{R}B} \| \omega_{\bar{R}} \otimes \sigma_B).$$

But $\omega_{\bar{R}} = \omega_M \otimes \omega_R = \rho_m \otimes \omega_R = \rho_m \otimes \Psi_R$.

So
$$I_h^\epsilon \geq \inf_{\sigma_B} D_h^\epsilon(\omega_{MRB} \| P_M \otimes \omega_R \otimes \sigma_B).$$

Now apply decoding $D_{RB\to M'}$. By DPI,
$$I_h^\epsilon \geq \inf_{\sigma_B} D_h^\epsilon(P_{MM'} \| P_M \otimes \mathcal{D}(\omega_R \otimes \sigma_B))$$
$$= \inf_{\sigma_B} D_h^\epsilon(P_{MM'} \| P_M \otimes \sigma_{M'}).$$

Now consider H.T. POVM element
$$\Pi = \sum_m |m\rangle\langle m| \otimes |m\rangle\langle m|. \quad \text{(Comparator)}$$

This obeys $\mathrm{tr}(\Pi P_{MM'}) = 1 - P_e \geq 1 - \epsilon$.

Also, $P_M = \frac{\mathbb{1}}{K}$, so
$$\mathrm{tr}(\Pi P_M \otimes \sigma_{M'}) = \sum_m \frac{1}{K} \mathrm{tr}(|m\rangle\langle m| \otimes |m\rangle\langle m| \sigma_{M'})$$
$$= \sum_m \frac{1}{K} \mathrm{tr}(|m\rangle\langle m| \sigma_{M'})$$
$$= \frac{1}{K} \mathrm{tr}(\mathbb{1} \sigma_{M'}) = \frac{1}{K}.$$

Thus $D_h^\epsilon \geq -\log \frac{1}{K} = \log K$.

Putting this together, for any (R, ϵ) code, we found
$$R \leq I_h^\epsilon(N). \checkmark$$

For $C_E(N)$, same exact logic.

- Choose $\rho_{XA} = \frac{1}{K} \sum_m |m\rangle\langle m| \otimes \rho_m$.
- Then $\chi_h^\epsilon(N) \geq D_h^\epsilon(\rho_{XB} \| \rho_X \otimes \omega_B)$
$$\geq \inf_{\sigma_B} D_h^\epsilon(\omega_{XB} \| \omega_X \otimes \sigma_B)$$
- Apply decoder,
$$\geq \inf_{\sigma_B} D_h^\epsilon(P_{XX'} \| P_X \otimes \sigma_{X'})$$
- Use Comparator $\Pi = \sum_m |m\rangle\langle m| \otimes |m\rangle\langle m|$,
$$\geq -\log \mathrm{tr}(\Pi P_X \otimes \sigma_{X'}) = \log K. \checkmark$$

That's it.

Next Time: Metaconverse / Gallager / 2nd Divergence
Asymptotic Limit.

Lecture 7

Today:

- Meta-converse
- Generalized Divergences
- Asymptotic Limit.

Generalized Divergence

Defn: A function $\mathbb{D}(\rho\|\sigma)$ states ρ and $\sigma \geq 0$ from the same system, is a generalized divergence if

1. $\mathbb{D}$ takes values in $\mathbb{R} \cup \{\infty\}$.
2. $\mathbb{D}(\rho\|\sigma) \geq d_0 \quad \forall \rho, \sigma \in \{\text{states}\}$, where d_0 is system-ind const.
3. $D(\rho\|S) \geq D(N(\rho)\|N(S))$ for all states ρ and operators $S \geq 0$.

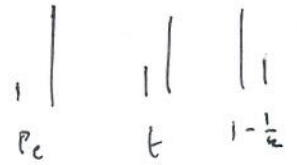
Examples:

- $D(\rho\|\sigma) = \mathrm{tr}(\rho\log\rho - \rho\log\sigma)$ ("Umegaki" RE)
- $D_h^\varepsilon(\rho\|\sigma)$ (Hyp Test RE)
- $d(\rho,\sigma) = \frac{1}{2}\|\rho-\sigma\|_1$ (Trace Distance)
- $d(\rho,\sigma) = \sqrt{1 - F(\rho,\sigma)^2}$ (Purified Distance/Fidelity)
- $D_\alpha(\rho\|\sigma) = \frac{1}{\alpha-1}\log \mathrm{tr}(\rho^\alpha\sigma^{1-\alpha})$, $\alpha\in(0,2)$ (Petz-Renyi RE)
- $\widetilde{D}_\alpha(\rho\|\sigma) = \frac{1}{\alpha-1}\log \mathrm{tr}\left[\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\,\sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha\right]$, $\alpha\in(\frac{1}{2},\infty)$. (Sand. Petz RE)
- $D_{\max}(\rho\|\sigma) = \inf_{\rho\le\lambda\sigma} \lambda$ (Max RE)

Some Preparatory Stuff

Fact. Let $0 < P_e < \epsilon < 1 - \frac{1}{K}$. Then

$$D\left(\begin{Bmatrix} P_e \\ 1-P_e \end{Bmatrix} \middle\| \begin{Bmatrix} 1-\frac{1}{K} \\ \frac{1}{K} \end{Bmatrix}\right) \geq D\left(\begin{Bmatrix} \epsilon \\ 1-\epsilon \end{Bmatrix} \middle\| \begin{Bmatrix} 1-\frac{1}{K} \\ \frac{1}{K} \end{Bmatrix}\right).$$



$\Rightarrow$ New Dist is more similar.

Proof. Exercise: Show $\exists$ classical channel Λ s.th. $\begin{Bmatrix} P_e \\ 1-P_e \end{Bmatrix} \xrightarrow{\Lambda} \begin{Bmatrix} \epsilon \\ 1-\epsilon \end{Bmatrix}$ and $\begin{Bmatrix} 1-\frac{1}{K} \\ \frac{1}{K} \end{Bmatrix} \to \begin{Bmatrix} 1-\frac{1}{K} \\ \frac{1}{K} \end{Bmatrix}$.

Fact. $\sup_{P_{XA}} D(\omega_{XB} \| \omega_X \otimes \omega_B) \geq$

$$\sup_{P_{XA}} \inf_{\sigma_B} D(\omega_{XB} \| \omega_X \otimes \sigma_B).$$

Proof. This is obvious.

Fact. Let $\Pi = \sum_m |m\rangle\langle m| \otimes |m\rangle\langle m|$.

Then $\mathrm{tr}\left(\Pi \frac{\mathbb{1}}{K} \otimes \sigma\right) = \frac{1}{K}$.

Proof. $\mathrm{tr}\left(\Pi \frac{\mathbb{1}}{K} \otimes \sigma\right) = \frac{1}{K} \mathrm{tr}\left(\sum_m |m\rangle\langle m| \sigma\right) = \frac{1}{K} \mathrm{tr}|\sigma| = \frac{1}{K}$.

Defn. (Comparator Channel).
For a state $\rho_{MM'}$, the comparator channel is $Q(\rho) = \mathrm{tr}(\Pi\rho)|0\rangle\langle 0| + \mathrm{tr}((\mathbb{1}-\Pi)\rho)|1\rangle\langle 1|$, where $\Pi = \sum_m |m\rangle\langle m| \otimes |m\rangle\langle m|$.

Now we are ready for the Metaconverse.

Lemma. (Metaconverse). For any channel N and $0 < \epsilon < 1 - \frac{1}{K}$, if $\exists$ a code with $P_e \leq \epsilon$ for K messages, then

$$\sup_{P_{XA}} \inf_{\sigma_B} D(\omega_{XB} \| \omega_X \otimes \sigma_B)_{\omega = N_{A\to B}(\rho)}$$
$$\geq D\left(\begin{Bmatrix} \epsilon \\ 1-\epsilon \end{Bmatrix} \middle\| \begin{Bmatrix} 1-\frac{1}{K} \\ 1/K \end{Bmatrix}\right)$$

for any guessing divergence.

Lemma. (EA Metaconverse).

$$\sup_{\rho_{RA}\,\sigma_B} \inf_{\omega = N_{A\to B}(\rho)} D\left(\omega_{RB} \,\|\, \omega_R \otimes \sigma_B\right) \geq D\left(\left\{\begin{matrix} t \\ 1-t \end{matrix}\right\} \,\|\, \left\{\begin{matrix} 1-\frac{1}{n} \\ 1/n \end{matrix}\right\}\right).$$

Proof. Plan First:

Let $\rho_{XA} = \frac{1}{K}\sum_m |m\rangle\langle m| \otimes \rho_m$, with ρ_m from $(\xi, K)_{code}$.

Then after channel, $N_{A\to B}$,

$$\omega_{XB} = \frac{1}{K}\sum_m |m\rangle\langle m| \otimes N(\rho_m).$$

Now apply decoder $D_{B\to M'}$ and comparator Q,

$$D(\omega_{XB}) = \frac{1}{K}\sum_m |m\rangle\langle m| \otimes D\left(N|\rho_m\right)$$

$$= \frac{1}{K}\sum_{mm'} P_{m'|m}\, |m\rangle\langle m| \otimes |m'\rangle\langle m'|$$

$$Q\circ D(\omega_{XB}) = (1-P_e)|0\rangle\langle 0| + P_e|1\rangle\langle 1|$$

For the other state,

$$\mathcal{D}(\omega_X \otimes \sigma_B) = \omega_X \otimes \mathcal{D}(\sigma_B) = \frac{\mathbb{1}}{K} \otimes \mathcal{D}(\sigma_B)$$

$$Q \circ \mathcal{D}(\omega_X \otimes \sigma_B) = Q\left(\frac{\mathbb{1}}{K} \otimes \sigma_B\right) = \frac{1}{K}|0\rangle\langle 0| + \left(1 - \frac{1}{K}\right)|1\rangle\langle 1|.$$

Thus by monotonicity,

[illegible]

$$\mathbb{D}(\omega_{XB} \| \omega_X \otimes \sigma_B) \geq \mathbb{D}(Q \circ \mathcal{D}(\omega_{XB}) \| Q \circ \mathcal{D}(\omega_X \otimes \sigma_B))$$

$$= \mathbb{D}\left(\begin{Bmatrix} 1-p_e \\ p_e \end{Bmatrix} \middle\| \begin{Bmatrix} 1/K \\ 1-1/K \end{Bmatrix}\right)$$

$$\geq \mathbb{D}\left(\begin{Bmatrix} 1-\epsilon \\ \epsilon \end{Bmatrix} \middle\| \begin{Bmatrix} 1/K \\ 1-1/K \end{Bmatrix}\right). ✓$$

For EA, do the same thing but start from

$$\rho_{RA} = \frac{1}{K}\sum_m |m\rangle\langle m| \otimes \mathcal{E}_m(\Phi_{R'A})$$

using existence of the code. (Absorb $R'M = R$.) ✓

Apply to different ID

Relative Entropy

$$D\left(\left\{\begin{matrix}\epsilon \\ 1-\epsilon\end{matrix}\right\} \middle\| \left\{\begin{matrix}\frac{1-\epsilon}{K} \\ \frac{1}{K}\end{matrix}\right\}\right) = -h_2(\epsilon) - \epsilon \log \frac{K-1}{K} - (1-\epsilon)\log\frac{1}{K}$$

$$= -h_2(\epsilon) + \epsilon \log \frac{K}{K-1} + (1-\epsilon)\log K$$

$$\geq (1-\epsilon)\log K - h_2(\epsilon).$$

Thus,

$$(1-\epsilon)\log K - h_2(\epsilon) \leq \sup_{P_{XA}} I(X:A)_\omega = \chi(\mathcal{N})$$

$$R \leq \frac{\chi(\mathcal{N}) + h_2(\epsilon)}{1-\epsilon} \quad \text{for any } (R,\epsilon) \text{ code.}$$

$$C_\epsilon(\mathcal{N}) \leq \frac{\chi(\mathcal{N}) + h_2(\epsilon)}{1-\epsilon}$$

Similarly,

$$C_\epsilon^{EA}(\mathcal{N}) \leq \frac{I(\mathcal{N}) + h_2(\epsilon)}{1-\epsilon}.$$

(Only thing that changes is $\sup_{P_{RA}}$.)

Sandwiched Renyi

You end up with

$$\tilde{D}_\alpha\left(\left\{ \begin{matrix} \epsilon \\ 1-\epsilon \end{matrix} \right\} \middle\| \left\{ \begin{matrix} 1-\frac{1}{K} \\ 1/K \end{matrix} \right\}\right) \geq \log K + \frac{\alpha}{\alpha-1} \log(1-\epsilon).$$

$$C_\epsilon(N) \leq \chi_\alpha(N) - \frac{\alpha}{\alpha-1} \log(1-\epsilon)$$

$(\alpha > 1)$. Nice form.

[illegible]

Note ... Since here dists are classical, $\tilde{D}_\alpha = D_\alpha$ which is usual Renyi divergence:

$$\log \left\langle \left(\frac{p}{q}\right)^s \right\rangle^{1/s} = \frac{1}{s} \log \left\langle \left(\frac{p}{q}\right)^s \right\rangle$$

$$= \frac{1}{s} \log \sum \frac{p^{s+1}}{q^s}$$

$$= \frac{1}{\alpha-1} \log \sum \frac{p^\alpha}{q^{\alpha-1}} \quad (s = \alpha - 1)$$

$$= D_\alpha(p \| q).$$

Hypotesis Testing

Choose $M = \{\Pi_i^0\}$. Then $\mathrm{tr}(M\{\rho_{1-\epsilon}^\epsilon\}) = 1-\epsilon$ is valid for

$$D_n^\epsilon(\{\rho_{1-\epsilon}^\epsilon\}\|\{\sigma_{1/n}^{1-\frac{1}{n}}\}) = \sup_{\substack{M \\ \mathrm{tr}M\rho \geq 1-\epsilon}} -\log(\mathrm{tr}(M\sigma)) \overset{\text{Neyman-Pearson}}{\geq} \mathrm{tr}(\{\Pi_i^0\}\{\sigma_{1/n}^{1-\frac{1}{n}}\})$$

$$= -\log\frac{1}{K} = \log K.$$

$\Rightarrow \quad C_\epsilon(N) \leq \chi_n^\epsilon(N)$

$$C_\epsilon^{EA}(N) \leq I_n^\epsilon(N)$$

As we already had. ✓

Asymptotic Limit.

Converse.

We found from M.C.,

$$C_\epsilon(N) \le \frac{\chi(N) + h_2(\epsilon)}{(1-\epsilon)} .$$

For $\epsilon = 0$,

$$C_0(N) \le \chi(N) + h_2(\epsilon)$$

Multiple uses,

$$\frac{C_0(N^{\otimes n})}{n} \le \frac{1}{n}\chi(N^{\otimes n}) + \frac{h_2(\epsilon)}{n}$$

Let $N \to \infty$,

$$C(N) \le \lim_{n\to\infty} \frac{1}{n}\chi(N^{\otimes n}) . \checkmark$$

Achievability. (Wang-Renner.)

Fix $\epsilon > \epsilon' > 0$. Choose a k and consider $N^{\otimes k}$.

For any k, we have a code where

$$\frac{R(k)}{k} \geq \frac{1}{k} \chi_h^{\epsilon'}(N^{\otimes k}) - \frac{1}{k} \log \frac{4\epsilon}{(\epsilon-\epsilon')^2}$$

Now, repeating $N^{\otimes k}$ many times, $(N^{\otimes k})^{\otimes n}$, we have

$$\lim_{n\to\infty} \frac{1}{n} \frac{R(k)}{k} \geq \lim_{n\to\infty} \frac{1}{n} \frac{1}{k} \chi_h^{\epsilon'}\left((N^{\otimes k})^{\otimes n}\right) - \frac{1}{n} \frac{1}{k} \log \frac{4\epsilon}{(\epsilon-\epsilon')}$$

$$\to \frac{1}{k} \lim_{n\to\infty} \frac{1}{n} \chi_h^{\epsilon'}\left((N^{\otimes k})^{\otimes n}\right).$$

But A.E.P. told us that $\frac{1}{n} D_h^{\epsilon}(\rho^{\otimes n} \| \sigma^{\otimes n}) \to D(\rho\|\sigma)$, for all ϵ. So

$$\frac{1}{n} \chi_h^{\epsilon}(N^{\otimes n}) = \frac{1}{n} \sup_{\rho_{XA^n}} D_h^{\epsilon}\left(I_X \otimes N_{A'}^{\otimes n}(\rho_{XA^n}) \,\middle\|\, I_X \otimes N_A^{\otimes n}(\rho_X \otimes \rho_{A^n})\right)$$

$$\geq \frac{1}{n} \sup_{(\rho_{XA})^{\otimes n}} \left[n \right]$$

$$= \frac{1}{n} \sup_{\rho_{XA}} D_h^{\epsilon}\left(I \otimes N(\rho)^{\otimes n} \,\|\, I \otimes N(\rho \otimes \rho)^{\otimes n}\right)$$

$$\xrightarrow{n=\infty} \sup_{\rho_{XA}} D(I \otimes N(\rho) \| I \otimes N(\rho \otimes \rho))$$

$$= \chi(N).$$

So going back,

$$\lim_{n\to\infty} \frac{1}{n} \frac{R(k)}{K} \geq \frac{1}{K} \lim_{n\to\infty} \frac{1}{n} \chi_h^{\epsilon'}\left((N^{\otimes k})^{\otimes n}\right)$$

$$\geq \frac{1}{K} \chi(N).$$

This holds for any ϵ, ϵ', We have

For any $\epsilon > \epsilon'$, for any K,

$$\lim_{n\to\infty} \frac{C_\epsilon\left((N^{\otimes k})^{\otimes n}\right)}{nK} \geq \frac{1}{K} \chi(N^{\otimes k}).$$

Thus, taking the limit in this way,

$$\lim_{K\to\infty} \lim_{n\to\infty} \frac{C_\epsilon\left((N^{\otimes k})^{\otimes n}\right)}{nK} \geq \lim_{K\to\infty} \frac{1}{K} \chi(N^{\otimes k}).$$

Taking the limit $\epsilon' \to 0$, then $\epsilon \to 0$,

$$C(N) \geq \lim_{K\to\infty} \frac{1}{K} \chi(N^{\otimes k}).$$

Thus, $$\boxed{C(N) = \lim_{K\to\infty} \frac{1}{K} \chi(N^{\otimes k})}.$$

The HSW theorem.

Similarly, $C^{EA}(N) = I(N)$.

Lecture 8

Today:

- Entropy too

Recall: Generalized Divergences

$\mathbb{D}(\rho\|\sigma)$ s.t.

$\mathbb{D}(\rho\|\sigma) \in \mathbb{R} \cup \{\infty\}$

$\mathbb{D}(\rho\|\sigma) \geq d_0$

$\mathbb{D}(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) \leq \mathbb{D}(\rho\|\sigma)$.

Examples

- GRE
- Hyp. Test. RE
- Trace Distance
- Purified distance (Fidelity)
- Petz-Renyi
- Sandwiched Renyi
- Max

RE as Parent Quantity

For QRE

$$D(\rho\|\sigma) = \mathrm{tr}(\rho\log\rho - \rho\log\sigma)$$

we have

- Entropy

$$H(\rho) = -D(\rho\|\mathbb{1})$$
$$= -\mathrm{tr}\,\rho\log\rho$$

- Conditional Entropy

$$H(A|B)_\rho = -D(\rho_{AB}\|\mathbb{1}_A\otimes\rho_B)$$
$$= H(A,B)_\rho - H(B)_\rho$$

- Mutual Information

$$I(A:B)_\rho = D(\rho_{AB}\|\rho_A\otimes\rho_B).$$

But also,

$$I(A:B)_\rho = \inf_{\sigma_B} D(\rho_{AB} \| \rho_A \otimes \sigma_B)$$

$$H(A|B)_\rho = -\inf_{\sigma_B} D(\rho_{AB} \| \mathbb{1}_A \otimes \sigma_B)$$

Proof.

$$D(\rho_{AB} \| \rho_{AB} \otimes \sigma_B) = \underbrace{-S(\rho_{AB})}_{} + \operatorname{tr} \rho_{AB} \log(\rho_A \otimes \sigma_B)$$

$$= -S(\rho_{AB}) + \operatorname{tr} \rho_{AB} \log\left[(\rho_A \otimes \mathbb{1})(\mathbb{1} \otimes \sigma_B)\right]$$

$$= -S(\rho_{AB}) + \operatorname{tr}\left(\rho_{AB}\left(\log(\rho_A \otimes \mathbb{1}) + \log(\mathbb{1} \otimes \sigma_B)\right)\right.$$

$$= -S(\rho_{AB}) + S(\rho_A) - \operatorname{tr} \rho_B \log \sigma_B$$

$$\geq -S(\rho_{AB}) + S(\rho_A) + S(\rho_B)$$

$$= I(A:B)_\rho .$$

Because

$$D(\rho_B \| \sigma_B) = \operatorname{tr} \rho_B \log \rho_B - \operatorname{tr} \rho_B \log \sigma_B \geq 0 .$$

Similarly for conditional.

So for generalized information quantities there are usually two versions, regular (↑) and infimized (↓). (Note K&W Book uses a ↓).

Smoothing.

Another way to modify info quantities is by smoothing.

Defn. Fidelity. (Wilde Books Defn)

$$F(\rho,\sigma) = \max_{\psi,\phi} |\langle\psi|\phi\rangle|^2$$

where ψ,ϕ are purifications of ρ,σ.

Fact. $F(\rho,\sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1^2$

Defn. Purified Distance

$$P(\rho,\sigma) = \sqrt{1 - F(\rho,\sigma)}.$$

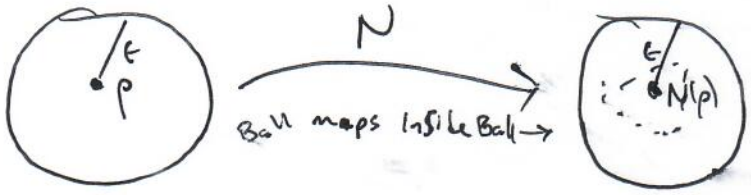
Defn. Smoothed Divergence.

$$\mathbb{D}^\epsilon(\rho,\sigma) = \inf_{\substack{\rho' \text{ s.t.} \\ P(\rho',\rho) \le \epsilon}} D(\rho'\|\sigma).$$

Lemma. If $\mathbb{D}$ is a divergence, then the smoothed $\mathbb{D}^\varepsilon$ is fco.

Proof. We only need to use the fact that the smoothing balls B_ε are defined wrt a divergence. (We have specified $P(\rho,\sigma)$ but it doesn't matter.) Only thing to check is DPI.

Let $\mathbb{D}(\rho\|\sigma) \stackrel{*}{=} \mathbb{D}(N(\rho)\|(N(\sigma)) \;\forall\, \rho,\sigma$.



Then

$$\mathbb{D}^\varepsilon(N\rho)\|N(\sigma)) = \inf_{\rho' \in B_\varepsilon(N(\rho))} \mathbb{D}(\rho'\|N(\sigma))$$

$$\leq \inf_{\rho' \in N(B_\varepsilon(\rho))} \mathbb{D}(\rho'\|N(\sigma))$$

$$= \inf_{\rho' \in B_\varepsilon(\rho)} \mathbb{D}(N(\rho')\|N(\sigma))$$

$$\leq \inf_{\rho' \in B_\varepsilon(\rho)} \mathbb{D}(\rho'\|\sigma)$$

$$= \mathbb{D}^\varepsilon(\rho\|\sigma).$$

So for any generalized divergence we have

- Entropy

$$H_{\mathbb{D}}(\rho) = -\mathbb{D}(\rho \| \mathbb{1})$$

- Conditional Entropy

$$H_{\mathbb{D}}^{\uparrow}(A|B) = -\mathbb{D}(\rho_{AB} \| \mathbb{1}_A \otimes \rho_B)$$

$$H_{\mathbb{D}}^{\downarrow}(A|B) = -\inf_{\sigma_B} \mathbb{D}(\rho_{AB} \| \mathbb{1}_A \otimes \sigma_B)$$

- Mutual Info

$$I_{\mathbb{D}}^{\uparrow}(A:B) = \mathbb{D}(\rho_{AB} \| \rho_A \otimes \rho_B)$$

$$I_{\mathbb{D}}^{\downarrow}(A:B) = \inf_{\sigma_B} \mathbb{D}(\rho_{AB} \| \rho_A \otimes \sigma_B)$$

As well as ϵ-smoothed versions.

Why smooth and minimize?

(Note Hmin $\leftrightarrow$ Dmax)

Thm. Generalized Mutual Info is monotone under local channels.

$$I^{\uparrow}_{\mathbb{D}}(A:B)_{\rho_{AB}} \geq I^{\uparrow}_{\mathbb{D}}(A:B)_{(N_A\otimes N'_B)\rho_{AB}}$$

$$\left(\downarrow \qquad \geq \qquad \downarrow \right).$$

Proof. ~~Recall ρ_{AB}, σ_B, (including $\sigma_B = \rho_B$),~~

If $\rho'_{AB} = N_A \otimes N'_B(\rho_{AB})$ then $\rho'_A = N_A(\rho_A)$, $\rho'_B = N'_B(\rho_B)$.

By DPI

$$\mathbb{D}(\rho'_{AB} \| \rho'_A \otimes \rho'_B) = \mathbb{D}(N_A\otimes N'_B(\rho_{AB}) \| N_A\otimes N_B(\rho_A\otimes\rho_B))$$
$$\leq \mathbb{D}(\rho_{AB}\|\rho_A\otimes\rho_B). \checkmark$$

For minimized

$$\inf_{\sigma_B} \mathbb{D}(\rho'_{AB}\|\rho'_A\otimes\sigma_B) \leq \inf_{\sigma_B} \mathbb{D}(\rho'_{AB}\|\rho'_A\otimes N_B(\sigma_B))$$
$$= \inf_{\sigma_B} \mathbb{D}(N_A\otimes N_B(\rho_{AB}) \| N_A\otimes N_B(\rho_A\otimes\sigma_B))$$
$$\leq \inf_{\sigma_B} \mathbb{D}(\rho_{AB}\|\rho_A\otimes\sigma_B). \checkmark$$

Thm. Suppose V_A is a channel s.t. $V_A(\mathbb{1}_A) = \mathbb{1}_A$.

Then $H^{\uparrow}_{\mathbb{D}}(A|B)_{V_A\otimes N_B(\rho)} \geq H^{\uparrow}_{\mathbb{D}}(A|B)_\rho$.

Proof. Left as exercise.

Some particular relations

We defined

$$D_\alpha(\rho\|\sigma) = \frac{1}{\alpha-1}\log \operatorname{tr}\rho^\alpha\sigma^{1-\alpha}$$

$$\widetilde{D}_\alpha(\rho\|\sigma) = \frac{1}{\alpha-1}\log \operatorname{tr}\left(\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\,\sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha\right)$$

$$D(\rho\|\sigma) = \operatorname{tr}(\rho\log\rho - \rho\log\sigma)$$

$$D_{\max}(\rho\|\sigma) = \inf_{\rho\le\lambda\sigma}\lambda$$

Aside, think of $D_{\max}$ as

$$\sum p_i\log\frac{p_i}{q_i} \le \log\max_i\frac{p_i}{q_i} = \log\max_i d_i \qquad \lambda_i = \frac{p_i}{q_i}$$

$$= \inf_{\substack{\lambda \\ p_i\le\lambda q_i}}\lambda.$$

As defined, $D_\alpha, \widetilde{D}_\alpha$ make sense for $\alpha\in(0,1)\cup(1,\infty)$.

But, fact.

$$D_0(\rho\|\sigma) = \lim_{\alpha\to0} D_\alpha(\rho\|\sigma) = D_H^0(\rho\|\sigma)$$

$$\widetilde{D}_\infty(\rho\|\sigma) = \lim_{\alpha\to\infty}\widetilde{D}_\alpha(\rho\|\sigma) = D_{\max}(\rho\|\sigma)$$

$$D_1 = \widetilde{D}_1 = D \quad (\lim_{\alpha\to1}).$$

Fact. D_α is G.D. for $\alpha \in [0, 2]$.
$\tilde{D}_\alpha$ is G.D. for $\alpha \in [\frac{1}{2}, \infty]$.

Lemma. D_{max} and D_{max}^ϵ are G.D.'s.

Proof. Since $\lambda\sigma \geq \rho$, $\lambda \operatorname{tr}\sigma \geq \operatorname{tr}\rho \Rightarrow \lambda \geq 1$.
Thus $D_{max} \geq 0$. ✓ $D_{max} \in \mathbb{R} \cup \{\infty\}$. ✓
Monotonicity:
Let $\lambda\sigma - \rho \geq 0$.
Then $N(\lambda\sigma - \rho) = \lambda N(\sigma) - N(\rho) \geq 0$ by CP.
Thus $\lambda N(\sigma) \geq N(\rho)$.
So $\inf_{\rho \leq \lambda\sigma} \log\lambda \geq \inf_{N(\rho) \leq \lambda N(\sigma)} \log\lambda$.

Fact. Araki-Lieb-Thirring inequality, for $A, B \geq 0$ states

$$\operatorname{tr}\left[(ABA)^\alpha\right] \leq \operatorname{tr}\left[A^\alpha B^\alpha A^\alpha\right] \quad (\alpha \geq 1)$$

$$\operatorname{tr}\left[(ABA)^\alpha\right] \geq \operatorname{tr}\left[A^\alpha B^\alpha A^\alpha\right] \quad (\alpha \leq 1).$$

Applying with $\rho = B$, $\sigma^{\frac{1-\alpha}{2\alpha}} = A$,

$$D_\alpha(\rho\|\sigma) \leq \tilde{D}_\alpha(\rho\|\sigma) \quad \forall \alpha.$$

• Show the "Entropy Zoo".

Semi-Definite Programming

• Khatri + Wilde 2.4.

• A class of linear optimization problems where you optimize over the cone of semi-definite matrices.

• Can be efficiently numerically solved.

• Arises in many quantum info problems.

Def. Given a Hermiticity preserving superoperator Φ and Hermitian operators A, B, an SDP consists of two optimization problems

Primal SDP

maximize $\operatorname{tr}(AX)$

subject to $\Phi(X) \leq B$

$X \geq 0.$

Dual SDP

minimize $\operatorname{tr}(BY)$

subject to $\Phi^\dagger(Y) \geq A$

$Y \geq 0$

Let

$$S(\Phi, A, B) = \sup_{\substack{X \geq 0 \\ \Phi(X) \leq B}} \operatorname{tr}(AX)$$

$$\hat{S}(\Phi, A, B) = \inf_{\substack{Y \geq 0 \\ \Phi^\dagger(Y) \geq A}} \operatorname{tr}(BY)$$

Prop. $S(\Phi, A, B) \leq \hat{S}(\Phi, A, B)$ (Weak Duality)

Strong Inequality: Often they are equal (see book).

Examples:

These are tricky examples, they're not obvious. See notes.

- $D_n^\varepsilon(\rho\|\sigma)$
- $I_n^\varepsilon(A:B)_\rho^{\downarrow}$
- $H_{\min}(A|B)_\rho$
- Trace Norm

Useful: Evaluate efficiently. Convert min's $\leftrightarrow$ max's.

Max-Min Inequality

$$\inf_x \sup_y F(x,y) \geq \sup_y \inf_x F(x,y)$$

Smallest grant $\geq$ biggest dwarf

(Sometimes equal ... minimax theorems).

- Homework Q's?

Lecture 9

Today

- Continue SDP (Comment: Why smooth & infimize?)
- AEP

SDP

(See Previous Notes).

AEP (Asymptotic Equipartition Property).

Many divergences converge to $D(\rho\|\sigma)$ in the large N asymptotic limit.

Theorem. (Quantum Stein's Lemma).

For any ρ, σ states, any $\epsilon \in (0,1)$,

$$\lim_{n\to\infty} \frac{1}{n} D_h^\epsilon(\rho^{\otimes n}\|\sigma^{\otimes n}) = D(\rho\|\sigma).$$

Proof. Use Renyi bounds

$$D_n^\varepsilon(\rho\|\sigma) \geq D_\alpha(\rho\|\sigma) - \frac{\alpha}{\alpha-1}\log\varepsilon \qquad (\alpha\in(0,1))$$

$$D_n^\varepsilon(\rho\|\sigma) \leq \widetilde{D}_\alpha(\rho\|\sigma) - \frac{\alpha}{\alpha-1}\log(1-\varepsilon) \qquad (\alpha\in(1,\infty))$$

Firstone, see HW.

Second follows because of DPI.

$$D_\alpha(\rho\|\sigma) \geq D_\alpha(N(\rho)\|N(\sigma))$$

where $N(\rho) = \mathrm{tr}(\Lambda\rho)|0\rangle\langle 0| + \mathrm{tr}((\mathbb{1}-\Lambda)\rho)|1\rangle\langle 1|$ is the channel of a hypothesis test.

Now use that

$$D_\alpha(\rho^{\otimes n}\|\sigma^{\otimes n}) = \frac{1}{\alpha-1}\log \underbrace{\mathrm{tr}\left(\rho^{\otimes n}\sigma^{(1-\alpha)\otimes n}\right)}$$

$$= \mathrm{tr}\left((\rho^\alpha)^{\otimes n}(\sigma^{1-\alpha})^{\otimes n}\right)$$

$$= \mathrm{tr}\left(\rho^\alpha\sigma^{1-\alpha}\right)^{\otimes n}$$

$$= \left(\mathrm{tr}\left(\rho^\alpha\sigma^{1-\alpha}\right)\right)^n$$

$$= n\, D_\alpha(\rho\|\sigma) \quad \text{(Additivity)}$$

So

$$\lim_{n\to\infty}\frac{1}{n} D_n^\varepsilon(\rho\|\sigma) \geq D_\alpha(\rho\|\sigma) - \frac{1}{n}\cdot\frac{\alpha}{\alpha-1}\log\varepsilon$$

$$\leq \widetilde{D}_\alpha(\rho\|\sigma) - \cancel{X}$$

Limit as $\alpha\to 1$ gives proof.

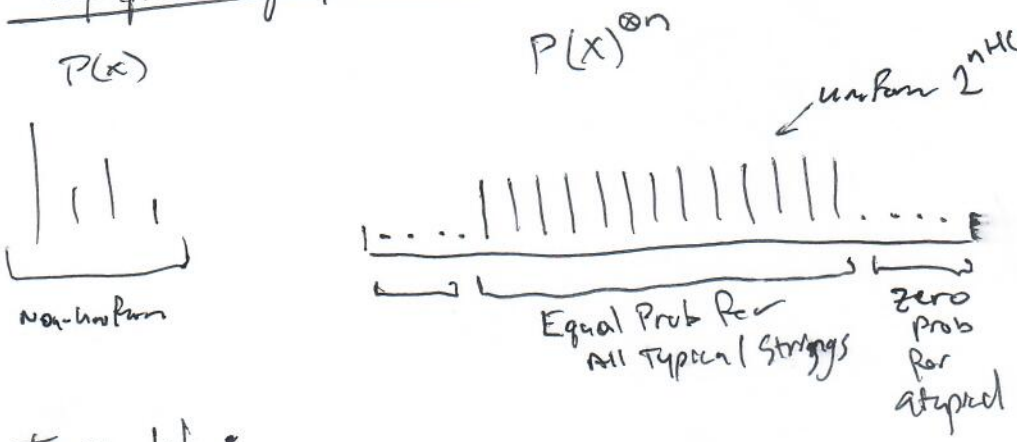
Another Term. For $t \geq 0$,

$$\lim_{n\to\infty} \frac{1}{n} D_{\max}(\rho^{\otimes n} \| \sigma^{\otimes n}) = D(\rho\|\sigma).$$

Pf. Another fancy x-proof (see notes).

- But why does this happen?

"Asymptotic Equipartition"



Typicality:

Non uniform dist $\xrightarrow{\otimes n}$ Uniform Distribution on typical subspace.

For quantum state w/ eigenbasis

$$\rho = \sum_x p(x) |x\rangle\langle x|$$

$$\rho^{\otimes n} = \sum_{x_1,x_2\ldots} p(x_1)p(x_2)\ldots |x_1,x_2\ldots\rangle\langle x_1,x_2\ldots|$$

$$= \sum_{x_n} p(x_n) |x_n\rangle\langle x_n| \quad \left(\begin{array}{c}\text{mixture of}\\ \text{strings}\end{array}\right)$$

Project into typical subspace

$$\Pi_n = \sum_{\substack{\text{typical} \\ x_n}} |x_n\rangle\langle x_n|$$

Then $\rho^{\otimes n} \simeq \Pi_n \rho^{\otimes n} \Pi_n \simeq \frac{\Pi_n \rho^{\otimes n} \Pi_n}{\mathrm{tr}(\Pi_n \rho^{\otimes n} \Pi_n)}$ b/c

you've barely gotten rid of any probability.

(Gentle Measurement Lemma.)

So always

$$\rho^{\otimes n} \simeq \sum_{\substack{\text{typical} \\ x_n}} \frac{|x_n\rangle\langle x_n|}{2^{nH(\rho)}}$$

is essentially <u>uniform</u> on typical subspace.

- So the question is, how do these entropies look to <u>uniform</u> distributions?

Suppose that

$$\rho = \frac{\Pi}{\mathrm{tr}\Pi} = \frac{\Pi}{W} \qquad \sigma = \frac{\Pi'}{W'} = \frac{\Pi'}{\mathrm{tr}\Pi'}.$$

Let's assume $\Pi \leq \Pi'$ (support), otherwise everything ∞.

QRE

$$D(\rho\|\sigma) = \mathrm{tr}\left(\rho \log \frac{\Pi'}{W} - \rho \log \frac{\Pi'^C}{W'}\right)$$

Restrict to support Π, everywhere else zero,
on support Π, $\Pi' = \Pi = \Pi$.

$$D(\rho\|\sigma) = \mathrm{tr}\left(\rho \log \frac{\Pi}{W} - \rho \log \frac{\Pi}{W'}\right)$$

$$= -\mathrm{tr}\,\rho \log W + \mathrm{tr}\,\rho \log W'$$

$$\boxed{D(\rho\|\sigma) = \log \frac{W'}{W}}$$ For uniform dists.

Renyi (The ρ, σ commute so either one.)

$$D_\alpha(\rho\|\sigma) = \frac{1}{\alpha - 1} \log \mathrm{tr}\left(\rho^\alpha \sigma^{1-\alpha}\right)$$

$$= \frac{1}{\alpha - 1} \log \mathrm{tr}\left(\frac{\Pi^\alpha}{W^\alpha} \frac{\Pi^{1-\alpha}}{W'^{1-\alpha}}\right) \quad \text{[Restrict to Support]}$$

$$= \frac{1}{\alpha - 1} \log \frac{\mathrm{tr}\,\Pi}{W^\alpha} \frac{1}{W'^{1-\alpha}}$$

$$= \frac{1}{\alpha - 1} \log \frac{W'^{1-\alpha}}{W^{1-\alpha}}$$

$$\boxed{D_\alpha(\rho\|\sigma) = \log \left|\frac{W'}{W}\right|}$$

Hyp. Test.

$$D_\epsilon^\epsilon(\rho\|\sigma) = -\log \inf_{\mathrm{tr}(M\frac{\Pi'}{w'})\geq 1-\epsilon} \mathrm{tr}\left(M\frac{\Pi'}{w'}\right)$$

For feasible M, $\mathrm{tr}(M\rho) \geq w(1-\epsilon)$. But $\Pi' \geq \Pi$, so

$$\mathrm{tr}\frac{M\Pi'}{w'} \geq \frac{\mathrm{tr}(M\Pi)}{w'} \geq \frac{w}{w'}(1-\epsilon).$$

On the other hand, Π is feasible ($\mathrm{tr}(\rho\Pi)=1$), and

$$\mathrm{tr}(\Pi\sigma) = \mathrm{tr}\frac{\Pi\Pi'}{w'} = \frac{\mathrm{tr}(\Pi)}{w'} = \frac{w}{w'}.$$

So

$$\log\frac{w'}{w} \leq D_\epsilon^\epsilon(\rho\|\sigma) \leq \log\frac{w'}{w} - \log(1-\epsilon)$$

($\log\frac{w'}{w}$: scales like n; $-\log(1-\epsilon)$: const)

In asymptotic setting, $\frac{w_n'}{w_n} = \left(\frac{w'}{w}\right)^n = \left(\frac{2^{H(\sigma)}}{2^{H(\rho)}}\right)^n$,

dividing by n gives

$$\log\left(\frac{2^{H(\sigma)}}{2^{H(\rho)}}\right) \leq \frac{1}{n} D_n^\epsilon(\rho\|\sigma) \leq \log\left(\frac{2^{H(\sigma)}}{2^{H(\rho)}}\right) - \frac{1}{n}\log(1-\epsilon)$$

$$\frac{1}{n} D_n^\epsilon(\rho\|\sigma) \to D(\rho\|\sigma).$$

Max.

$$D_{\max}(\rho\|\sigma) = \inf_{\rho \le \lambda\sigma} \log \lambda.$$

Restrict to support (for other states all $\lambda \ge 0$ valid).

$$\frac{\pi}{w} \le \lambda \frac{\pi'}{w'} \Rightarrow \lambda \ge \frac{w'}{v}. \quad \text{Also } \lambda = \frac{w'}{v} \checkmark$$

So $D_{\max}(\rho\|\sigma) = \log \frac{w'}{v}$

Fidelity.

$$F(\rho,\sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1^2 = \left(\mathrm{tr}\left|\frac{\pi}{w^{1/2}}\frac{\pi'}{w'^{1/2}}\right|\right)^2$$

$$= \left(\mathrm{tr}\left|\frac{\pi}{\text{overlap}}\right|\right)^2 = \left(\frac{w'^{1/2}}{w^{1/2}}\right)^2 = \frac{w'}{w}$$

$$-\log F(\rho,\sigma) = \log \frac{w'}{v}$$

(Note: This is $\widetilde{D}_{\frac{1}{2}}(\rho\|\sigma)$).

*Not Rigorous: we assumed π, π' commute, etc.

Summary. (Roughly Speaking)

- These all converge when applied to uniform ρ, σ, up to constants (independent of dimension of the inputs)!
- Asymptotic limit makes everything uniform!

With extra time:

- Resource Theatres
- Norms & Distances
- Questions

Lecture 10

Today

- Compression
- State Merging

Smooth Conditional Min/Max Entropy

$$H_{min}(A|B)_\rho \equiv -\inf_{\sigma_B} D_\infty(\rho_{AB} \| \mathbb{1}_A \otimes \sigma_B)$$

$$H_{max}(A|B)_\rho \equiv -\inf_{\sigma_B} \tilde{D}_{1/2}(\rho_{AB} \| \mathbb{1}_A \otimes \sigma_B)$$

$$H^\epsilon_{min}(A|B)_\rho \equiv -\inf_{\sigma_B} D^\epsilon_\infty(\rho_{AB} \| \mathbb{1}_A \otimes \sigma_B)$$

$$H^\epsilon_{max}(A|B)_\rho \equiv \inf_{\tilde{\rho} \in B_\epsilon(\rho)} H_{max}(\tilde{\rho}) .$$

↗ Breaks the pattern.

For any pure $|\psi_{ABR}\rangle$,

$$H^\epsilon_{max}(A|B)_{\rho_{AB}} = -H^\epsilon_{min}(A|R)_{\rho_{AR}} .$$

Intuition...

$$H(AB) - H(B) = H(R) - H(AR)$$
$$= -(H(AR) - H(R)) .$$

What are these? Trivial B system...

$$H_{\min}(A)_{\rho_A} = -D_\infty(\rho_A \| \mathbb{I}_A)$$

$$= -\log \max_{\substack{\lambda \\ \rho \le \lambda \mathbb{I}}} \lambda$$

$$= -\log \|\rho\|_\infty \quad \leftarrow \text{Largest E.V.}$$

~~$H_{\min}^\varepsilon(A)_{\rho_A} \le \inf_{\tilde\rho \in B_\varepsilon(\rho)} \log\|\tilde\rho\|_\infty$ ← largest E.V.~~

$$H_{\min}^\varepsilon(A)_{\rho_A} = -\inf_{\tilde\rho \in B_\varepsilon(\rho)} \log\|\tilde\rho\|_\infty \quad \leftarrow \text{smallest largest E.V.}$$

$$= \sup_{\tilde\rho \in B_\varepsilon(\rho)} \left(-\log\|\tilde\rho\|_\infty\right)$$

$$H_{\max}(A)_\rho = -\tilde D_{1/2}(\rho \| \mathbb{I})$$

$$= 2\log F(\rho, \mathbb{I})$$

$$= 2\log \operatorname{tr}\sqrt{\rho}$$

$$\le \log(\operatorname{rank}(\rho))$$

$$= -D_0(\rho\|\mathbb{I}) \quad \text{[Alt. Renyi Def.]}$$

$$* \operatorname{tr}\sqrt{\rho} = \sum_{k=1}^{r} \sqrt{\lambda_k}$$

$$\le \sqrt{\sum_{k=1}^{r} 1}\sqrt{\sum_{k=1}^{r}\lambda_k}$$

$$= \sqrt{r}$$

$$H_{\max}^\varepsilon(A)_\rho = \inf_{\tilde\rho \in B_\varepsilon(\rho)} 2\log \operatorname{tr}\sqrt{\tilde\rho}$$

$$\le \inf_{\tilde\rho \in B_\varepsilon(\rho)} \log \operatorname{rank}\tilde\rho. \quad \leftarrow \text{Compression}$$

Entanglement: Schmidt Coefficients

Bipartite pure entanglement is very simple.

- Every $|\psi_{AB}\rangle$ has Schmidt decomposition

$$|\psi_{AB}\rangle = \sum_{k=1}^{r} \sqrt{\lambda_k}\, |k\,k\rangle_{AB}$$

($\sqrt{\lambda_k}$: Schmidt coeffs!)

- The Schmidt basis is also the reduced density matrix (ρ_A or ρ_B) eigenbasis, and $\{\lambda_k\}$ are reduced density matrix E.V.s.

$$\rho_A = \sum_{k=1}^{r} \lambda_k |k\rangle_A\langle k|, \quad \rho_B = \sum_{k=1}^{r} \lambda_k |k\rangle_B\langle k|.$$

The Schmidt coeffs $\{\lambda_k\}$ completely determine the entanglement structure.

↳ Asymptotic Distillable Entanglement $H(\{\lambda_k\}) = S(\rho_A)$

↳ Asymptotic Entanglement cost $S(\rho_A)$

↳ LOCC interconvertibility.

Interconvertibility: (Nielsen Thm).

Thm. It is possible to transform $|\psi\rangle \longrightarrow |\varphi\rangle$ using LOCC if and only if the ψ schmidt coeffs $\{\lambda_k^\psi\}$ are majorized by the φ schmidt coeffs $\{\lambda_k^\varphi\}$.

Let p majorize q. Then p is more concentrated.

Def. p majorizes q if

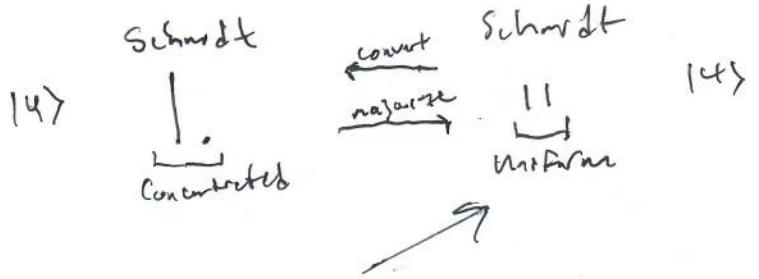
$$\sum_{i=1}^{k} p_i^{\downarrow} \geq \sum_{i=1}^{k} q_i^{\downarrow} \quad \forall k,$$

where $p_i^{\downarrow}, q_i^{\downarrow}$ are ordered large ($i=1$) to small.

- (It means: Fastest-growing cumulative distribution is always bigger.
  - (original order doesn't matter.)
  - → can be accomplished with "bilocal equivalent" operations!

Product State $|\psi_A\rangle \otimes |\psi_B\rangle$

Bell Pair $|\psi_{AB}^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$



Uniform coeffs $\Rightarrow$ More Entangled

LOCC decreases Entanglement

New $|\psi\rangle$ $\xrightarrow{\text{majorizes}}$ Old $|\psi\rangle$.

("add majorized by new")

Cor. LOCC (Pure $\to$ Pure) cannot increase Schmidt rank.

Pf. Schmidt Rank = # nonzero λ_k.

Let $|\psi\rangle$ have Schmidt rank r_ψ and $|\phi\rangle$, r_ϕ.

If ~~$r_\psi > r_\phi$ then~~ we need

$$\sum_{i=1}^{k} \lambda_k^\psi \le \sum_{i=1}^{k} \lambda_k^\phi \quad \forall k.$$

But for $k = r_\psi < r_\phi$, this is

$$1 \le \sum_{i=1}^{r_\psi} \lambda_n^\phi < 1$$

which is false!

Cor. Let $|\psi_{AB}\rangle$ be pure state and let $|\Phi_K\rangle$ be maximally entangled state of Schmidt rank K. Then

$$|\psi_{AB}\rangle \xrightarrow{\text{LOCC}} |\Phi_K\rangle$$

if and only if $\lambda_{max} \leq \frac{1}{K}$, where λ_{max} is biggest schmidt coeff.

That is, from $|\psi_{AB}\rangle$ you can 1-shot distill $\log K = -\log \lambda_{max} = H_{\min}(\rho_A)$ perfect Bell pairs.

Exercise. Use this to prove asymptotic distillation rate is $H(\rho_A)$.

Proof. For all ℓ,

$$\sum_{i=1}^{\ell} \lambda_i \leq \ell \lambda_{max} \leq \ell \frac{1}{K} = \sum_{i=1}^{\ell} \frac{1}{K}.$$

if $\lambda_{max} \leq \frac{1}{K}$.

LOCC: Kraus Operators form Conditional chans of local ops.

$$K_{ij} = (\mathbb{1}_A \otimes K^B_{j|i})(K^A_i \otimes \mathbb{1}_B)$$

↑ for each fixed i, $\{K^B_{j|i}\}_j$ is a set of Kraus operators. ("cond. channel on measurement outcome from A")

* Extending on long LO/LOCC SEPP/PPT tangent

Lemma. A product operator $A \otimes B$ acting on a pure state $|\psi\rangle$ cannot increase its Schmidt rank.

Pf. Schmidt decomp is really SVD.

$$|\psi\rangle = \sum_{ij} \underbrace{Y_{ij}}_{\text{↖ } \lambda_k = \text{Sing. Vals.}} |i\rangle \otimes |j\rangle$$

$r_\psi = \text{Rank } Y$

But

$$A \otimes B |\psi\rangle = \sum_{\substack{i,j \\ i',j'}} Y_{ij} A_{i'i} B_{j'j} |i'\rangle \otimes |j'\rangle$$

$$= \sum_{ij} (AYB^T)_{ij} |i\rangle |j\rangle.$$

And $\text{rank}(AYB^T) \le \text{rank}(Y)$ by lin. alg.

Cor. If $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ such that $r_{\psi_i} \leq K$, and Λ is LOCC, then

$$\Lambda(\rho) = \sum_j q_j |\phi_j\rangle\langle\phi_j|$$

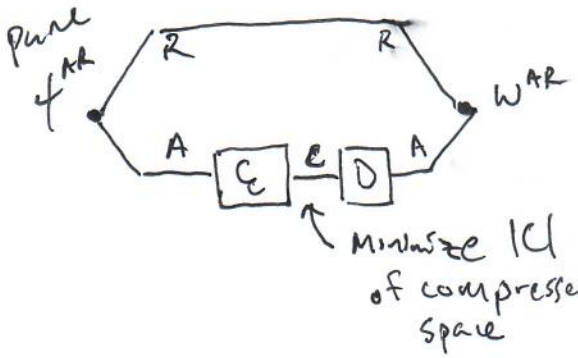
where $r_{\phi_j} \leq K$.

pf. Kraus operators are strings of products.

Finally :).

Now on to compression.

Compression



One-shot:
- Allow error ϵ.

Want $w^{AR} \approx \psi^{AR}$

Def. A compression scheme is a code $(\mathcal{E}^{A\to C}, D^{C\to A})$. It maps

$$w_{AR} = (\mathcal{E}^{A\to C} \otimes \mathrm{id}^{R\to R})(\psi_{AR}).$$

We say it has error ϵ if $P(w,\psi) \le \epsilon$ in the purified distance, and rate $R = \log|C|$. The ϵ-one-shot compressibility of ψ is

$$\ell_\epsilon(\psi) = \min_{\substack{(\mathcal{E},D)\\ \text{error}\le\epsilon}} R .$$

We'll show the lower bound is the smooth max entropy.

Thm. (Compressibility limit):

$$\ell_\epsilon(\psi) \geq H_{\max}^\epsilon(p^A)$$

(Compare to Schumacher $\ell(\psi) = S(\rho_A)$).

Proof.

- The intermediate state $\rho_{CR} = \mathcal{E}(\psi)$ has E.B.

$$\rho_{CR} = \sum_i p_i |\psi_i\rangle\langle\psi_i|$$

where $|\psi_i\rangle$ has S.R. $\leq$ Kl. Therefore

- So does

$$\omega_{AR} = \sum_j q_j |\xi_j\rangle\langle\xi_j|$$

with $|\xi_j\rangle$ S.R. $\leq |C|$, by earlier lemma, since $D^{R\to A}(\psi_{RA})$ is local.

- But since $P = \sqrt{1-F} \leq \epsilon$

$$1-\epsilon^2 \leq F(\psi,\omega) = \langle\psi|\omega|\psi\rangle = \sum p_c \langle\psi|\xi_i\rangle$$

$$\leq \max_{\substack{\text{S.R.}\leq|C| \\ \psi}} F(\psi,\psi')$$

$$= \max_{\substack{\sigma^A \\ \text{rank}\,\sigma \leq |C|}} F(\rho^A,\sigma^A) \quad \left[\text{B/c } \text{Rank}\,\sigma_A = \text{S.R.}\,\psi,\right]$$

Thus $\exists\ \sigma$, rank $\leq |\mathcal{C}|$, with

$$P(\rho^A, \sigma) \leq \epsilon \quad (F(\rho^A, \sigma) \geq 1 - \epsilon^2).$$

- Thus,

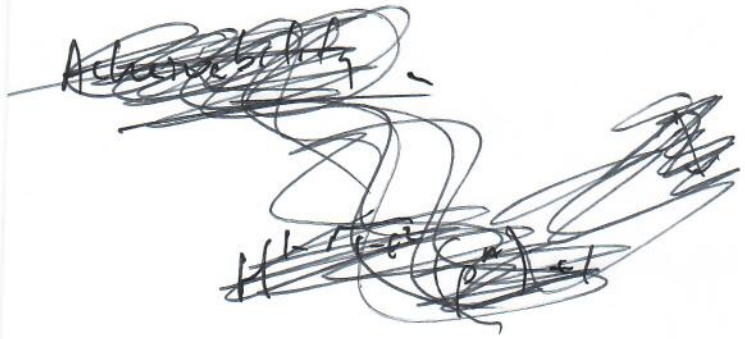
$$H_{\max}^{\epsilon}(\rho) = \min_{\tilde{\rho} \in B_\epsilon(\rho)} H_{\max}(\tilde{\rho}) \leq H_{\max}(\sigma)$$

And earlier we saw

$$H_{\max}(\sigma) = 2 \log \operatorname{tr} \sqrt{\sigma} \leq \log \operatorname{rank} \sigma$$

$$\leq \log |\mathcal{C}| \ \checkmark$$

So ( ) $\leq$ . . .



Achievability

Consider $H_H^\delta(\rho) = \log \min_{\mathrm{tr}M\rho \geq 1-\delta} \mathrm{tr}M$.

- If M is s.t. $\mathrm{tr}M\rho \geq 1-\delta$, then also

$$M' = \sum_n |n\rangle\langle k|\, M\, |k\rangle\langle l|$$

(ρ eigenbasis; "pinching"; "dephasing"; inside about pinching?)

always $\mathrm{tr}(M\rho) = \mathrm{tr}\left[M' \sum_n |n\rangle\langle n|\rho|k\rangle\langle l|\right] = \mathrm{tr}(M'\rho)$.

- So we only need to consider M' that are diagonal in ρ eig. basis.

- Let $\rho = \sum_{k=1}^{d} \lambda_k |k\rangle\langle k|$ with $\lambda_0 \geq \lambda_1 \geq \cdots$.

For fixed $\mathrm{tr}M$, optimal M' is

$$M' = \sum_i \mu_i |i\rangle\langle i|, \quad \mu_i = \underbrace{1\ 1\ 1\ 1\ 1}_{r \text{ terms}}\ 0\ 0\ 0 \cdots$$

Let P projector onto this subspace for optimal M.

$$\mathrm{tr}M \leq \mathrm{tr}P \leq \mathrm{tr}(M) + 1$$

Now compress to this subspace

$$C(\rho) = P\rho P + \sigma(\mathbb{1}-P)\rho(\mathbb{1}-P).$$

Decompress w/ (unit) isometry.

$D \circ C(\rho) \approx \rho$ above

• The output

$$\omega_{AR} = D_c(\psi_{AR}) \geq \psi_{AR}(P\otimes\mathbb{1}_R)\psi_{AR}(P\otimes\mathbb{1}_R)$$

so

$$\begin{aligned}
\operatorname{tr}\psi_\omega &\geq \operatorname{tr}(\psi(\mathbb{1}\otimes P)\psi(\mathbb{1}\otimes P)) \\
&= \sum |\langle\psi|P\otimes\mathbb{1}|\psi\rangle|^2 \\
&= |\operatorname{tr}[\psi(P\otimes\mathbb{1})]|^2 \\
&= (\operatorname{tr}(P\rho_A))^2 \\
&\geq (\operatorname{tr}(M\rho_A))^2 \\
&\geq (1-\delta)^2.
\end{aligned}$$

Let $1-\epsilon^2 = (1-\delta)^2$ to get purified distance bound.

Thus $\exists C$ with $|c| \leq \operatorname{tr}P \leq \operatorname{tr}\sigma 1$.

$$\boxed{l_t(\psi) \leq H_h^{1-\sqrt{1-\epsilon^2}}(\rho_A) + 1}.$$

Also a similar lower bound

$$H_h^{\epsilon^2(2-\epsilon^2)}(\rho_A) \leq l_t(\psi).$$

Lecture 11

Today:

- Compression
- State merging / QC

Compression:

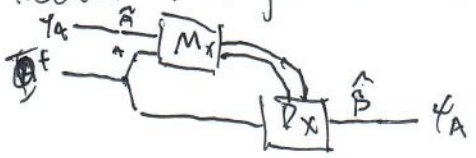
↳ Finish previous lecture.

Compression ⟶ Teleportation.

- Compressibility $k_\epsilon(t)$ is also the number of ebits you need to teleport t from $A \to B$!

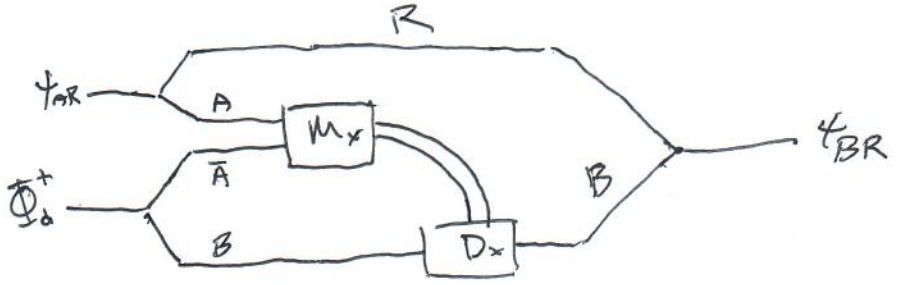
~~After protocol you have k ebits for teleportation~~
~~q state available~~

- Recall teleportation.



M_x = Bell basis measurement
P_x = Pauli correction

- You can do the same for Qdits + RCS...

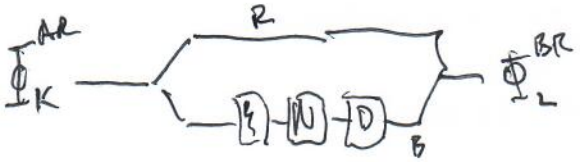


Schmidt rank (#ebits) of $\Phi_d^{\otimes}$ is $= |A|$ for this to work.

$$\psi_{AR} \xrightarrow[\text{to } \tilde{\psi}]{\text{compress}} \tilde{\psi} \xrightarrow[\#\text{ebits}]{\text{Teleport}} \psi'_{AR}.$$

- Quantum Communication =

Entanglement Distillation / Transmission using N + Teleportation (Perfect Channel)



[Picture Above]

Entanglement Transmission (Quantum Capacity)

Quantum Part

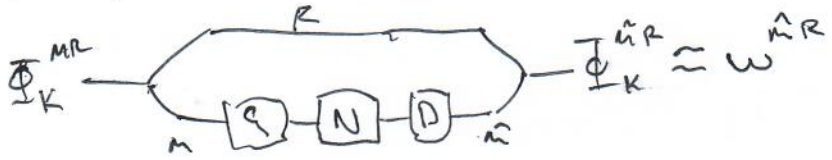
generates state ψ_{ABE}

State Merging
- Distillation
- Teleportation

LOCC

1. Q Comm (Entanglement Transmission)

Def. An ϵ-one-shot quantum error correcting code is a pair of encoding/decoding maps



Error: $P_{err} = P(\Phi^{MR}, \omega^{\hat{m}R})$

Rate: $R = \log K$

Capacity: $Q_\epsilon(N) = \max_{P_e \le \epsilon} R$.

* Many variants! See K+W for comparisons.

A general result:

Thm. $Q_\epsilon(N) \ge \max_{\phi_{AA'}} H_{min}^{\epsilon^2/5}(A|E)_\psi - 2\log\frac{5}{\epsilon^2}$

where $\psi_{ABE} = V_N^{A'\to BE}(\phi)$ is Stinespring of N.

Conditional min-entropy $\Rightarrow I(A>B)$ asymptotically.

* But not 1-shot tight.

* Converse: Adapt Meta-converse

Proof?

- See notes: uses postprocessing by S.M., then teleport.
- Also see Kew for other protocols.
- Notes proof uses

$$Q_E(N) \geq \max_{\phi_{AB}} \left(-\tilde{I}(A)\right)$$

↖ state merging capacity of ψ_{ABE}.

Examples:

(1) Identity $\mathbb{I}_A$

- Clearly $R = \log|A|$ achievable/zero error.
- Also $F(\mathcal{T}, \omega) \leq \frac{|A|}{k}$ by the S. Rank argument.

$$\log|A| \leq Q_\epsilon(I_A) \leq \log|A| - \log(1-\epsilon^2)$$

(2) Entanglement Breaking (Separable output)

- output is a mixture of S. Rank 1.

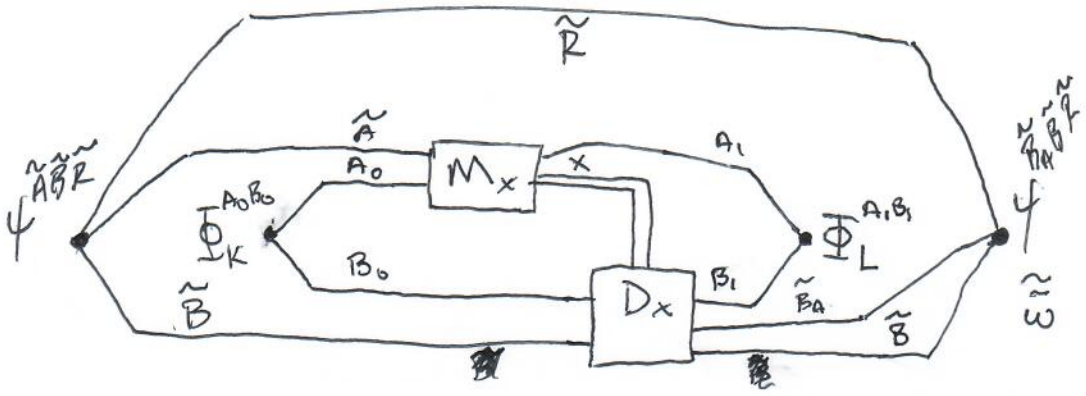
$$F(\mathcal{T}, \omega) \leq \frac{1}{k}.$$

(overlap of reduced density matrices).

(3) Anti-degradable

state merging $\longrightarrow$ No cloning $\longrightarrow$ Low Capacity

Quantum State Merging



Def. A <u>Q.S.M. scheme</u> for pure $\psi^{\tilde{A}\tilde{B}\tilde{R}}$ is a joint LOCC operation

$$\Lambda : \tilde{A}A_0 \otimes \tilde{B}B_0 \longrightarrow A_1 \otimes B_1 \tilde{B}_A \tilde{B}$$

with error ϵ,

$$F\left(\omega_{A_1B_1\tilde{B}_A\tilde{B}\tilde{R}}\,,\ \Phi_L^{A_1B_1} \otimes \psi^{\tilde{B}_A\tilde{B}\tilde{R}}\right) \geq 1-\epsilon$$

where

$$\omega_{A_1B_1\tilde{B}_A\tilde{B}\tilde{R}} \equiv \Lambda\left(\Phi_K^{A_0B_0} \otimes \psi^{\tilde{B}_A\tilde{B}\tilde{R}}\right)$$

with rate (entanglement cost)

$$R = \log K - \log L\,.$$

*What does CQCC mean if spaces change?

The smallest achievable rate

$$\tilde{R}_t(\psi) = \inf_{\text{Perfect}} R$$

is the merging cost.

* Different entropies

- Gives meaning to negative conditional entropies.

+Rate ↙ Need to input Entanglement (Teleportation)

-Rate ↘ Merge + get entanglement as output (distillation).

Examples:

- Note: By teleporting A using $\tilde{\Phi}_{AA'}$ & generating nothing, $\tilde{R}_t(\psi) \leq \log |A|$.

Examples ...

~~ψ_{AR}~~

(1) $\psi_{ABR} \equiv \psi_{AR}$. (Teleportation)

- Compress + Teleport!

  $\tilde{\ell}_E \leq \ell_E$

- Schmidt Rank Argument for similar converse as ℓ_E.

(2) $\psi_{ABR} = \psi_{AB}$ (Distillation)

- Bob creates ψ_{AB} locally
- Use argument ψ_{AB} to Alice
- Entanglement generated from ψ_{AB} is $H^{\epsilon}_{min}(\rho_A)$ using smoothed version of arguments from before

(3) GHZ

$|\psi_{ABR}\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$

$R \simeq 0$ ...

Teleportation / Distillation cancel out.

The general picture:

Thm. *Note after 6.

$$H_{max}^{\varepsilon}(A|B) \leq \tilde{I}_t(t) \leq -H_{min}^{\varepsilon^2/6}(A|R)_\psi + 2\log\frac{5}{\varepsilon^2}$$

(Recall Minimum duality)

Proof?

All relies on decoupling:

$$\rho_{A_1\tilde{R}}^{\kappa} \simeq \tau_{A_1} \otimes \rho_{\tilde{R}}$$

After Alice's measurement.

↳ Necessary & sufficient

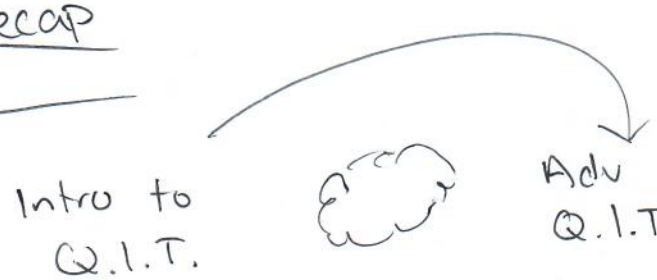
↳ Achieved using Randomness.

Lecture 12

Today:

- Recap
- Misc

Recap



"one-shot" Quantum Shannon Theory

Lecture 2
- Classical Information Theory
  - Compression: $H(p)$ = min bits
  - Noisy Channel Coding Thm
    - ↳ Fundamental Limits on Communication
  - Relative Entropy is King
    - ↳ Monotonicity, Non-Negativity, Joint Convexity, Chain Rule
  - Information Can be Quantified!

- Entropy
  - ↳ Axiomatic Derivation of $H(p) = -\sum p \log p$ (2)
  - ↳ How many bits? (Variable length coding). Kraft-McMillan.
  - ↳ Uncertainty
- Typicality → Typical strings $-\log p = nH(p)$ w/ prob. 1.
- Info in a book?

Lecture 3

- Classical Channels & Prob. Theory Notation.
- Markov Chains, Data Processing
- Noisy Channel Coding Theorem
  - At what rate can channel reliably send messages!

  =

  Channel Capacity
- Proof used
  - ↳ Random Coding
  - ↳ Jointly Typical Decoding

* Lecture 4

- Quantum Channels

- Classical Comm over Quantum Channels.
  - Plain Codes
    - Achievability: [redacted]
      - Random Coding
      - SQRT Measurement Decoding
        - ↳ Hayashi/Nagaoka Lemma
      - Hypothesis Testing RE (Wang/Renner)
    - ↳ Proved ϵ-one-shot capacity
      - ~ ϵ-hypothesis χ quantity.
        (CQ mutual info across channel).
  - EA Codes
    - Achievability:
      - Position Based Coding
      - SQRT meas / [redacted] H-N Lemma
      - Hyp testing @ each position
    - ↳ Proved ϵ-one-shot capacity
      - ~ ϵ-hyp testing mutual info
- Converses: Proved general meta-converses based on Monotonicity & Comparator, using generalized divergences.

- Asymptotic Limit
  - ↳ Asymptotic Equipartition property
    i.e. Quantum Typicality.
  - ↳ Proved ϵ-one-shot capacities for $n \to \infty$ become
    $\chi_{reg}(N)$, $I(N)$.
  - ↳ Using AEP for Hyp Testing

- The Zoo of Entropies
  - Generalized Divergences
    - ↳ Entropy, Mutual Info, Conditional
    - ↳ Monotonicity rules all.
    - ↳ ϵ-Smoothing
      σ_B-Infimizing
    - ↳ α-Renyi's & minimax
  - SDP's / minimax / Duality
  - Entropy Zoo

- Quantum Compression & Communication
  - Smallest Hilbert space ρ can be compressed to with ϵ-error.
    - ϵ-one-shot compressibility $\sim H_{max}^{\epsilon}(\rho)$
      - Duality of ϵ-smooth min/max entropies
        - Asymptotically: AEP $\rightarrow H(\rho)$ = Schumacher Compression

  - Bottlenecks based on the Schmidt rank
  - Majorization & LOCC-interconvertibility (Entanglement Theory)
    - what is LOCC?
  - Q. Comm = Entanglement Distillation + teleportation
    - one-Shot Distillation H_{min}^{ϵ}
    - Quantum State Merging $I(A\rangle B) = -H(A|B)$

↳ ϵ-one-shot Quantum capacity
in terms of entanglement transmission
↳ Achievability w/ Q.S.M. → $I(A\rangle B)$

↳ Method for Q. Comm:
- Keep info out of (Sleeping) Environment
- Decouple Alice from Env.
- Distribute Ent. Then use LOCC.

A few priciples:

- Information can be quantified!
  Which entropic quantity fits the task?
- Relative Entropy Monotonicity!
  ↳ Know whether your proof is specific, or for the particular quantity -
  ↳ What property are you using!
- Revisit Entropy Zoo PDF

Finally, a little bit of

- Resource Theory
- Entanglement / Correlations

Resource Theory:

- Free States ($\mathcal{S}$)
- Free Operations ($\mathcal{F}$)
  - ↳ Leave free states invariant
- Relative Entropy Measures
  - $E_R(\rho) = \inf_{\sigma \in \mathcal{S}} D(\rho \| \sigma)$
  - Resource monotone!

Ex. Purity.

Free State: Maximally Mixed $\tau = \frac{\mathbb{1}}{d}$

Free Ops: Unital

$\Phi(\tau) = \tau$.

RE of Purity $= D(\rho \| \frac{\mathbb{1}}{d})$

$= \log d - H(\rho)$.

Ex.

Resource: Total Correlations

Free States: Product States

$\rho_A \otimes \rho_B$

Free Operations: Local Operations

$\Lambda_A \otimes \Lambda_B$

RE of correlations:

$$E(\rho) = \inf_{\sigma_A, \sigma_B} D(\rho_{AB} \| \sigma_A \otimes \sigma_B)$$

$$= D(\rho_{AB} \| \rho_A \otimes \rho_B)$$

= Mutual Info!

Ex.

Resource: Entanglement

Free States: Separable States

Free OPs: LOCC

RE: $E(\rho) = \inf_{\text{other}} D(\rho \| \sigma)$

RE of entanglement!

Entanglement & Quantum Correlations

- Product State
  $\rho = \rho_A \otimes \rho_B$
- Classically Correlated State (CC)
  $\rho = \sum_{ij} p_{ij} |i\rangle\langle i| \otimes |j\rangle\langle j|$ (Local O.N. Bases)
  (classical prob dist)
- Separable State
  $\rho = \sum_k p_k \rho_k^A \otimes \rho_k^B$
  $\rightarrow \rho = \frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 11|)$
  is SEP but not CC.
- Entangled State
  ρ = Not Separable
- Bell Nonlocal State

- Mutual Info
- RE of Quantumness/Discord
- RE of Entanglement

- Discord: Difference in mutual info
  ↳ Different Operational Tasks!
  For Comms we care that Ent. is King.