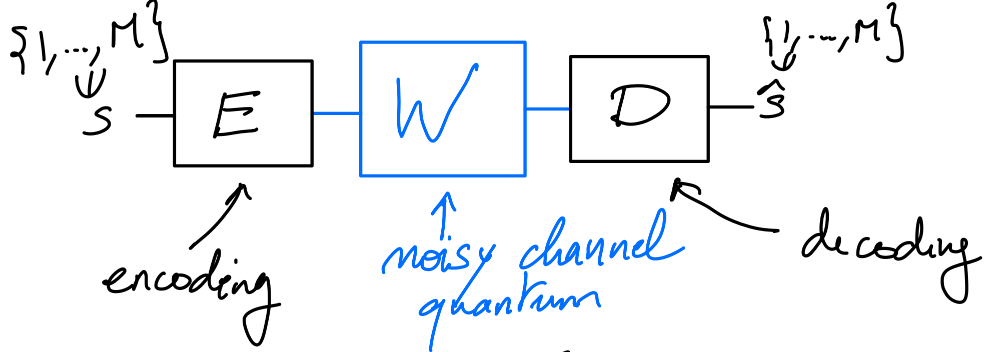


Objective : * $\mathbb{P}\{s \neq \hat{s}\}$ small

* M large.

1. Classical - quantum channels.

Input of W is classical

finite set

Def: A classical - quantum channel W with input space $\mathcal{X}$ and output space B is a collection $\{W_x\}_{x \in \mathcal{X}}$ of density operators W_x acting on B.

← Hilbert space

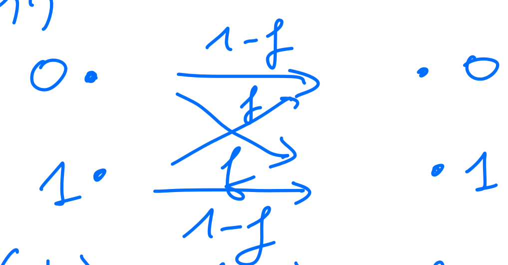
Ex: • Classical channel : $\{W(y|x)\}_{x\in\mathcal{X}, y\in\mathcal{Y}}$

$W(y|x) =$ probability output y for input x.

For example : Binary symmetric channel flip probability f

$\in [0,1]$

$\mathcal{X} = \{0,1\}$ $\qquad$ $\mathcal{Y} = \{0,1\}$



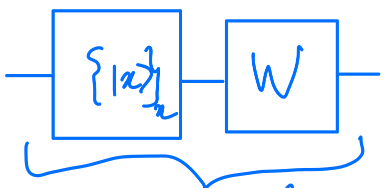
$W(0|0) = W(1|1) = 1-f, \quad W(0|1) = W(1|0) = f.$

Can see it as a classical-quantum channel with output B a Hilbert space of dimension $|\mathcal{Y}|$

$$W_x = \sum_{y \in \mathcal{Y}}' W(y|x)\, |y\rangle\langle y|.$$

where $\{|y\rangle : y \in \mathcal{Y}\}$ is a fixed orthonormal basis.

- $W_0 = |0\rangle\langle 0|$ and $W_1 = |+\rangle\langle +|$ $\quad |+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$
- Can see W as a quantum channel that starts by measuring in a basis $\{|x\rangle\}_{x \in \mathcal{X}}$ followed by preparation

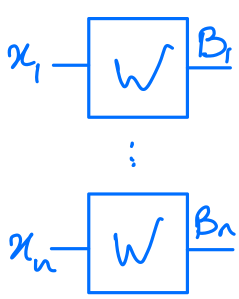


quantum channel $\mathcal{W}$ satisfying $\mathcal{W}(|x\rangle\langle x|) = W_x$ for $x \in \mathcal{X}$.

and $\mathcal{W}(|x\rangle\langle x'|) = 0$ for $x \neq x'$

- Given W and $n \geq 1$ integer, can define $W^{\otimes n}$: input $\mathcal{X}^n$ and output $B^{\otimes n}$

$$(W^{\otimes n})_{x_1 \cdots x_n} = W_{x_1} \otimes W_{x_2} \otimes \cdots \otimes W_{x_n}.$$



$[M] := \{1, \dots, M\}$.

Def: An M-code (E, D) for W is given by

- $E : [M] \to \mathcal{X}$ encoding function
- Decoding is a POVM $\{D_s\}_{s \in [M]}$ on B.

Ex: For a classical channel, we may assume $\{D_s\}$ are diagonal $D_s = \mathrm{diag}(D_s(y) : y \in \mathcal{Y})$

$D_s(y)$ = Probability of decoding to s when seeing y.

POVM condition: $\sum_s D_s(y) = 1$.

Def: The error probability $p_{err}(E, D)$ of an M-code for W is defined by

$$p_{err}(E, D) = 1 - \underbrace{\frac{1}{M} \sum_{s \in [M]} \mathrm{Tr}(D_s W_{E(s)})}_{\text{probability of successful decoding}}.$$

If $p_{err}(E, D) \leq \varepsilon$, we say that (E, D) is an (M, ε)-code

Remark: Used a uniform prior on $[M]$, another natural choice is

$$p_{err,max}(E, D) = \max_{s \in [M]} 1 - \mathrm{Tr}(D_s W_{E(s)})$$

p_{err} and $p_{err,max}$ are related (see TD)

$x \in \{0,1\}^{10} \longrightarrow y \in \{0,1\}^{10}$ flip each bit w.p. $1/4$

$M = 2$. $E(1) = \overbrace{0 \dots 0}^{10 \text{ bits}}$ $E(2) = \overbrace{1 \dots 1}^{10 \text{ bits}}$ $D_0 = \sum_{y : |y| \leq 5} |y\rangle\langle y|$, 31/01/2023

Ex: • $\dim B = |\mathcal{X}|$. $W_x = |x\rangle\langle x|$.

$(|\mathcal{X}|, 0)$ code given by

$E(s) = |s\rangle\langle s|$ (identifying $\mathcal{X}$ with $[|\mathcal{X}|]$)

$D_s = |s\rangle\langle s|$.

$$\frac{1}{|\mathcal{X}|} \sum_{s} \mathrm{Tr}\left(|s\rangle\langle s| \cdot |s\rangle\langle s|\right) = 1.$$

• Let $\rho \in S(B)$ and $W_x = \rho$ for all $x \in \mathcal{X}$.

(useless channel, output does not depend on input)

For any choice of (E, D), we have

$$\sum_{s \in [M]} \mathrm{Tr}(D_s \rho) = 1. \Rightarrow p_{err}(E,D) = 1 - \frac{1}{M}$$

Question: Fixed ε, largest M for which there exists an (M, ε)-code for W?

$$M^{opt}(W, \varepsilon) = \max\{M : \exists\ (M,\varepsilon)\text{-code for } W\}.$$

Objective: Characterize $M^{opt}(W, \varepsilon)$ in terms of "simple" properties of W.

Important special case: • $W^{\otimes n}$ with large n.

• ε small

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \underbrace{\frac{\log_2 M^{opt}(W^{\otimes n}, \varepsilon)}{n}}_{\text{number of bits transmitted per channel use}} = ?$$

Intuition: $M^{opt}(W,\varepsilon)$ should be given by a correlation measure between input and output of W.

Given a probability measure P_X on $\mathcal{X}$ let

$$\rho_{XB} = \sum_{x\in\mathcal{X}} P_X(x)\, |x\rangle\langle x| \otimes W_x \quad \text{cq-state}$$

Recall we write $\rho_X = \mathrm{Tr}_B\, \rho_{XB}$ and $\rho_B = \mathrm{Tr}_X\, \rho_{XB}$.

To characterize $M^{opt}(W,\varepsilon)$ need:

* upper bound (called converse)
* lower bound (called achievability)

Converse

Th: If there exists an (M,ε) code for W then

$$\log M \leq \sup_{P_X} \underbrace{D_H^{\varepsilon}(\rho_{XB} \| \rho_X \otimes \rho_B)}_{\text{correlation measure}}$$

Rk: Channel W is arbitrary, "one-shot" entropy measure expected. For W iid, D_H^{ε} will become a relative entropy D.

Proof: Consider an (M,ε) code (E,D).

Let $C = \{x\in\mathcal{X} : \exists s\in[M] : E(s)=x\}$

and define $P_X(x) = 1/|C|$ for $x\in C$ and $P_X(x)=0$ $x\notin C$

Then

$$\rho_{XB} = \frac{1}{|C|}\sum_{x\in C}' |x\rangle\langle x|\otimes W_x.$$

$$\mathrm{Tr}_A(\rho_{XB}) = \frac{1}{|C|}\sum_{x\in C}\mathrm{Tr}_B(|x\rangle\langle x|\otimes W_x) = \frac{1}{|C|}\sum_{x\in C}|x\rangle\langle x|.$$

and

$$F = \frac{|C|}{M}\sum_{x\in C}' |x\rangle\langle x|\otimes\Big(\sum_{s:E(s)=x}' D_s\Big)$$

As $\{D_s\}$ is a POVM, $0\le F\le I$.

$$\mathrm{Tr}(F\rho_{XB}) = \frac{1}{M}\sum_{x\in C}' \mathrm{Tr}\Big(\Big(\sum_{s:E(s)=x}' D_s\Big)W_x\Big)$$

$$= \frac{1}{M}\sum_{s=1}^{M}{}' \mathrm{Tr}(D_s W_{E(s)})$$

$$\geq 1-\varepsilon.$$

by the fact that (E,D) is an (M,ε)-code.

On the other hand,

$$\mathrm{Tr}(F\rho_X\otimes\rho_B) = \frac{|C|}{M}\mathrm{Tr}\Big(\Big(\sum_{x\in C}' |x\rangle\langle x|\otimes\Big(\sum_{s:E(s)=x}' D_s\Big)\Big)(\rho_X\otimes\rho_B)\Big)$$

$$\rho_X = \frac{1}{|C|}\sum_{x\in C}' |x\rangle\langle x|$$

$$= \frac{|C|}{M}\sum_{x\in C}' \mathrm{Tr}\Big(\Big(|x\rangle\langle x|\otimes\sum_{s:E(s)=x} D_s\Big)\Big(\frac{|x\rangle\langle x|}{|C|}\otimes\rho_B\Big)\Big)$$

$$= \frac{1}{M}\mathrm{Tr}\Big(\Big(\sum_{x\in C}'\sum_{s:E(s)=x}' D_s\Big)\rho_B\Big)$$

$$= \frac{1}{M}$$

$$\sup_{0\le F\le I}\{-\log\mathrm{Tr}(F\rho_X\otimes\rho_B) : \mathrm{Tr}(F\rho_{XB})\ge 1-\varepsilon\}$$

So $D_H^{\varepsilon}(\rho_{XB}\,\|\,\rho_X\otimes\rho_B) \geq \log M$ ∎

Achievability

Th: For any $\varepsilon \in (0,1)$
and any $\varepsilon' \in (0,\varepsilon)$ and $c>0$] tunable parameters for the bound.
and M satisfying:

(*) $$\log M \leq \sup_{P_X} D_H^{\varepsilon'}\left(\rho_{XB} \| \rho_X \otimes \rho_B\right) - \underbrace{\log \frac{2+c+c^{-1}}{\varepsilon-(1+c)\varepsilon'}}_{\text{Think of this as a small error term}}$$

then exists an (M,ε)-code.

Not the same ε but can choose it arbitrarily close to ε.

Rk: * Achievability statement matches converse up to error terms that are "small" in many settings of interest.

* Proof uses the probabilistic method: does not give an explicit (E,D) that is an (M,ε) code but rather we choose (E,D) at random and show that on average, it has an error probability $\leq \varepsilon$.

Proof: Let $\varepsilon' < \varepsilon$, $c>0$, P_X distribution on $\mathcal{X}$.
By definition, there exists $F \in \mathrm{Pos}(X \otimes B)$ s.t.

$$\mathrm{Tr}(F \rho_{XB}) \geq 1-\varepsilon'$$

and $$\mathrm{Tr}(F \rho_X \otimes \rho_B) = 2^{-D_H^{\varepsilon'}(\rho_{XB} \| \rho_X \otimes \rho_B)}.$$

Will construct (E,D) from F.

$* \; E : [M] \to \mathcal{X}$ (M should satisfy $\circledast$)
choose $E(s)$ random with distribution P_X.
independently for every $s \in [M]$.

$* \; \{D_s\}_{s \in [M]}$ a POVM.
Want to use the test F distinguishing ρ_{XB} from $\rho_X \otimes \rho_B$.

operator on B.

Simple to see that we may assume $F = \sum_{x \in \mathcal{X}} |x\rangle\langle x| \otimes F_x$

Would like to set $D_s = F_{E(s)}$
$\longrightarrow$ does not work as $\sum_s F_{E(s)} \neq I$ in general.

$v_x \geq 0$

$\longrightarrow$ we have to normalize it.

$\frac{v_x}{\sum_{x'} v_{x'}}$

Let $\Lambda = \sum_{s \in [M]} F_{E(s)} \geq 0$.

$$D_s = \Lambda^{-1/2} F_{E(s)} \Lambda^{-1/2}$$

$\left[\Lambda^{-1/2} = \sum_{i : \lambda_i \neq 0} \lambda_i^{-1/2} |e_i\rangle\langle e_i| \text{ for } \Lambda = \sum_i \lambda_i |e_i\rangle\langle e_i| \right]$

Note that $D_s \geq 0$
and $\sum_s D_s = \Lambda^{-1/2} \sum_s F_{E(s)} \Lambda^{-1/2} = \Lambda^{-1/2} \Lambda \Lambda^{-1/2} = I$.

Rk: This construction is sometimes called pretty-good measurement.

Compute error probability
For a fixed s it is given by:

$$1 - \mathrm{Tr}\big(\underbrace{D_s W_{E(s)}}_{\text{decode } s \text{ when } s \text{ is transmitted}}\big) = \mathrm{Tr}\big((I - D_s) W_{E(s)}\big)$$
$$= \mathrm{Tr}\big((I - \Lambda^{-1/2} F_{E(s)} \Lambda^{-1/2}) W_{E(s)}\big).$$

Operator inequality (Hayashi-Nagaoka)

For any $c>0$, $0\le S\le I$, $0\le T$,

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

↑ ie difference is a positive semidefinite operator.

Elementary fact: For $A\le B$ and $W\ge 0$

$$\mathrm{Tr}(AW)\le\mathrm{Tr}(BW)$$

Use Hayashi-Nagaoka + fact:

↳ with $S=F_{E(s)}$ and $T=\sum'_{s'\neq s}F_{E(s')}$.

$$\mathrm{Tr}\left((I-\Lambda^{-1/2}F_{E(s)}\Lambda^{-1/2})W_{E(s)}\right)\le(1+c)\mathrm{Tr}\left((I-F_{E(s)})W_{E(s)}\right)+(2+c+c^{-1})\sum'_{s'\neq s}\mathrm{Tr}\left(F_{E(s')}W_{E(s)}\right)$$

So

$$\begin{aligned}P_{err}(E,D)&=\frac{1}{M}\sum'_{s\in[M]}\left(1-\mathrm{Tr}(D_sW_{E(s)})\right)\\&\le\frac{1}{M}\sum'_{s\in[M]}\Big[(1+c)\,\mathrm{Tr}\left((I-F_{E(s)})W_{E(s)}\right)\\&\qquad+(2+c+c^{-1})\sum'_{s'\neq s}\mathrm{Tr}\left(F_{E(s')}W_{E(s)}\right)\Big]\end{aligned}$$

Rk: we did not use our choice for E so far, we now use it by computing the expectation over the choice of E.

$$\mathbb{E}\{P_{err}(E,D)\}\le\frac{1}{M}\sum'_{s}\Big[(1+c)\underbrace{\mathbb{E}\{\mathrm{Tr}((I-F_{E(s)})W_{E(s)})\}}_{1}+(2+c+c^{-1})\underbrace{\sum'_{s'\neq s}\mathbb{E}\{\mathrm{Tr}(F_{E(s')}W_{E(s)})\}}_{2}\Big]$$

↗ over the randomness in choice of E.

① $\mathbb{E}\left\{ \mathrm{Tr}\left((I - F_{E(s)}) W_{E(s)}\right)\right\} = \sum_{x \in \mathcal{X}}' \mathbb{P}\{E(s) = x\} \, \mathrm{Tr}\left((I - F_x) W_x\right)$

$= \sum_{x \in \mathcal{X}}' P_X(x) \, \mathrm{Tr}\left((I - F_x) W_x\right)$

$= 1 - \mathrm{Tr}\left(\left(\sum_x' |x\rangle\langle x| \otimes F_x\right) \cdot \sum_x' P_X(x) |x\rangle\langle x| \otimes W_x\right)$

$= 1 - \underbrace{\mathrm{Tr}\left(F \rho_{XB}\right)}_{\geq 1 - \varepsilon'}$ by assumption

$\leq \varepsilon'.$

② $\mathbb{E}\left\{ \mathrm{Tr}\left(F_{E(s')} W_{E(s)}\right)\right\} = \sum_{x, x' \in \mathcal{X}}' \mathbb{P}\{E(s) = x, E(s') = x'\} \, \mathrm{Tr}\left(F_{x'} W_x\right)$

$= \sum_{x, x'}' P_X(x) P_X(x') \, \mathrm{Tr}\left(F_{x'} W_x\right)$

$= \mathrm{Tr}\left(\left(\sum_{x'}' |x'\rangle\langle x'| \otimes F_{x'}\right)\left(\left(\sum_{x'}' P_X(x') |x'\rangle\langle x'|\right) \otimes \left(\sum_x' P_X(x) W_x\right)\right)\right)$

$= \mathrm{Tr}\left(F \; \rho_X \otimes \rho_B\right)$

$= 2^{-D_H^{\varepsilon'}(\rho_{XB} \| \rho_X \otimes \rho_B)}$

As a result

$$\mathbb{E}\{ p_{err}(E, D)\} \leq (1 + c)\varepsilon' + (2 + c + c^{-1}) \underbrace{(M - 1)}_{\text{because we sum over } s' \neq s.} 2^{-D_H^{\varepsilon'}(\rho_{XB} \| \rho_X \otimes \rho_B)}$$

Using condition on M in (∗)

$\leq \varepsilon$

$\Rightarrow$ There exists (E, D) s.t. $p_{err}(E, D) \leq \varepsilon$ ∎

This characterization of $\log M^{opt}(W,\varepsilon)$ is very general.

Important special case where we can evaluate the expression more explicitly: Memoryless channel $W^{\otimes n}$.

<u>Def</u>: Let W be a cq channel.
The classical capacity $C(W)$ of W is defined by

$$C(W) := \lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{\log M^{opt}(W^{\otimes n},\varepsilon)}{n}$$

optimal rate for transmitting information.

<u>Corollary</u>: For any cq channel W

$$\sup_{P_X} I(X:B)_{\rho_{XB}} \le C(W) \le \sup_n \frac{1}{n} \sup_{P_{X_1\dots X_n}} I(X_1\dots X_n : B_1\dots B_n)_{\rho_{X_1\dots X_n B_1\dots B_n}}$$

where $\rho_{X_1\dots X_n B_1\dots B_n} = \sum_{x_1\dots x_n} P_{X_1\dots X_n}(x_1\dots x_n)\, W_{x_1}\otimes W_{x_2}\otimes\dots\otimes W_{x_n}$

Notation: $X^n := X_1\dots X_n$ $\quad B^n = B_1\dots B_n$

<u>Proof</u>:

We have

$$\lim_{\substack{\varepsilon\to 0\\ n\to\infty}} \frac{1}{n}\sup_{P_{X^n}} D_H^{\varepsilon/3}(\rho_{X^nB^n}\|\rho_{X^n}\otimes\rho_{B^n}) - \underbrace{\frac{1}{n}\log\left(\frac{4}{\varepsilon/3}\right)}_{\to 0} \le C(W) \le \lim_{\substack{\varepsilon\to 0\\ n\to\infty}} \frac{1}{n}\sup_{P_{X^n}} D_H^{\varepsilon}(\rho_{X^nB^n}\|\rho_{X^n}\otimes\rho_{B^n})$$

We should evaluate $\lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{1}{n} \sup_{P_{X^n}} D_H^\varepsilon\left(\rho_{X^nB^n} \| \rho_{X^n} \otimes \rho_{B^n}\right) =: \alpha$

- $\alpha \geq \sup_{P_X} I(X:B)_{\rho_{XB}}$ Let P_X achieve the sup.

Choose $P_{X^n} = P_X \otimes P_X \cdots \otimes P_X$ ($X_1 \dots X_n$ independent distribution P_X).

$$\alpha \geq \lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{D_H^\varepsilon\left(\rho_{XB}^{\otimes n} \| \rho_X^{\otimes n} \otimes \rho_B^{\otimes n}\right)}{n}$$

Stein lemma:

$$= D(\rho_{XB} \| \rho_X \otimes \rho_B)$$
$$= I(X:B)_{\rho_{XB}}.$$

- $\alpha \leq \sup_n \sup_{P_{X^n}} \frac{1}{n} I(X^n : B^n)_{\rho_{X^nB^n}}$

In the converse part of Stein's lemma, we showed

$$D_H^\varepsilon(\rho \| \sigma) \leq \frac{D(\rho\|\sigma) + 1}{1-\varepsilon}$$

$$\alpha \leq \lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{1}{n} \sup_{P_{X^n}} \frac{D(\rho_{X^nB^n} \| \rho_{X^n} \otimes \rho_{B^n}) + 1}{1-\varepsilon}$$

$$= \lim_{n\to\infty} \frac{1}{n} \sup_{P_{X^n}} I(X^n : B^n)_{\rho_{X^nB^n}}$$

$$= \sup_n \frac{1}{n} \sup_{P_{X^n}} I(X^n : B^n)_{\rho_{X^nB^n}}$$

$f(n) := \sup_{P_{X^n}} I(X^n : B^n)$ is superadditive (ex) $f(n+m) \geq f(n) + f(m)$. + Fekete lemma.

<u>Rk</u>: Actually easy to see $C(W) = \sup_n \frac{1}{n} \sup_{P_{X^n}} I(X^n : B^n)_{\rho_{X^nB^n}}$.

How to compute $\underbrace{\sup_n \frac{1}{n} \sup_{P_{X^n}} I(X^n : B^n)_{\rho_{X^n B^n}}}_{= \sup_{P_X} I(X:B)}$?

For $\underline{\text{cq channels}} \longrightarrow$ $= \sup_{P_X} I(X:B)$, ie $f(n)$ additive

$\underline{\text{Lemma:}}$ For any n

$$\frac{1}{n} \sup_{P_{X^n}} I(X^n : B^n) = \sup_{P_X} I(X:B)$$

$\underline{\text{Proof}}$: • $\geq$ simple (always true, not only cq channels)

• $\leq$ Let P_{X^n} be arbitrary $\qquad \rho_{X^n B^n} = \sum_{x^n} P_{X^n}(x^n) |x^n\rangle\langle x^n| \otimes W_{x^n}$

$$I(X^n : B^n) = H(B^n) - H(B^n | X^n)$$

* $H(B^n) \leq \sum_{i=1}^n H(B_i)$ (subadditivity).

* $H(B^n | X^n) = \sum_{x_1 \ldots x_n} P_{X^n}(x_1 \ldots x_n) H(B^n)_{W_{x_1} \otimes W_{x_2} \otimes \cdots \otimes W_{x_n}}$

Property of von Neumann entropy:

conditional entropy = average of entropy of conditional state.

$$= \sum_{x_1 \ldots x_n} P_{X^n}(x_1 \ldots x_n) \underbrace{\left(H(B)_{W_{x_1}} + H(B)_{W_{x_2}} + \cdots + H(B)_{W_{x_n}} \right)}_{\text{entropy of a product state is sum of entropies.}}$$

$$= \sum_{i=1}^n \sum_{x_i} P_{X_i}(x_i) H(B)_{W_{x_i}}$$

$$= \sum_{i=1}^n H(B_i | X_i)$$

So $I(X^n : B^n) \leq \sum_{i=1}^{n} H(B_i) - H(B_i|X_i)$

$$= \sum_{i=1}^{n} I(X_i : B_i)$$

$$\leq n \cdot \sup_{P_X} I(X:B)_{\rho_{XB}} \quad \blacksquare .$$

Th (Shannon theorem for cq channels)

The capacity of a cq channel is given by

$$C(W) = \sup_{P_X} I(X:B)_{\rho_{XB}}$$

$$\rho_{XB} = \sum_x P_X(x) |x\rangle\langle x| \otimes W_x$$

Rk: • If $W_x = W$ for all $x \Rightarrow C(W) = 0$

Surprisingly, converse also true

$$C(W) = 0 \iff W_x = W \ \forall x$$

This is surprising, just using "repetition" will not work.

→ See TD.

2. General quantum channels.

Now $\mathcal{W}: L(A) \to L(B)$ quantum channel.

Very similar definitions:

Def: An M-code (E,D) for $\mathcal{W}$ is given by
- $E: [M] \to S(A)$ encoding function
- Decoding is a POVM $\{D_s\}_{s\in[M]}$ on B.

Def: The error probability $p_{err}(E,D)$ of an M-code for $\mathcal{W}$ is defined by

$$p_{err}(E,D) = 1 - \underbrace{\frac{1}{M}\sum_{s\in[M]} \mathrm{Tr}\left(D_s \mathcal{W}(E(s))\right)}_{\text{probability of successful decoding.}}$$

If $p_{err}(E,D) \le \varepsilon$, we say that (E,D) is an (M,ε)-code

Looking back at the proofs for cq channels, we see that it suffices to optimize over choice of $\{\sigma_A^x\}_{x\in\mathcal{X}} \in S(A)$ and consider the corresponding cq channel $W_x = \mathcal{W}(\sigma_A^x)$

We then define (as before) for $\{P_X(x), \sigma_A^x\}_{x\in\mathcal{X}}$

This is called an ensemble

$$\rho_{XB} = \sum_{x\in\mathcal{X}} P_X(x)\, |x\rangle\langle x| \otimes \mathcal{W}(\sigma_A^x)$$

Th: Any (M,ε)-code for $\mathcal{W}$ satisfies

$$\log M \leq \sup_{\sigma_A^x} \sup_{P_X} D_H^{\varepsilon}(\rho_{XB} \| \rho_X \otimes \rho_B)$$

and there exists and (M,ε)-code for $\mathcal{W}$

$$\log M \geq \sup_{\sigma_A^x} \sup_{P_X} D_H^{\varepsilon'}(\rho_{XB} \| \rho_X \otimes \rho_B) - \log\left(\frac{2+c+c^{-1}}{\varepsilon-(1+c)\varepsilon'}\right)$$

Basically the same proof. (good exercice to redo it yourself)

Rk: we take supremum over arbitrarily large X but in many cases can bound it.

Important special case: $\mathcal{W}^{\otimes n}$.

Def: The classical capacity $C(\mathcal{W})$ of a quantum channel $\mathcal{W}$ is defined as:

same def as for cq channels

$$C(\mathcal{W}) := \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{\log M^{opt}(\mathcal{W}^{\otimes n}, \varepsilon)}{n}$$

Same as before: using Stein lemma $\frac{1}{n} D_H^{\varepsilon} \to D$.

Notation: $\chi(\mathcal{W}) := \sup_{\{\sigma_A^x, P_X(x)\}} \underbrace{D(\rho_{XB} \| \rho_X \otimes \rho_B)}_{I(X:B)_{\rho_{XB}}}$

where $\rho_{XB} = \sum_x P_X(x) |x\rangle\langle x| \otimes \mathcal{W}(\sigma_A^x)$

Rk: In literature $\chi(\mathcal{W})$ is called the Holevo information of $\mathcal{W}$. See [Wilde, Ch 13] or [Watrous, Ch 8] for properties.

* The Holevo information of an ensemble $\{P_X(x), \sigma_A^x\}$ also commonly denoted

$$\underbrace{I(X:A)_{\rho_{XA}}}_{\overset{\shortparallel}{\chi(\{P_X(x),\sigma_A^x\})}} \text{ for } \rho_{XA} = \sum_x P_X(x)\, |x\rangle\langle x| \otimes \sigma_A^x$$

With this notation, for a cq channel W: $C(W) = \sup_{P_X} \chi(\{P_X(x), W_x\})$

Th (Holevo-Schumacher-Westmorland, HSW)

Let $\mathcal{W}$ be a quantum channel

$$C(\mathcal{W}) = \lim_{n\to\infty} \frac{1}{n}\chi(\mathcal{W}^{\otimes n}) = \sup_n \frac{1}{n}\chi(\mathcal{W}^{\otimes n})$$

Proof is the same as what we did in cq case.

Question: Is χ additive under tensor product ?

i.e. $\chi(\mathcal{W}^{\otimes n}) \overset{?}{=} n\,\chi(\mathcal{W})$

Note that $\chi(\mathcal{W}^{\otimes n}) \geq n\,\chi(\mathcal{W})$ is simple, follows from the fact that $D(\rho\otimes\rho \| \sigma\otimes\sigma) = 2D(\rho\|\sigma)$.

Answer: . NO in general, i.e., there exists channels $\mathcal{W}$ s.t. $\chi(\mathcal{W}^{\otimes 2}) > 2\chi(\mathcal{W})$.

This means that optimal choices of state $\sigma_{A_1A_2}^x$ will be entangled

Construction in [Hastings, 2009] by choosing $\mathcal{W}$ "random" and a very involved analysis. [See book Alice & Bob meet Banach]

- But there are families of channels for which additivity can be proved.

Research question:

Consider the amplitude damping channel $\mathcal{A}_\gamma$ parameter $\gamma \in [0,1)$

$$\begin{pmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{pmatrix} \xrightarrow{\mathcal{A}_\gamma} \begin{pmatrix} \rho_{00} + \gamma \rho_{11} & \sqrt{1-\gamma}\,\rho_{01} \\ \sqrt{1-\gamma}\,\rho_{10} & (1-\gamma)\rho_{11} \end{pmatrix}$$

$C(\mathcal{A}_\gamma)$ unknown.

[Simplistic model for decay of 2-level atom due to spontaneous emission of photon]