# Lecture 7 : Quantum Shannon Theory

## von Neumann entropy

$\hat{\rho} \in \mathcal{D}(\mathcal{H})$

$$S(\hat{\rho}) = -\mathrm{Tr}(\hat{\rho} \log \hat{\rho})$$

Spectral decomp. $\hat{\rho} = \sum_a \lambda_a |\varphi_a\rangle\langle\varphi_a|$

$\lambda_a \in [0,1]$

$\sum_a \lambda_a = 1$

$$S(\hat{\rho}) = -\sum_a \lambda_a \log \lambda_a$$

$$= H(A)$$

where $A = \{ a, \; p(a) = \lambda_a \}$

Revision

# Properties of von Neumann entropy

① $\hat{\rho} = |\psi\rangle\langle\psi| \Rightarrow S(\hat{\rho}) = 0$

② Unitary $\hat{U}$

$$S(\hat{U}\hat{\rho}\hat{U}^{+}) = S(\hat{\rho})$$

③ $\dim \mathcal{H} = d$

$S(\hat{\rho}) \leq \log d$, with

equality for max. mixed state

④ $S(\hat{\rho}_A) = S(\hat{\rho}_B)$

for $\hat{\rho}_A = \mathrm{Tr}_A(|\psi\rangle\langle\psi|_{AB})$

$\hat{\rho}_B = \mathrm{Tr}_B(|\psi\rangle\langle\psi|_{AB})$

New

⑤ Subadditivity

$$S(\hat{\rho}_{AB}) \leq S(\hat{\rho}_A) + S(\hat{\rho}_B)$$

Equality iff $\hat{\rho}_{AB} = \hat{\rho}_A \otimes \hat{\rho}_B$

⑥ Q. mutual information

$$I(\hat{\rho}_A ; \hat{\rho}_B) = S(\hat{\rho}_A) + S(\hat{\rho}_B) - S(\hat{\rho}_{AB})$$

$I(\hat{\rho}_A ; \hat{\rho}_B) = 0$ for

$\hat{\rho}_{AB} = \hat{\rho}_A \otimes \hat{\rho}_B$

⑦ Triangle inequality

$$S(\hat{\rho}_{AB}) \geq |S(\hat{\rho}_A) - S(\hat{\rho}_B)|$$

Proof of 7

Consider $|\psi\rangle_{ABC}$ s.t.

$$\hat{\rho}_{AB} = \mathrm{Tr}_C\left( |\psi\rangle\langle\psi|_{ABC} \right)$$

$$\begin{aligned} S(\hat{\rho}_A) &= S(\hat{\rho}_{BC}) \\ &\leq S(\hat{\rho}_B) + S(\hat{\rho}_C) \quad 5 \\ &= S(\hat{\rho}_B) + S(\hat{\rho}_{AB}) \end{aligned}$$

$$S(\hat{\rho}_A) - S(\hat{\rho}_B) \leq S(\hat{\rho}_{AB})$$

Exchange A & B for the other way

cf. Shannon entropy

$$H(XY) \geq H(X), H(Y)$$

For quantum states information can be encoded in the correlations between subsystems.

e.g. $|\psi\rangle_{AB} = \frac{1}{\sqrt{2}}\left(|00\rangle_{AB} + |11\rangle_{AB}\right)$

$\hat{\rho}_A = \hat{\rho}_B = \frac{1}{2}\left(|0\rangle\langle 0| + |1\rangle\langle 1|\right)$

$S(|\psi\rangle\langle\psi|_{AB}) = 0$

$S(\hat{\rho}_A) = S(\hat{\rho}_B) = \frac{1}{2}$

can also define the conditional entropy

$S(\hat{\rho}_A | \hat{\rho}_B) = S(\hat{\rho}_{AB}) - S(\hat{\rho}_B)$

$H(X|Y) = H(XY) - H(Y) \geq 0$

For the state above $S(\hat{\rho}_A | \hat{\rho}_B) = -\frac{1}{2}$ !

# Schumacher Compression

Consider a density operator

$$\hat{\rho} = \sum_x p(x) |\psi(x)\rangle\langle\psi(x)|$$

This is the analogue of a classical random variable $X = \{x, p(x)\}$.

Suppose we sample from the ensemble $\hat{\rho}$ $n$ times.

The resulting density matrix is $\hat{\rho}^{\otimes n} = \hat{\rho} \otimes \cdots \otimes \hat{\rho}$

Can we compress this to a state on fewer than $n$ subsystems?

Consider the spectral decomposition

of $\hat{\rho} = \sum_a \lambda_a |\varphi_a\rangle\langle\varphi_a|$

orthonormal

In this basis

$$\hat{\rho}^{\otimes n} = \lambda_a^n \left(|\varphi_a\rangle\langle\varphi_a|\right)^{\otimes n} + \dots$$

Define the <u>$\delta$-typical subspace</u>

<u>$\Lambda$</u> as the subspace of $\mathcal{H}$ spanned

by eigenvectors of $\hat{\rho}^{\otimes n}$ w/

eigenvalues $\mu$ satisfying

$$2^{-n(S(\hat{\rho})-\delta)} \geq \mu \geq 2^{-n(S(\hat{\rho})+\delta)}$$

As the $|\varphi_a\rangle$ are orthogonal

we can model the state $\hat{\rho}^{\otimes n}$

as the classical distribution

$X^n$ where $X = \{ a , \lambda_a \}$

We know that for any $\varepsilon, \delta > 0$

$\exists N$ s.t. for all $n \geq N$

$$2^{-n(H(X)-\delta)} \leq \underbrace{\lambda_{a_1} \lambda_{a_2} \cdots \lambda_{a_n}}_{\mu} \leq 2^{-n(H(X)+\delta)}$$

w/ probability $1-\varepsilon$.

We also know that $H(X) = S(\hat{\rho})$

ie w/ probability at least $1-\varepsilon$, $\hat{\rho}^{\otimes n}$ lies in the $\delta$-typical subspace.

Or equivalently

$$\mathrm{Tr}\left(\hat{\Pi}_{typ} \hat{\rho}^{\otimes n}\right) \geq 1 - \varepsilon$$

$$\hat{\Pi}_{typ} = \sum_{i \in \Lambda} |i\rangle\langle i|$$

Projector onto $\delta$-typical subspace

We can bound the dimension of $\Lambda$, $\dim(\Lambda) = N_{typ}$

$$N_{typ} 2^{-n(S(\hat{\rho})+\delta)} \leq 1$$

$$N_{typ} \leq 2^{n(S(\hat{\rho})+\delta)}$$

$$N_{typ} 2^{-n(S(\hat{\rho})-\delta)} \geq 1-\varepsilon$$

$$N_{typ} \geq (1-\varepsilon) 2^{n(S(\hat{\rho})-\delta)}$$

$$(1-\varepsilon) 2^{n(S(\hat{\rho})-\delta)} \leq \dim(\Lambda) \leq 2^{n(S(\hat{\rho})+\delta)}$$

Our compression strategy is to only encode the $\delta$-typical subspace.

First we make a projective measurement that projects the state onto $\Lambda$ or $\Lambda^\perp$ $\leftarrow$ complement

$$\hat{\Pi}_{typ} = \sum_{i \in \Lambda} |i\rangle\langle i|$$

$$\hat{\Pi}^{\perp}_{typ} = \hat{I} - \hat{\Pi}_{typ}$$

w/ probability $1-\varepsilon$ we obtain

$$\hat{\rho}' = \frac{\hat{\Pi}_{typ}\, \hat{\rho}^{\otimes n}\, \hat{\Pi}_{typ}}{1-\varepsilon}$$

which is entirely supported on $\Lambda$.

w/ probability $\varepsilon$ the compression fails.

We can then compress

$$\hat{U} \hat{\rho}' \hat{U}^{+}$$

$$= \hat{\rho}'_A \otimes |\underline{0}\rangle\langle\underline{0}|_B$$

where $\dim(A) = \dim(\Lambda)$

and $\dim(B) = \dim(\Lambda^{\perp})$

$$= \dim(\mathcal{H}^{\otimes n}) - \dim(\Lambda)$$

Now suppose we want to undo the compression

We can describe the state

after decompression as

$$(1-\varepsilon)\,\hat{\rho}' + \varepsilon\,\hat{\rho}_\varepsilon = \hat{\rho}_{out}$$

$$= \hat{\Pi}_{typ}\,\hat{\rho}^{\otimes n}\,\hat{\Pi}_{typ} + \varepsilon\,\hat{\rho}_\varepsilon$$

Let us compute the fidelity of

this w/ the initial state

$$|\psi(\underline{x})\rangle = |\psi(x_1)\rangle \otimes \dots \otimes |\psi(x_n)\rangle$$

$$\hat{\rho}^{\otimes n} = \sum_{\underline{x}} p(\underline{x})\,|\psi(\underline{x})\rangle\langle\psi(\underline{x})|$$

For a given message $\underline{x}$

the fidelity is

$$F(\underline{x}) = \left| \langle \psi(\underline{x}) | \hat{\Pi}_{typ} | \psi(\underline{x}) \rangle \right|^2 + \varepsilon \underbrace{\langle \psi(\underline{x}) | \hat{\rho}_{junk} | \psi(\underline{x}) \rangle}_{\geq 0}$$

$$\geq \left| \langle \psi(\underline{x}) | \hat{\Pi}_{typ} | \psi(\underline{x}) \rangle \right|^2$$

The average fidelity is

$$F = \sum_{\underline{x}} p(\underline{x}) F(\underline{x})$$

$$\geq \sum_{\underline{x}} p(\underline{x}) \left| \langle \psi(\underline{x}) | \hat{\Pi}_{typ} | \psi(\underline{x}) \rangle \right|^2$$

Note that for $z \in \mathbb{R}$

$$(z-1)^2 \geq 0 \qquad z^2 \geq 2z - 1$$

$$F \geq \sum_{\underline{x}} p(\underline{x}) \left[ 2 \langle \psi(\underline{x}) | \hat{\Pi}_{typ} | \psi(\underline{x}) \rangle - 1 \right]$$

$$= 2 \sum_{\underline{x}} p(\underline{x}) \langle \psi(\underline{x}) | \hat{\Pi}_{typ} | \psi(\underline{x}) \rangle - \overbrace{\sum_{\underline{x}} p(\underline{x})}^{1}$$

$$= 2 \, \mathrm{Tr}\left( \hat{\rho}^{\otimes n} \hat{\Pi}_{typ} \right) - 1$$

$$\geq 2(1-\varepsilon) - 1 = 1 - 2\varepsilon$$

$\Rightarrow$ We can compress the state to a state on $n S(\hat{\rho})$ subsystems with negligible loss in fidelity ($\varepsilon$ can be arbitrarily close to 0).

Shannon & Schumacher compression seem very similar but there is a difference.

Consider again

$$\hat{\rho} = \sum_x p(x) \, |\psi(x)\rangle\langle\psi(x)|$$

$$X = \{ x, p(x) \}$$

The Holevo bound gives

$$H(X) - \sum_x p(x) \overbrace{S(|\psi(x)\rangle\langle\psi(x)|)}^{0}$$

$$\leq S(\hat{\rho})$$

$$H(X) \leq S(\hat{\rho})$$

equality when $\langle\psi(x)|\psi(y)\rangle = \delta_{xy}$

Now consider $n$ samples drawn from the ensemble $\hat{\rho}$.

We can use Shannon to compress the classical description $X^n$ using $H(X) + o(1)$ bits per letter.

But we can use Schumacher compression to compress the state using $S(\hat{\rho}) + o(1)$ qubits per letter.

As $S(\hat{\rho}) \geq H(X)$ quantum compression is more efficient.

# Capacities of quantum channels

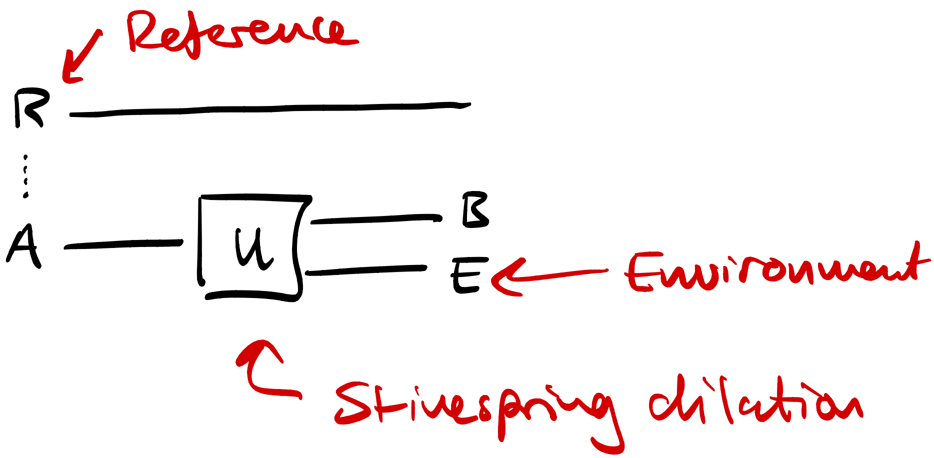
This is a much harder problem than in the classical case.

We still don't know the capacity of the depolarizing channel,
$\mathcal{D}(\hat{\rho}) = (1-p)\hat{\rho} + p\frac{\hat{I}}{d}$, for example.

We don't have time to go into this topic in depth, but here is one counter-intuitive result.

$$A - \boxed{\mathcal{N}} - B$$

can be written as




$|\psi\rangle_{RA}$ any state such that

$$\mathrm{Tr}_R\left( |\psi\rangle\langle\psi|_{RA} \right) = \hat{\rho}_A .$$

$$I_c(R\rangle B) = \frac{1}{2}\left( I(R;B) - I(R;E) \right)$$

$$\| $$

$$-S(R|B) = S(B) - S(RB)$$

$$= S(B) - S(E)$$

RBE is pure

The quantum analogue of Shannon's noisy coding theorem is

$$Q\left(\mathcal{N}^{A\to B}\right) = \lim_{n\to\infty} \max_{A^n} \frac{1}{n} I_c\left(R^n \rangle B^n\right)$$

input distribution

where $I_c$ is the coherent information and is defined

$$I_c(R\rangle B) = \frac{1}{2}\left(I(R;B) - I(R;E)\right)$$

Note that the input state

$\hat{\rho}_{A^n}$ can be entangled.

In general the coherent information

can be superadditive and

so $\max_{A} I_c(R \rangle B)$ is only

a lower bound on the capacity.

# Superactivation of quantum channels

$\exists$ q. channels $\mathcal{N}_1$, $\mathcal{N}_2$ s.t.

$$Q(\mathcal{N}_1) = Q(\mathcal{N}_2) = 0$$

quantum capacity

$$Q(\mathcal{N}_1 \otimes \mathcal{N}_2) > 0$$

So for a given channel we cannot bound the capacity using a single use of the channel as we could in the classical case. We instead may need to consider arbitrarily large tensor powers of the channel!