

Quantum information theory

Lectures 25-27

Quantum entropy

Classical information and Shannon entropy

Random variable: $X, x_j, p_j = p(x_j) = P_X(x_j)$

Shannon entropy: $H(\vec{p}) = H(X) = - \sum_j p(x_j) \log p(x_j)$

Ignorance and information $0 \leq H \leq \log N$

Typical sequences and block coding

Yes-no questions and variable-length coding

Relative entropy:

$H(\vec{p} \parallel \vec{q}) \rightarrow \infty$ if $p_j \neq 0 = q_j$ for some j

$$H(\vec{p} \parallel \vec{q}) = \sum_j p_j \log \frac{p_j}{q_j} = -H(\vec{p}) - \sum_j p_j \log q_j \geq 0$$

\uparrow
iff $\vec{p} = \vec{q}$

Two random variables: X, Y

Conditional entropy: $H(X|Y) = \sum_y p(y) \left(- \sum_x p(x|y) \log p(x|y) \right)$

$$= \sum_{x,y} p(x,y) \log p(x|y)$$

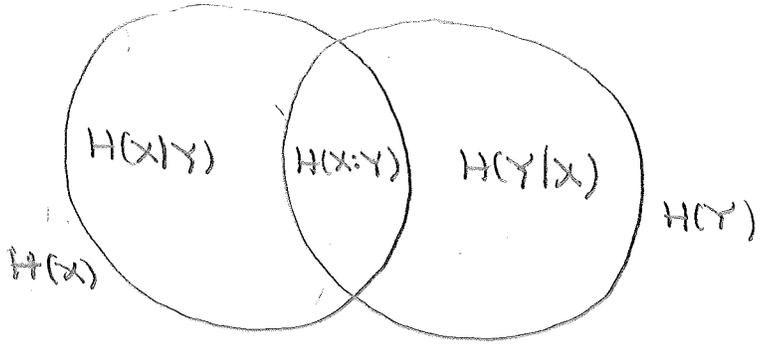
Mutual information:

$$H(X:Y) = H(X) - H(X|Y) = \sum_{x,y} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right)$$

$$= H(\vec{p}_{x,y} \parallel \vec{p}_x \vec{p}_y) \geq 0$$

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

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



$H(X:Y) \leq H(X), H(Y)$

Quantum information and von Neumann entropy

Quantum state: $\rho = \sum_j \lambda_j |e_j\rangle\langle e_j|$ \rightarrow eigendecomposition

von Neumann entropy:

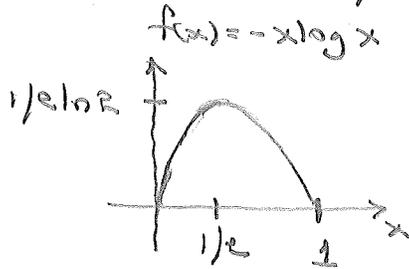
$$S(\rho) = H(\vec{\lambda}) = - \sum_j \lambda_j \log \lambda_j = - \text{tr}(\rho \ln \rho)$$

- Why?
- ① Schumacher compression and block coding
 - ② Accessible info
 - ⋮

$$0 \leq S(\rho) \leq \log D$$

iff $\rho = |\psi\rangle\langle\psi|$ iff $\rho = I/D$

① ODOF inequality:



$$f'(x) = -\log(x)$$

$$f''(x) = -\frac{1}{x \ln 2} < 0$$

$$g_j = \langle f_j | \rho | f_j \rangle = \sum_k \underbrace{|\langle f_j | e_k \rangle|^2}_{P_{jk}} \lambda_k$$

$$H(\vec{g}) \geq H(\vec{\lambda}) = S(\rho)$$

iff $\vec{g} = \vec{\lambda}$

doubly stochastic matrix

$$1 = \sum_j P_{jk} = \sum_k P_{jk}$$

Any concave function $f(x)$: $f(\vec{g}) = \sum_j f(g_j) \geq f(\vec{\lambda}) = \text{tr}(f(\rho))$

$$f(\vec{g}) = \sum_j f\left(\sum_k |\langle f_j | e_k \rangle|^2 \lambda_k\right)$$

$$\geq \sum_k |\langle f_j | e_k \rangle|^2 f(\lambda_k)$$

concavity of f and double stochasticity of P_{jk}

$$\geq \sum_k f(\lambda_k) \underbrace{\sum_j |\langle f_j | e_k \rangle|^2}_1$$

$$= f(\vec{\lambda}) = \text{tr}(f(\rho))$$

↓

If $g_j = \sum_k D_{jk} p_k$

1
DS

then $f(\vec{g}) \geq f(\vec{p})$

Direct proof for H:

$$H(\vec{q}) - H(\vec{\lambda}) = - \sum_k q_k \log q_k + \sum_j \lambda_j \log \lambda_j$$

Use $\log x = \frac{\ln x}{\ln 2} \leq \frac{x-1}{\ln 2}$

$$\begin{aligned} & \sum_{j,k} \frac{1}{\ln 2} \sum_{j,k} |K_{f_n}(e_j)|^2 \lambda_j \left(1 - \frac{q_k}{\lambda_j}\right) \\ &= \frac{1}{\ln 2} \sum_{j,k} |K_{f_n}(e_j)|^2 \lambda_j - \sum_{j,k} |K_{f_n}(e_j)|^2 q_k \\ &= \frac{1}{\ln 2} \left(\sum_k q_k - \sum_k q_k \right) \\ &= 0 \end{aligned}$$

use ΔS

$$\begin{aligned} &= - \sum_{j,k} |K_{f_n}(e_j)|^2 \lambda_j \log \frac{q_k}{\lambda_j} \\ &= \sum_{j,k} |K_{f_n}(e_j)|^2 \lambda_j \\ & \quad \times \log \left(\frac{|K_{f_n}(e_j)|^2 \lambda_j}{|K_{f_n}(e_j)|^2 q_k} \right) \\ & \geq 0 \end{aligned}$$

This is a relative entropy

Concavity:

$$S(\mu p_1 + (1-\mu) p_2) \geq \mu S(p_1) + (1-\mu) S(p_2)$$

$0 < \mu < 1$
iff $p_1 \neq p_2$

$$\begin{aligned} &\Rightarrow S\left(\sum_j p_j p_j\right) \geq \sum_j p_j S(p_j) \\ &\text{or } \chi = S(p) - \sum_j p_j S(p_j) \geq 0 \end{aligned}$$

↑ Holevo quantity

Any concave function $f(x)$: $\text{tr}(f(p)) \geq \mu \text{tr}(f(p_1)) + (1-\mu) \text{tr}(f(p_2))$

$$\begin{aligned} \text{tr}(f(p)) &= \sum_j f(\lambda_j) = \sum_j f(\langle e_j | p | e_j \rangle) \\ &= \sum_j f(\mu \langle e_j | p_1 | e_j \rangle + (1-\mu) \langle e_j | p_2 | e_j \rangle) \end{aligned}$$

concavity of $f \longrightarrow \geq \mu \sum_j f(\langle e_j | p_1 | e_j \rangle) + (1-\mu) \sum_j f(\langle e_j | p_2 | e_j \rangle)$

ODOP inequality $\rightarrow \geq u \operatorname{tr}(f(\rho_1)) + (1-u) \operatorname{tr}(f(\rho_2))$ ④

③ Ensemble info inequality: $\rho = \sum_i p_i \rho_i$

$$H(\vec{p}) \stackrel{?}{\geq} S(\rho) - \sum_i p_i S(\rho_i) \stackrel{?}{=} \chi \stackrel{?}{\geq} 0$$

\downarrow iff ρ_i commute \downarrow iff ρ_i same

② $\rho = \sum_{\alpha} p_{\alpha} |\psi_{\alpha}\rangle \langle \psi_{\alpha}| = \sum_j \lambda_j |e_j\rangle \langle e_j|$
 Ensemble decomposition

HJW theorem: $\sqrt{p_{\alpha}} |\psi_{\alpha}\rangle = \sum_j V_{\alpha j} \sqrt{\lambda_j} |e_j\rangle = \sum_{\beta} V_{\alpha \beta} \sqrt{\lambda_{\beta}} |e_{\beta}\rangle$
 \uparrow unitary matrix

$\Rightarrow p_{\alpha} = \sum_{\beta} |V_{\alpha \beta}|^2 \lambda_{\beta}$
 \uparrow doubly stochastic matrix

$\therefore H(\vec{p}) \stackrel{?}{\geq} H(\vec{\lambda}) = S(\rho)$
 \downarrow iff eigendecomposition

⑥ $\rho = \sum_i p_i \rho_i = \sum_{j,k} p_j \lambda_{kj} |e_{kj}\rangle \langle e_{kj}|$

② $\Rightarrow H(\text{joint distribution } \lambda_{kj} p_j) \geq S(\rho)$

"
 $H(\vec{p}) + \underbrace{\sum_j p_j \left(- \sum_k \lambda_{kj} \log \lambda_{kj} \right)}_{S(\rho)} = H(\vec{p}) + \sum_j p_j S(\rho_j)$

④ POVM inequality: POVM $\{E_a\}$
 $g_a = \text{tr}(p E_a)$

$$H(\vec{g}) + \sum_a g_a \log(\text{tr}(E_a)) = - \sum_a g_a \log\left(\frac{g_a}{\text{tr}(E_a)}\right) \geq S(p)$$

$$g_a = \sum_j \lambda_j \underbrace{\langle e_j | E_a | e_j \rangle}_{= P_{aj}}$$

$$\sum_j P_{aj} = \text{tr}(E_a)$$

Let $f(x) = -x \log x$

$$H(\vec{g}) = \sum_a f(g_a) = \sum_a f\left(\sum_j P_{aj} \lambda_j\right)$$

$$= \sum_a f\left(\sum_j \frac{P_{aj}}{\text{tr}(E_a)} \lambda_j \text{tr}(E_a)\right)$$

Concavity of f
 in average over j
 (distribution $P_{aj}/\text{tr}(E_a)$) \rightarrow

$$\geq \sum_{a,j} \frac{P_{aj}}{\text{tr}(E_a)} f(\lambda_j \text{tr}(E_a))$$

$$= - \sum_{a,j} \frac{P_{aj}}{\text{tr}(E_a)} \lambda_j \text{tr}(E_a) \log(\lambda_j \text{tr}(E_a))$$

$$= - \underbrace{\sum_{a,j} P_{aj} \lambda_j \log \lambda_j}_{= - \sum_j \lambda_j \log \lambda_j} - \underbrace{\sum_{a,j} P_{aj} \lambda_j \log(\text{tr}(E_a))}_{= \sum_a g_a \log(\text{tr}(E_a))}$$

$$= - \sum_j \lambda_j \log \lambda_j = H(\vec{\lambda})$$

$$= \sum_a g_a \log(\text{tr}(E_a))$$

$$= S(p) - \sum_a g_a \log(\text{tr}(E_a))$$

⑤ Quantum relative entropy:

$$S(\rho \parallel \sigma) = \text{tr}(\rho \log \rho) - \text{tr}(\rho \log \sigma) = -S(\rho) - \text{tr}(\rho \log \sigma)$$

Like a distance, but not symmetric

⑥ $\rho = \sum_j p_j |e_j\rangle\langle e_j|, \quad \sigma = \sum_j q_j |f_j\rangle\langle f_j|$

$$S(\rho \parallel \sigma) = -S(\rho) - \text{tr}(\rho \log \sigma)$$

$$= -H(\vec{p}) - \sum_j \underbrace{\langle f_j | \rho | f_j \rangle}_{\equiv r_j} \log q_j$$

$$= -H(\vec{p}) - \sum_j r_j \log r_j + \sum_j r_j \log(q_j / r_j)$$

$$S(\rho \parallel \sigma) = \underbrace{-S(\rho)}_{-H(\vec{p})} + \underbrace{-\text{tr}(\rho \log \sigma)}_{H(\vec{r}) + H(\vec{r} \parallel \vec{q})} \geq 0$$

$\underbrace{\hspace{10em}}_{\geq 0 \text{ (odop)}} \quad \underbrace{\hspace{10em}}_{\geq 0} \quad \underbrace{\hspace{10em}}_{\text{iff } \rho = \sigma}$

$S(\rho \parallel \sigma) \rightarrow \infty$ if $r_j \neq 0 = q_j$ for some j

⑦ Unitary invariance

$$\chi = S(\rho) - \sum_j p_j S(p_j) = \sum_j p_j \underbrace{(\text{tr}(p_j \log p_j) - \text{tr}(p_j \log \rho))}_{S(p_j \parallel \rho)}$$

$$\chi = \sum_j p_j S(p_j \parallel \rho)$$

Two systems: ρ_{AB} ρ_A ρ_B
 $S(\rho_{AB}) = S(A, B)$ $S(\rho_A) = S(A)$ $S(\rho_B) = S(B)$

① If ρ_{AB} is pure, $S(\rho_{AB}) = 0$ and $S(\rho_A) = S(\rho_B)$
Use Schmidt decomposition

② Subadditivity: $S(A, B) \leq S(A) + S(B)$
iff $\rho_{AB} = \rho_A \otimes \rho_B$

③ $\rho_A = \sum_j p_j |e_j\rangle\langle e_j|$
 $\rho_B = \sum_k q_k |f_k\rangle\langle f_k|$ } Eigen decompositions

$\langle e_j, f_n | \rho_{AB} | e_j, f_n \rangle = r_{jk}$ $\sum_k r_{jk} = p_j$
 $\sum_j r_{jk} = q_k$

$S(A, B) \leq H(\vec{r}) \leq H(\vec{p}) + H(\vec{q}) = S(\rho_A) + S(\rho_B)$
↑
ODOP
ineq. ↑
classical
Subadditivity

OR

④ $0 \leq S(\rho_{AB} \| \rho_A \otimes \rho_B) = -S(\rho_{AB}) - \underbrace{\text{tr}(\rho_{AB} \log(\rho_A \otimes \rho_B))}_{S(A) + S(B)}$

⑧ Araki-Lieb (triangle inequality)

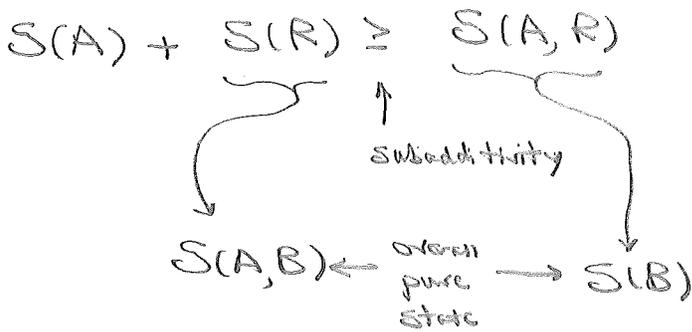
$S(A, B) \geq |S(A) - S(B)|$

P_{AB} is a mixture of orthogonal purifications of P_A or P_B

iff $P_{AB} = \sum_j P_j |\psi_j\rangle\langle\psi_j|$, with

$\text{tr}_A(|\psi_j\rangle\langle\psi_j|) = P_B \delta_{jk} \quad [S(A, B) = S(A) - S(B)]$
 or $\text{tr}_B(|\psi_j\rangle\langle\psi_j|) = P_A \delta_{jk} \quad [S(A, B) = S(B) - S(A)]$

⑨ Let $|\mathbb{I}_{RAB}\rangle$ be a purification of P_{AB} .



Here purification works because it changes joint entropies to single-system entropies & vice versa.

OR

⑩ $P_{AB} = \sum_j \lambda_j |e_j\rangle\langle e_j| \leftarrow$ eigendecomposition

$P_A = \sum_j \lambda_j P_j^A, \quad P_B = \sum_j \lambda_j P_j^B$

$S(P_j^A) = S(P_j^B)$

$P_j^A = \text{tr}_B(|e_j\rangle\langle e_j|)$ are \perp

$S(P_{AB}) = H(\vec{\lambda}) \geq S(P_A) - \sum_j \lambda_j S(P_j^A) \geq 0$

$S(P_{AB}) = H(\vec{\lambda}) \geq S(P_B) - \sum_j \lambda_j S(P_j^B) \geq 0$

$P_j^B = \text{tr}_A(|e_j\rangle\langle e_j|)$ are same

$S(P_{AB}) \geq S(P_A) - S(P_B)$

Quantum conditional entropy

$$S(A|B) = S(A, B) - S(B) \begin{cases} \geq -S(A), & S(B) \geq S(A) \\ \geq S(A) - 2S(B), & S(A) \geq S(B) \end{cases}$$

$$S(B|A) = S(A, B) - S(A)$$

Could we use negativity of these as entanglement criteria?

Quantum mutual information

$$S(A:B) = S(A) + S(B) - S(A, B) \geq 0$$

Subadditivity

$$S(A:B) \leq 2S(A), 2S(B)$$

$$\begin{matrix} \uparrow & \uparrow \\ S(A) \geq S(B) & S(B) \geq S(A) \end{matrix}$$

Too correlated to be classical

Quantum conditional entropy and relative entropy

$$S(\rho_{AB} \| \frac{I}{d} \otimes \rho_B) = -S(\rho_{AB}) - \text{tr}(\rho_{AB} \log(\frac{I}{d} \otimes \rho_B))$$

$$\log\left(\sum_{j,k} \frac{1}{d} |e_j\rangle\langle e_j| \otimes |f_k\rangle\langle f_k|\right)$$

$$= \sum_{j,k} \log\left(\frac{\lambda_k}{d}\right) |e_j\rangle\langle e_j| \otimes |f_k\rangle\langle f_k|$$

$$= I \otimes \left(\sum_k \log \lambda_k |f_k\rangle\langle f_k| - I \log d\right)$$

$$= I \otimes \log \rho_B - I \otimes I \log d$$

$$= -S(\rho_{AB}) - \text{tr}(\rho_B \log \rho_B) + \log d$$

$$= -S(A, B) + S(B) + \log d$$

$$S(A|B) = -S(\rho_{AB} \| (I/d) \otimes \rho_B) + \log d$$

④ Another proof of concavity of S

$$\rho = \sum_i p_i \rho_i$$

Define $\rho_{AB} = \sum_i p_i \rho_i \otimes |f_i\rangle\langle f_i|$ ←

extension
(not a purification)
B stores $|f_i\rangle$ record
which state ρ_i
applies to A

$$\rho_A = \rho \quad \rho_B = \sum_i p_i |f_i\rangle\langle f_i|$$

$$S(A, B) = H(\vec{p}) + \sum_i p_i S(\rho_i) \leftarrow \text{why?}$$

$$S(A) = S(\rho)$$

$$S(B) = H(\vec{p})$$

Subadditivity $\Rightarrow \sum_i p_i S(\rho_i) \leq S(\rho)$

↓
Extensions allow us
to get the ensemble
info + the average
ensemble entropy into
our equations.

⑤ Another proof of the ensemble information inequality

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$$

$|\Phi_{AB}\rangle = \sum_i \sqrt{p_i} |\psi_i\rangle \otimes |f_i\rangle$ is a purification of ρ

$$\rho_A = \rho, \quad \rho_B = \sum_{j,k} \sqrt{p_j p_k} \langle\psi_j|\psi_k\rangle |f_j\rangle\langle f_k|$$

$$\Rightarrow \langle f_j | \rho_B | f_j \rangle = \sum_i p_i |f_i\rangle\langle f_i|$$

$$H(\vec{p}) \geq S(\rho_B) = S(\rho_A)$$

↑
ODOP ineq.

Three systems:

Strong subadditivity (SS):

$$S(A, B, C) + S(B) \leq S(A, B) + S(B, C)$$

OR

$$S(A|B, C) \leq S(A|B) \leftarrow \begin{matrix} \text{Trivial} \\ \text{Classically} \end{matrix}$$

Proof:
N&C App 6
and 11.4.1

Mother of all properties: use purifications and extensions

$$P_{AB} \rightarrow |\Phi_{ABC}\rangle: 0 + S(B) \leq S(A, B) + S(A) \quad \text{Araki-Lieb}$$

$$P_{AC} \rightarrow |\Phi_{ABC}\rangle: 0 + S(A, C) \leq S(C) + S(A) \quad \text{Sub-additivity}$$

SS \Rightarrow concavity of conditional quantum entropy

$$P_{AB} = \sum_j P_j P_j^{AB}$$

Extend to $P_{ABC} = \sum_j P_j P_j^{AB} \otimes |f_j\rangle\langle f_j|$

Apply to $S(A, B, C) + S(A) \leq S(A, B) + S(A, C)$

$$\downarrow$$

$$= H(P) + \sum_j P_j S(P_j^{AB})$$

$$\downarrow$$

$$P_{AC} = \sum_j P_j P_j^A \otimes |f_j\rangle\langle f_j|$$

$$= H(P) + \sum_j P_j S(P_j^A)$$

$$\Rightarrow \underbrace{S(A, B) - S(A)}_{S(B|A)} \geq \sum_j P_j \underbrace{(S(P_j^{AB}) - S(P_j^A))}_{S_j(B|A)}$$

$$S(B|A) \geq \sum_j p_j S_j(B|A)$$

Concavity of
quantum conditional
entropy

→ Rewrite as

$$\underbrace{S(\rho_A) - \sum_j p_j S(\rho_j^A)}_{\chi_A} \leq \underbrace{S(\rho_{AB}) - \sum_j p_j S(\rho_j^{AB})}_{\chi_{AB}}$$

Marginalization property of Holevo quantity

Note: $H(\vec{p}) \geq \chi_A$, $H(\vec{p}) \geq \chi_{AB}$ are ensemble info inequalities.

The marginalization property establishes that $H(\vec{p}) \geq \chi_{AB} \geq \chi_A$

Entanglement inequality: Specialize marginalization of χ to case where ρ_j^{AB} is a pure-state ensemble.

$$\sum_j p_j S(\rho_j^A) \geq S(A) - S(A, B) = -S(B|A)$$

↓
If $S(A) > S(A, B)$, then ρ_{AB} cannot be written as a mixture of product states.

An equivalent form of strong subadditivity:

$$\left. \begin{matrix} P_{ABC} \\ P_{RBC} \end{matrix} \right\} \rightarrow |\Phi_{RABC}\rangle$$

$$0 \geq \underbrace{S(A,B,C)}_{S(R)} + S(B) - \underbrace{S(A,B)}_{S(R,C)} - S(B,C)$$

$\rightarrow S(A) + S(B) \leq S(A,C) + S(B,C)$ <p style="text-align: center;">OR</p> $S(C A) + S(C B) \geq 0 \iff S(A:C) + S(B:C) \leq 2S(C)$	<p style="text-align: center;">Monogamy of entanglement</p>
---	---

$S(A|C) + S(B|C)$ can be negative: $|\psi_{ABC}\rangle = |\psi_A\rangle \otimes \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$

$S(A,C) = 1$
 $S(A) = 0, S(C) = 1$
 $S(A|C) = 0, S(C|A) = 1$
 $S(B,C) = 0$
 $S(B) = S(C) = 1$
 $S(B|C) = S(C|B) = -1$

Properties equivalent to SS (see pages A-C)

① Monotonicity of relative entropy: $S(p||\sigma_A) \leq S(p_{AB}||\sigma_{AB})$

② Double convexity of relative entropy:

$$S(p||\sigma) \leq \sum_j p_j S(p_j||\sigma_j),$$

$$p = \sum_j p_j p_j$$

$$\sigma = \sum_j p_j \sigma_j$$

① & ② are good distinguishability properties for relative entropy.

③ Concavity of conditional entropy.

Related properties:

Contractivity of relative entropy:

$$S(\rho' \| \sigma') = S(Q(\rho) \| Q(\sigma))$$

$$S(\rho' \| \sigma') \leq S(\rho \| \sigma)$$

$$\rho' = Q(\rho), \sigma' = Q(\sigma), Q = \text{tr}_B \otimes \mathbb{I}_A$$

$$= S(\text{tr}_A(u \rho \otimes |0\rangle\langle 0| u^\dagger) \| \text{tr}_A(u \sigma \otimes |0\rangle\langle 0| u^\dagger))$$

monotonicity
of relative
entropy

$$\leq S(u \rho \otimes |0\rangle\langle 0| u^\dagger \| u \sigma \otimes |0\rangle\langle 0| u^\dagger)$$

$$= S(\rho \otimes |0\rangle\langle 0| \| \sigma \otimes |0\rangle\langle 0|)$$

unitary
invariance
of relative
entropy

$$= S(\rho \| \sigma)$$

⑥ Since $\chi = \sum_i p_i S(\rho_i \| \rho)$, we get

$$\chi_A \leq \chi_{AB} \text{ from monotonicity of relative entropy}$$

$$\chi' \leq \chi \text{ from contractivity of relative entropy}$$

⑦ Since $S(A|B) = -S(\rho_{AB} \| (\mathbb{I}_A \otimes \rho_B))$, we get

$$S(A'|B') \geq S(A|B) \text{ from contractivity of relative entropy.}$$

Holevo bound

classical into down a noiseless quantum channel

only "noise" is due to quantum measurement

Input: Alice sends ρ_x with probability p_x

Output: Bob measures POVM E_y

$$p_{y|x} = \text{tr}(E_y \rho_x)$$

Holevo bound: $H(X:Y) \leq S(p) - \sum_x p_x S(\rho_x) = \chi$

iff ρ_x commute
maximum value of this mutual information is called the accessible information

Proof: Let Q denote the quantum system. Introduce a reference system R and an ancilla A that record the input and output

Initial state: $\rho_{RQA} = \left(\sum_x p_x |x\rangle\langle x| \otimes \rho_x \right) \otimes |0\rangle\langle 0|$
 $\rho_{RQ} \quad \rho_A = \sum_x p_x |x\rangle\langle x|$

Final state: Quantum operation \mathcal{Q} on Q and A such that

$$\rho(\sigma \otimes |0\rangle\langle 0|) = \sum_y \underbrace{\sqrt{E_y} \otimes U_y (\sigma \otimes |0\rangle\langle 0|) \sqrt{E_y} \otimes U_y^\dagger}_{\text{Krans operators}}$$

$$= \sum_y \underbrace{\sqrt{E_y} \sigma \sqrt{E_y} \otimes |y\rangle\langle y|}_{P_{Y|X} \sigma_Y}$$

makes an extension of ensemble $P_{Y|X} \sigma_Y$

$$P'_{RQA} = \rho \otimes \rho(P_{RQA})$$

$$= \sum_{x,y} P_x |x\rangle\langle x| \otimes \sqrt{E_y} P_x \sqrt{E_y} \otimes |y\rangle\langle y|$$

$$P'_{AA} = \sum_{x,y} P_{y|x} P_x |x\rangle\langle x| \otimes |y\rangle\langle y|$$

classical correlation of X and Y

$$P'_R = \sum_x P_x |x\rangle\langle x| = P_R$$

Our guess: $H(X:Y) = S(R:A) \stackrel{(\leq)}{=} S(R:Q) = \chi$

↑
obvious

$$S(R:Q) = \underbrace{S(R)}_{H(P)} + \underbrace{S(Q)}_{S(P)} - \underbrace{S(R,Q)}_{H(P) + \sum_x P_x S(P_x)}$$

$$\begin{aligned} S(R:Q) &= S(R) + S(Q) - S(R,Q) \\ &= S(R) + \underbrace{S(Q,A) - S(R,Q,A)}_{-S(R|Q,A)} \\ &= \underbrace{S(R) + S(Q,A) - S(R,Q,A)}_{S(R:Q,A)} \end{aligned}$$

because A is initially uncorrelated

conditional entropy increases under quantum operation $\rightarrow \geq S(R') - S(R'|Q',A')$

$$\begin{aligned} SS &\rightarrow \geq S(R') - S(R'|A') \\ &= S(R':A') \end{aligned}$$

Properties equivalent to SS

- ① Monotonicity of relative entropy
- ② Double convexity of relative entropy
- ③ Concavity of conditional entropy

$$\textcircled{1} \Rightarrow \textcircled{2}: \quad \rho = \sum_j p_j \rho_j, \quad \sigma = \sum_j p_j \sigma_j$$

$$\rho_{AB} = \sum_j p_j \rho_j \otimes |f_j\rangle\langle f_j|, \quad \rho_B = \sum_j p_j |f_j\rangle\langle f_j| = \sigma_B$$

$$\sigma_{AB} = \sum_j p_j \sigma_j \otimes |f_j\rangle\langle f_j|, \quad \sigma_B = \sum_j p_j |f_j\rangle\langle f_j| = \sigma_B$$

$$\log \rho_{AB} = \mathbb{I} \otimes \log \rho_B + \sum_j p_j \log \rho_j \otimes |f_j\rangle\langle f_j|$$

$$\log \sigma_{AB} = \mathbb{I} \otimes \log \rho_B + \sum_j p_j \log \sigma_j \otimes |f_j\rangle\langle f_j|$$

$$\text{So } S(\rho_{AB} \| \sigma_{AB}) = -S(\rho_{AB}) - \text{tr}(\rho_{AB} \log \sigma_{AB})$$

$$= -H(\vec{p}) - \sum_j p_j S(\rho_j)$$

$$+ H(\vec{p}) - \sum_j p_j \text{tr}(\rho_j \log \sigma_j)$$

$$= \sum_j p_j S(\rho_j \| \sigma_j)$$

$$\textcircled{1} \Rightarrow S(\rho \| \sigma) \leq S(\rho_{AB} \| \sigma_{AB}) = \sum_j p_j S(\rho_j \| \sigma_j)$$

② \Rightarrow ①: The completely depolarizing channel can be written as $\mathcal{Q} = \mathbb{I}/d = \frac{1}{d^2} \sum_{\alpha} U_{\alpha} \otimes U_{\alpha}^{\dagger}$, where the U_{α} are a set of d^2 orthogonal unitary operators. Thus

$$\frac{1}{d^2} \sum_{\alpha} U_{\alpha} \rho U_{\alpha}^{\dagger} = \mathcal{Q}(\rho) = \frac{1}{d} \mathbb{I}(\rho) = \frac{\mathbb{I}}{d} \text{tr}(\rho) = \frac{\mathbb{I}}{d}$$

So

$$\frac{1}{d^2} \sum_{\alpha} I \otimes U_{\alpha} \rho_{AB} I \otimes U_{\alpha}^{\dagger} = \mathcal{D} \otimes \rho_{AB} = \text{tr}_B(\rho_{AB}) \otimes \frac{I}{d} = \rho_A \otimes \frac{I}{d}$$

$$\Rightarrow S(\rho_A \| \sigma_A) = S(\rho_A \otimes (I/d) \| \sigma_A \otimes (I/d))$$

$$\xrightarrow{\text{convexity of relative entropy}} \leq \frac{1}{d^2} \sum_{\alpha} S(I \otimes U_{\alpha} \rho_{AB} I \otimes U_{\alpha}^{\dagger} \| I \otimes U_{\alpha} \sigma_{AB} I \otimes U_{\alpha}^{\dagger})$$

convexity of relative entropy

$$= \frac{1}{d^2} \sum_{\alpha} S(\rho_{AB} \| \sigma_{AB})$$

unitary invariance of relative entropy

$$= S(\rho_{AB} \| \sigma_{AB})$$

$$\textcircled{2} \Rightarrow \textcircled{3}: S(A|B) = -S(\rho_{AB} \| (I/d) \otimes \rho_B) + \log d$$

$$\textcircled{3} \Rightarrow \text{SS: Define } T(\rho_{ABC}) = S(A) + S(B) - S(A,C) - S(B,C) \\ = -S(C|A) - S(C|B)$$

$$\rho_{ABC} = \sum_j \lambda_j |e_j\rangle \langle e_j|$$

$$T(\rho_{ABC}) = -S(C|A) - S(C|B)$$

$$\leq \sum_j \lambda_j (-S_j(C|A) - S_j(C|B))$$

concavity of conditional entropy

$$= \sum_j \lambda_j \left(S_j(A) + S_j(B) - \underbrace{S_j(A,C)}_{S_j(B)} - \underbrace{S_j(B,C)}_{S_j(A)} \right)$$

$$= 0$$

SS \Rightarrow ②: Already shown on pp. 11-12.

③

I haven't been able to find a proof that ③ or SS implies ① or ②, so it remains possible that ① and ② are strictly stronger than ③ and SS.

Schumacher compression.

i.i.d. source: $\rho_j, \rho_j = |\psi_j\rangle\langle\psi_j|$

$$\rho = \sum_j \rho_j \rho_j = \left(\begin{array}{l} \text{unconditioned source} \\ \text{density operator} \end{array} \right)$$

Key is to work with the eigenstates and their eigenvalues rather than the source states and their probabilities.

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

↑ x ↑ i.i.d. random variable X with probability p(x)
eigendecomposition

$$H(\vec{p}) = S(p)$$

$$\rho^{\otimes N} = \sum_{x_1, \dots, x_N} \overbrace{p(x_1) \dots p(x_N)}^{p(x_1, \dots, x_N) = p(x)} |x_1, \dots, x_N\rangle\langle x_1, \dots, x_N|$$

E-typical states: A state $|x_1, \dots, x_N\rangle$ is E-typical if the sequence x_1, \dots, x_N is E-typical, i.e., if

$$\left| \frac{1}{N} \log \left(\frac{1}{p(x_1, \dots, x_N)} \right) - \underset{\substack{\uparrow \\ H(\vec{p})}}{S(p)} \right| \leq \epsilon.$$

Let $T(N, \epsilon)$ be the subspace spanned by the E-typical sequences, and $P(N, \epsilon)$ be the projector onto $T(N, \epsilon)$.

The AEP (typical sequences theorem) can be translated directly to a typical subspace theorem.

Typical Subspace Theorem.

(i) For any $\epsilon, \delta > 0$, there exists N_0 such that for all $N \geq N_0$, $\text{tr}(P(N, \epsilon) \rho^{\otimes N}) \geq 1 - \delta$.

Proof: $\text{tr}(P(N, \epsilon) \rho^{\otimes N}) = \sum_{x \in \text{typical}} p(x_1, \dots, x_N) \geq 1 - \delta$
↑
AEP

(ii) $(1-\delta) 2^{N(S(p) - \epsilon)} \leq |T(N, \epsilon)| = \text{tr}(P(N, \epsilon)) \leq 2^{N(S(p) + \epsilon)}$
↑
(dimension of $T(N, \epsilon)$) = (# of ϵ -typical sequences)

(iii) Let S_N be any subspace of $\mathcal{H}^{\otimes N}$ of dimension $\leq NR$, where $R < S(p)$. Given any $\delta > 0$, there exists N_0 such that for all $N \geq N_0$,

$$\text{tr}(S_N \rho^{\otimes N}) \leq \delta.$$

Proof: $\text{tr}(S_N \rho^{\otimes N}) \leq \sum_{\substack{R \text{ largest} \\ \text{probabilities} \\ p(x)}} p(x) \leq \delta$
↑
AEP

uses lemma on next page

Lemma: Let $\vec{\lambda} = (\lambda_1, \dots, \lambda_D)$, with $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_D$.

$$\max \left(\vec{\lambda} \cdot \vec{g} = \sum_j \lambda_j g_j \mid 0 \leq g_j \leq 1, j=1, \dots, D \text{ and } \sum_{j=1}^D g_j = N \right) = \sum_{j=1}^N \lambda_j.$$

Proof: $\vec{\lambda} \cdot \vec{g}$ does not decrease whenever we decrement g_j by δ and increase $g_k, k \leq j$, by δ , consistent with $g_j \geq 0$ and $g_k \leq 1$. So the maximum occurs when

$$g_j = \begin{cases} 1, & j=1, \dots, N, \\ 0, & j=N+1, \dots, D. \end{cases}$$

Lemma: If $H = H^\dagger$ has eigendecomposition $H = \sum_{j=1}^D \lambda_j |e_j\rangle\langle e_j|$, where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_D$, and Π is an N -dimensional projector, then

$$\left(\begin{array}{l} \text{Sum of } N \\ \text{Smallest} \\ \text{eigenvalues} \end{array} \right) = \sum_{j=D+1-N}^D \lambda_j \stackrel{\text{tr}(H\Pi)}{\approx} \sum_{j=1}^N \lambda_j = \left(\begin{array}{l} \text{Sum of } N \\ \text{largest} \\ \text{eigenvalues} \end{array} \right)$$

attained by $\Pi = \sum_{j=D+1-N}^D |e_j\rangle\langle e_j|$

attained by $\Pi = \sum_{j=1}^N |e_j\rangle\langle e_j|$

Proof: $\text{tr}(H\Pi) = \sum_j \lambda_j \underbrace{\langle e_j | \Pi | e_j \rangle}_{g_j} \leq \sum_{j=1}^N \lambda_j$
 $0 \leq g_j \leq 1, \sum_{j=1}^D g_j = \text{tr}(\Pi) = N$

$$\underbrace{\text{tr}(H(1-\Pi))}_{(D-N)\text{-dimensional projector}} \leq \sum_{j=1}^{D-N} \lambda_j \implies \text{tr}(H\Pi) \geq \text{tr}(H) - \sum_{j=1}^{D-N} \lambda_j = \sum_{j=D-N+1}^D \lambda_j$$

i.i.d. source: $\rho_j, \rho_j = |\psi_j\rangle\langle\psi_j|$

$$\rho = \sum_j P_j \rho_j$$

$$\rho^{\otimes N} = \sum_{j_1, \dots, j_N} \underbrace{P_{j_1} \dots P_{j_N}}_{P_j} \underbrace{|\psi_{j_1}\rangle \dots |\psi_{j_N}\rangle}_{\rho_j}$$

Compression scheme of "rate" R :

Coding $\mathcal{C}^N: \mathcal{H}^{\otimes N} \rightarrow \mathbb{R}^{NR}$ -dimensional subspace

Decoding \mathcal{D}^N

$$\rho_j = |\psi_{j_1}\rangle \dots |\psi_{j_N}\rangle \langle\psi_{j_1}| \dots \langle\psi_{j_N}|$$

$$= |j\rangle\langle j|$$

Reliability:

(average fidelity) = $\overline{F^N} = \sum_j P_j F(\rho_j, \mathcal{D}^N \circ \mathcal{C}^N(\rho_j)) \leq F(\rho^{\otimes N}, \mathcal{D}^N \circ \mathcal{C}^N(\rho^{\otimes N}))$

NBC use entanglement fidelity ↑ double concavity of fidelity

Schumacher compression theorem:

(i) $R > S(\rho)$: For any $\delta > 0$, there exists N_0 such that for all $N \geq N_0$, $\overline{F^N} \geq 1 - \delta$.
 It is easy to push through the proof for entanglement fidelity, even though compressing half of entangled pairs is a harder task.

Proof: Let $R = S(\rho) + \epsilon$, $\epsilon > 0$. The ϵ_T typical subspace

$T(N, \epsilon)$ can be thought of as all subspace of the \mathbb{R}^{NR} -dimensional subspace.

Coding: $\mathcal{C}^N = \underbrace{P(N, \epsilon) \circ P(N, \epsilon)}_{\text{projection into typical subspace}} + \sum_{\alpha} |0\rangle\langle\alpha| \otimes |\alpha\rangle\langle 0|$

↑
orthonormal basis in atypical subspace

↑
any state in typical subspace

Decoding: \mathcal{D}^N is unit superoperator on $T(N, \epsilon)$.

$$\begin{aligned}
 \mathcal{C}^N(\rho_J) &= \underbrace{P(N, \epsilon) \rho_J P(N, \epsilon)}_{P(N, \epsilon) |J\rangle\langle J| P(N, \epsilon)} + |0\rangle\langle 0| \sum_{\alpha} \langle \alpha | \rho_J | \alpha \rangle \\
 &= \text{tr}(\rho_J \bar{P}(N, \epsilon)) \\
 &\quad \downarrow \qquad \qquad \qquad \downarrow \\
 &= \langle J | \bar{P}(N, \epsilon) | J \rangle
 \end{aligned}$$

$$\mathcal{C}^N(\rho_J) = P(N, \epsilon) |J\rangle\langle J| P(N, \epsilon) + |0\rangle\langle 0| \langle J | \bar{P}(N, \epsilon) | J \rangle$$

$$\mathcal{A}^N \circ \mathcal{C}^N(\rho_J) = \mathcal{C}^N(\rho_J)$$

$$\begin{aligned}
 \bar{F}^N &= \sum_J P_J F(\rho_J, \mathcal{A}^N \circ \mathcal{C}^N(\rho_J)) \\
 &= \langle J | \mathcal{A}^N \circ \mathcal{C}^N(\rho_J) | J \rangle^{1/2} = \left(\langle J | P(N, \epsilon) | J \rangle^2 \right. \\
 &\quad \left. + | \langle J | 0 \rangle |^2 \langle J | \bar{P}(N, \epsilon) | J \rangle \right)^{1/2} \\
 &\geq \langle J | P(N, \epsilon) | J \rangle
 \end{aligned}$$

$$\bar{F}^N \geq \sum_J P_J \langle J | P(N, \epsilon) | J \rangle = \text{tr}(\rho^{\otimes N} P(N, \epsilon)) \geq 1 - \delta$$

If we use entanglement fidelity, we have

$$\begin{aligned}
 F_e(\rho^{\otimes N}, \mathcal{A}^N \circ \mathcal{C}^N) &= (\rho^{\otimes N} | \mathcal{A}^N \circ \mathcal{C}^N | \rho^{\otimes N}) \\
 &= (\rho^{\otimes N} | \mathcal{C}^N | \rho^{\otimes N}) \\
 &= |\text{tr}(\rho^{\otimes N} P(N, \epsilon))|^2 + \sum_{\alpha} |\text{tr}(\rho^{\otimes N} |0\rangle\langle \alpha|)|^2
 \end{aligned}$$

The entanglement fidelity result implies the average fidelity result because $\frac{1}{N} \sum_{e=1}^N F_e \geq 1 - \delta$

$$\begin{aligned} &\geq |\text{tr}(P^{\otimes N} P(N, e))|^2 \\ &\geq (1 - \delta')^2 \\ &\geq 1 - 2\delta' \\ &= 1 - \delta \quad (\delta = 2\delta') \end{aligned}$$

This square is the expression of how compressing half of entanglement pairs is a harder task.

(ii) $R < S(p)$: Given any $\delta > 0$, there exists N_0 such that $F_e(p^{\otimes N}, \mathcal{D}^{N_0}(\mathcal{C}^N)) \geq 1 - \delta$ for all $N \geq N_0$.

Proof: Coding $\mathcal{C}^N = \sum_{\alpha} C_{\alpha} C_{\alpha}^{\dagger}$
 ↙ maps to 2^{NR} -dimensional subspace S_N , projector P_N

Decoding: $\mathcal{D}^N = \sum_{\beta} D_{\beta} D_{\beta}^{\dagger}$
 ↙ maps S_N to a subspace $S_{N\beta}$ of dimensionality $\leq \dim(S_N) = 2^{NR}$, projector $P_{N\beta}$.

$$\begin{aligned} F_e(p^{\otimes N}, \mathcal{D}^N(\mathcal{C}^N)) &= (p^{\otimes N} | \mathcal{D}^N(\mathcal{C}^N) | p^{\otimes N}) \\ &= \sum_{\alpha, \beta} |\text{tr}(P^{\otimes N} D_{\beta} C_{\alpha})|^2 \end{aligned}$$

↗ can put P_N here
 ↘ can put $P_{N\beta}$ here

$$= \sum_{\alpha, \beta} |\text{tr}(\sqrt{P^{\otimes N}} P_{N\beta} D_{\beta} C_{\alpha} \sqrt{P^{\otimes N}})|^2$$

$$\begin{aligned}
& \leq \text{tr}(\sqrt{p_{\alpha\beta}^N} D_{\alpha\beta} \sqrt{p_{\alpha\beta}^N}) \\
& \quad \uparrow \\
& \quad \times \text{tr}(\sqrt{p_{\alpha\beta}^N} C_{\alpha}^{\dagger} D_{\beta}^{\dagger} D_{\beta} C_{\alpha} \sqrt{p_{\alpha\beta}^N}) \\
& \quad \text{Schwarz inequality} \\
& \leq \sum_{\alpha,\beta} \underbrace{\text{tr}(p_{\alpha\beta}^N D_{\alpha\beta})}_{\leq \delta} \underbrace{\text{tr}(p_{\alpha\beta}^N C_{\alpha}^{\dagger} D_{\beta}^{\dagger} D_{\beta} C_{\alpha})}_{P_{\alpha\beta}} \\
& \leq \delta \sum_{\alpha,\beta} \underbrace{\text{tr}(p_{\alpha\beta}^N C_{\alpha}^{\dagger} D_{\beta}^{\dagger} D_{\beta} C_{\alpha})}_1 \\
& = \delta
\end{aligned}$$

Let's try to push through the same proof using average fidelity.

$$\begin{aligned}
(\overline{FN})^2 &= \left(\sum_{\alpha} P_{\alpha} F_{\alpha} \right)^2 \\
&\leq \sum_{\alpha} P_{\alpha} F_{\alpha}^2 \\
&= \sum_{\alpha} P_{\alpha} F(p_{\alpha}^N, \mathcal{C}^N(p_{\alpha})) \\
& \quad \langle \mathcal{J} | \mathcal{C}^N(p_{\alpha}) | \mathcal{J} \rangle = \sum_{\alpha,\beta} |\langle \mathcal{J} | D_{\beta} C_{\alpha} | \mathcal{J} \rangle|^2 \\
& \quad \quad \quad \uparrow \quad \quad \uparrow \\
& \quad \quad \quad P_{\alpha\beta} \quad P_{\alpha} \\
&= \sum_{\alpha} P_{\alpha} \sum_{\alpha,\beta} |\langle \mathcal{J} | P_{\alpha\beta} D_{\beta} C_{\alpha} | \mathcal{J} \rangle|^2 \\
& \leq \langle \mathcal{J} | P_{\alpha\beta} | \mathcal{J} \rangle \underbrace{\langle \mathcal{J} | C_{\alpha}^{\dagger} D_{\beta}^{\dagger} D_{\beta} C_{\alpha} | \mathcal{J} \rangle}_{P_{\alpha\beta}}
\end{aligned}$$

$$\leq \sum_{\beta} P_{\beta} \langle J | P_{N\beta} | J \rangle P_{\beta} | J \rangle$$

We're stuck. If there is only one β (unitary decoding) or if all D_{β} map to some subspace S'_N , then we have

$$\begin{aligned} (\overline{FN})^2 &\leq \sum_{\beta} P_{\beta} \langle J | P'_N | J \rangle \underbrace{\sum_{\beta} P_{\beta} | J \rangle}_{\mathbb{1}} \\ &= \text{tr}(P^{\otimes N} P'_N) \\ &\leq \delta \end{aligned}$$

It seems clear we should be able to do the general case, but I don't know how.