QUANTUM INFORMATION THEORY:

ANURAN MAKUR

Presentation Notes: "Lower Bounds on the Probability of Error for Classical and Classical-Quantum Channels," by Marco Dalai

1 Rough Trajectory:

Intro. {
- Basic quantum mechanics
- Classical-quantum channel
- Classical channel and basic definitions

- Lovász's approach in a quantum light: umbrella bound
- The Quantum sphere-packing bound
- Relationships between fundamental quantities

2 Basic Quantum Mechanics:

- The state of a quantum system can be encoded in a finite-dimensional Hilbert space eg: $\mathbb{C}^d$
- Dirac bra-ket notation: The state is written as a column vector: $|x\rangle \leftarrow$ ket

  $|x\rangle^H = \langle x| \leftarrow$ bra $\qquad \langle x|y\rangle = x^H y \leftarrow$ inner product (bra-ket)

  $|x\rangle\langle y| = xy^H \longleftarrow$ outer prouduct (rank 1 matrix)

Physically, if $\langle x|y\rangle = 0$, then $|x\rangle$ and $|y\rangle$ are distinguishable states. eg: spin of electrons $(+\frac{1}{2}, -\frac{1}{2})$

* - Superposition principle: If a quantum system can be in 2 distinguishable states $|x\rangle, |y\rangle$, then it can also be in any linear combination of the states: $\alpha|x\rangle + \beta|y\rangle$.

We restrict $\langle x|x\rangle = \langle y|y\rangle = 1$ and $|\alpha|^2 + |\beta|^2 = 1$.

To observe a quantum system, we must make a measurement.

eg: Let $|\psi\rangle = \alpha|x\rangle + \beta|y\rangle$. Measure: $\langle\psi|x\rangle = \alpha^*\langle x|x\rangle = \alpha^* \Rightarrow |\langle\psi|x\rangle|^2 = |\alpha|^2$

private world of quantum system — inner product — probability of being in state $|x\rangle$

Measuring $|\psi\rangle$ immediately changes its state to $|x\rangle$ wp $|\alpha|^2$ and $|y\rangle$ wp $|\beta|^2$.

This is a manifestation of the observer effect.

- Tensor spaces: If we have 2 quantum systems (each with state space $\mathcal{H}$), then the joint state space is $\mathcal{H}\otimes\mathcal{H}$. eg: n quantum systems $\underset{(\text{qubit})}{\mathbb{C}^2} \rightarrow \mathbb{C}^{2^n}$ state space.

  → Entanglement: Suppose we have 2 quantum systems in states $|\psi_1\rangle = \alpha_1|x\rangle + \beta_1|y\rangle$ and $|\psi_2\rangle = \alpha_2|x\rangle + \beta_2|y\rangle$. Their joint state is:

  $$|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle = \alpha_1\alpha_2|xx\rangle + \alpha_1\beta_2|xy\rangle + \beta_1\alpha_2|yx\rangle + \beta_1\beta_2|yy\rangle$$

  Superposition principle $\Rightarrow |\psi\rangle$ can be any state $|\psi\rangle = a|xx\rangle + b|xy\rangle + c|yx\rangle + d|yy\rangle$, where $|a|^2+|b|^2+|c|^2+|d|^2 = 1$.

  So, a joint system can be in a state s.t. the individual systems do not have a well-defined state! eg: Einstein-Podolsky-Rosen (EPR) photons showed by Bell, Greenburger-Horne-Zeilinger (GHZ) state

  ↳ This along with high (exponential) dim of joint state space is the strength of quantum computation.

  ↳ No good low rate upper bound on $E(R)$ because measurements are entangled.

* - Linearity principle: Isolated quantum system undergoes linear evolution.

  → Schrödinger picture: $|\psi(t)\rangle = U(t,t_0)|\psi(t_0)\rangle$

  state at time $t > t_0$ — unitary operator — state at time t_0

  → Compare w/ Classical Probability: finite-state Markov chain

  $P_n = W^n P_0$

  n-step distribution — stoch. matrix — initial dist.

Benefit of quant. view [
∴ Classical preserves $\mathcal{L}_1$-norm, quantum preserves $\mathcal{L}_2$-norm.
So, a quantum perspective allows us to work on the smooth unit sphere instead of the simplex.

- **Density Operators:** These are $n\times n$ Hermitian positive semi-definite matrices with unit trace in an n-dim quantum state space.

  eg: A s.t. $A^H = A$, $A \succeq 0$, $\mathrm{tr}(A) = 1$.

  → Pure-state: If $|\psi\rangle$ is the state of a quantum system, then $\rho = |\psi\rangle\langle\psi|$ is a rank-1 pure-state density operator.

  → Mixed-state: Suppose we have states $|\psi_1\rangle, \dots, |\psi_n\rangle$ w.p. $p_1, \dots, p_n$ respectively. A mixed-state is: $\rho = \sum_{i=1}^{n} p_i |\psi_i\rangle\langle\psi_i|$. ← useful in info. theory as we have prob. dist. over signals (source symbols)

* Why use density operators? Let $|\psi\rangle = \sum_{i=1}^{n} \sqrt{p_i}\, |\psi_i\rangle$ be the mixed state.

- Suppose $\{\psi_i\}$ is an ⊥-normal basis. Given $|\psi\rangle$, we know the actual state we don't care or know about the distinguishable states $|\psi_i\rangle$. When we do not know $|\psi\rangle$, but only know p_i for $|\psi_i\rangle$, we use ρ.
- In the form $\rho = \sum_{i=1}^{n} p_i |\psi_i\rangle\langle\psi_i|$, we can take an EVD of ρ to find $\{\psi_i\}$ upto $e^{j\theta}$-factor. ← doesn't matter in quant. mech.

→ ∴ Density operators capture all information about state.

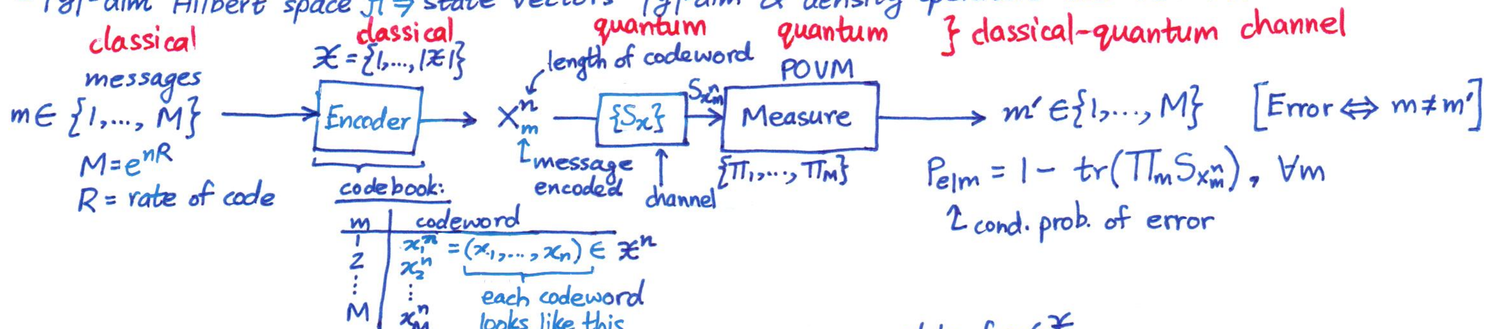
→ Observe: $|\langle\psi|\psi_i\rangle|^2 = \mathrm{tr}(\rho \cdot |\psi_i\rangle\langle\psi_i|)$ for ⊥-normal $\{\psi_i\}$

↑ represents probability of ρ being in state $|\psi_i\rangle\langle\psi_i|$.

(We have augmented vector states to matrix states & use the matrix inner product.)

③ Classical-Quantum Channel:

- Positive operator valued measurement (POVM) ← generalization of von Neumann measurement
- $|\mathcal{Y}|$-dim Hilbert space $\mathcal{H}$ ⇒ state vectors $|\mathcal{Y}|$-dim & density operators are $|\mathcal{Y}|\times|\mathcal{Y}|$ matrices



$P_{e|m} = 1 - \mathrm{tr}(\Pi_m S_{x_m^n})$, $\forall m$

↑ cond. prob. of error

- For each $x \in \mathcal{X}$, there is an associated density operator S_x ← state of $x\in\mathcal{X}$. So, with $X_m^n = (x_1, \dots, x_n)$ we have the associated density operator $S_{x_m^n} = S_{x_1} \otimes \cdots \otimes S_{x_n}$ (n-fold tensor product space $\mathcal{H}^{|\mathcal{Y}|^n}$)
- POVM: Collection of M Hermitian positive semidefinite matrices $\{\Pi_1, \dots, \Pi_M\}$ (each $|\mathcal{Y}|^n \times |\mathcal{Y}|^n$) ← projection to subspace corresp. to message M

  s.t. $\sum_{i=1}^{M} \Pi_i \preceq I$. ← Löwner partial order

  eg: $|\mathcal{Y}|^n = e^{nR} = M$. Then let $\sum_{i=1}^{M} \Pi_i = I$ and let $\Pi_i = |\psi_i\rangle\langle\psi_i|$, where $\{|\psi_i\rangle\}$ is an ⊥-normal basis of $\mathcal{H}^{|\mathcal{Y}|^n}$. If $S_{x_i^n} = |\psi_i\rangle\langle\psi_i|$, then $\mathrm{tr}(\Pi_i S_{x_i^n}) = 1$ and $\mathrm{tr}(\Pi_i S_{x_j^n}) = 0$ for $i \neq j$. So, $P_{i|i} = 1$ and $P_{i|j} = 0$.
- Each Π_i is a projection matrix onto a subspace $S_i \subseteq \mathcal{H}^{|\mathcal{Y}|^n}$. When $\Pi_i \Pi_j = 0$, $i \neq j$ and $\sum_{i=1}^{M} \Pi_i = I$, the S_i are orthogonal subspaces & $S_1 \oplus \cdots \oplus S_M = \mathcal{H}^{|\mathcal{Y}|^n}$. In general, the direct sum of S_i may be a subspace of $\mathcal{H}^{|\mathcal{Y}|^n}$ (e.g. $M < |\mathcal{Y}|^n$); this is why we use $\sum_{i=1}^{M} \Pi_i \preceq I$.
- S_i is the subspace corresponding to message i. Hence, probability message m' is decoded given m is transmitted is: $P_{m'|m} = \mathrm{tr}(\Pi_{m'} S_{x_m^n}) \Rightarrow P_{e|m} = 1 - \mathrm{tr}(\Pi_m S_{x_m^n})$ } Prob. of error given m sent.

- Pure-state channel: Each S_x, $x \in \mathcal{X}$ is rank-1 i.e. $S_x = |\psi_x\rangle\langle\psi_x|$.
- In general, S_x are mixed state and $S_{x_m^n}$ are certainly mixed state if we use a prob. dist. on $\mathcal{X}$. eg: $S_{x_m^n} = S_{x_1} \otimes \cdots \otimes S_{x_n}$, $S_{x_i} = \sum_x P_X(x) S_x$, $\forall i$ in random coding.

④ Classical Channel and Basic Definitions:

messages $m \in \{1,\dots,M\}$, $M = e^{nR}$, R = rate of code

$\mathcal{X} = \{1,\dots,|\mathcal{X}|\}$

$$m \longrightarrow \boxed{\text{Encode}} \longrightarrow X_m^n \xrightarrow{\ \mathcal{X}\ } \boxed{\underset{\text{DMC}}{W}} \xrightarrow{\ \mathcal{Y}\ } Y^n \longrightarrow \boxed{\text{Decoder}} \longrightarrow \underset{\{1,\dots,M\}}{m'}$$

Encode: codebook (same as quantum case)

X_m^n: n-length codeword

Decoder: $\{\mathcal{Y}_1^n,\dots,\mathcal{Y}_M^n\}$ ← $\mathcal{Y}_{m'}^n$ is the set of output seq. mapped to m'

[Error $\Leftrightarrow m \neq m'$]

$P_{e|m} = 1 - \sum_{y^n \in \mathcal{Y}_m^n} W_{x_m^n}(y^n)$

W represents the conditional distributions: $W_x(y) = P_{Y|X}(y|x), \forall x, y$.

- Observations: For each $x \in \mathcal{X}$, W_x is analogous to the density operator S_x.
  Let $W_{x^n}(y^n) \triangleq P_{Y^n|X^n}(y^n|x^n) = \prod_{i=1}^{n} W_{x_i}(y_i)$ [memoryless]. Clearly, $\underline{W_{x^n} = W_{x_1} \otimes \cdots \otimes W_{x_n}}$, where $x^n = (x_1,\dots,x_n)$.
  vectors
  tensor space in quantum case

- From Classical to Quantum and back: (2 ways to change classical to quantum)

This quant. channel is NOT equiv. to the classical channel, but it has the same C_0 & confusability graph.

1. Pure-state: useful to understand Lovász framework in quantum light i.e. combinatorial framework
   For each $x \in \mathcal{X}$ in classical channel, let $|\psi_x\rangle = [\sqrt{W_x(1)} \cdots \sqrt{W_x(|\mathcal{Y}|)}]^T$ be the state vector in the quantum channel. So, $S_x = |\psi_x\rangle\langle\psi_x|$.
   ✱ Notice how square roots are introduced in $|\psi_x\rangle$ to make W_x normalize in $\mathcal{L}_2$ rather than $\mathcal{L}_1$.
2. Mixed-state: useful to understand sphere-packing bound in quantum light i.e. probabilistic framework
   For each $x \in \mathcal{X}$ in classical channel, let $S_x = \begin{bmatrix} W_x(1) & & 0 \\ & \ddots & \\ 0 & & W_x(|\mathcal{Y}|) \end{bmatrix}$ in the quantum channel.
   Let $\Pi_i = \begin{bmatrix} * & & 0 \\ & \ddots & \\ 0 & & * \end{bmatrix} \begin{cases} 1, \text{if } y^n \in \mathcal{Y}_i^n \\ 0, \text{ow} \end{cases}$ $\forall i \in \{1,\dots,M\}$.

This quant. channel IS equiv. to the classical channel in capacity sense.

Thm: If matrices A, B commute, then A, B are jointly diagonalizable. → common property used in blind source separation in (sensor) array processing eg: JADE algorithm

Hence, if the density operators S_x of a quantum channel pairwise commute, then the S_x are diagonal in some basis, the Π_i are also diagonal in the same basis, and the quantum channel reduces to a classical channel.

- Basic Definitions: (Classical and Quantum)
  - Let $P_{e,max} \triangleq \max_m P_{e|m}$ and $P_{e,max}^{(n)}(R) \triangleq$ smallest $P_{e,max}$ among all n-length codes with rate $\geq R$.
    Shannon's channel coding thm: If $R < C$, then $\exists$ sequence of codes s.t. $\lim_{n\to\infty} P_{e,max}^{(n)}(R) = 0$. (achievability)
  - C: capacity of channel, $C = \underbrace{\max_{P_X} I(P_X; P_{Y|X})}_{\text{classical case}}$
  - For rates $C_0 < R < C$, $P_{e,max}^{(n)}(R)$ has exponential decay in n. (C_0 will be defined)
    Reliability function: $\boxed{E(R) \triangleq \limsup_{n\to\infty} -\frac{1}{n}\log\left(P_{e,max}^{(n)}(R)\right)}$. ← lower bound on $P_{e,max}(R)$ is upper bound on $E(R)$
    So, $\underline{P_{e,max}(R) \doteq e^{-nE(R)}}$ (when limit exists).

classical:

- Gallager's Random Coding Bound: $\boxed{E(R) \geq E_r(R)}$,
  prove using Hölder's ineq. on ML decoding analysis

$$\boxed{\begin{aligned} E_r(R) &= \sup_{0\leq\rho\leq 1}\{E_0(\rho) - \rho R\} \text{ and} \\ E_0(\rho) &= \max_P E_0(\rho, P), \\ E_0(\rho, P) &= -\log \sum_y \left(\sum_x P(x) W_x(y)^{\frac{1}{1+\rho}}\right)^{1+\rho} \end{aligned}}$$

- (Shannon-Gallager-Berlekamp) Sphere-packing bound:
  $\boxed{E(R) \leq E_{sp}(R)}$, $\boxed{E_{sp}(R) = \sup_{\rho \geq 0}\{E_0(\rho) - \rho R\}}$
  goes to ∞ for $R < R_\infty$
  same $E_0(\rho) \Rightarrow$ bounds tight ($E(R)$ known) for R s.t. sup over $\rho \geq 0$ is the same as sup over $0 \leq \rho \leq 1$.

quantum:

- Random Coding Bound only exists for pure state channels: $\boxed{E_0(\rho, P) = -\log\left(\mathrm{tr}\left(\sum_x P(x) S_x^{\frac{1}{1+\rho}}\right)^{1+\rho}\right)}$
  everything else is same as above
- Sphere-packing bound for general mixed-state channels will be derived.

- **Zero-error capacity:** $\boxed{C_0 \triangleq \sup\{R : P_{e,\max}^{(n)}(R) = 0 \text{ for some } n\}}$ ← communication with no prob. of error

Clearly, $E(R) = \infty$ for $0 \le R < C_0$. (Recall: $P_{e,\max}(R) \doteq e^{-nE(R)}$)

→ Classical: $C_0 > 0 \Leftrightarrow \forall x_m^n, x_{m'}^n$, the cond. dist.s $W_{x_m^n}$ and $W_{x_{m'}^n}$ have disjoint supports.

$\Leftrightarrow \forall x_m^n, x_{m'}^n, \exists i$ s.t. $W_{x_{m,i}}(y) W_{x_{m',i}}(y) = 0, \forall y$ (have disjoint supports), where $x_{m,i}$ and $x_{m',i}$ are the ith symbols of x_m^n and $x_{m'}^n$, respectively.

$\forall x_m^n, x_{m'}^n;$ $\langle W_{x_m^n}, W_{x_{m'}^n}\rangle = 0$

$= \prod_{i=1}^n \langle W_{x_{m,i}}, W_{x_{m',i}}\rangle$

$\Leftrightarrow \forall x_m^n, x_{m'}^n \; \exists i$ s.t. $\langle W_{x_{m,i}}, W_{x_{m',i}}\rangle = 0$

→ Quantum: $C_0 > 0 \Leftrightarrow \forall m \neq m'$, $tr(\Pi_m S_{x_m^n}) = 1$ and $tr(\Pi_m S_{x_{m'}^n}) = 0$ (i.e. $P_{m|m} = 1$ and $P_{m|m'} = 0$)

$\Leftrightarrow \forall m \neq m'$, $S_{x_m^n} \perp S_{x_{m'}^n}$ i.e. $tr(S_{x_m^n} S_{x_{m'}^n}) = 0$ $\left[tr(S_{x_m^n} S_{x_{m'}^n}) = \prod_{i=1}^n tr(S_{x_{m,i}} S_{x_{m',i}})\right]$

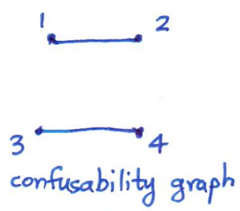
$\Leftrightarrow \forall m \neq m', \exists i$ s.t. $tr(S_{x_{m,i}} S_{x_{m',i}}) = 0$

- **Confusability graph:**

→ For $x_1, x_2 \in \mathcal{X}$, we say x_1, x_2 are not confusable if $\begin{cases} W_{x_1} \text{ and } W_{x_2} \text{ have disjoint supports (classical).} \\ S_{x_1} \text{ and } S_{x_2} \text{ satisfy } tr(S_{x_1} S_{x_2}) = 0 \text{ (quantum).} \end{cases}$

→ The confusability graph has vertices $\mathcal{X}$, and edges $(x_1, x_2) \Leftrightarrow x_1, x_2$ confusable.

eg: $\mathcal{X} = \{1,2,3,4\}$, $\mathcal{Y} = \{0,1\}$. Let $W_x(y) = \begin{cases} 1, & y=0 \\ 0, & y=1 \end{cases}$ for $x = 1,2$, and $W_x(y) = \begin{cases} 0, & y=0 \\ 1, & y=1 \end{cases}$ for $x = 3,4$.



Observe: C_0 depends only on the confusability graph.

⇒ Finding or bounding C_0 is a COMBINATORIAL problem.

classical

- **Shannon's bound:** $C_0 \le C_{FB}$, where C_{FB} is the zero-error capacity with feedback.] clever combinatorial trick

$C_{FB} = R_\infty$ ← value below which $E_{sp}(R)$ diverges ⇒ $\boxed{C_0 \le C_{FB} = R_\infty}$ ← Since $E(R) = \infty$ for $R < C_0$, and $E_{sp}(R) \ge E(R)$, the value when $E_{sp}(R)$ diverges is a bound on C_0

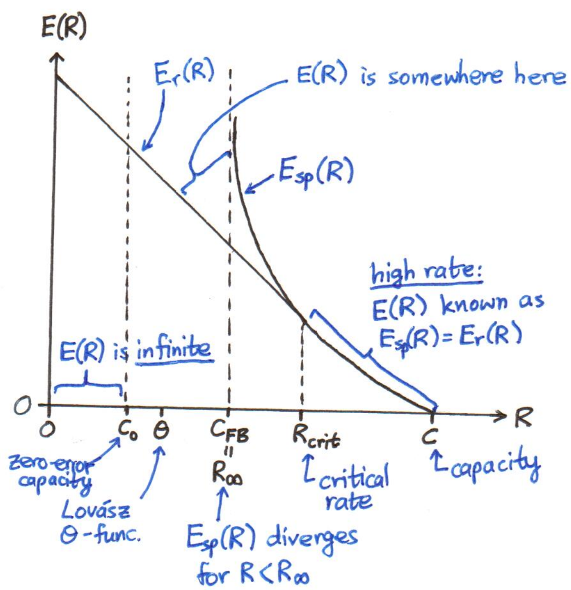
- **Lovász theta-function:** tighter bound on C_0 thant C_{FB}

$\boxed{C_0 \le \theta}$, where $\boxed{\theta = \min_{\{u_x\}} \min_{c : \|c\| = 1} \max_{x \in \mathcal{X}} \log\left(\frac{1}{|\langle u_x | c\rangle|^2}\right)}$

all sets of unit norm vectors s.t. $u_x \perp u_{x'}$ if x, x' cannot be confused

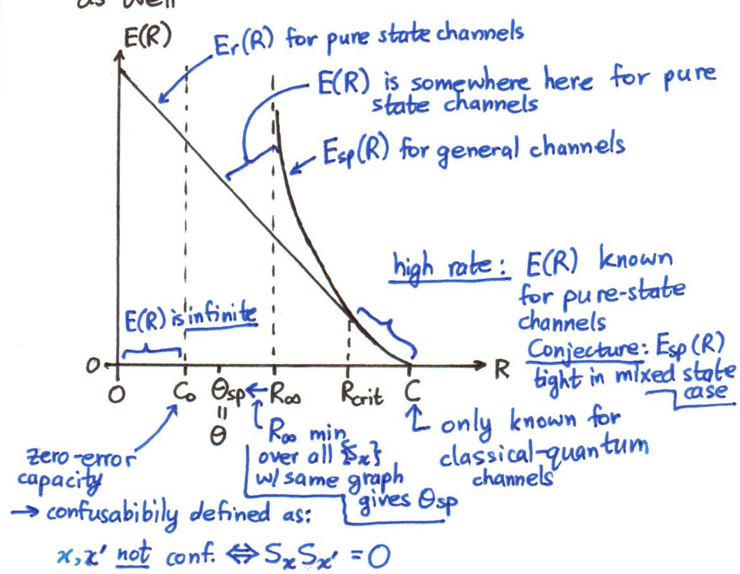
combinatorial proof (as we will see)

- **Summary:** (classical)



- **Summary:** (quantum)

- no $E_{sp}(R)$ yet as quantum Chernoff bound is recent
- $E_r(R)$ exists only for pure-state channels
- Author creates an analogous picture here as well



⑤ Lovász's Approach in a quantum light: Umbrella bound

- Goal: Obtain Lovász's θ-bound as a consquence of bounding $E(R)$ for classical DMC
  - $\hookrightarrow$ extension of Lovász's approach
  - $\hookrightarrow$ objective is not to get tight bound on $E(R)$
- Large deviations: (Classical)

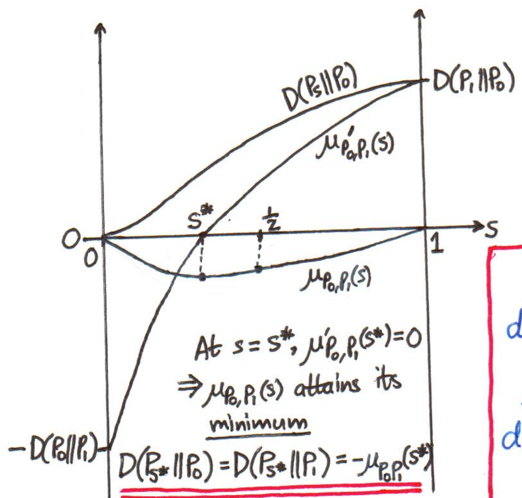
Consider 2 distributions P_0, P_1 on $\mathcal{Y}$.

Let the geometric mean be: $P_s(y) = \dfrac{P_0^{1-s}(y) P_1^{s}(y)}{e^{\mu_{P_0,P_1}(s)}}$, $0 \le s \le 1$. ← normalization

$$P_s(y) = \underbrace{P_0(y)}_{\text{base distribution}} \exp\Big[\overset{\text{natural parameter}}{s}\, \underbrace{\log\Big(\frac{P_1(y)}{P_0(y)}\Big)}_{\text{natural statistic}} - \underbrace{\mu_{P_0,P_1}(s)}_{\text{log-partition function}}\Big] \leftarrow \text{(regular) linear exponential family}$$

$$\mu_{P_0,P_1}(s) = \log\Big(\sum_y P_0^{1-s}(y) P_1^{s}(y)\Big), \quad 0 \le s \le 1 \quad \Rightarrow \mu'_{P_0,P_1}(s) = \mathbb{E}_{P_s}\Big[\log\Big(\frac{P_1(Y)}{P_0(Y)}\Big)\Big]$$

$\&\ \mu''_{P_0,P_1}(s) \ge 0$ ← Fisher information



From the figure, clearly:
$\mu_{P_0,P_1}(s^*) \le \mu_{P_0,P_1}(\tfrac{1}{2})$

Also true that:
$\mu_{P_0,P_1}(s^*) \ge 2\mu_{P_0,P_1}(\tfrac{1}{2})$

Chernoff distance:
$$d_C(P_0,P_1) \triangleq -\min_{0\le s\le 1} \mu_{P_0,P_1}(s)$$
Bhattacharyya distance:
$$d_B(P_0,P_1) \triangleq -\mu_{P_0,P_1}(\tfrac{1}{2}) = -\log\Big(\sum_y \sqrt{P_0(y)P_1(y)}\Big)$$

Analogously in the quantum case, we have:
For density operators A, B:
$$\mu_{A,B}(s) = \log\big(\mathrm{tr}(A^{1-s}B^{s})\big), \quad 0 \le s \le 1$$
$$d_C(A,B) \triangleq -\min_{0\le s\le 1} \mu_{A,B}(s)$$
$$d_B(A,B) \triangleq -\mu_{A,B}(\tfrac{1}{2}) = -\log\big(\mathrm{tr}(\sqrt{A}\sqrt{B})\big)$$
$$d_B(A,B) \le d_C(A,B) \le 2 d_B(A,B)$$

We have: $\boxed{d_B(P_0,P_1) \le d_C(P_0,P_1) \le 2 d_B(P_0,P_1)}$ (from μ relations above)

If we do Binary Hypothesis testing in Neyman-Pearson formulation:

$H_0: Y_1^n \sim$ iid P_0 $\quad H_1: Y_1^n \sim$ iid P_1 $\longrightarrow$ log-likelihood ratio test w/ γ threshold

P_0, P_s, P_1, $\mathcal{L} = \{P : \mathbb{E}_P[\log(\frac{P_1}{P_0}(Y))] = \gamma\}$

Then,
$$\left.\begin{aligned} &\lim_{n\to\infty} -\frac{1}{n}\log(P_{e|H_0}) = D(P_s\|P_0) \text{ for some } s \text{ depending on } \gamma \\ &\lim_{n\to\infty} -\frac{1}{n}\log(P_{e|H_1}) = D(P_s\|P_1) \text{ for some } s \text{ depending on } \gamma \end{aligned}\right\} \lim_{n\to\infty} -\frac{1}{n}\log(P_e) = \min\{D(P_s\|P_1), D(P_s\|P_0)\}$$

→ Hence, P_e decays fastest when $D(P_s\|P_1) = D(P_s\|P_0) = d_C(P_0,P_1)$: $\boxed{\lim_{n\to\infty} -\frac{1}{n}\log(P_e) = d_C(P_0,P_1)}$.

- Quantum view: (DMC) For each $x \in \mathcal{X}$, we have cond. dist. W_x. Def: $|\psi_x\rangle = [\sqrt{W_x(1)} \cdots \sqrt{W_x(|\mathcal{Y}|)}]^T$. ← state of $x \in \mathcal{X}$

For codeword $x_m^n = (x_1, \dots, x_n)$, $\sqrt{W_{x_m^n}(y^n)} = \prod_{i=1}^{n} \sqrt{W_{x_i}(y_i)} \Rightarrow |\psi_m\rangle = |\psi_{x_1}\rangle \otimes \cdots \otimes |\psi_{x_n}\rangle$ ← state of x_m^n (mth codeword)

Note: $C_0 > 0 \Leftrightarrow \exists x, x', x \ne x'$ s.t. $\langle \psi_x | \psi_{x'} \rangle = 0$. $\Rightarrow$ Codes exist s.t. $\langle \psi_m | \psi_{m'} \rangle = 0$ for some $m \ne m'$.

$$d_B(W_{x_m^n}, W_{x_{m'}^n}) = -\log\Big(\sum_{y^n} \sqrt{W_{x_m^n}(y^n) W_{x_{m'}^n}(y^n)}\Big) = -\log(\langle \psi_m | \psi_{m'} \rangle)$$
← This intuition changes Lovász framework to hyp. testing framework

- Binary Hypothesis testing between m & m': $-\log(P_e) = d_C(W_{x_m^n}, W_{x_{m'}^n}) + o(n)$ [see above]

$\Rightarrow -\log(P_e) \le 2 d_B(W_{x_m^n}, W_{x_{m'}^n}) + o(n) \Rightarrow -\log(P_e) \le -2\log(\langle \psi_m | \psi_{m'} \rangle) + o(n)$

For a fixed code, $-\log(P_{e,\max}) \le \min_{m\ne m'} -2\log(\langle \psi_m | \psi_{m'} \rangle) + o(n) \Rightarrow -\log(P_{e,\max}) \le -2\log\big(\max_{m\ne m'} \langle \psi_m | \psi_{m'} \rangle\big) + o(n)$

$P_{e,\max} \ge$ Prob. of error of in hyp. test between 2 codewords

- IDEA: $\boxed{\text{Upper bound } E(R) \text{ by lower bounding } \max_{m\ne m'} \langle \psi_m | \psi_{m'} \rangle}$
- Value of a representation: For $\rho \ge 1$, $\boxed{\Gamma(\rho) \triangleq \{\{\tilde{\psi}_x\}, x \in \mathcal{X} : \langle \tilde{\psi}_x | \tilde{\psi}_x \rangle = 1, \forall x \text{ and } |\langle \tilde{\psi}_x | \tilde{\psi}_{x'} \rangle| \le \langle \psi_x | \psi_{x'} \rangle^{\frac{1}{\rho}}, \forall x, x'\}}$

$\{\tilde{\psi}_x\}$ ← tilted vectors ($\perp$-normal representation of degree ρ)

$$\boxed{V(\{\tilde{\psi}_x\}) \triangleq \min_{f : \|f\| = 1} \max_{x \in \mathcal{X}} -\log(|\langle \tilde{\psi}_x | f \rangle|^2)}$$
← value of $\{\tilde{\psi}_x\}$ ← Chebyshev/minmax of Bhattacharyya distance (inner prod. corresp. to graph structure)

optimal f^* is called HANDLE

- Theta function: $\boxed{\Theta(\rho) \triangleq \min_{\{\tilde{\psi}_x\}\in T(\rho)} V(\{\tilde{\psi}_x\}) = \min_{\{\tilde{\psi}_x\}\in T(\rho)} \min_{f:\|f\|=1} \max_{x\in\mathcal{X}} -\log(|\langle\tilde{\psi}_x|f\rangle|^2)}$ ← Find pure state channel s.t. minmax Bhat. dist. is smallest

UMBRELLA BOUND

* **Theorem:** For a DMC, given any $\rho \geq 1$, $\boxed{E(R) \leq 2\rho\Theta(\rho) \text{ for } R > \Theta(\rho)}$.

→ min-max formulation of $\Theta(\rho)$

Proof: (Idea: Construct auxiliary state f close to all possible states $\tilde{\psi}_x$ with any sequence x^n.)

Consider optimal $\{\tilde{\psi}_x\}$ and f for $\Theta(\rho)$. For $x^n = (x_1, \dots, x_n)$, $|\tilde{\psi}_{x^n}\rangle = |\tilde{\psi}_{x_1}\rangle \otimes \cdots \otimes |\tilde{\psi}_{x_n}\rangle$ and we have:

$(1 \geq)\ |\langle\tilde{\psi}_{x^n}|f^{\otimes n}\rangle|^2 = \prod_{i=1}^n |\langle\tilde{\psi}_{x_i}|f\rangle|^2 \geq e^{-n\Theta(\rho)}$ ← small $\Theta(\rho) \Rightarrow$ tighter bound, because $|\langle\tilde{\psi}_x|f\rangle|^2 \geq e^{-\Theta(\rho)}, \forall x$.

Aside {

Lovász bound: $1 = \|f^{\otimes n}\|_2^2 \geq \sum_m |\langle\tilde{\psi}_m|f^{\otimes n}\rangle|^2 \geq M e^{-n\Theta(\rho)}$, for $\underbrace{\{\tilde{\psi}_m\}\ \perp\text{-normal}}_{\text{corresponds to } C_0>0}$ [observe how idea comes into play]

$\Rightarrow M \leq e^{n\Theta(\rho)} \Rightarrow R \leq \Theta(\rho)$. — rate of code on $\{\tilde{\psi}_m\}$ channel

If $R > \Theta(\rho)$, $M > e^{n\Theta(\rho)}$, $\exists \tilde{\psi}_m, \tilde{\psi}_{m'}$ s.t. $|\langle\tilde{\psi}_m|\tilde{\psi}_{m'}\rangle|^2 > 0$. Then $|\langle\psi_m|\psi_{m'}\rangle|^2 \geq |\langle\tilde{\psi}_m|\tilde{\psi}_{m'}\rangle|^{2\rho} > 0$, & $C_0 = 0$.

Hence, $C_0 \leq \Theta(\rho)$.
}

Let $\Psi \triangleq \frac{1}{\sqrt{M}}[|\tilde{\psi}_1\rangle \cdots |\tilde{\psi}_M\rangle]$. Then, $\langle f^{\otimes n}|\Psi\Psi^H|f^{\otimes n}\rangle \geq e^{-n\Theta(\rho)}$.

$\Rightarrow \lambda_{max}(\Psi^H\Psi) \geq e^{-n\Theta(\rho)}$ ← λ denotes eigenvalue

Thm: For a square matrix A, $\lambda_{max}(A) \leq \max_i \sum_j |A_{ij}|$. ← Corollary of the beautiful Gershgorin Circle Theorem used in numerical linear algebra.

$\Rightarrow e^{-n\Theta(\rho)} \leq \max_m \frac{1}{M}\sum_{m'} |\langle\tilde{\psi}_m|\tilde{\psi}_{m'}\rangle|$

$\Rightarrow \frac{Me^{-n\Theta(\rho)} - 1}{M-1} \leq \max_m \frac{1}{M-1}\sum_{m' \neq m} \langle\psi_m|\psi_{m'}\rangle^{\frac{1}{\rho}} \underset{[\text{Jensen}]}{\overset{[\rho \geq 1]}{\leq}} \max_m \left(\frac{1}{M-1}\sum_{m'\neq m}\langle\psi_m|\psi_{m'}\rangle\right)^{\frac{1}{\rho}}$

Observe that: $\max_{m\neq m'} \langle\psi_m|\psi_{m'}\rangle \geq \max_m \frac{1}{M-1}\sum_{m'\neq m}\langle\psi_m|\psi_{m'}\rangle \geq \left(\frac{Me^{-n\Theta(\rho)} - 1}{M-1}\right)^\rho \geq \left(e^{-n\Theta(\rho)} - e^{-nR}\right)^\rho$

If $R > \Theta(\rho)$, $e^{-n\Theta(\rho)}$ dominates [Laplace Principle] $\Rightarrow -\frac{1}{n}\log\left(\max_{m\neq m'}\langle\psi_m|\psi_{m'}\rangle\right) \leq \rho\Theta(\rho) + o(1)$

$\Rightarrow \boxed{E(R) = \lim_{n\to\infty} -\frac{1}{n}\log(P_{e,max}) \leq 2\rho\Theta(\rho)}$ (from earlier expression) [Q.E.D.]

- Remarks:
  - Def: Cut-off rate of channel → limiting rate for sequencial decoding practicality of error-correcting codes → now rendered obsolete due to LDPC and Turbo codes

    $\boxed{R_{cut} \triangleq \max_P -\sum_{x,x'} P(x)P(x')\langle\psi_x|\psi_{x'}\rangle}$

    For $\rho = 1$, WLOG $\tilde{\psi}_x = \psi_x, \forall x \in \mathcal{X}$. Since $\tilde{\psi}_x \geq 0$, we can choose optimal $f \geq 0$. Let $f = \sqrt{Q}$, where Q is a probability distribution.

    $\boxed{\Theta(1) = \min_f \max_x -\log(|\langle\psi_x|f\rangle|^2) = \min_Q \max_x -2\log\left(\sum_y \sqrt{Q(y)W_x(y)}\right) \overset{\text{Csiszár}}{=} R_{cut}}$

    ↑ over all prob. dist.s

Quantum interpretation of Lovász Θ-bound {

  - As $\rho \to \infty$, the constraint $|\langle\tilde{\psi}_x|\tilde{\psi}_{x'}\rangle| \leq \langle\psi_x|\psi_{x'}\rangle^{\frac{1}{\rho}}$ becomes $|\langle\tilde{\psi}_x|\tilde{\psi}_{x'}\rangle| = 0$ if $\langle\psi_x|\psi_{x'}\rangle = 0$.

    Hence, $T(\infty)$ = the set of all pure state quantum channels with the same confusability graph as the channel W.

    $C_0 \leq \Theta(\rho) \Rightarrow C_0 \leq \lim_{\rho\to\infty}\Theta(\rho) = \Theta$ ← tightest bound when $\rho \to \infty$ } eg: $C_0 = \Theta = 0$ for complete confusability graph (min over all $\{\tilde{\psi}_x\}$); $C_0 = \Theta = \log(|\mathcal{X}|)$ for empty conf. graph (min over $\{\tilde{\psi}_x\}$ disappears)
}

  - Using sphere-packing bound techniques in conjunction with Lovász's technique, an umbrella bound for a general classical-quantum channel can be derived.

    For classical-quantum channel $\mathcal{C}$ with density operators $S_x, x \in \mathcal{X}$, for $\rho \geq 1$ let $T(\rho) \triangleq \{\{\tilde{S}_x\}: tr(\sqrt{\tilde{S}_x}\sqrt{\tilde{S}_{x'}}) \leq tr(|\sqrt{S_x}\sqrt{S_{x'}}|)^{2/\rho}\}$. For auxiliary channel $\tilde{\mathcal{C}} \in T(\rho)$, let $\tilde{E}_{sp}(R)$ be the sphere-packing bound (to appear next).

    Thm: $\boxed{E(R) \leq E_{spu}(R)}$, where $\boxed{E_{spu}(R) = \inf_{\rho\geq 1, \tilde{\mathcal{C}}\in T(\rho)} \rho(\tilde{E}_{sp}(R) + R)}$.

6 The Quantum Sphere-Packing Bound:

- We first need a quantum analog of the tight Chernoff-type bound used in the classic Shannon-Gallager-Berlekamp paper.

eg: $A = |x\rangle\langle x|$, $B = |y\rangle\langle y|$, $\langle x|y\rangle = 0 \Rightarrow A, B$ disjoint
So, $\Pi = |y\rangle\langle y|$.

* **Lemma:** For (mixed or pure-state) density operators A, B with non-disjoint supports, let Π be the projection (measurement) operator (corresp. to B) for the binary hypothesis test. Let $P_{e|A} = tr(\Pi A)$ and $P_{e|B} = tr((I-\Pi)B)$, and let $\mu(s) = \mu_{A,B}(s) = \log(tr(A^{1-s}B^s))$.
Then for $0 < s < 1$, either $P_{e|A} > \frac{1}{8}\exp\left[\mu(s) - s\mu'(s) - s\sqrt{2\mu''(s)}\right]$ ← exponent has form of classical KL-divergence: $D(P_s\|P_0) = s\mu'(s) - \mu(s)$

or $P_{e|B} > \frac{1}{8}\exp\left[\mu(s) + (1-s)\mu'(s) - (1-s)\sqrt{2\mu''(s)}\right]$.

* BOTH cond. P_e's cannot decay too fast.

Proof: (Idea: Use Quantum Neyman-Pearson Lemma to get optimal test Π (like classical LRT) and then solve simplified problem using Nussbaum-Szkoła mapping.)

$A = \sum_i \alpha_i |a_i\rangle\langle a_i|$ and $B = \sum_j \beta_j |b_j\rangle\langle b_j|$ [Spectral decomposition] $\{a_i\}, \{b_j\}$ ⊥-normal bases

Quantum N-P Lemma: Π orthogonal $\Rightarrow \Pi(I-\Pi) = 0$ or $\Pi = \Pi^2$

$\Rightarrow \Pi = \sum_j \Pi|b_j\rangle\langle b_j|\Pi$ and $I - \Pi = \sum_i (I-\Pi)|a_i\rangle\langle a_i|(I-\Pi)$

$\Rightarrow P_{e|A} = tr(\Pi A) = \sum_{i,j} \alpha_i |\langle a_i|\Pi|b_j\rangle|^2$ and $P_{e|B} = tr((I-\Pi)B) = \sum_{i,j} \beta_j |\langle a_i|(I-\Pi)|b_j\rangle|^2$

For $\eta_0, \eta_1 > 0$, we have: $\eta_0 P_{e|A} + \eta_1 P_{e|B} \geq \frac{1}{2}\sum_{i,j} \min(\eta_0 \underbrace{\alpha_i |\langle a_i|b_j\rangle|^2}_{Q_0(i,j)}, \eta_1 \underbrace{\beta_j |\langle a_i|b_j\rangle|^2}_{Q_1(i,j)})$ (after algebra)

effective to go from quantum → classical prob.

This is the classical proof for test between Q_0, Q_1.

Nussbaum-Skota mapping: $Q_0(i,j) = \alpha_i |\langle a_i|b_j\rangle|^2$, $Q_1(i,j) = \beta_j |\langle a_i|b_j\rangle|^2$ ← valid probability distributions over (i,j)

For this mapping, $\mu_{A,B}(s) = \log(tr(A^{1-s}B^s)) = \log\left(\sum_{i,j} Q_0(i,j)^{1-s} Q_1(i,j)^s\right) = \mu_{Q_0,Q_1}(s)$ ← same log-partition functions

Exponentially tilt to get $Q_s(i,j) = \frac{Q_0^{1-s}(i,j)Q_1^s(i,j)}{e^{\mu_{Q_0,Q_1}(s)}} = e^{-\mu(s)} Q_0^{1-s}(i,j) Q_1^s(i,j)$, $\mu(s) = \mu_{Q_0,Q_1}(s)$

Let $\mathcal{Z}_s \triangleq \left\{(i,j) : \left|\log\left(\frac{Q_1(i,j)}{Q_0(i,j)}\right) - \mu'(s)\right| \leq \sqrt{2\mu''(s)}\right\}$, where $\mu'(s) = \mathbb{E}_{Q_s}[\log(Q_1/Q_0)]$, $\mu''(s) = VAR_{Q_s}[\log(Q_1/Q_0)]$

For $(i,j) \in \mathcal{Z}_s$,
$Q_s(i,j) \leq Q_0(i,j) \underbrace{\exp[\mu(s) - s\mu'(s) - s\sqrt{2\mu''(s)}]^{-1}}_{\eta_0}$ and $Q_s(i,j) \leq Q_1(i,j) \underbrace{\exp[\mu(s) + (1-s)\mu'(s) - (1-s)\sqrt{2\mu''(s)}]^{-1}}_{\eta_1}$

This choice is arbitrary.

[Chebyshev]
$\frac{1}{2} < \sum_{(i,j)\in\mathcal{Z}_s} Q_s(i,j) \leq \sum_{i,j} \min(\eta_0 Q_0(i,j), \eta_1 Q_1(i,j))$

Hence, $\eta_0 P_{e|A} + \eta_1 P_{e|B} > \frac{1}{4} \Rightarrow P_{e|A} > \frac{1}{8}\eta_0^{-1}$ or $P_{e|B} > \frac{1}{8}\eta_1^{-1}$. ← trade-off between $P_{e|A}$ & $P_{e|B}$ [Q.E.D.]

---

* **Theorem:** (Sphere-packing bound) Let $S_1, \dots, S_{|\mathcal{X}|}$ be density operators for a general classical-quantum channel and $E(R)$ be its reliabity function. Then,

$\forall R > 0, \forall 0 < \epsilon < R$, $\boxed{E(R) \leq E_{sp}(R - \epsilon)}$, where $\boxed{E_{sp}(R) = \sup_{\rho \geq 0} E_0(\rho) - \rho R,\ E_0(\rho) = \max_P E_0(\rho, P)}$,

and $\boxed{E_0(\rho, P) = -\log\left(tr\left(\left(\sum_x P(x) S_x^{\frac{1}{1+\rho}}\right)^{1+\rho}\right)\right)}$.

for rigour

Proof: (Idea: low rate → P_e dominated by worst pair of codewords, high rate → bound $P_{e|m}$ due to "bulk of competitors".
So, bound $P_{e,max}$ by looking at hypothesis test between $S_{x_m^n}$ and dummy density operator F_n.
F_n represents the "bulk of competitors". Author shows $\exists m, F_n$ s.t.
$P_{m|F} = tr(\Pi_m F_n)$ is small $\Rightarrow F_n$ (bulk of competitors m') are distinguishable from $S_{x_m^n}$ to some degree)

Proof cont'd:

every rate R code contains such a subcode as no. of codewords exp. in n ($e^{nR}=M$) but no. of compositions is poly(n)

WLOG assume the code is a constant composition code i.e. all codewords x_m^n, $m\in\{1,\dots,M\}$, have the same empirical distribution P.

Let F_n be a density operator in $\mathcal{H}^{\otimes n}$. ← The whole proof constructs F_n

For any $m\in\{1,\dots,M\}$, consider hypothesis test between $S_{x_m^n}$ and F_n:

From Lemma, $P_{e|m} = \mathrm{tr}\big(\underbrace{(I-\Pi_m)}_{\to F \text{ represents all other } m'\neq m} S_{x_m^n}\big) > \frac{1}{8}\exp\left[\mu(s)-s\mu'(s)-s\sqrt{2\mu''(s)}\right]$

or $P_{e|F_n} = \mathrm{tr}(\Pi_m F_n) > \frac{1}{8}\exp\left[\mu(s)+(1-s)\mu'(s)-(1-s)\sqrt{2\mu''(s)}\right]$

$\mu(s) = \log\left(\mathrm{tr}\left(S_{x_m^n}^{1-s}F_n^{s}\right)\right)$

$\forall m$, $P_{e,max}\geq P_{e|m}$ and $\sum_{m=1}^{M}\Pi_m \preceq I \Rightarrow \exists m$ s.t. $\mathrm{tr}(\Pi_m F_n)\leq \frac{1}{M} = e^{-nR}$. Fix this m & let $x^n = x_m^n$.

$\Rightarrow P_{e,max} > \frac{1}{8}\exp\left[\mu(s)-s\mu'(s)-s\sqrt{2\mu''(s)}\right]$ or $R < -\frac{1}{n}\left[\mu(s)+(1-s)\mu'(s)-(1-s)\sqrt{2\mu''(s)}-\log(8)\right]$ ← *trade-off between R and $P_{e,max}$

Remark: $\mu(s)$ & $\mu'(s)$ grow linearly in n but $\sqrt{\mu''(s)}$ will grow like $\sqrt{n}$, so it will not matter in first order behaviour.

Want result to depend on P, not $x^n = x_m^n$. Let $F_n = F^{\otimes n}$.

Then $\mu(s) = \log\left(\mathrm{tr}\left(S_{x^n}^{1-s}F_n^{s}\right)\right) = n\sum_{x\in\mathcal{X}}P(x)\mu_{S_x,F}(s) \Rightarrow \mu'(s) = n\sum_{x\in\mathcal{X}}P(x)\mu'_{S_x,F}(s)$, $\mu''(s) = n\sum_{x\in\mathcal{X}}P(x)\mu''_{S_x,F}(s)$.

$\sqrt{\mu''(s)} \propto \sqrt{n}$

Let $R_n(s,P,F) \triangleq -\frac{1}{n}\left[\mu(s)+(1-s)\mu'(s)-(1-s)\sqrt{2\mu''(s)}-\log(8)\right]$.

$\Rightarrow -\frac{1}{n}\log(P_{e,max}) < -\frac{1}{1-s}\sum_x P(x)\mu_{S_x,F}(s) - \frac{s}{1-s}R_n(s,P,F) + \frac{1}{n}\left(2s\sqrt{2\mu''(s)}+\frac{\log(8)}{1-s}\right)$ or $R < R_n(s,P,F)$

Remark: Can we use Nussbaum-Skota mapping to make the quantum problem classical? Unfortunately, no. Q_0 and Q_1 depend on both S_x and F. Even if F is fixed, changing x to get different S_x would change both Q_0 and Q_1. In classical proof, both distributions cannot change with x.

This section is not drawn from classical

Construct F: Let $A(s,P) \triangleq \sum_x P(x)S_x^{1-s}$ & let $P_s \triangleq \arg\min_P \mathrm{tr}\left(A(s,P)^{\frac{1}{1-s}}\right)$ for $0<s<1$.

Intuition: F_s "close" to all S_x $\big[\underbrace{-\mu_{S_x F_s}(s)}_{\text{distance}} \leq (1-s)E_0\left(\frac{s}{1-s}\right), \forall x$ eg: $s=\frac{1}{2}$, min s

Define $A_s = A(s,P_s)$.

Holevo: $\mathrm{tr}\left(S_x^{1-s}A_s^{s/1-s}\right)\geq \mathrm{tr}\left(A_s^{\frac{1}{1-s}}\right)$, $\forall x$ w/ equality for x s.t. $P_s(x)>0$.

Let $\boxed{F_s \triangleq \frac{A_s^{1/1-s}}{\mathrm{tr}(A_s^{1/1-s})}}$. Then, $\mu_{S_x,F_s}(s) = \log\left(\mathrm{tr}\left(S_x^{1-s}A_s^{s/1-s}\right)\right) - s\log\left(\mathrm{tr}\left(A_s^{1/1-s}\right)\right) \geq (1-s)\log\left(\mathrm{tr}\left(A_s^{1/1-s}\right)\right) = -(1-s)E_0\left(\frac{s}{1-s}\right)$

P_s is optimal

$\Rightarrow -\frac{1}{n}\log(P_{e,max}) < E_0\left(\frac{s}{1-s}\right) - \frac{s}{1-s}R_n(s,P,F_s) + \frac{2s\sqrt{2}}{\sqrt{n}}\sqrt{\sum_x P(x)\mu''_{S_x,F_s}(s)} + \frac{\log(8)}{(1-s)n}$ or $R < R_n(s,P,F_s)$, where

still arbitrary P, not P_s

$R_n(s,P,F_s) = -\sum_x P(x)\left[\mu_{S_x,F_s}(s)+(1-s)\mu'_{S_x,F_s}(s)\right] + \frac{1}{\sqrt{n}}(1-s)\sqrt{2\sum_x P(x)\mu''_{S_x,F_s}(s)} + \frac{1}{n}\log(8)$

By compactness argument, $\exists$ sequence of codes with length n, rate R_n, composition P_n s.t. $P_n\to P$, $R_n\to R$, $-\frac{1}{n}\log(P_{e,max}^{(n)})\to E(R)$ as $n\to\infty$.

Since $\mu_{S_x,F_s}(s)$ is nonpositive and convex for $s\in(0,1)$: $\mu_{S_x,F_s}(s)+(1-s)\mu'_{S_x,F_s}(s)\leq \mu_{S_x,F_s}(1^-) = 0$ ← A differentiable convex function lies above all its tangents.

$\Rightarrow R_n(s,P_n,F_s)\geq 0$. $R_n(s,P_n,F_s)$ is also continuous in $s\in(0,1)$.

3 Cases: ① $R_n > R_n(s,P_n,F_s)$, $\forall s\in(0,1)$ ② $R_n < R_n(s,P_n,F_s)$, $\forall s\in(0,1)$ ③ $R_n = R_n(s,P_n,F_s)$ for some $s\in(0,1)$.

① Suppose Case ① is satisfied infinitely often for n. Focus on subsequence s.t. $R_n(s,P_n,F_s) < R_n$, $\forall n$. Then, $-\frac{1}{n}\log(P_{e,max}) < E_0\left(\frac{s}{1-s}\right) - \frac{s}{1-s}R_n(s,P_n,F_s) + \underbrace{\frac{2s\sqrt{2}}{\sqrt{n}}\sqrt{\sum_x P(x)\mu''_{S_x,F_s}(s)}}_{} + \frac{\log(8)}{(1-s)n}$.

Fix $s\in(0,1)$ and let $n\to\infty$: $E(R)\leq E_0\left(\frac{s}{1-s}\right) - \frac{s}{1-s}\underbrace{R_n(s,P_n,F_s)}_{\geq 0} \leq E_0\left(\frac{s}{1-s}\right)$

$\therefore E(R)\leq E_0\left(\frac{s}{1-s}\right)$, $\forall s\in(0,1)$ $\Rightarrow E(R)\leq E_0(\rho)$, $\rho\geq 0$ $\left[\rho = \frac{s}{1-s}\right]$

So, $E(R)\leq E_0(0) - 0.R = 0 \Rightarrow \boxed{E(R)\leq E_{sp}(R)}$.

→ Cases ② & ③ similar.

[Q.E.D.]

7 Relationships between Fundamental Quantities:

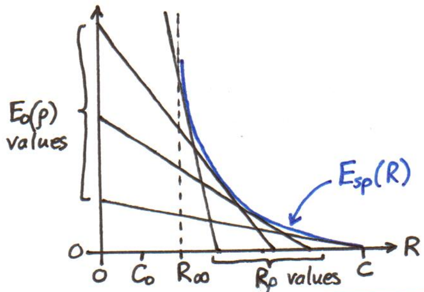
- Rényi divergence: $\boxed{D_\alpha(U,V) \triangleq \frac{1}{\alpha-1}\mu_{U,V}(1-\alpha) = \frac{1}{\alpha-1}\log\sum_z U(z)^\alpha V(z)^{1-\alpha}}$ (when $\alpha \to 1$, $D_\alpha \to D_{KL}$.)

(Quantum case) $\boxed{\underbrace{D_\alpha}_{\text{order}}(A\|B) \triangleq \frac{1}{\alpha-1}\log\left(tr\left(A^\alpha B^{1-\alpha}\right)\right)}$

- The sphere-packing bound: $E_{sp}(R) = \sup_{\rho \geq 0} E_0(\rho) - \rho R$, $E_0(\rho) = \max_P -\log\left(tr\left(\sum_x P(x) S_x^{\frac{1}{1+\rho}}\right)^{1+\rho}\right)$

We work with the quantum version as classical is special case

$E_{sp}(R)$ is the upper envelope of lines $E_0(\rho) - \rho R$. The R-axis intercept of these lines is $\boxed{R_\rho \triangleq \frac{E_0(\rho)}{\rho}}$.



As $\rho \to 0$, the gradient of $E_0(\rho) - \rho R$ is 0
$\Rightarrow R_0 = C$.

As $\rho \to \infty$, the gradient of $E_0(\rho) - \rho R$ is ∞
$\Rightarrow R_\infty$ is where the sphere packing bound diverges

Note: $C_0 \leq R_\infty$ because R_∞ is the smallest R-value when $E_{sp}(R) < \infty \Rightarrow E(R) < \infty$ and $E(R) = \infty$ for $R < C_0$.

- Information Radii: Recall $\boxed{C = \min_{Q \in \mathcal{P}_\mathcal{Y}} \max_{x \in \mathcal{X}} D(W_x\|Q)}$ in classical case. [Gallager's Capacity-Red. Thm shows this is C.]

It turns out, the right measure to look at in quantum information is Rényi divergence. The entire classical formulation can be done using Rényi divergence.

Classical: $\boxed{R_\rho = \min_{Q \in \mathcal{P}_\mathcal{Y}} \max_{x \in \mathcal{X}} D_\alpha(W_x\|Q), \ \alpha = \frac{1}{1+\rho}}$ ← due to Csiszár

As $\rho \to 0$ $(\alpha \to 1)$, $R_0 = C$, as $D_\alpha \to D_{KL}$.

As $\rho \to \infty$ $(\alpha \to 0)$, $\boxed{R_\infty = \min_{Q \in \mathcal{P}_\mathcal{Y}} \max_{x \in \mathcal{X}} -\log\left(\sum_{y: W_x(y) > 0} Q(y)\right)}$.

Quantum: Thm: For a classical-quantum channel with states $S_x, x \in \mathcal{X}$ and $\rho > 0$, $\boxed{R_\rho = \min_F \max_{x \in \mathcal{X}} D_\alpha(S_x\|F)}$ for $\boxed{\alpha = \frac{1}{1+\rho}}$, where $\min_F$ is over all density operators.

Proof Sketch: Start with $R_\rho = \frac{E_0(\rho)}{\rho}$. Use Hölder's inequality with Schatten norms to get:

$$R_\rho = \frac{1}{\alpha-1}\log\left(\min_P \max_{\|B\|_{\frac{1}{1-\alpha}} \leq 1} tr\left(\sum_x P(x) S_x^\alpha B\right)\right).$$

Use von Neumann's minmax theorem. ∎

Remark: If S_x pairwise commute, then optimal F is diagonal in same basis as S_x. So, we recover the classical R_ρ.

- Connection to Lovász Θ-function:

From the quantum formulation, as $\rho \to \infty$ $(\alpha \to 0)$, $\boxed{R_\infty = \min_F \max_x -\log\left(tr\left(S_x^0 F\right)\right)}$.

Assume $S_x = |\psi_x\rangle\langle\psi_x|$ are pure states. Restrict $F = |f\rangle\langle f|$ to be rank 1 density operator.

Then $tr(S_x^0 F) = |\langle\psi_x|f\rangle|^2$ and $\underbrace{V(\{\psi_x\})}_{\text{value of representation } \{\psi_x\}} = \min_{f: \|f\|=1} \max_{x \in \mathcal{X}} -\log\left(|\langle\psi_x|f\rangle|^2\right) = R_{\infty,\text{constrained}}$.

Hence, we have: $\boxed{\underbrace{C_0 \leq R_\infty}_{\text{obvious}} \leq V(\{\psi_x\}) \Rightarrow C_0 \leq \Theta(\rho) = \min_{\{\tilde{\psi}_x\} \in \Gamma(\rho)} V(\{\tilde{\psi}_x\})}$.

Best bound on C_0 is given by min R_∞ over all quantum channels with same confusability graph.

analogous to Θ-func. (Lovász) → $\boxed{\Theta_{sp} \triangleq \min_{\substack{\{S_x\}: \{S_x\} \\ \text{density operators} \\ \text{s.t. } S_x S_{x'} = 0 \text{ if} \\ (x,x') \text{ not connected} \\ \text{in conf. graph}}} R_\infty(\{S_x\}) = \min_{\{S_x\}} \underbrace{\min_F \max_x -\log\left(tr\left(S_x^0 F\right)\right)}_{\text{like min max Bhattacharyya dist. which comes from Rényi-divergence}}}$

plays role of value in Lovász Θ-function

projector onto support of S_x

Clearly, $\boxed{C_0 \leq \Theta_{sp} \leq \underbrace{\Theta}_{\text{Lovász-}\Theta\text{ func.}} = \lim_{\rho \to \infty} \Theta(\rho) \leq \Theta(\rho)}$.

Remark: $C_0 \leq \Theta_{sp}$ can also be proved using exactly Lovász's argument using general density operators as handles. This does not provide connection to $E_{sp}(R)$.

Indeed, it turns out that $\Theta_{sp} = \Theta$.

Hence, $\boxed{C_0 \leq \Theta_{sp} = \Theta}$.

Implication: Minimizing R_∞ over all $\{S_x\}$, F (density operators) gives the same result as minimizing R_∞ over all $\{\psi_x\}$, f (pure state vectors), which we call $R_{\infty,\text{constrained}}$. We conclude that $E_{sp}(R)$ bound on classical-quantum channels (general/mixed) gives same Lovàsz Θ bound to C_0, but pure state channels suffice to give the bound. We also see that $\exists$ a pure state channel whose optimizing F is rank 1. This is not true in general for $R_\infty(\{\psi_x\})$.

- Classical and Pure-state channels:

Consider classical-quantum channels with pure states $S_x = |\psi_x\rangle\langle\psi_x|$. Then $S_x^{\frac{1}{1+\rho}} = S_x$.

Let $\bar{S}_P = \sum_x P(x)|\psi_x\rangle\langle\psi_x|$. ← mixed state generated by dist. P over S_x.

$E_0(\rho, P) = -\log\left(\text{tr}\left(\bar{S}_P^{1+\rho}\right)\right) = -\log\left(\sum_i \underbrace{\lambda_i}_{\text{e-vals}}(\bar{S}_P)^{1+\rho}\right)$

$\Rightarrow \boxed{R_\infty = -\log\left(\min_P \lambda_{max}(\bar{S}_P)\right)}$

Observe: For general S_x,

$E_0(\rho, P) = -\log\left(\text{tr}\left(\left(\sum_x P(x) S_x^{\frac{1}{1+\rho}}\right)^{1+\rho}\right)\right)$

$R_\rho = \frac{\max_P E_0(\rho, P)}{\rho} = \frac{-\log\left(\min_P\left(\text{tr}\left(\left(\sum_x P(x) S_x^{\frac{1}{1+\rho}}\right)^{1+\rho}\right)\right)\right)}{\rho}$

$\lim_{\rho\to\infty} R_\rho = \underbrace{-\log\left(\min_P \lambda_{max}\left(\sum_x P(x) S_x^0\right)\right)}_{\text{eigenvalue problem}} = R_\infty$

From "Connection to Lovász Θ-function", we have:

$R_\infty = \min_F \max_x -\log(\text{tr}(S_x^0 F))$

$= \min_F \max_x -\log(\langle\psi_x|F|\psi_x\rangle)$

↳ looks like $\Theta(1)$ in Lovász section if we restrict F to be rank 1

- Fact: Optimizing F^* is rank 1 if $\lambda_{max}(\bar{S}_{P^*})$ has multiplicity 1 for optimizing P^*.
- Fact: For pure-state channels $\{\psi_x\}$ constructed by $|\psi_x\rangle = \sqrt{W_x}$ from classical channel, there is always an F^* with rank 1, and for $\{\psi_x\}$

R_∞ of pure state channel $\longrightarrow \boxed{R_\infty = \Theta(1) = R_{cut}}$ ← cutoff rate of classical channel.

(Proof uses KKT due to technical issues with applying von Neumann's theorem.)

- Final Remarks:

- The sphere-packing bound $E_{sp}(R)$ produces the quantity R_ρ (quantum case).
- R_ρ has an information radius-type description using quantum Rényi divergence which is used for the development
- $\lim_{\rho\to 0} R_\rho = C$ & $\lim_{\rho\to\infty} R_\rho = R_\infty$
- Minimizing R_∞ over all $\{S_x\}$ channels gives Θ ← Lovász Θ-function.
- This relates C, R_∞, Θ and C_0 in one unified framework in a quantum light.
- Key point: The Θ_{sp} bound on C_0 is obtained using quantum probability. The Θ bound on C_0 is obtained using combinatorics. A combinatorial argument (of Lovász) shows $\Theta_{sp} = \Theta$. So, the quantum probability view gives the same result as the combinatorial view. In this regard, quantum probability unifies the divergence between combinatorial and probabilistic techniques in information theory.