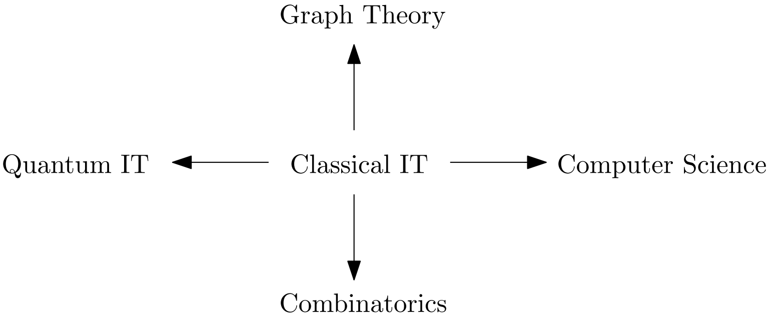
Channel reliability: from ordinary to zero-error capacity

Marco Dalai

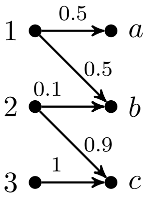
Department of Information Engineering
University of Brescia - Italy

Why I like it


Graph Theory
Quantum IT
Classical IT
Computer Science
Combinatorics

- **Discrete channel** W ($\mathcal{X}, \mathcal{Y}$ finite)

$$x \in \mathcal{X} \longrightarrow \boxed{W_x(y)} \longrightarrow y \in \mathcal{Y}$$



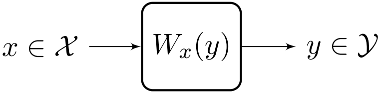
- **Memoryless extension** $\boldsymbol{W}$

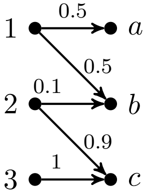
$$\boldsymbol{x} = (x_1, \ldots, x_n) \longrightarrow \boxed{\boldsymbol{W}_{\boldsymbol{x}}(\boldsymbol{y})} \longrightarrow \boldsymbol{y} = (y_1, \ldots, y_n)$$

$$\boldsymbol{W}_{\boldsymbol{x}}(\boldsymbol{y}) = \prod_i W_{x_i}(y_i)$$

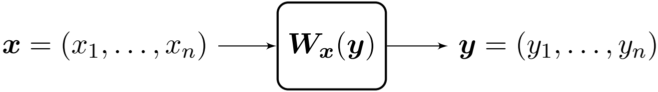
Classical DMCs

- **Discrete channel** W ($\mathcal{X}, \mathcal{Y}$ finite)





- **Memoryless extension** $\boldsymbol{W}$



$$\boldsymbol{W_x}(\boldsymbol{y}) = \prod_i W_{x_i}(y_i)$$

Code and Error Probability

- **Code**: M codewords $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\} \subset \mathcal{X}^n$
- **Decoder**: M disjoint decision regions $\{\boldsymbol{\mathcal{Y}}_1, \ldots, \boldsymbol{\mathcal{Y}}_M\} \subseteq \mathcal{Y}^n$ (here: maximum likelyhood decoder)
- **Probability of error** given message m

$$\mathsf{P}_{\mathrm{e}|m} = \sum_{\boldsymbol{y} \notin \boldsymbol{\mathcal{Y}}_m} \boldsymbol{W}_{\boldsymbol{x}_m}(\boldsymbol{y}) = \boldsymbol{W}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m})$$

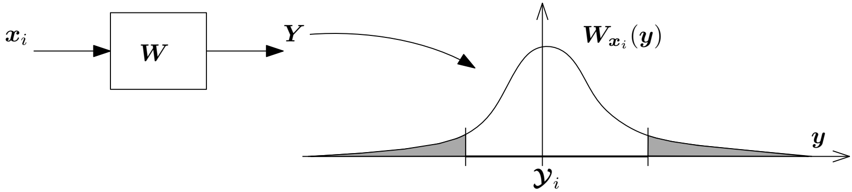
Code and Error Probability

- **Code**: M codewords $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\} \subset \mathcal{X}^n$
- **Decoder**: M disjoint decision regions $\{\boldsymbol{\mathcal{Y}}_1, \ldots, \boldsymbol{\mathcal{Y}}_M\} \subseteq \mathcal{Y}^n$ (here: maximum likelyhood decoder)
- **Probability of error** given message m

$$\mathsf{P}_{\mathrm{e}|m} = \sum_{\boldsymbol{y} \notin \boldsymbol{\mathcal{Y}}_m} \boldsymbol{W}_{\boldsymbol{x}_m}(\boldsymbol{y}) = \boldsymbol{W}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m})$$

- **Code**: M codewords $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\} \subset \mathcal{X}^n$
- **Decoder**: M disjoint decision regions $\{\boldsymbol{\mathcal{Y}}_1, \ldots, \boldsymbol{\mathcal{Y}}_M\} \subseteq \mathcal{Y}^n$ (here: maximum likelyhood decoder)
- **Probability of error** given message m

$$\mathsf{P}_{\mathrm{e}|m} = \sum_{\boldsymbol{y} \notin \boldsymbol{\mathcal{Y}}_m} \boldsymbol{W}_{\boldsymbol{x}_m}(\boldsymbol{y}) = \boldsymbol{W}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m})$$



- **Maximum error probability**

$$\mathsf{P}_{\mathrm{e,max}} = \max_{m} P_{e|m}$$

- **Optimal codes**

$$\mathsf{P}^{(n)}_{\mathrm{e,max}}(R) = \min_{\mathcal{C}} P_{e,\max}$$

where the minimum is over codes of length n and rate at least R

- **Channel capacity**

$$C = \sup\left\{R \,:\, \limsup_{n\to\infty} \mathsf{P}^{(n)}_{\mathrm{e,max}}(R) = 0\right\}$$

- **Maximum error probability**

$$\mathsf{P}_{\mathrm{e,max}} = \max_m P_{e|m}$$

- **Optimal codes**

$$\mathsf{P}^{(n)}_{\mathrm{e,max}}(R) = \min_{\mathcal{C}} P_{e,\max}$$

where the minimum is over codes of length n and rate at least R

- **Channel capacity**

$$C = \sup\left\{ R \,:\, \limsup_{n\to\infty} \mathsf{P}^{(n)}_{\mathrm{e,max}}(R) = 0 \right\}$$

- **Maximum error probability**

$$\mathsf{P}_{\mathrm{e,max}} = \max_{m} P_{e|m}$$

- **Optimal codes**

$$\mathsf{P}^{(n)}_{\mathrm{e,max}}(R) = \min_{\mathcal{C}} P_{e,\max}$$

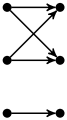
where the minimum is over codes of length n and rate at least R

- **Channel capacity**

$$C = \sup\left\{R \,:\, \limsup_{n\to\infty} \mathsf{P}^{(n)}_{\mathrm{e,max}}(R) = 0\right\}$$

- **Zero-error capacity**

$$C_0 = \sup\{R : \mathsf{P}_{\text{e,max}}^{(n)}(R) = 0 \text{ for some } n\}.$$

Example:  $\implies$ $C_0 = 1$ [bit/ch. use]

- **Reliability function**:

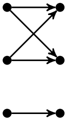
$$E(R) = \limsup_{n\to\infty} -\frac{1}{n} \log \mathsf{P}_{\text{e,max}}^{(n)}(R)$$

that is,

$$\mathsf{P}_{\text{e,max}}^{(n)}(R) \approx e^{-nE(R)} \qquad C_0 < R < C$$

- **Zero-error capacity**

$$C_0 = \sup\{R : \mathsf{P}^{(n)}_{\mathrm{e,max}}(R) = 0 \text{ for some } n\}.$$

Example:  $\Longrightarrow$ $C_0 = 1$ [bit/ch. use]

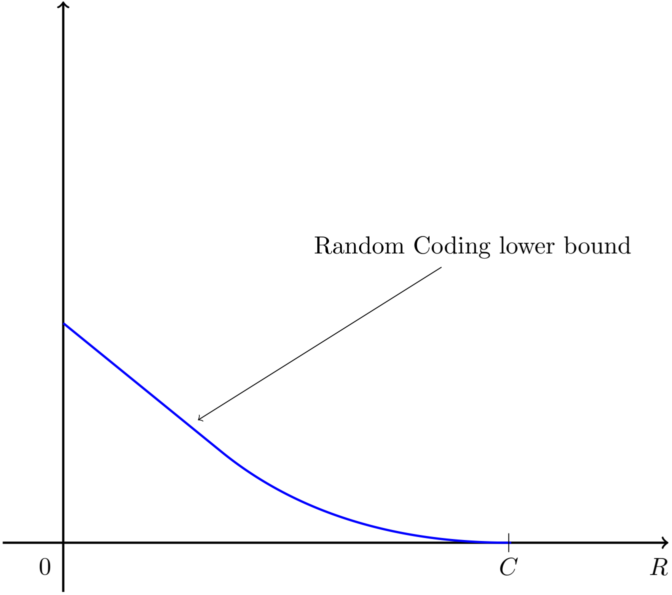
- **Reliability function**:

$$E(R) = \limsup_{n\to\infty} -\frac{1}{n} \log \mathsf{P}^{(n)}_{\mathrm{e,max}}(R)$$

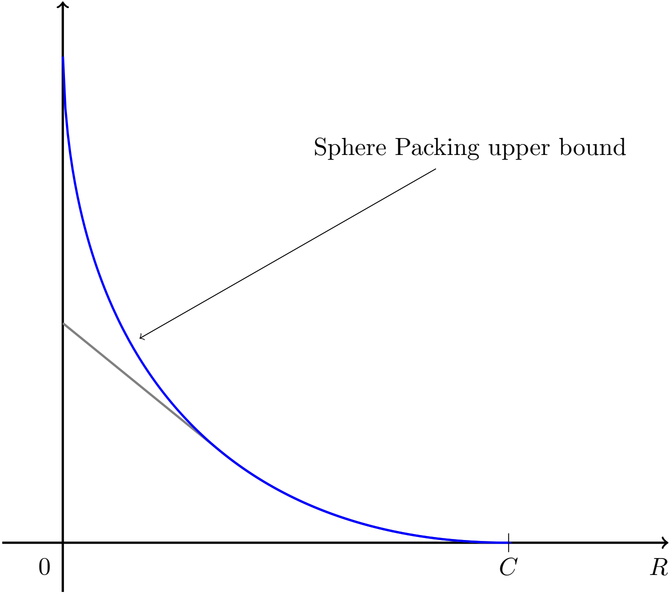
that is,

$$\mathsf{P}^{(n)}_{\mathrm{e,max}}(R) \approx e^{-nE(R)} \qquad C_0 < R < C$$

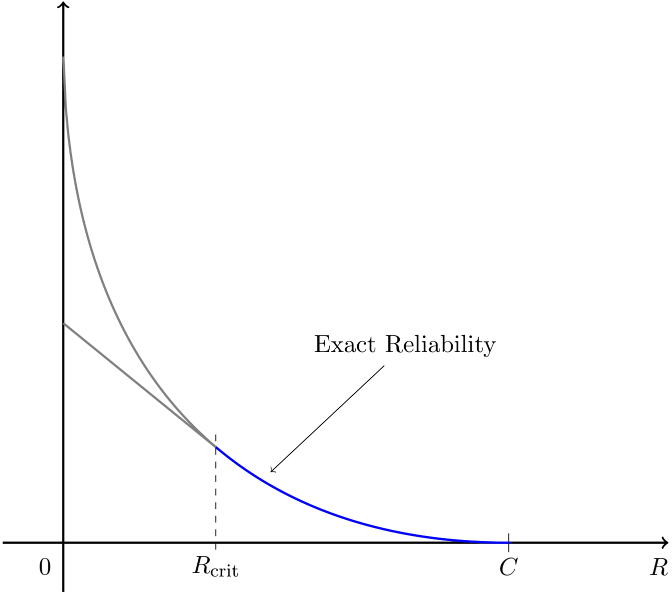
Bounds on $E(R)$: typical case with $C_0 = 0$



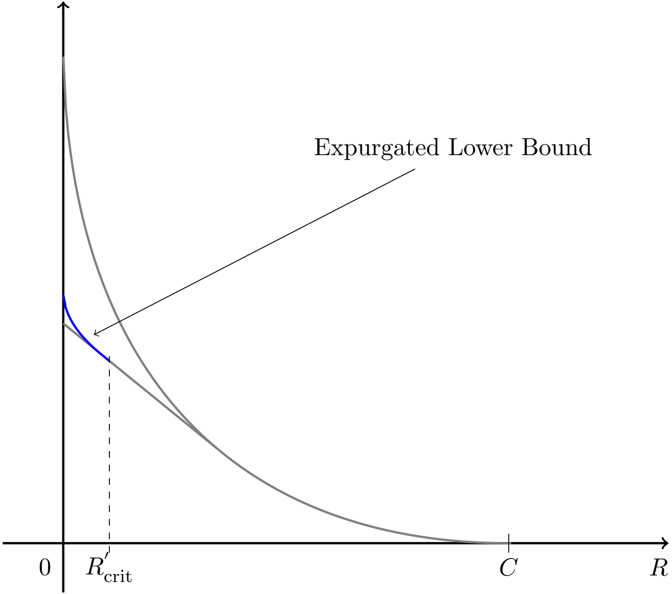
Bounds on $E(R)$: typical case with $C_0 = 0$



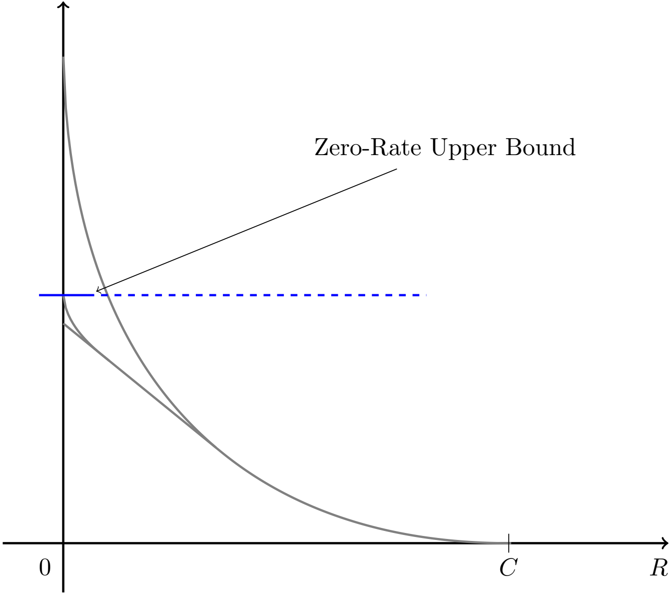
Bounds on $E(R)$: typical case with $C_0 = 0$



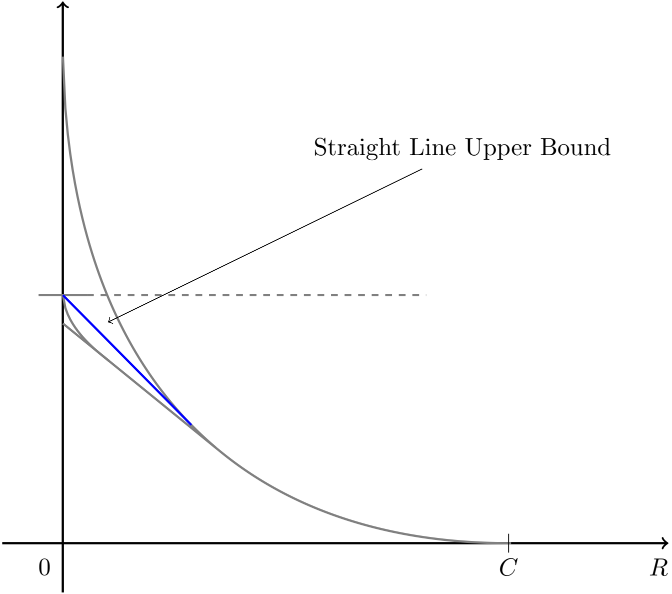
Bounds on $E(R)$: typical case with $C_0 = 0$



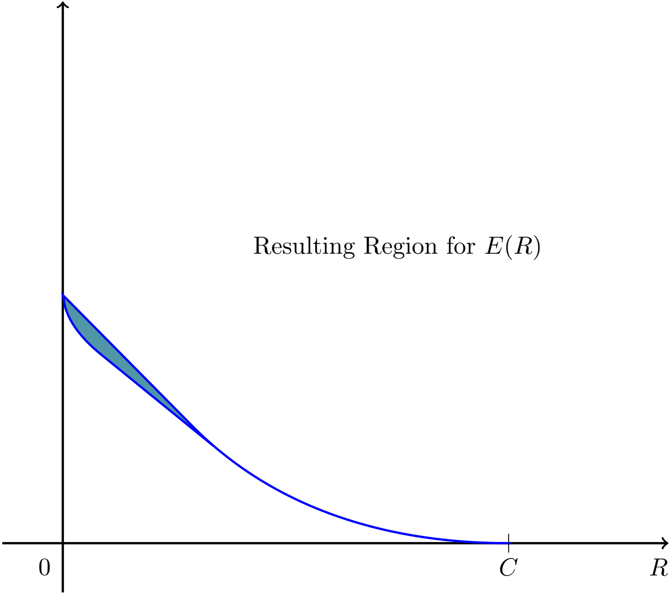
Bounds on $E(R)$: typical case with $C_0 = 0$



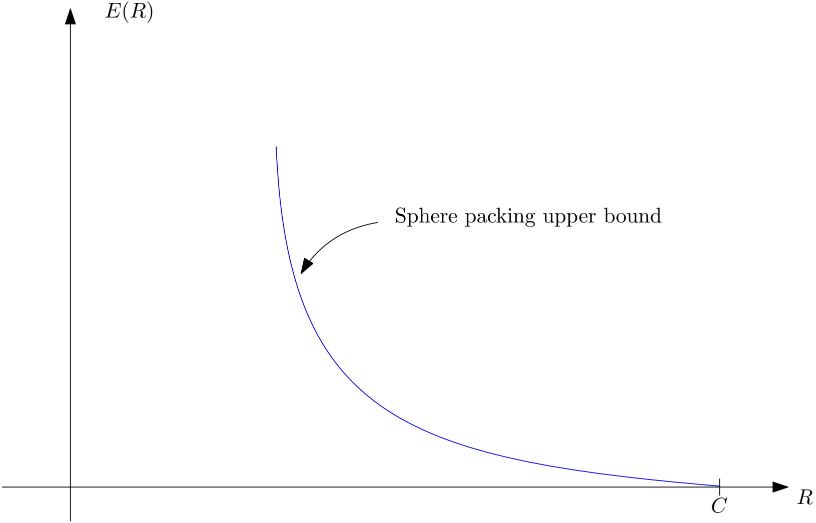
Bounds on $E(R)$: typical case with $C_0 = 0$



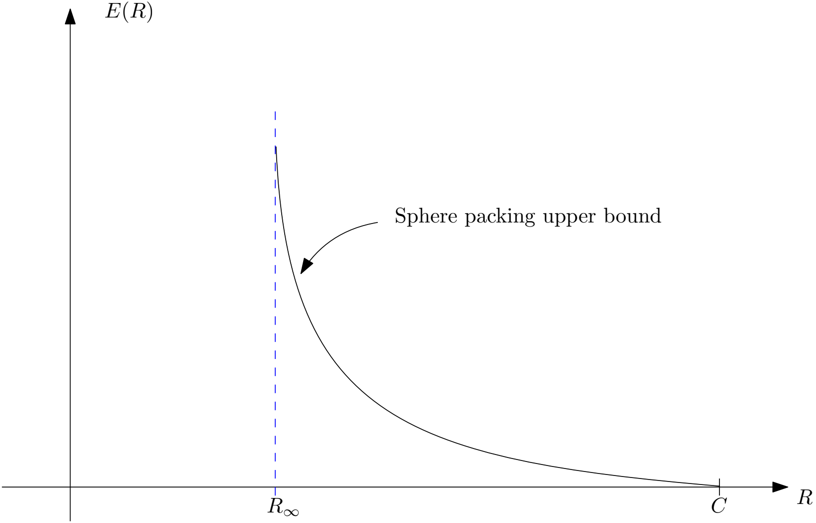
Bounds on $E(R)$: typical case with $C_0 = 0$



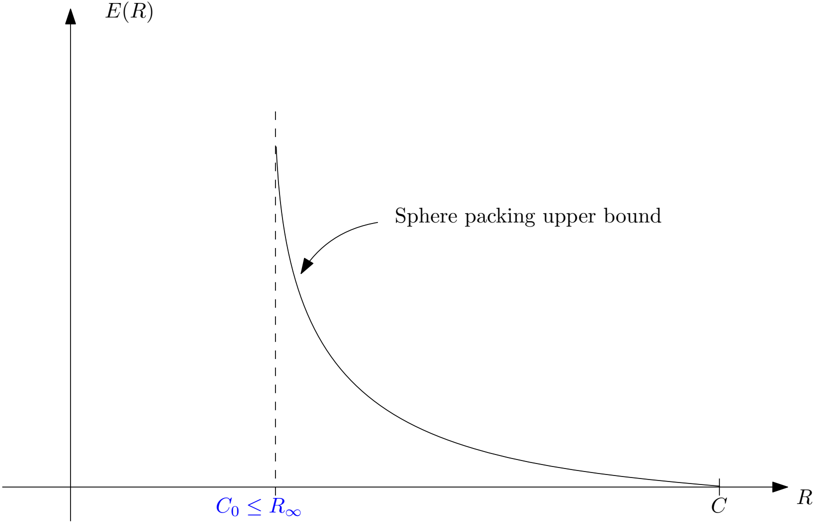
General case: $C_0 > 0$



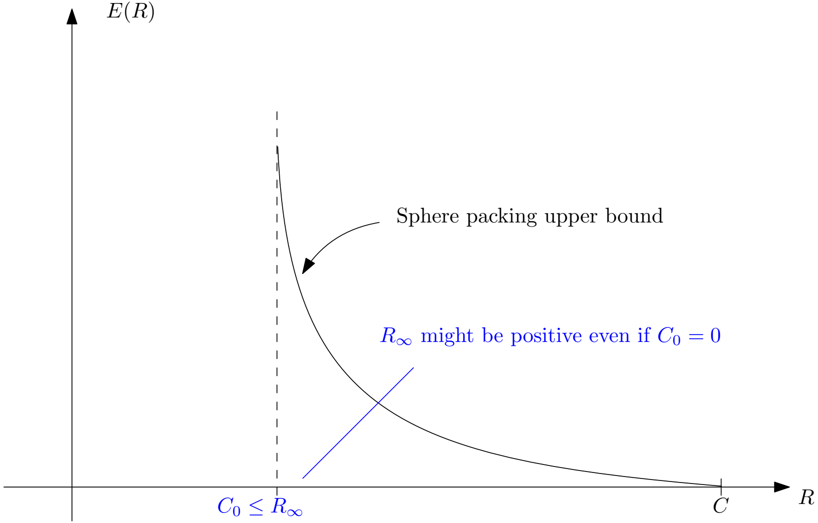
General case: $C_0 > 0$



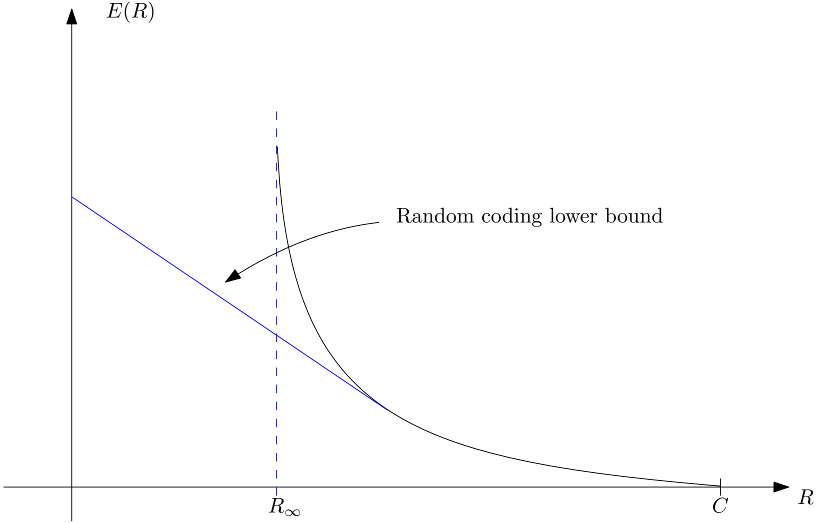
General case: $C_0 > 0$



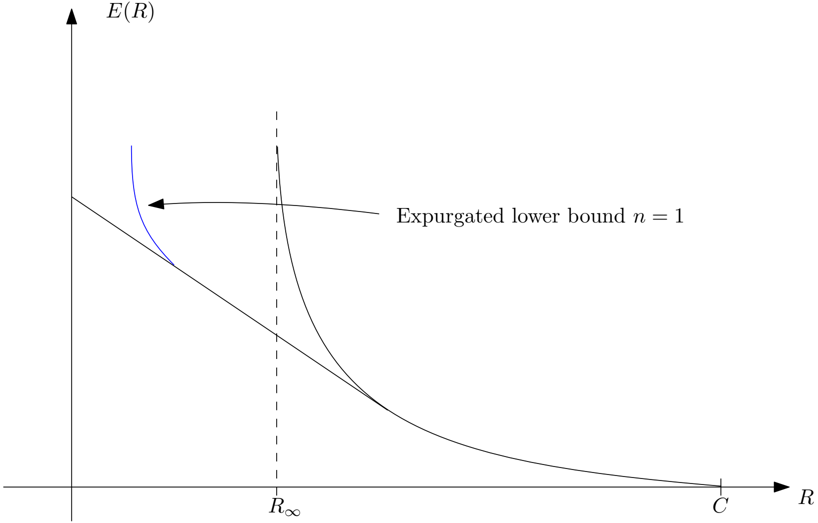
General case: $C_0 > 0$



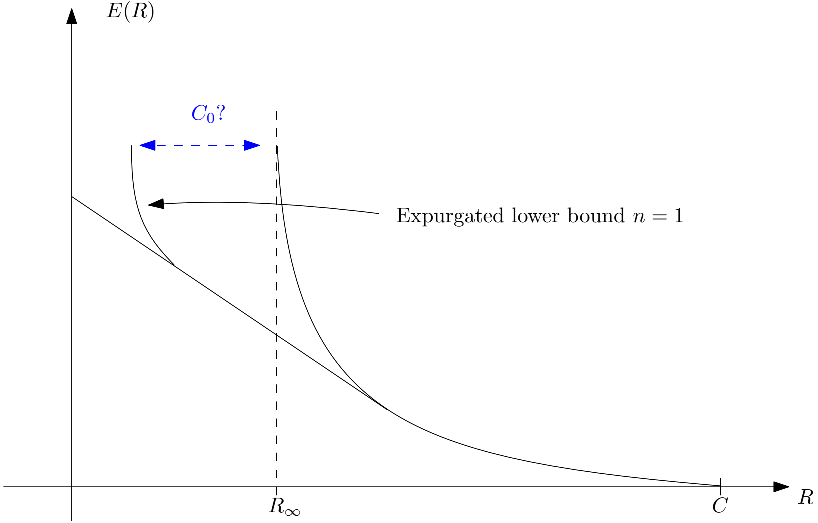
General case: $C_0 > 0$



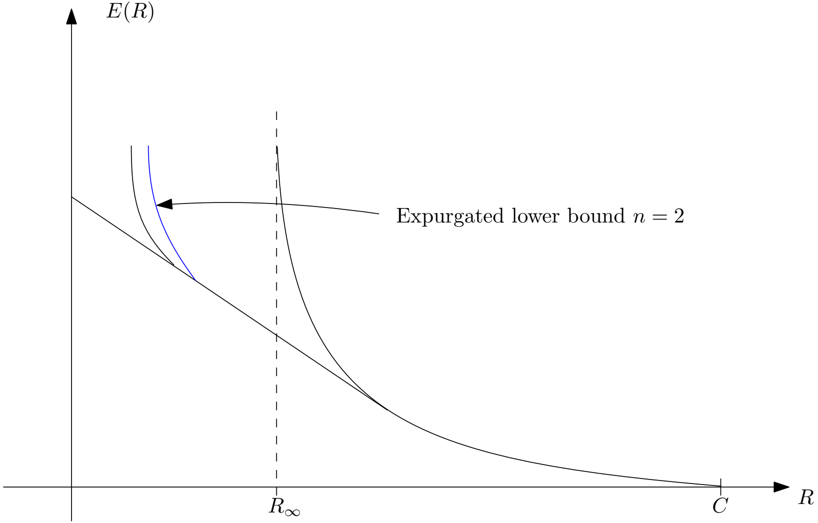
General case: $C_0 > 0$



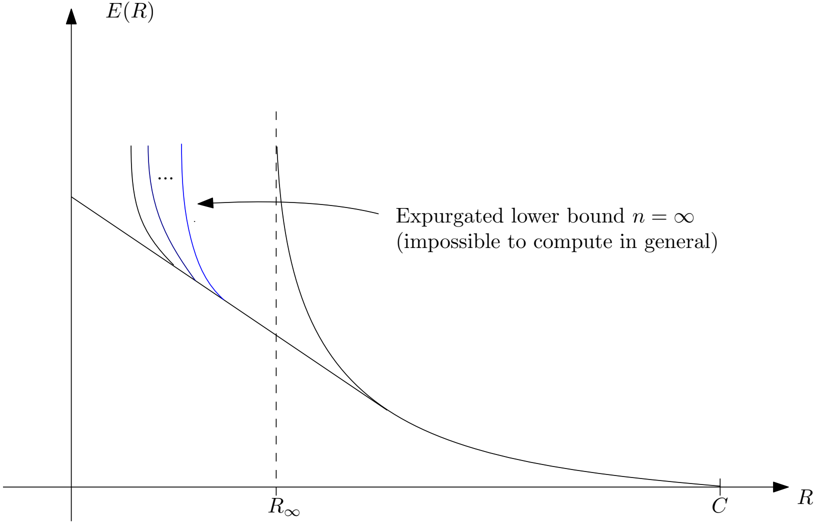
General case: $C_0 > 0$



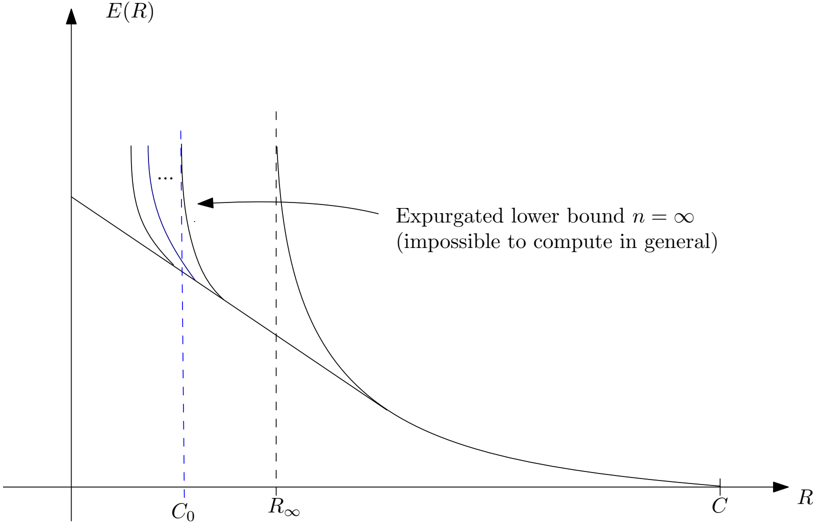
General case: $C_0 > 0$



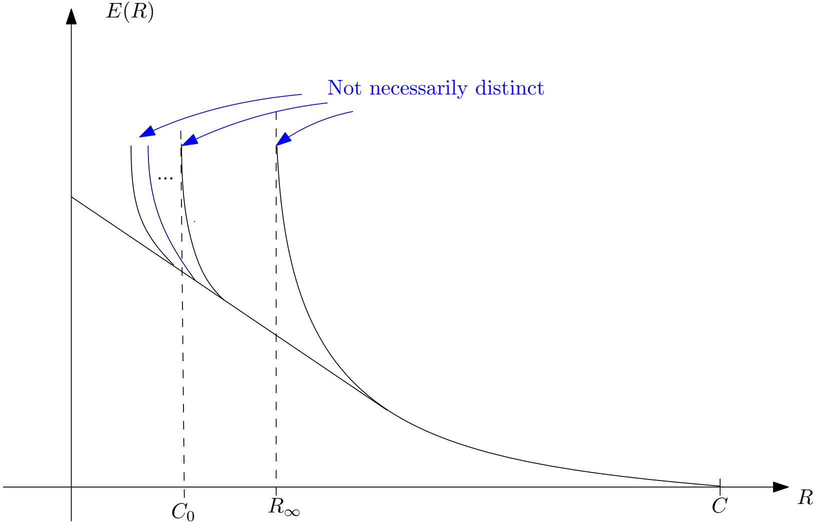
General case: $C_0 > 0$



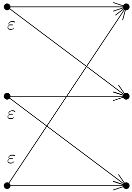
General case: $C_0 > 0$

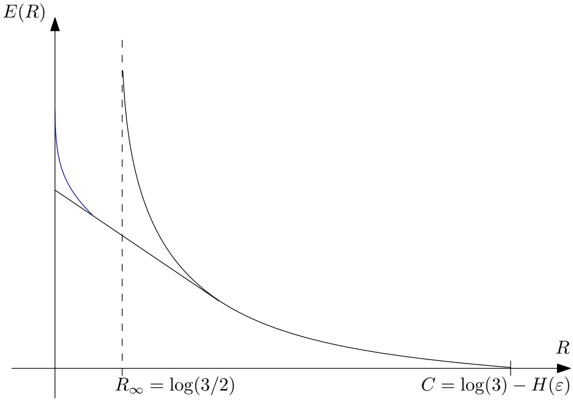


General case: $C_0 > 0$

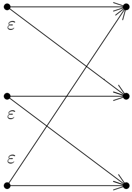


Example: typewriter channels

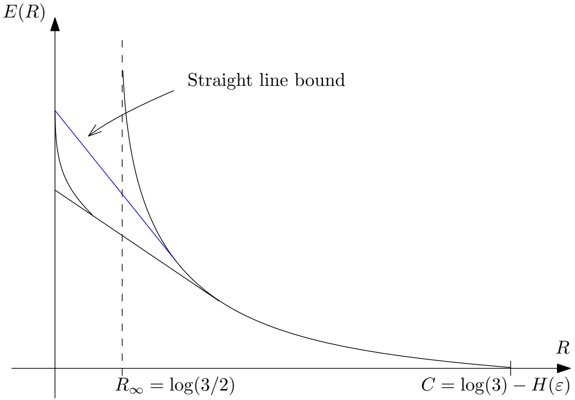




Note: $C_0 = 0$



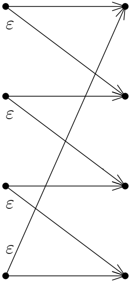
ε
ε
ε

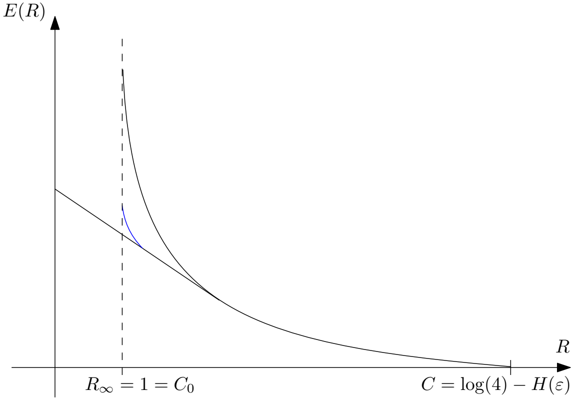


E(R)
Straight line bound
R
$R_\infty = \log(3/2)$
$C = \log(3) - H(\varepsilon)$

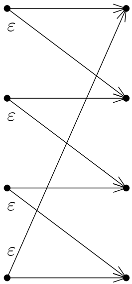
Note: $C_0 = 0$

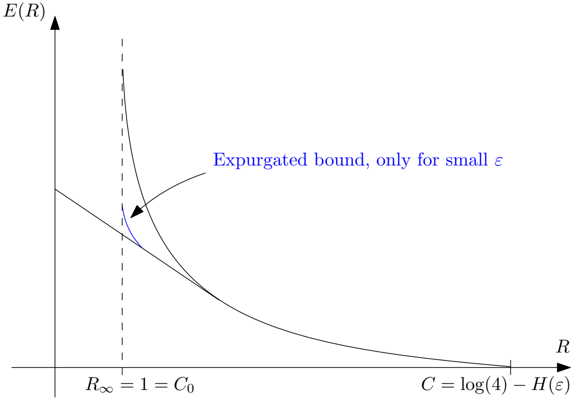
Example: typewriter channels

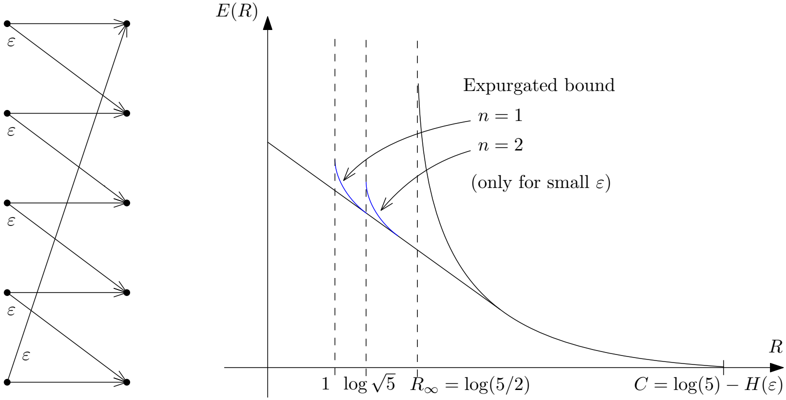




Example: typewriter channels

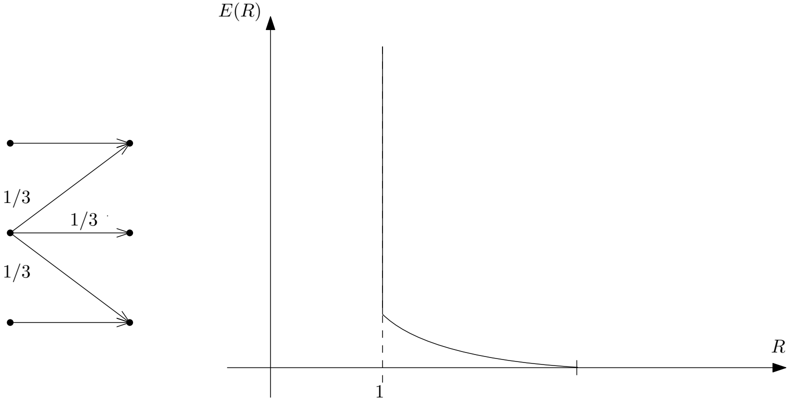






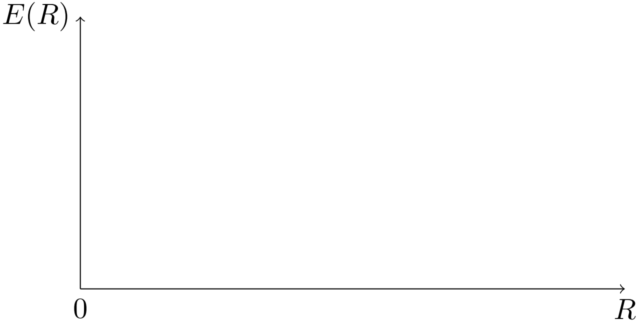
Note: $C_0 = \log\sqrt{5}$ (see later)

Example: strange case



Note: exact reliability!

Some More Facts



Shannon, 1948

$$C = \max_P \sum_{x,y} P(x)W_x(y) \log \frac{W_x(y)}{\sum_{x'} P(x')W_{x'}(y)},$$

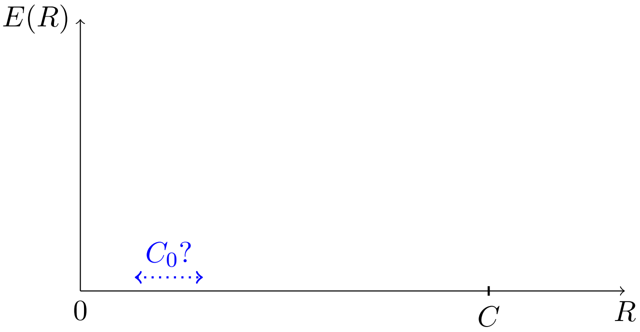
Some More Facts

...later noticed to be an **information radius**

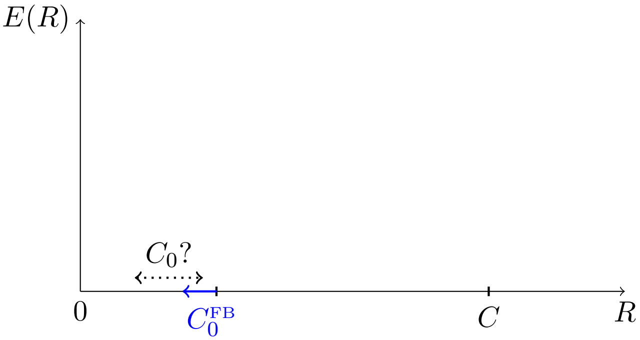
$$C = \min_Q \max_x D(W_x||Q)$$

where $D(Q_1||Q_2) = \sum_y Q_1(y) \log \frac{Q_1(y)}{Q_2(y)}$ is the Kullback-Leibler divergence

Some More Facts

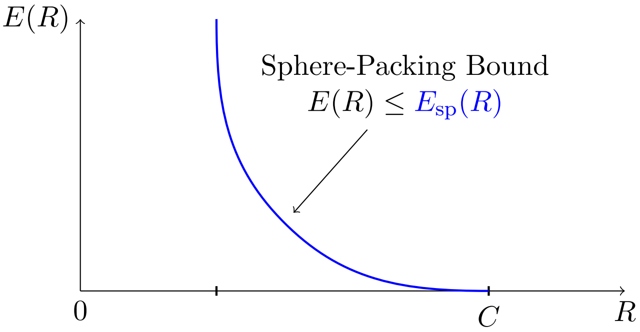


Shannon, 1956



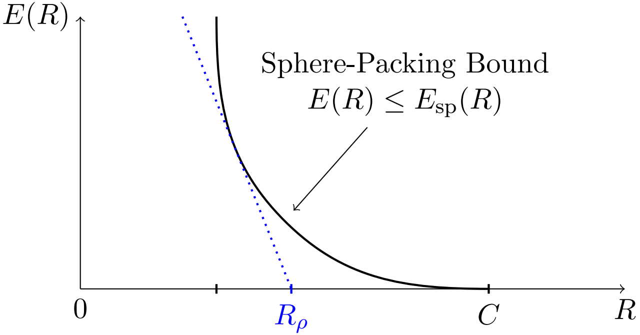
Shannon, 1956 (combinatorial)
Upper bounded by the zero-error capacity with feedback

$$C_0^{\mathrm{FB}} = \max_P \left[-\log \max_y \sum_{x: W_x(y)>0} P(x) \right]$$



Fano, 1961 - Shannon, Gallager and Berlekamp, 1967
(probabilistic)

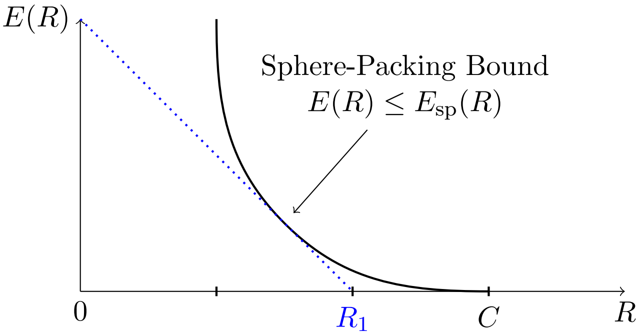
$$E_{\mathrm{sp}}(R) = \sup_{\rho \geq 0} \max_P \left[-\log \sum_y \left(\sum_x P(x) W_x(y)^{1/(1+\rho)} \right)^{1+\rho} - \rho R \right]$$



Also

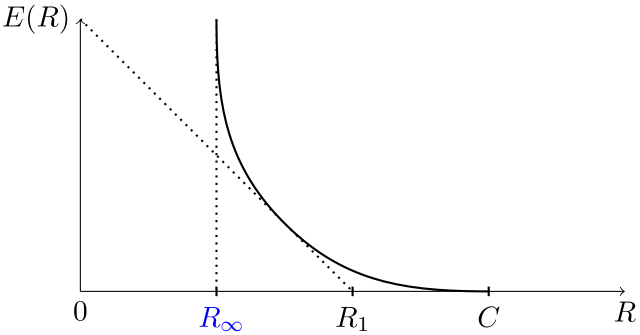
$$R_\rho = \min_Q \max_x D_\alpha(W_x || Q), \quad \alpha = 1/(1+\rho)$$

$D_\alpha(Q_1||Q_2) = \frac{1}{\alpha-1} \log \sum_y Q_1(y)^\alpha Q_2(y)^{1-\alpha}$ is the Rényi divergence



Cutoff rate

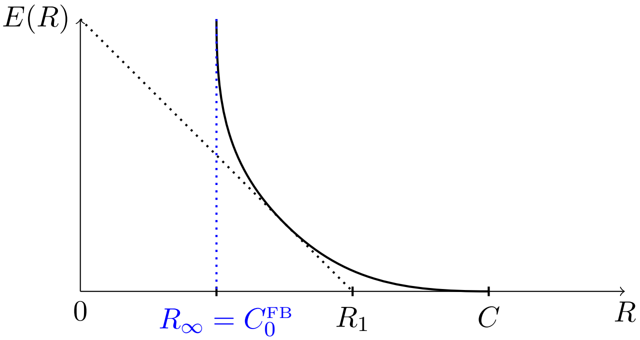
$$R_1 = \min_Q \max_x \log \frac{1}{\left(\sum_y \sqrt{W_x(y)Q(y)}\right)^2}$$



R_∞ rate

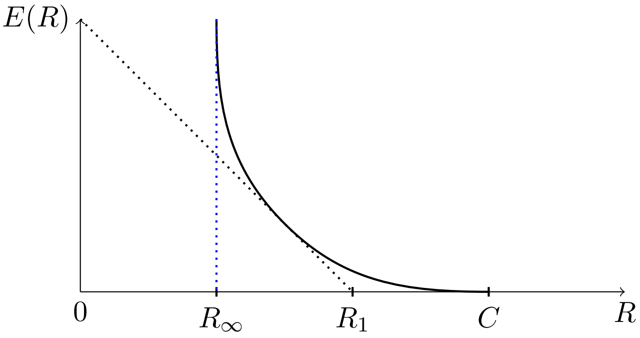
$$R_\infty = \min_Q \max_x \log \frac{1}{\sum_{y:W_x(y)>0} Q(y)}$$

Some More Facts

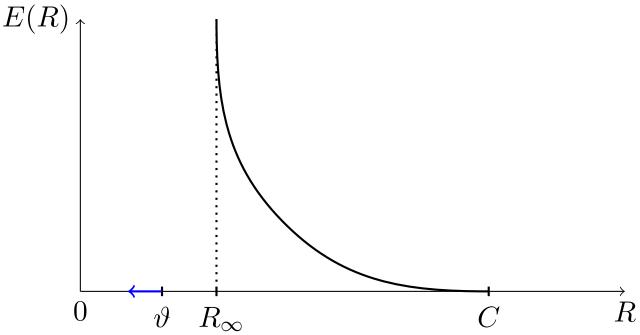


- $E_{\mathrm{sp}}(R)$ gives $C_0 \leq R_\infty$
- So we have both $C_0 \leq C_0^{\mathrm{FB}}$ and $C_0 \leq R_\infty$

Some More Facts



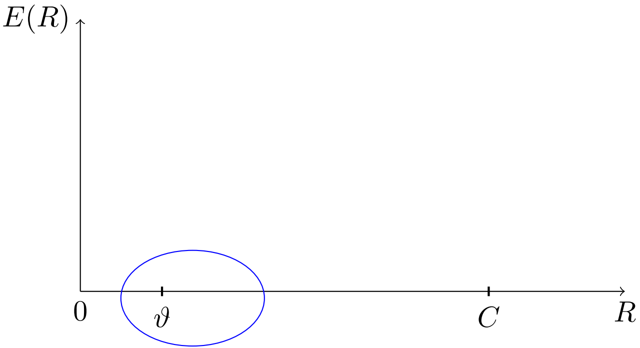
- $E_{\mathrm{sp}}(R)$ gives $C_0 \le R_\infty$
- So we have both $C_0 \le C_0^{\mathrm{FB}}$ and $C_0 \le R_\infty$
- It turns out that $R_\infty = C_0^{\mathrm{FB}}$ (whenever $C_0 > 0$)
- Same bound for C_0 using combinatorial or probabilistic approaches
- We can then minimize R_∞ over auxiliary channels $\tilde{W}$



Lovász, 1979

- New bound: $C_0 \leq \vartheta$
- Using *geometric representations of graphs*
- Combinatorial, apparently no connection with probability

Some More Facts



Lovász, 1979

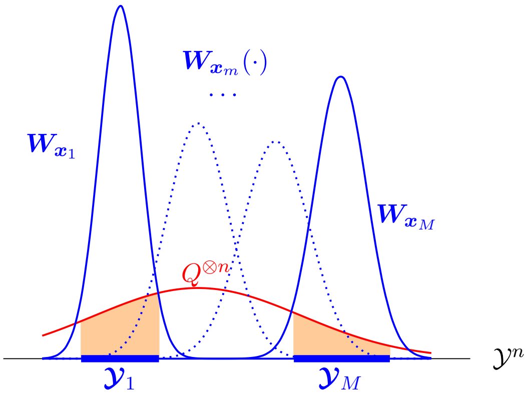
- New bound: $C_0 \leq \vartheta$
- Using *geometric representations of graphs*
- Combinatorial, apparently no connection with probability
- **Goal**: better understanding of the R_∞ vs ϑ

Sphere-Packing Bound: Sketch of Proof

Binary hypothesis testing: compare $Q^{\otimes n}$ with $\boldsymbol{W}_{\boldsymbol{x}_m}$

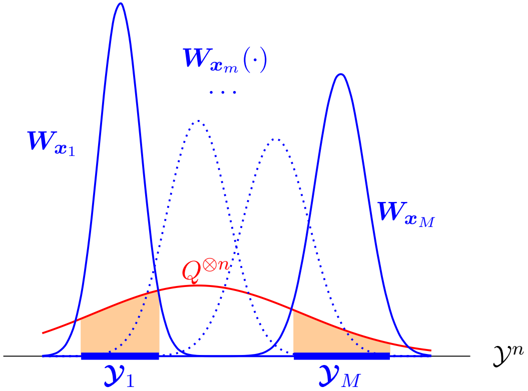
- The decision regions $\boldsymbol{\mathcal{Y}}_1, \ldots, \boldsymbol{\mathcal{Y}}_M$ are disjoint
- $Q^{\otimes n}(\boldsymbol{\mathcal{Y}}_m) \leq 1/M$ for at least one m, since $\int Q^{\otimes n} = 1$
- $\boldsymbol{W}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m}) \geq e^{-n(E_{\text{sp}}(R)+o(1))}$ using Neyman-Pearson/Chernoff

Binary hypothesis testing: compare $Q^{\otimes n}$ with $\boldsymbol{W}_{\boldsymbol{x}_m}$



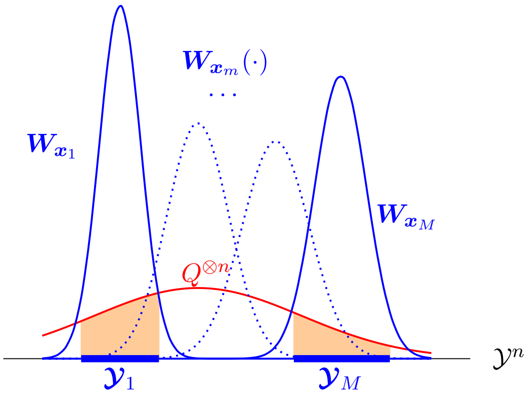
- The decision regions $\boldsymbol{\mathcal{Y}}_1, \ldots, \boldsymbol{\mathcal{Y}}_M$ are disjoint
- $Q^{\otimes n}(\boldsymbol{\mathcal{Y}}_m) \leq 1/M$ for at least one m, since $\int Q^{\otimes n} = 1$
- $\boldsymbol{W}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m}) \geq e^{-n(E_{\text{sp}}(R)+o(1))}$ using Neyman-Pearson/Chernoff

Binary hypothesis testing: compare $Q^{\otimes n}$ with $\boldsymbol{W}_{\boldsymbol{x}_m}$



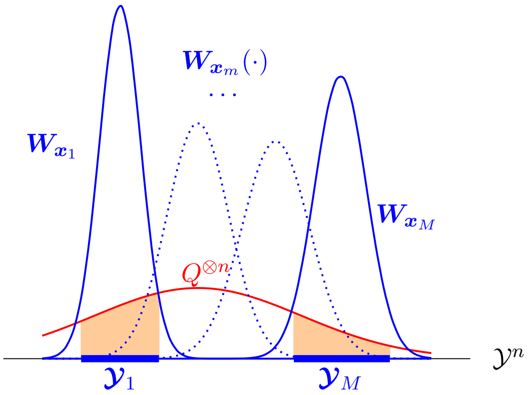
- The decision regions $\boldsymbol{\mathcal{Y}}_1, \ldots, \boldsymbol{\mathcal{Y}}_M$ are disjoint
- $Q^{\otimes n}(\boldsymbol{\mathcal{Y}}_m) \leq 1/M$ for at least one m, since $\int Q^{\otimes n} = 1$
- $\boldsymbol{W}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m}) \geq e^{-n(E_{\text{sp}}(R)+o(1))}$ using Neyman-Pearson/Chernoff

Binary hypothesis testing: compare $Q^{\otimes n}$ with $\boldsymbol{W}_{\boldsymbol{x}_m}$



- The decision regions $\boldsymbol{\mathcal{Y}}_1, \ldots, \boldsymbol{\mathcal{Y}}_M$ are disjoint
- $Q^{\otimes n}(\boldsymbol{\mathcal{Y}}_m) \leq 1/M$ for at least one m, since $\int Q^{\otimes n} = 1$
- $\boldsymbol{W}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m}) \geq e^{-n(E_{\text{sp}}(R)+o(1))}$ using Neyman-Pearson/Chernoff

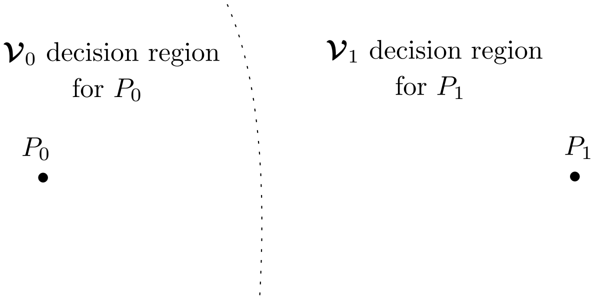
Binary hypothesis testing: compare $Q^{\otimes n}$ with $\boldsymbol{W}_{\boldsymbol{x}_m}$



- The decision regions $\boldsymbol{\mathcal{Y}}_1, \ldots, \boldsymbol{\mathcal{Y}}_M$ are disjoint
- $Q^{\otimes n}(\boldsymbol{\mathcal{Y}}_m) \leq 1/M$ for at least one m, since $\int Q^{\otimes n} = 1$
- $\boldsymbol{W}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m}) \geq e^{-n(E_{\text{sp}}(R)+o(1))}$ using Neyman-Pearson/Chernoff

BHT between distributions P_0 and P_1 over $\mathcal{V}$ from n i.i.d. samples

- Two decision regions



- Error probabilities

$$\mathsf{P}_{\mathrm{e}|0} = \sum_{v \in \boldsymbol{\mathcal{V}}_1} P_0(v)\,, \qquad \mathsf{P}_{\mathrm{e}|1} = \sum_{v \in \boldsymbol{\mathcal{V}}_0} P_1(v)$$

Binary Hypothesis Testing (Rényi form)

Error exponents in BHT between P_0 and P_1 with n i.i.d. samples

$$\frac{1}{n}\log \mathsf{P}_{\mathrm{e}|0} = \mu(s) - s\mu'(s) + o(1)$$
$$\frac{1}{n}\log \mathsf{P}_{\mathrm{e}|1} = \mu(s) + (1-s)\mu'(s) + o(1)$$

where $0 < s < 1$,

$$\mu(s) = \log \sum_{v\in\mathcal{V}} P_0(v)^{1-s} P_1(v)^s$$
$$= -sD_{1-s}(P_0\|P_1)$$

and $D_\alpha(P\|Q)$ is the Rényi divergence

$$D_\alpha(P\|Q) = \frac{1}{\alpha-1}\log\sum_{v\in V} P^\alpha(v)Q^{1-\alpha}(v)$$

Note:

$$\lim_{\alpha\to 1} D_\alpha(P\|Q) = \sum_{v\in V} P(v)\log\frac{P(v)}{Q(v)} =: D_{\mathrm{KL}}(P\|Q)$$

Error exponents in BHT between P_0 and P_1 with n i.i.d. samples

$$\frac{1}{n}\log \mathsf{P}_{e|0} = \mu(s) - s\mu'(s) + o(1)$$
$$\frac{1}{n}\log \mathsf{P}_{e|1} = \mu(s) + (1-s)\mu'(s) + o(1)$$

where $0 < s < 1$,

$$\mu(s) = \log \sum_{v\in\mathcal{V}} P_0(v)^{1-s} P_1(v)^s$$
$$= -sD_{1-s}(P_0\|P_1)$$

and $D_\alpha(P\|Q)$ is the Rényi divergence

$$D_\alpha(P\|Q) = \frac{1}{\alpha-1}\log\sum_{v\in V} P^\alpha(v) Q^{1-\alpha}(v)$$

Note:

$$\lim_{\alpha\to 1} D_\alpha(P\|Q) = \sum_{v\in V} P(v)\log\frac{P(v)}{Q(v)} =: D_{\mathrm{KL}}(P\|Q)$$

Binary Hypothesis Testing (Rényi form)

Error exponents in BHT between P_0 and P_1 with n i.i.d. samples

$$\frac{1}{n}\log \mathsf{P}_{\mathrm{e}|0} = \mu(s) - s\mu'(s) + o(1)$$
$$\frac{1}{n}\log \mathsf{P}_{\mathrm{e}|1} = \mu(s) + (1-s)\mu'(s) + o(1)$$

where $0 < s < 1$,

$$\mu(s) = \log \sum_{v\in\mathcal{V}} P_0(v)^{1-s} P_1(v)^s$$
$$= -sD_{1-s}(P_0\|P_1)$$

and $D_\alpha(P\|Q)$ is the Rényi divergence

$$D_\alpha(P\|Q) = \frac{1}{\alpha-1}\log\sum_{v\in V} P^\alpha(v) Q^{1-\alpha}(v)$$

Note:

$$\lim_{\alpha\to 1} D_\alpha(P\|Q) = \sum_{v\in V} P(v)\log\frac{P(v)}{Q(v)} =: D_{\mathrm{KL}}(P\|Q)$$

Interpretation: Shannon-Gallager-Berlekamp, 1967

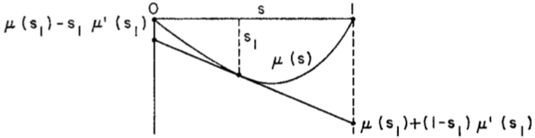
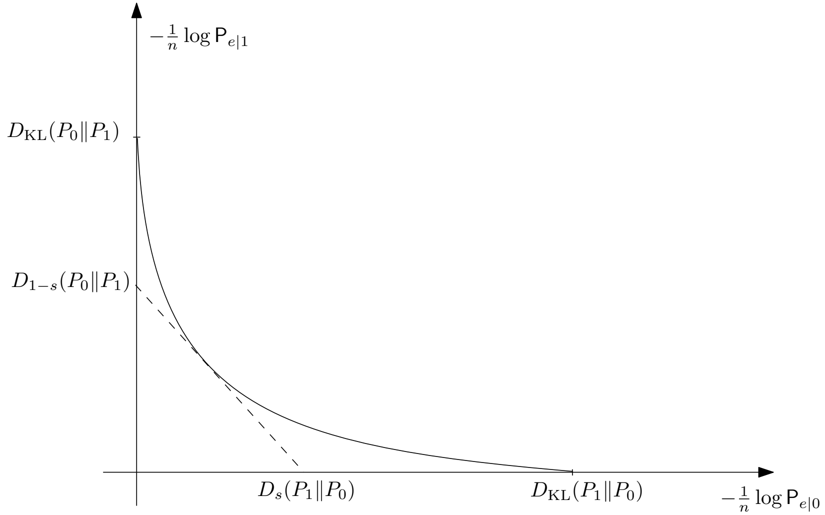


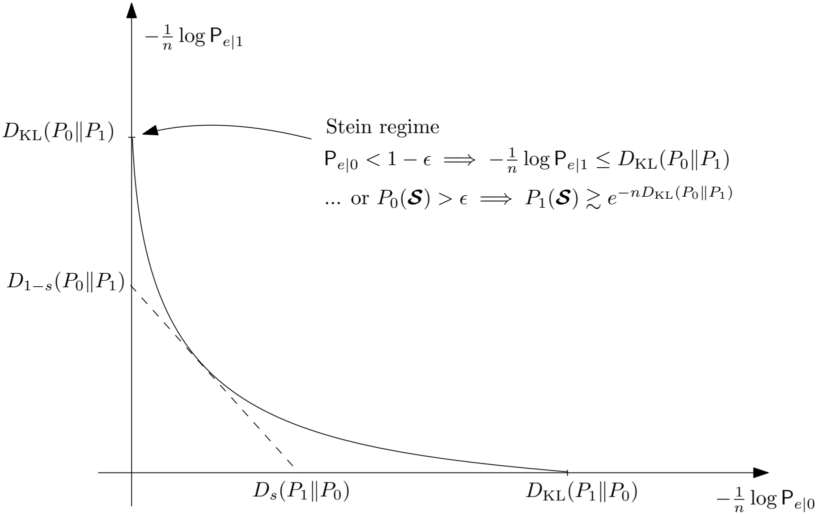
FIG. 6. Geometric interpretation of the exponents $\mu(s) - s\mu'(s)$ and $\mu(s) + (1 - s)\mu'(s)$.

Figure 6 gives a graphical interpretation of the terms $\mu(s) - \mu'(s)$ and $\mu(s) + (1 - s)\mu'(s)$. It is seen that they are the endpoints, at 0 and 1, of the tangent at s to the curve $\mu(s)$. As s increases, the tangent see-saws around, decreasing $\mu(s) - s\mu'(s)$ and increasing $\mu(s) + (1 - s)\mu'(s)$. In the special case where $\mu(s)$ is a straight line, of course, this see-sawing does not occur and $\mu(s) - s\mu'(s)$ and $\mu(s) + (1 - s)\mu'(s)$ do not vary with s.

Another graphical representation

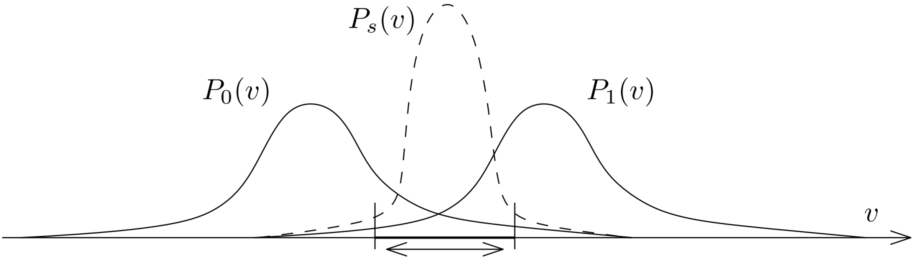


Another graphical representation



Key role played by the tilted mixture P_s

$$P_s(v) = \frac{P_0(v)^{1-s}P_1(v)^s}{\sum_{v'} P_0(v')^{1-s}P_1(v')^s} \implies \frac{P_0(v)}{P_s(v)} = e^{\mu(s)} e^{-s \log \frac{P_1(v)}{P_0(v)}}.$$



High probability under P_s

$$\frac{1}{n} \log \frac{P_1(\boldsymbol{v})}{P_0(\boldsymbol{v})} \approx \mu'(s) = \mathsf{E}_{P_s}\left[\log \frac{P_1(V)}{P_0(V)}\right]$$

Alternative expression (more popular)

$$-\frac{1}{n}\log \mathsf{P}_{\mathrm{e}|0} = D_{\mathrm{KL}}(P_s \| P_0) + o(1)$$
$$-\frac{1}{n}\log \mathsf{P}_{\mathrm{e}|1} = D_{\mathrm{KL}}(P_s \| P_1) + o(1)$$

- Very simple and intuitive: probabilities that P_0 and P_1 generate P_s-like sequences
- Directly uses the Stein regime in P_0 vs P_s and P_1 vs P_s
- Note (for later): this does *not* work in the quantum setting

Standard procedure for DMCs

- Given code with $M = e^{nR}$ codewords
- Group codewords by empirical "compositions" (or "type", empirical frequency of symbols in the codeword)
- At most $n^{|\mathcal{X}|} = e^{o(n)}$ groups
- At least one group contains $e^{n(R-o(1))}$ codewords
- Bound probability of error for this subcode
- So, we can assume all codewords have same composition, say P

Constant composition

Standard procedure for DMCs

- Given code with $M = e^{nR}$ codewords
- Group codewords by empirical "compositions" (or "type", empirical frequency of symbols in the codeword)
- At most $n^{|\mathcal{X}|} = e^{o(n)}$ groups
- At least one group contains $e^{n(R-o(1))}$ codewords
- Bound probability of error for this subcode
- So, we can assume all codewords have same composition, say P

Standard procedure for DMCs

- Given code with $M = e^{nR}$ codewords
- Group codewords by empirical “compositions” (or “type”, empirical frequency of symbols in the codeword)
- At most $n^{|\mathcal{X}|} = e^{o(n)}$ groups
- At least one group contains $e^{n(R-o(1))}$ codewords
- Bound probability of error for this subcode
- So, we can assume all codewords have same composition, say P

Constant composition

Standard procedure for DMCs

- Given code with $M = e^{nR}$ codewords
- Group codewords by empirical "compositions" (or "type", empirical frequency of symbols in the codeword)
- At most $n^{|\mathcal{X}|} = e^{o(n)}$ groups
- At least one group contains $e^{n(R-o(1))}$ codewords
- Bound probability of error for this subcode
- So, we can assume all codewords have same composition, say P

Constant composition

Standard procedure for DMCs

- Given code with $M = e^{nR}$ codewords
- Group codewords by empirical "compositions" (or "type", empirical frequency of symbols in the codeword)
- At most $n^{|\mathcal{X}|} = e^{o(n)}$ groups
- At least one group contains $e^{n(R-o(1))}$ codewords
- Bound probability of error for this subcode
- So, we can assume all codewords have same composition, say P

Constant composition

Standard procedure for DMCs

- Given code with $M = e^{nR}$ codewords
- Group codewords by empirical "compositions" (or "type", empirical frequency of symbols in the codeword)
- At most $n^{|\mathcal{X}|} = e^{o(n)}$ groups
- At least one group contains $e^{n(R-o(1))}$ codewords
- Bound probability of error for this subcode
- So, we can assume all codewords have same composition, say P

Back to sphere packing: MIT Proof

- BHT between output distribution $\boldsymbol{W}_{\boldsymbol{x}_m}$ and auxiliary $\boldsymbol{Q} = Q^{\otimes n}$
- Use $\mathcal{Y}_m$ as decision region for $\boldsymbol{W}_{\boldsymbol{x}_m}$
- $M = e^{nR}$ codewords; for at least one m, $\boldsymbol{Q}(\mathcal{Y}_m) \leq 1/M$ and so

$$-\frac{1}{n}\log \mathsf{P}_{\mathrm{e}|\boldsymbol{Q}} \geq R$$

- But for the optimal test

$$-\frac{1}{n}\log \mathsf{P}_{\mathrm{e}|\boldsymbol{W}_{\boldsymbol{x}_m}} = -\mu(s) + s\mu'(s) + o(1)$$

$$-\frac{1}{n}\log \mathsf{P}_{\mathrm{e}|\boldsymbol{Q}} = -\mu(s) - (1-s)\mu'(s) + o(1)$$

where

$$\mu(s) = \sum_x P(x)\left[\log \sum_{y\in\mathcal{Y}} W_x(y)^{1-s}Q(y)^s\right].$$

Back to sphere packing: MIT Proof

- BHT between output distribution $\boldsymbol{W}_{\boldsymbol{x}_m}$ and auxiliary $\boldsymbol{Q} = Q^{\otimes n}$
- Use $\boldsymbol{\mathcal{Y}}_m$ as decision region for $\boldsymbol{W}_{\boldsymbol{x}_m}$
- $M = e^{nR}$ codewords; for at least one m, $\boldsymbol{Q}(\boldsymbol{\mathcal{Y}}_m) \leq 1/M$ and so

$$-\frac{1}{n} \log \mathsf{P}_{\mathrm{e}|\boldsymbol{Q}} \geq R$$

- But for the optimal test

$$-\frac{1}{n} \log \mathsf{P}_{\mathrm{e}|\boldsymbol{W}_{\boldsymbol{x}_m}} = -\mu(s) + s\mu'(s) + o(1)$$

$$-\frac{1}{n} \log \mathsf{P}_{\mathrm{e}|\boldsymbol{Q}} = -\mu(s) - (1-s)\mu'(s) + o(1)$$

where

$$\mu(s) = \sum_x P(x) \left[\log \sum_{y \in \mathcal{Y}} W_x(y)^{1-s} Q(y)^s \right].$$

Back to sphere packing: MIT Proof

- BHT between output distribution $\boldsymbol{W}_{\boldsymbol{x}_m}$ and auxiliary $\boldsymbol{Q} = Q^{\otimes n}$
- Use $\boldsymbol{\mathcal{Y}}_m$ as decision region for $\boldsymbol{W}_{\boldsymbol{x}_m}$
- $M = e^{nR}$ codewords; for at least one m, $\boldsymbol{Q}(\boldsymbol{\mathcal{Y}}_m) \leq 1/M$ and so

$$-\frac{1}{n}\log \mathsf{P}_{\mathrm{e}|\boldsymbol{Q}} \geq R$$

- But for the optimal test

$$-\frac{1}{n}\log \mathsf{P}_{\mathrm{e}|\boldsymbol{W}_{\boldsymbol{x}_m}} = -\mu(s) + s\mu'(s) + o(1)$$
$$-\frac{1}{n}\log \mathsf{P}_{\mathrm{e}|\boldsymbol{Q}} = -\mu(s) - (1-s)\mu'(s) + o(1)$$

where

$$\mu(s) = \sum_x P(x)\left[\log \sum_{y\in\mathcal{Y}} W_x(y)^{1-s} Q(y)^s\right].$$

Back to sphere packing: MIT Proof

- BHT between output distribution $\boldsymbol{W}_{\boldsymbol{x}_m}$ and auxiliary $\boldsymbol{Q} = Q^{\otimes n}$
- Use $\boldsymbol{\mathcal{Y}}_m$ as decision region for $\boldsymbol{W}_{\boldsymbol{x}_m}$
- $M = e^{nR}$ codewords; for at least one m, $\boldsymbol{Q}(\boldsymbol{\mathcal{Y}}_m) \leq 1/M$ and so

$$-\frac{1}{n}\log \mathsf{P}_{\mathrm{e}|\boldsymbol{Q}} \geq R$$

- But for the optimal test

$$-\frac{1}{n}\log \mathsf{P}_{\mathrm{e}|\boldsymbol{W}_{\boldsymbol{x}_m}} = -\mu(s) + s\mu'(s) + o(1)$$
$$-\frac{1}{n}\log \mathsf{P}_{\mathrm{e}|\boldsymbol{Q}} = -\mu(s) - (1-s)\mu'(s) + o(1)$$

where

$$\mu(s) = \sum_x P(x)\left[\log \sum_{y\in\mathcal{Y}} W_x(y)^{1-s}Q(y)^s\right].$$

So,

$$-\frac{1}{n}\log \mathsf{P}_{\mathrm{e}|\boldsymbol{W}_{\boldsymbol{x}_m}} \leq \sup_{0<s<1}\left[E_0(s,P) - \frac{s}{1-s}(R-\epsilon)\right] + o(1)$$

where

$$\begin{aligned} E_0(s,P) &= \min_Q\left[\frac{1}{s-1}\sum_x P(x)\log\sum_y W_x(y)^{1-s}Q(y)^s\right] \\ &= \min_Q\left[\frac{s}{1-s}\sum_x P(x)D_{1-s}(W_x\|Q)\right] \\ &= \frac{s}{1-s}I_{1-s}(P,W), \end{aligned}$$

where $I_\alpha(P,W)$ is Csiszár's version of α-mutual information.

Auxiliary Q - Auxiliary V

The optimal Q is such that

$$Q(y) = \sum_x P(x) V_x(y)$$

if we define $V_x(y)$ as

$$V_x(y) = \frac{W_x^{1-s}(y) Q^s(y)}{\sum_{y'} W_x^{1-s}(y') Q^s(y')}.$$

This channel V is such that

$$\begin{aligned} I(P, V) &= \sum_x P(x) D(V_x \| Q) \\ &= R - \epsilon \end{aligned}$$

Sphere packing: Haroutunian's proof

- Consider an auxiliary channel V such that $I(P,V) < R$
- Converse: original coding scheme incurs $\mathsf{P}_\mathrm{e} > \epsilon$ on V
- For at least one codeword m, $\boldsymbol{V}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m}) > \epsilon$.
- Stein Lemma

$$\boldsymbol{W}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m}) \gtrsim e^{-nD(V\|W|P)}$$

- Optimizing over V

$$\frac{1}{n}\log\frac{1}{\mathsf{P}_{\mathrm{e}|\boldsymbol{W}_{\boldsymbol{x}_m}}} \leq \inf_{V:I(P,V)<R} D(V\|W|P)(1+o(1)).$$

- Optimal V induces

$$Q(y) = \sum_x P(x)V_x(y)$$

optimal for MIT procedure.

Sphere packing: Haroutunian's proof

- Consider an auxiliary channel V such that $I(P, V) < R$
- Converse: original coding scheme incurs $\mathsf{P}_\mathrm{e} > \epsilon$ on V
- For at least one codeword m, $\boldsymbol{V}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m}) > \epsilon$.
- Stein Lemma

$$\boldsymbol{W}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m}) \gtrsim e^{-nD(V\|W|P)}$$

- Optimizing over V

$$\frac{1}{n}\log\frac{1}{\mathsf{P}_{\mathrm{e}|\boldsymbol{W}_{\boldsymbol{x}_m}}} \leq \inf_{V:I(P,V)<R} D(V\|W|P)(1+o(1)).$$

- Optimal V induces

$$Q(y) = \sum_x P(x)V_x(y)$$

optimal for MIT procedure.

Sphere packing: Haroutunian's proof

- Consider an auxiliary channel V such that $I(P, V) < R$
- Converse: original coding scheme incurs $\mathsf{P}_\mathrm{e} > \epsilon$ on V
- For at least one codeword m, $\boldsymbol{V}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m}) > \epsilon$.
- Stein Lemma
$$\boldsymbol{W}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m}) \gtrsim e^{-nD(V\|W|P)}$$
- Optimizing over V
$$\frac{1}{n}\log\frac{1}{\mathsf{P}_{\mathrm{e}|\boldsymbol{W}_{\boldsymbol{x}_m}}} \leq \inf_{V:I(P,V)<R} D(V\|W|P)(1+o(1)).$$
- Optimal V induces
$$Q(y) = \sum_x P(x)V_x(y)$$
optimal for MIT procedure.

Sphere packing: Haroutunian's proof

- Consider an auxiliary channel V such that $I(P,V) < R$
- Converse: original coding scheme incurs $\mathsf{P}_\mathrm{e} > \epsilon$ on V
- For at least one codeword m, $\boldsymbol{V}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m}) > \epsilon$.
- Stein Lemma

$$\boldsymbol{W}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m}) \gtrsim e^{-nD(V\|W|P)}$$

- Optimizing over V

$$\frac{1}{n}\log\frac{1}{\mathsf{P}_{\mathrm{e}|\boldsymbol{W}_{\boldsymbol{x}_m}}} \leq \inf_{V:I(P,V)<R} D(V\|W|P)(1+o(1)).$$

- Optimal V induces

$$Q(y) = \sum_x P(x)V_x(y)$$

optimal for MIT procedure.

Sphere packing: Haroutunian's proof

- Consider an auxiliary channel V such that $I(P, V) < R$
- Converse: original coding scheme incurs $\mathsf{P}_\mathrm{e} > \epsilon$ on V
- For at least one codeword m, $\boldsymbol{V}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m}) > \epsilon$.
- Stein Lemma

$$\boldsymbol{W}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m}) \gtrsim e^{-nD(V\|W|P)}$$

- Optimizing over V

$$\frac{1}{n}\log\frac{1}{\mathsf{P}_{\mathrm{e}|\boldsymbol{W}_{\boldsymbol{x}_m}}} \leq \inf_{V:I(P,V)<R} D(V\|W|P)(1+o(1)).$$

- Optimal V induces

$$Q(y) = \sum_x P(x)V_x(y)$$

optimal for MIT procedure.

- Consider an auxiliary channel V such that $I(P,V) < R$
- Converse: original coding scheme incurs $\mathsf{P}_{\mathrm{e}} > \epsilon$ on V
- For at least one codeword m, $\boldsymbol{V}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m}) > \epsilon$.
- Stein Lemma

$$\boldsymbol{W}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m}) \gtrsim e^{-nD(V\|W|P)}$$

- Optimizing over V

$$\frac{1}{n}\log\frac{1}{\mathsf{P}_{\mathrm{e}|\boldsymbol{W}_{\boldsymbol{x}_m}}} \leq \inf_{V:I(P,V)<R} D(V\|W|P)(1+o(1)).$$

- Optimal V induces

$$Q(y) = \sum_x P(x)V_x(y)$$

optimal for MIT procedure.

Key difference

MIT proof

- Just a single $\boldsymbol{Q}$ and M decoding regions implies $\boldsymbol{Q}(\boldsymbol{\mathcal{Y}}_m) \leq 1/M$ for some m
- If $\boldsymbol{Q}(\boldsymbol{\mathcal{Y}}_m) \leq e^{-nR}$ then $\boldsymbol{W}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m})$ is at least $e^{-nE_{\text{sp}}(R)}$

Haroutunian

- Converse for V implies $\boldsymbol{V}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m}) > \epsilon$
- If $\boldsymbol{V}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m}) > \epsilon$ then $\boldsymbol{W}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m}) \gtrsim e^{-nD(V\|W|P)}$

Equivalent

- The optimal Q induces the optimal channel V
- The optimal channel V induces the optimal Q

Key difference

MIT proof

- Just a single $\boldsymbol{Q}$ and M decoding regions implies $\boldsymbol{Q}(\boldsymbol{\mathcal{Y}}_m) \leq 1/M$ for some m
- If $\boldsymbol{Q}(\boldsymbol{\mathcal{Y}}_m) \leq e^{-nR}$ then $\boldsymbol{W}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m})$ is at least $e^{-nE_{\mathrm{sp}}(R)}$

Haroutunian

- Converse for V implies $\boldsymbol{V}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m}) > \epsilon$
- If $\boldsymbol{V}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m}) > \epsilon$ then $\boldsymbol{W}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m}) \gtrsim e^{-nD(V\|W|P)}$

Equivalent

- The optimal Q induces the optimal channel V
- The optimal channel V induces the optimal Q

Key difference

MIT proof

- Just a single $\boldsymbol{Q}$ and M decoding regions implies $\boldsymbol{Q}(\boldsymbol{\mathcal{Y}}_m) \leq 1/M$ for some m
- If $\boldsymbol{Q}(\boldsymbol{\mathcal{Y}}_m) \leq e^{-nR}$ then $\boldsymbol{W}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m})$ is at least $e^{-nE_{\mathrm{sp}}(R)}$

Haroutunian

- Converse for V implies $\boldsymbol{V}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m}) > \epsilon$
- If $\boldsymbol{V}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m}) > \epsilon$ then $\boldsymbol{W}_{\boldsymbol{x}_m}(\overline{\boldsymbol{\mathcal{Y}}_m}) \gtrsim e^{-nD(V\|W|P)}$

Equivalent

- The optimal Q induces the optimal channel V
- The optimal channel V induces the optimal Q

- The zero-error capacity only depends on the *confusability* of symbols in the input alphabet $\mathcal{X}$
- Symbols x and x' confusable if $\exists y : W_x(y)W_{x'}(y) > 0$, or

$$\sum_y W_x(y)W_{x'}(y) > 0$$

- **Confusability graph**

Hence $C_0(W) = C(G)$ *(Graph Capacity)*

Zero-Error Capacity

- The zero-error capacity only depends on the *confusability* of symbols in the input alphabet $\mathcal{X}$
- Symbols x and x' confusable if $\exists y : W_x(y)W_{x'}(y) > 0$, or

$$\sum_y W_x(y)W_{x'}(y) > 0$$

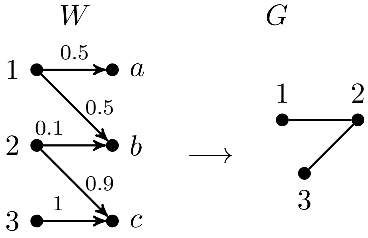
- **Confusability graph**

Hence $\quad C_0(W) \quad = \quad C(G) \quad$ *(Graph Capacity)*

- The zero-error capacity only depends on the *confusability* of symbols in the input alphabet $\mathcal{X}$
- Symbols x and x' confusable if $\exists y : W_x(y) W_{x'}(y) > 0$, or

$$\sum_y W_x(y) W_{x'}(y) > 0$$

- **Confusability graph**



Hence $C_0(W) = C(G)$ *(Graph Capacity)*

Graph Capacity

- Graph G
  - vertex set $V(G)$ (channel input symbols),
  - edge set $E(G)$ (pairs of distinct confusable symbols).
- $A \subseteq V(G)$ *independent set* if

$$x, x' \in A \implies x \nsim x'$$

- Independence number

$$\alpha(G) = \max\{|A| : A \subseteq V(G) \text{ independent set}\}$$

- Strong power G^n
  - $V(G^n) = V(G) \times V(G) \cdots \times V(G) = V(G)^n$
  - $\boldsymbol{x} \neq \boldsymbol{x'}$ connected in G^n if entrywise either equal or connected in G

$$(x_1, x_2, \ldots, x_n) \sim (x'_1, x'_2, \ldots, x'_n) \iff \forall\, i\,, x_i \sim x'_i \text{ or } x_i = x'_i$$

  i.e., confusable sequences.

So,

- $\alpha(G^n)$ is the largest size of an independent set in G^n or
- $\alpha(G^n)$ is the largest size of a zero-error code.

Graph Capacity

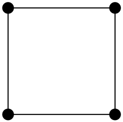
$$C(G) := \lim_{n\to\infty} \frac{1}{n} \log \alpha(G^n)$$

- $C(G)$ is highest asymptotic rate achievable with zero-error codes.
- Note: the limit exists due to Fekete's lemma since

$$\alpha(G^{n+m}) \geq \alpha(G^n)\alpha(G^m)$$

Three meaningful examples

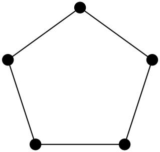
Square



$C(G) = 2$

(Shannon '56)

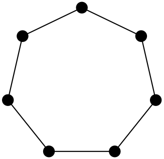
Pentagon



$C(G) = \log\sqrt{5}$

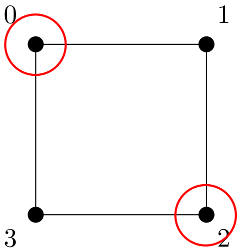
(Lovász '79)

Heptagon



$C(G)$

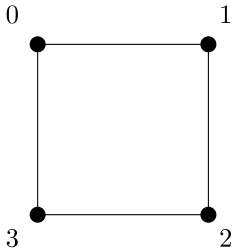
unknown



- Achievability:

$$\alpha(G) = 2 \implies \alpha(G^n) \geq 2^n \implies C(G) \geq 1 \text{ bit/ch. use}$$

- Converse
  - Each sequence symbol either in A or in B
  - 2^n "classes" of codewords
  - Codewords in each class are all confusable.
  - Pigeonhole principle: $\alpha(G^n) \leq 2^n$, so $C(G) \leq 1$

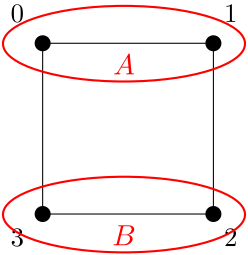


- Achievability:

$$\alpha(G) = 2 \implies \alpha(G^n) \geq 2^n \implies C(G) \geq 1 \text{ bit/ch. use}$$

- Converse
  - Each sequence symbol either in A or in B
  - 2^n "classes" of codewords
  - Codewords in each class are all confusable.
  - Pigeonhole principle: $\alpha(G^n) \leq 2^n$, so $C(G) \leq 1$

Square: easy case

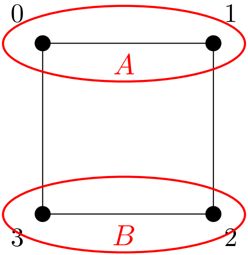


- Achievability:

$$\alpha(G) = 2 \implies \alpha(G^n) \geq 2^n \implies C(G) \geq 1 \text{ bit/ch. use}$$

- Converse
  - Each sequence symbol either in A or in B
  - 2^n "classes" of codewords
  - Codewords in each class are all confusable.
  - Pigeonhole principle: $\alpha(G^n) \leq 2^n$, so $C(G) \leq 1$

Square: easy case

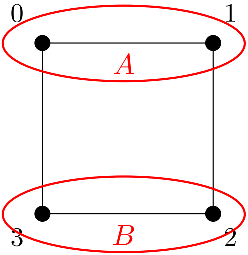


- Achievability:

$$\alpha(G) = 2 \implies \alpha(G^n) \geq 2^n \implies C(G) \geq 1 \text{ bit/ch. use}$$

- Converse
  - Each sequence symbol either in A or in B
  - 2^n "classes" of codewords
  - Codewords in each class are all confusable.
  - Pigeonhole principle: $\alpha(G^n) \leq 2^n$, so $C(G) \leq 1$

Square: easy case

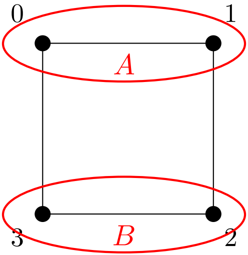


- Achievability:

$$\alpha(G) = 2 \implies \alpha(G^n) \geq 2^n \implies C(G) \geq 1 \text{ bit/ch. use}$$

- Converse
  - Each sequence symbol either in A or in B
  - 2^n "classes" of codewords
  - Codewords in each class are all confusable.
  - Pigeonhole principle: $\alpha(G^n) \leq 2^n$, so $C(G) \leq 1$

Square: easy case



- Achievability:

$$\alpha(G) = 2 \implies \alpha(G^n) \geq 2^n \implies C(G) \geq 1 \text{ bit/ch. use}$$

- Converse
  - Each sequence symbol either in A or in B
  - 2^n "classes" of codewords
  - Codewords in each class are all confusable.
  - Pigeonhole principle: $\alpha(G^n) \leq 2^n$, so $C(G) \leq 1$

More general

[¡+-¿] Using the same reasoning

- Clique: subset of $V(G)$ completely connected in G (independent set in $\bar{G}$)
- Assume G can be *covered* with k cliques
- Then $\alpha(G^n) \leq k^n$, and $C(G) \leq \log(k)$

Theorem

$$C(G) \leq \log \bar{\chi}(G)$$

where

$$\begin{aligned} \bar{\chi}(G) &= \text{clique covering number of } G \\ &= \text{minimum number of cliques to cover } G \\ &= \text{chromatic number of } \bar{G} \\ &=: \chi(\bar{G}) \end{aligned}$$

Extension to fractional covers

- A set of cliques $A_1, \ldots, A_k \subseteq V(G)$ is a factional cover of G with weights $\lambda_1, \lambda_2, \ldots \lambda_k$ if

$$\sum_{i:v\in A_i} \lambda_i \geq 1\,, \quad \forall v \in V(G)$$

- Fractional clique covering number

$$\bar{\chi}^*(G) = \min \sum_i \lambda_i$$

minimum over fractional clique covers ($\lambda_1, \lambda_2, \ldots \lambda_k$ = weights).

Theorem

$$C(G) \leq \log \bar{\chi}^*(G)$$

Extension to fractional covers

- A set of cliques $A_1, \ldots, A_k \subseteq V(G)$ is a factional cover of G with weights $\lambda_1, \lambda_2, \ldots \lambda_k$ if

$$\sum_{i:v\in A_i} \lambda_i \geq 1\,, \quad \forall v \in V(G)$$

- Fractional clique covering number

$$\bar{\chi}^*(G) = \min \sum_i \lambda_i$$

minimum over fractional clique covers ($\lambda_1, \lambda_2, \ldots \lambda_k =$ weights).

Theorem

$$C(G) \leq \log \bar{\chi}^*(G)$$

Extension to fractional covers

- A set of cliques $A_1, \ldots, A_k \subseteq V(G)$ is a factional cover of G with weights $\lambda_1, \lambda_2, \ldots \lambda_k$ if

$$\sum_{i:v \in A_i} \lambda_i \geq 1\,, \quad \forall v \in V(G)$$

- Fractional clique covering number

$$\bar{\chi}^*(G) = \min \sum_i \lambda_i$$

minimum over fractional clique covers ($\lambda_1, \lambda_2, \ldots \lambda_k =$ weights).

Theorem

$$C(G) \leq \log \bar{\chi}^*(G)$$

Proof

- Let $\lambda_1, \lambda_2, \ldots \lambda_k$ achieve $\bar{\chi}^*(G) = \sum_i \lambda_i$.
  Define a probability distribution q on cliques

$$q_i = \frac{\lambda_i}{\sum_j \lambda_j}$$

- If A is random clique $\sim q$ then

$$\mathsf{P}[v \in A] \geq \frac{1}{\sum_i \lambda_i}$$

- Pick random clique $\boldsymbol{A}$ in G^n as cartesian product of i.i.d. $\sim q$ cliques. Then,

$$\mathsf{P}[\boldsymbol{v} \in \boldsymbol{A}] \geq \left(\sum_i \lambda_i\right)^{-n}, \quad \forall \boldsymbol{v} \in V(G^n)$$

Proof

- Let $\lambda_1, \lambda_2, \ldots \lambda_k$ achieve $\bar{\chi}^*(G) = \sum_i \lambda_i$.
Define a probability distribution q on cliques

$$q_i = \frac{\lambda_i}{\sum_j \lambda_j}$$

- If A is random clique $\sim q$ then

$$\mathsf{P}[v \in A] \geq \frac{1}{\sum_i \lambda_i}$$

- Pick random clique $\boldsymbol{A}$ in G^n as cartesian product of i.i.d. $\sim q$ cliques. Then,

$$\mathsf{P}[\boldsymbol{v} \in \boldsymbol{A}] \geq \left(\sum_i \lambda_i\right)^{-n}, \quad \forall \boldsymbol{v} \in V(G^n)$$

Proof

- Let $\lambda_1, \lambda_2, \dots \lambda_k$ achieve $\bar{\chi}^*(G) = \sum_i \lambda_i$.
  Define a probability distribution q on cliques

$$q_i = \frac{\lambda_i}{\sum_j \lambda_j}$$

- If A is random clique $\sim q$ then

$$\mathsf{P}[v \in A] \geq \frac{1}{\sum_i \lambda_i}$$

- Pick random clique $\boldsymbol{A}$ in G^n as cartesian product of i.i.d. $\sim q$ cliques. Then,

$$\mathsf{P}[\boldsymbol{v} \in \boldsymbol{A}] \geq \left(\sum_i \lambda_i\right)^{-n}, \quad \forall \boldsymbol{v} \in V(G^n)$$

proof

- So

$$\mathsf{E}[|\mathcal{C} \cap \boldsymbol{A}|] \geq |\mathcal{C}| \cdot \left(\sum_i \lambda_i\right)^{-n}$$

- If $\mathcal{C}$ is a zero-error code

$$\begin{aligned} 1 &\geq \max_{A \text{ clique}} |\mathcal{C} \cap A| \\ &\geq \mathsf{E}[|\mathcal{C} \cap \boldsymbol{A}|] \\ &\geq |\mathcal{C}| \cdot \left(\sum_i \lambda_i\right)^{-n} \end{aligned}$$

- Hence,

$$\begin{aligned} \alpha(G^n) &\leq \left(\sum_i \lambda_i\right)^{n} \\ &= \bar{\chi}^*(G)^n \end{aligned}$$

proof

- So

$$\mathsf{E}[|\mathcal{C} \cap \boldsymbol{A}|] \geq |\mathcal{C}| \cdot \left(\sum_i \lambda_i\right)^{-n}$$

- If $\mathcal{C}$ is a zero-error code

$$\begin{aligned} 1 &\geq \max_{A\ clique} |\mathcal{C} \cap A| \\ &\geq \mathsf{E}[|\mathcal{C} \cap \boldsymbol{A}|] \\ &\geq |\mathcal{C}| \cdot \left(\sum_i \lambda_i\right)^{-n} \end{aligned}$$

- Hence,

$$\begin{aligned} \alpha(G^n) &\leq \left(\sum_i \lambda_i\right)^{n} \\ &= \bar{\chi}^*(G)^n \end{aligned}$$

proof

- So
$$\mathsf{E}[|\boldsymbol{\mathcal{C}} \cap \boldsymbol{A}|] \geq |\boldsymbol{\mathcal{C}}| \cdot \left(\sum_i \lambda_i\right)^{-n}$$
- If $\boldsymbol{\mathcal{C}}$ is a zero-error code
$$\begin{aligned} 1 &\geq \max_{A \text{ clique}} |\boldsymbol{\mathcal{C}} \cap A| \\ &\geq \mathsf{E}[|\boldsymbol{\mathcal{C}} \cap \boldsymbol{A}|] \\ &\geq |\boldsymbol{\mathcal{C}}| \cdot \left(\sum_i \lambda_i\right)^{-n} \end{aligned}$$
- Hence,
$$\begin{aligned} \alpha(G^n) &\leq \left(\sum_i \lambda_i\right)^{n} \\ &= \bar{\chi}^*(G)^n \end{aligned}$$

- Setting $\mathcal{X}_y = \{x : W_x(y) > 0\}$, $\mathcal{Y}_x = \{y : W_x(y) > 0\}$

$$R_\infty(W) = \log \min_Q \max_x \frac{1}{\sum_{y \in \mathcal{Y}_x} Q(y)}$$

- Setting $q(y) = \max_x \frac{Q(y)}{\sum_{y' \in \mathcal{Y}_x} Q(y')}$

$$R_\infty(W) = \log \min_q \sum_y q(y)$$

under constraints

$$q(y) \geq 0\,, \qquad \sum_{y \in \mathcal{Y}_x} q(y) \geq 1$$

- Like a fractional clique cover, every output symbol a clique on G.

Comparison with R_∞

- Setting $\mathcal{X}_y = \{x : W_x(y) > 0\}$, $\mathcal{Y}_x = \{y : W_x(y) > 0\}$

$$R_\infty(W) = \log \min_Q \max_x \frac{1}{\sum_{y\in\mathcal{Y}_x} Q(y)}$$

- Setting $q(y) = \max_x \frac{Q(y)}{\sum_{y'\in\mathcal{Y}_x} Q(y')}$

$$R_\infty(W) = \log \min_q \sum_y q(y)$$

under constraints

$$q(y) \geq 0\,, \quad \sum_{y\in\mathcal{Y}_x} q(y) \geq 1$$

- Like a fractional clique cover, every output symbol a clique on G.

Comparison with R_∞

- Setting $\mathcal{X}_y = \{x : W_x(y) > 0\}$, $\mathcal{Y}_x = \{y : W_x(y) > 0\}$

$$R_\infty(W) = \log \min_Q \max_x \frac{1}{\sum_{y\in\mathcal{Y}_x} Q(y)}$$

- Setting $q(y) = \max_x \frac{Q(y)}{\sum_{y'\in\mathcal{Y}_x} Q(y')}$

$$R_\infty(W) = \log \min_q \sum_y q(y)$$

under constraints

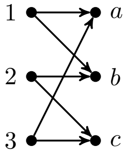
$$q(y) \geq 0\,, \quad \sum_{y\in\mathcal{Y}_x} q(y) \geq 1$$

- Like a fractional clique cover, every output symbol a clique on G.

Comparison with R_∞

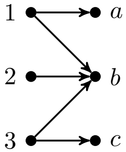
- Indeed R_∞ does not only depend on $G(W)$

W_1



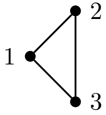
,

W_2



$\Longrightarrow$

$G(W_1) = G(W_2)$

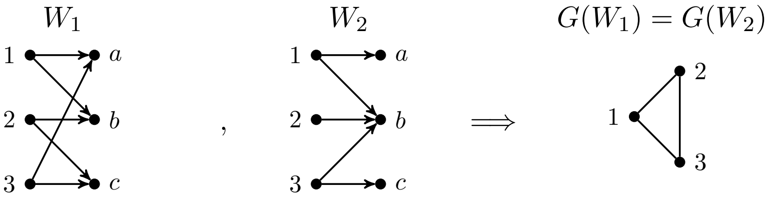


$R_\infty(W_1) = \log 3/2$ $\quad$ $R_\infty(W_2) = 0$ $\quad$ but $\quad$ $C(G) = 0$

- To bound $C(G)$ pick the most useful W with $G(W) = G$.
- That is, define one output symbol for each clique in G. Then

$$R_\infty(W_{\text{opt}}) = \log \bar{\chi}^*(G)$$

- Indeed R_∞ does not only depend on $G(W)$

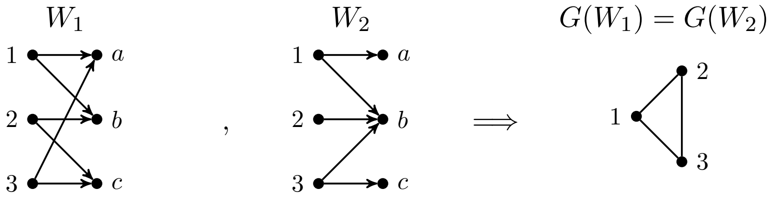


$R_\infty(W_1) = \log 3/2$ $\quad$ $R_\infty(W_2) = 0$ $\quad$ but $\quad$ $C(G) = 0$

- To bound $C(G)$ pick the most useful W with $G(W) = G$.
- That is, define one output symbol for each clique in G. Then

$$R_\infty(W_{\text{opt}}) = \log \bar{\chi}^*(G)$$

- Indeed R_∞ does not only depend on $G(W)$



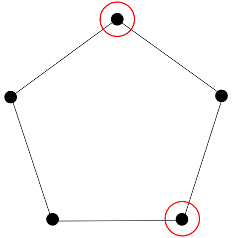
$$R_\infty(W_1) = \log 3/2 \qquad R_\infty(W_2) = 0 \quad \text{but} \qquad C(G) = 0$$

- To bound $C(G)$ pick the most useful W with $G(W) = G$.
- That is, define one output symbol for each clique in G. Then

$$R_\infty(W_{\text{opt}}) = \log \bar{\chi}^*(G)$$

Pentagon

G



$\alpha(G) = 2$

G^2

- Achievability:

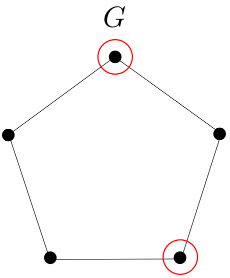
$$\alpha(G^2) = 5 \implies C(G) \geq \frac{1}{2} \log 5$$

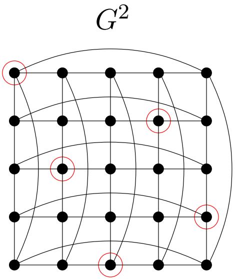
- Converse: fractional clique cover with 5 cliques, weight 1/2 each

$$\bar{\chi}^*(G) = 5/2 \implies C(G) \leq \log(5/2)$$

- Lovász (1979): $C(G) = \frac{1}{2} \log 5$

Pentagon





$\alpha(G) = 2$ but $\alpha(G^2) = 5$

- Achievability:

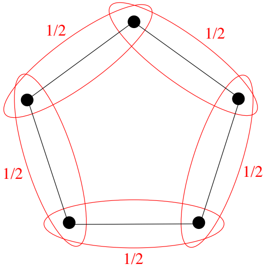
$$\alpha(G^2) = 5 \implies C(G) \geq \frac{1}{2} \log 5$$

- Converse: fractional clique cover with 5 cliques, weight 1/2 each

$$\bar{\chi}^*(G) = 5/2 \implies C(G) \leq \log(5/2)$$

- Lovász (1979): $C(G) = \frac{1}{2} \log 5$

Pentagon



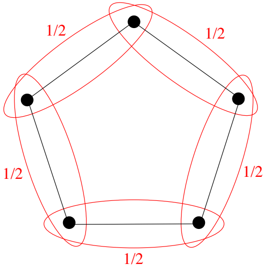
- Achievability:

$$\alpha(G^2) = 5 \implies C(G) \geq \frac{1}{2} \log 5$$

- Converse: fractional clique cover with 5 cliques, weight 1/2 each

$$\bar{\chi}^*(G) = 5/2 \implies C(G) \leq \log(5/2)$$

- Lovász (1979): $C(G) = \frac{1}{2} \log 5$



- Achievability:

$$\alpha(G^2) = 5 \implies C(G) \geq \frac{1}{2} \log 5$$

- Converse: fractional clique cover with 5 cliques, weight 1/2 each

$$\bar{\chi}^*(G) = 5/2 \implies C(G) \leq \log(5/2)$$

- Lovász (1979): $C(G) = \frac{1}{2} \log 5$

Lovász theta function

Lovász's idea

- Graph representation: map vertices x to unit norm $u_x \in \mathbb{R}^d$ so that

$$x \nsim x' \implies u_x \perp u_{x'}$$

- An independent set A is mapped to an orthonormal basis
- For any unit norm c and independent set A

$$1 \geq \|c\|^2 \geq \sum_{x \in A} |\langle u_x | c \rangle|^2 \geq |A| \min_x |\langle u_x | c \rangle|^2$$

- Take $\{u_x\}$ and c optimally; if

$$\theta(G) = \left(\max_{\{u_x\}, c} \min_x |\langle u_x | c \rangle|^2 \right)^{-1}$$

then

$$\alpha(G) \leq \theta$$

Tensorization

- Note $\langle a \otimes b | c \otimes d \rangle = \langle a|c\rangle\langle b|d\rangle$
- So, if $\{u_x\}$ representation of G used with c gives

$$\alpha(G) \leq \theta(G)$$

then $\{u_x\}^{\otimes n}$ representation of G^n used with $c^{\otimes n}$ gives

$$\alpha(G^n) \leq \theta(G)^n$$

- Taking $\lim \frac{1}{n} \log(\cdot)$

$$C(G) \leq \log \theta(G)$$

Tensorization

- Note $\langle a \otimes b | c \otimes d \rangle = \langle a|c \rangle \langle b|d \rangle$
- So, if $\{u_x\}$ representation of G used with c gives

$$\alpha(G) \leq \theta(G)$$

then $\{u_x\}^{\otimes n}$ representation of G^n used with $c^{\otimes n}$ gives

$$\alpha(G^n) \leq \theta(G)^n$$

- Taking $\lim \frac{1}{n} \log(\cdot)$

$$C(G) \leq \log \theta(G)$$

Tensorization

- Note $\langle a \otimes b | c \otimes d \rangle = \langle a|c\rangle\langle b|d\rangle$
- So, if $\{u_x\}$ representation of G used with c gives

$$\alpha(G) \leq \theta(G)$$

then $\{u_x\}^{\otimes n}$ representation of G^n used with $c^{\otimes n}$ gives

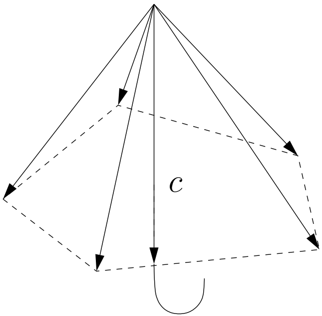
$$\alpha(G^n) \leq \theta(G)^n$$

- Taking $\lim \frac{1}{n} \log(\cdot)$

$$C(G) \leq \log \theta(G)$$

Lovász theta function: pentagon

Lovász's umbrella



$$\theta = \sqrt{5}$$

$$C(G) \leq \log \sqrt{5}$$

- **Orthonormal Representation**:
  A set of unit norm vectors $\{u_x\}$, $x \in \mathcal{X}$

$$x, x' \text{ not confusable} \quad \Longrightarrow \quad \langle u_x | u_{x'} \rangle = 0$$

- **Trivial Representation**: $u_x = \sqrt{W_x}$
- **Value** (log domain):

$$V(\{u_x\}) = \min_c \max_x \log \frac{1}{|\langle u_x | c \rangle|^2} \qquad (\|c\| = 1)$$

  c is the *handle*. Note: $|\langle u_x | c \rangle|^2 \geq e^{-V(\{u_x\})}$, $\forall x$
- **The bound**:

$$C_0 \leq V(\{u_x\})$$

- **Theta function** (log domain):

$$\vartheta = \min_{\{u_x\}} V(\{u_x\})$$

Lovász' Bound, channel interpretation

- **Orthonormal Representation**:
  A set of unit norm vectors $\{u_x\}$, $x \in \mathcal{X}$

$$x, x' \text{ not confusable} \quad \Longrightarrow \quad \langle u_x | u_{x'} \rangle = 0$$

- **Trivial Representation**: $u_x = \sqrt{W_x}$
- **Value** (log domain):

$$V(\{u_x\}) = \min_c \max_x \log \frac{1}{|\langle u_x | c \rangle|^2} \qquad (\|c\| = 1)$$

  c is the *handle*. Note: $|\langle u_x | c \rangle|^2 \geq e^{-V(\{u_x\})}$, $\forall x$

- **The bound**:

$$C_0 \leq V(\{u_x\})$$

- **Theta function** (log domain):

$$\vartheta = \min_{\{u_x\}} V(\{u_x\})$$

Lovász' Bound, channel interpretation

- **Orthonormal Representation**:
  A set of unit norm vectors $\{u_x\}$, $x \in \mathcal{X}$

$$x, x' \text{ not confusable} \quad \Longrightarrow \quad \langle u_x | u_{x'} \rangle = 0$$

- **Trivial Representation**: $u_x = \sqrt{W_x}$
- **Value** (log domain):

$$V(\{u_x\}) = \min_c \max_x \log \frac{1}{|\langle u_x | c \rangle|^2} \qquad (\|c\| = 1)$$

  c is the *handle*. Note: $|\langle u_x | c \rangle|^2 \geq e^{-V(\{u_x\})}$, $\forall x$

- **The bound**:

$$C_0 \leq V(\{u_x\})$$

- **Theta function** (log domain):

$$\vartheta = \min_{\{u_x\}} V(\{u_x\})$$

Lovász' Bound, channel interpretation

- **Orthonormal Representation**:
  A set of unit norm vectors $\{u_x\}$, $x \in \mathcal{X}$

$$x, x' \text{ not confusable} \quad \Longrightarrow \quad \langle u_x | u_{x'} \rangle = 0$$

- **Trivial Representation**: $u_x = \sqrt{W_x}$
- **Value** (log domain):

$$V(\{u_x\}) = \min_c \max_x \log \frac{1}{|\langle u_x | c \rangle|^2} \qquad (\|c\| = 1)$$

  c is the *handle*. Note: $|\langle u_x | c \rangle|^2 \geq e^{-V(\{u_x\})}$, $\forall x$
- **The bound**:

$$C_0 \leq V(\{u_x\})$$

- **Theta function** (log domain):

$$\vartheta = \min_{\{u_x\}} V(\{u_x\})$$

- **Orthonormal Representation**:
  A set of unit norm vectors $\{u_x\}$, $x \in \mathcal{X}$

$$x, x' \text{ not confusable} \quad \Longrightarrow \quad \langle u_x | u_{x'} \rangle = 0$$

- **Trivial Representation**: $u_x = \sqrt{W_x}$
- **Value** (log domain):

$$V(\{u_x\}) = \min_c \max_x \log \frac{1}{|\langle u_x | c \rangle|^2} \qquad (\|c\| = 1)$$

  c is the *handle*. Note: $|\langle u_x | c \rangle|^2 \geq e^{-V(\{u_x\})}$, $\forall x$
- **The bound**:

$$C_0 \leq V(\{u_x\})$$

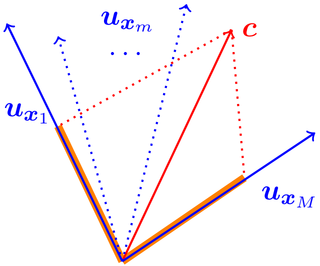
- **Theta function** (log domain):

$$\vartheta = \min_{\{u_x\}} V(\{u_x\})$$

Representation for $\boldsymbol{W}$

$$\text{Vectors} \quad \boldsymbol{x} = (x_1, \ldots, x_n) \longrightarrow \boldsymbol{u_x} = u_{x_1} \otimes \cdots \otimes u_{x_n}$$

$$\text{Handle } \ldots\ldots \qquad \boldsymbol{c} = c \otimes \cdots \otimes c$$



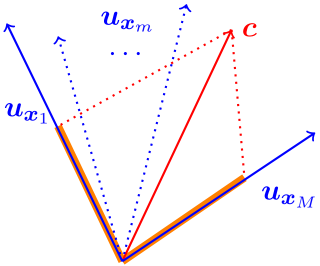
- For a zero-error code, the vectors $\boldsymbol{u}_{\boldsymbol{x}_m}$ are pairwise orthogonal
- $|\langle \boldsymbol{u}_{\boldsymbol{x}_m} | \boldsymbol{c} \rangle|^2 \leq 1/M$ for at least one m, because $\|\boldsymbol{c}\| = 1$
- $|\langle \boldsymbol{u}_{\boldsymbol{x}_m} | \boldsymbol{c} \rangle|^2 \geq e^{-nV(\{u_x\})}$ by definition of $V(\{u_x\})$
  Hence $M \leq e^{nV(\{u_x\})}$

Lovász' Bound, channel interpretation

Representation for $\boldsymbol{W}$

Vectors $\quad \boldsymbol{x} = (x_1, \ldots, x_n) \longrightarrow \boldsymbol{u_x} = u_{x_1} \otimes \cdots \otimes u_{x_n}$

Handle ...... $\quad \boldsymbol{c} = c \otimes \cdots \otimes c$

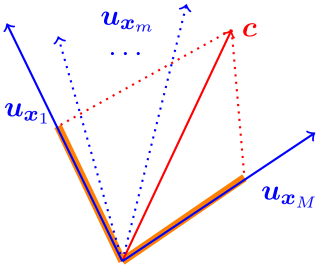


- For a zero-error code, the vectors $\boldsymbol{u}_{\boldsymbol{x}_m}$ are pairwise orthogonal
- $|\langle \boldsymbol{u}_{\boldsymbol{x}_m} | \boldsymbol{c} \rangle|^2 \leq 1/M$ for at least one m, because $\|\boldsymbol{c}\| = 1$
- $|\langle \boldsymbol{u}_{\boldsymbol{x}_m} | \boldsymbol{c} \rangle|^2 \geq e^{-nV(\{u_x\})}$ by definition of $V(\{u_x\})$
  Hence $M \leq e^{nV(\{u_x\})}$

Representation for $\boldsymbol{W}$

$$\text{Vectors} \quad \boldsymbol{x} = (x_1, \ldots, x_n) \longrightarrow \boldsymbol{u_x} = u_{x_1} \otimes \cdots \otimes u_{x_n}$$

$$\text{Handle} \ \ldots\ldots \quad \boldsymbol{c} = c \otimes \cdots \otimes c$$



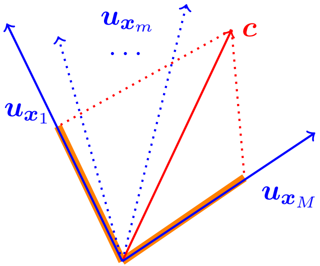
- For a zero-error code, the vectors $\boldsymbol{u}_{\boldsymbol{x}_m}$ are pairwise orthogonal
- $|\langle \boldsymbol{u}_{\boldsymbol{x}_m} | \boldsymbol{c} \rangle|^2 \leq 1/M$ for at least one m, because $\|\boldsymbol{c}\| = 1$
- $|\langle \boldsymbol{u}_{\boldsymbol{x}_m} | \boldsymbol{c} \rangle|^2 \geq e^{-nV(\{u_x\})}$ by definition of $V(\{u_x\})$
  Hence $M \leq e^{nV(\{u_x\})}$

Lovász' Bound, channel interpretation

Representation for $\boldsymbol{W}$

Vectors $\quad \boldsymbol{x} = (x_1, \ldots, x_n) \longrightarrow \boldsymbol{u_x} = u_{x_1} \otimes \cdots \otimes u_{x_n}$

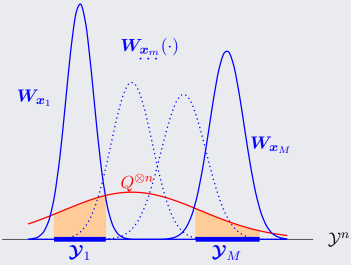
Handle ...... $\quad \boldsymbol{c} = c \otimes \cdots \otimes c$

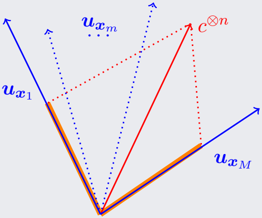


- For a zero-error code, the vectors $\boldsymbol{u}_{\boldsymbol{x}_m}$ are pairwise orthogonal
- $|\langle \boldsymbol{u}_{\boldsymbol{x}_m} | \boldsymbol{c} \rangle|^2 \leq 1/M$ for at least one m, because $\|\boldsymbol{c}\| = 1$
- $|\langle \boldsymbol{u}_{\boldsymbol{x}_m} | \boldsymbol{c} \rangle|^2 \geq e^{-nV(\{u_x\})}$ by definition of $V(\{u_x\})$
  Hence $M \leq e^{nV(\{u_x\})}$

Analogies

Sphere-Packing vs Lovász





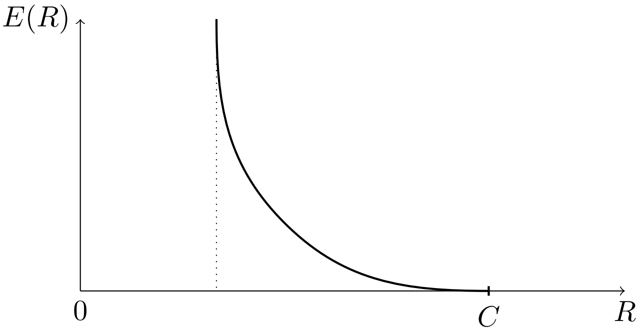
We note the following analogies

$$\sqrt{W_{x_m}} \leftrightarrow u_{x_m}$$

$$\sqrt{Q} \leftrightarrow c$$

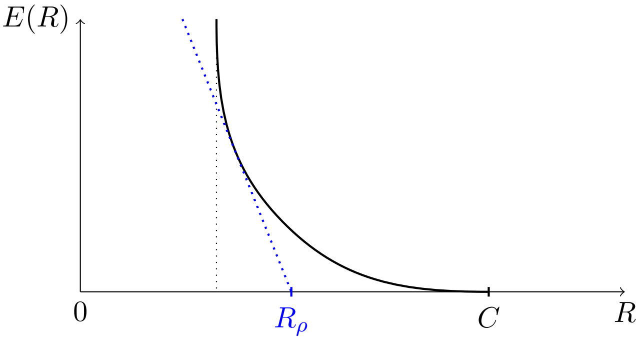
$$Q^{\otimes n}(\mathcal{Y}_m) \leftrightarrow |\langle u_{x_m}|c^{\otimes n}\rangle|^2$$

Sphere-Packing Bound as an Information Radius



What about min-max expressions?

Sphere-Packing Bound as an Information Radius

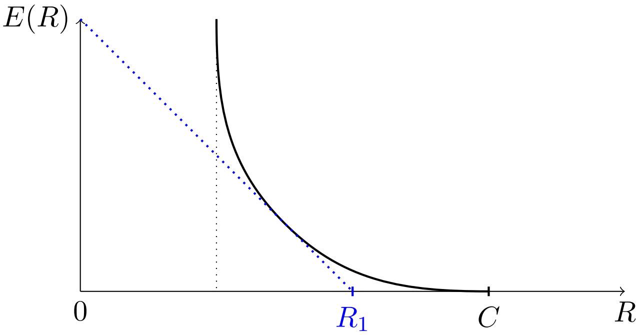


Remind

$$R_\rho = \min_Q \max_x D_\alpha(W_x||Q), \quad \alpha = 1/(1+\rho)$$

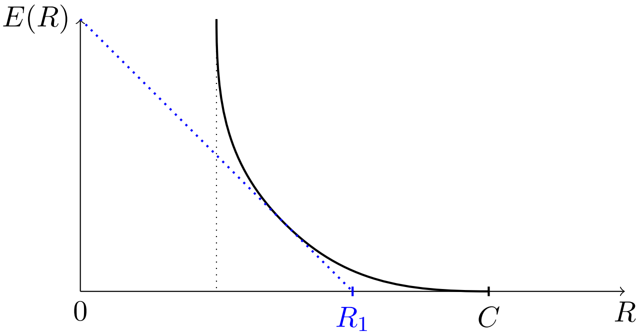
$D_\alpha(Q_1||Q_2) = \frac{1}{\alpha-1}\log\sum_y Q_1(y)^\alpha Q_2(y)^{1-\alpha}$ is the Rényi divergence

Sphere-Packing Bound as an Information Radius



Setting $\rho = 1$, **cutoff rate**:

$$R_1 = \min_Q \max_x \log \frac{1}{\left(\sum_y \sqrt{W_x(y)Q(y)}\right)^2}$$



Setting $\rho = 1$, **cutoff rate**:

$$R_1 = \min_Q \max_x \log \frac{1}{\left(\sum_y \sqrt{W_x(y)Q(y)}\right)^2}$$

$$= V(\{u_x\}) \qquad \text{if} \quad u_x = \sqrt{W(\cdot|x)}$$

Representations, values and cutoff rates

- So,

$$u_x = \sqrt{W_x} \quad \Longrightarrow \quad V(\{u_x\}) = \text{cutoff rate}$$

- If all u_x have non-negative components we always get the cutoff rate of some classical channel
- Lovász' optimal u_x can (often will!) have negative components.

Intuition (?)

Use wave functions of quantum physics to play the role of $\sqrt{W_x}$

Representations, values and cutoff rates

- So,

$$u_x = \sqrt{W_x} \quad \Longrightarrow \quad V(\{u_x\}) = \text{cutoff rate}$$

- If all u_x have non-negative components we always get the cutoff rate of some classical channel
- Lovász' optimal u_x can (often will!) have negative components.

Intuition (?)

Use wave functions of quantum physics to play the role of $\sqrt{W_x}$

Representations, values and cutoff rates

- So,

$$u_x = \sqrt{W_x} \quad \Longrightarrow \quad V(\{u_x\}) = \text{cutoff rate}$$

- If all u_x have non-negative components we always get the cutoff rate of some classical channel
- Lovász' optimal u_x can (often will!) have negative components.

Intuition (?)

Use wave functions of quantum physics to play the role of $\sqrt{W_x}$

Representations, values and cutoff rates

- So,

$$u_x = \sqrt{W_x} \quad \Longrightarrow \quad V(\{u_x\}) = \text{cutoff rate}$$

- If all u_x have non-negative components we always get the cutoff rate of some classical channel
- Lovász' optimal u_x can (often will!) have negative components.

Intuition (?)

Use wave functions of quantum physics to play the role of $\sqrt{W_x}$

Definition

- **Basic Idea**

$$W_x \text{ now density operator}$$

W_x is a positive semi-definite matrix with unit trace

- **Classical channels**: all w_x are diagonal

$$W_x = \begin{bmatrix} W_x(1) & 0 & \cdots & 0 \\ 0 & W_x(2) & \cdots & 0 \\ 0 & \cdots & \ddots & \end{bmatrix}$$

- **Pure-State Channel**: all W_x are rank-one matrices

$$W_x = |u_x\rangle\langle u_x|$$

Definition

- **Basic Idea**

$$W_x \text{ now density operator}$$

W_x is a positive semi-definite matrix with unit trace

- **Classical channels**: all w_x are diagonal

$$W_x = \begin{bmatrix} W_x(1) & 0 & \cdots & 0 \\ 0 & W_x(2) & \cdots & 0 \\ 0 & \cdots & \ddots & \end{bmatrix}$$

- **Pure-State Channel**: all W_x are rank-one matrices

$$W_x = |u_x\rangle\langle u_x|$$

Definition

- **Basic Idea**

$$W_x \text{ now density operator}$$

  W_x is a positive semi-definite matrix with unit trace

- **Classical channels**: all w_x are diagonal

$$W_x = \begin{bmatrix} W_x(1) & 0 & \cdots & 0 \\ 0 & W_x(2) & \cdots & 0 \\ 0 & \cdots & \ddots & \end{bmatrix}$$

- **Pure-State Channel**: all W_x are rank-one matrices

$$W_x = |u_x\rangle\langle u_x|$$

- **Memoryless extension:**

$$\boldsymbol{x} = (x_1, \ldots, x_n) \to \boldsymbol{W}_{\boldsymbol{x}} = W_{x_1} \otimes \cdots \otimes W_{x_n}$$

- **Code**: M codewords $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\} \subset \mathcal{X}^n$
- **Decoder**: a POVM, collection of M positive operators $\{\Pi_1, \ldots, \Pi_M\}$ (positive semi-definite matrices) such that

$$I - \sum_{m=1}^{M} \Pi_m \geq 0$$

- Classical deterministic case: Π_m diagonal $\{0,1\}$-valued matrix, indicator function of $\mathcal{Y}_m$
- **Probability of error**: $\mathsf{P}_{\mathsf{e}|m} = 1 - \mathrm{Tr}(\Pi_m \boldsymbol{W}_{\boldsymbol{x}_m})$
- **Capacities and reliability**: as before.

- **Memoryless extension:**

$$\boldsymbol{x} = (x_1, \ldots, x_n) \to \boldsymbol{W_x} = W_{x_1} \otimes \cdots \otimes W_{x_n}$$

- **Code**: M codewords $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\} \subset \mathcal{X}^n$
- **Decoder**: a POVM, collection of M positive operators $\{\Pi_1, \ldots, \Pi_M\}$ (positive semi-definite matrices) such that

$$I - \sum_{m=1}^{M} \Pi_m \geq 0$$

- Classical deterministic case: Π_m diagonal $\{0,1\}$-valued matrix, indicator function of $\mathcal{Y}_m$
- **Probability of error**: $\mathsf{P}_{\mathsf{e}|m} = 1 - \mathrm{Tr}(\Pi_m \boldsymbol{W}_{\boldsymbol{x}_m})$
- **Capacities and reliability**: as before.

- **Memoryless extension:**

$$\boldsymbol{x} = (x_1, \dots, x_n) \to \boldsymbol{W_x} = W_{x_1} \otimes \cdots \otimes W_{x_n}$$

- **Code**: M codewords $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_M\} \subset \mathcal{X}^n$
- **Decoder**: a POVM, collection of M positive operators $\{\Pi_1, \dots, \Pi_M\}$ (positive semi-definite matrices) such that

$$I - \sum_{m=1}^{M} \Pi_m \geq 0$$

- Classical deterministic case: Π_m diagonal $\{0, 1\}$-valued matrix, indicator function of $\mathcal{Y}_m$
- **Probability of error**: $\mathsf{P}_{\mathsf{e}|m} = 1 - \mathrm{Tr}(\Pi_m \boldsymbol{W}_{\boldsymbol{x}_m})$
- **Capacities and reliability**: as before.

- **Memoryless extension:**

$$\boldsymbol{x} = (x_1, \ldots, x_n) \to \boldsymbol{W_x} = W_{x_1} \otimes \cdots \otimes W_{x_n}$$

- **Code**: M codewords $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\} \subset \mathcal{X}^n$
- **Decoder**: a POVM, collection of M positive operators $\{\Pi_1, \ldots, \Pi_M\}$ (positive semi-definite matrices) such that

$$I - \sum_{m=1}^{M} \Pi_m \geq 0$$

- Classical deterministic case: Π_m diagonal $\{0, 1\}$-valued matrix, indicator function of $\boldsymbol{\mathcal{Y}}_m$
- **Probability of error**: $\mathsf{P}_{\mathsf{e}|m} = 1 - \mathrm{Tr}(\Pi_m \boldsymbol{W}_{\boldsymbol{x}_m})$
- **Capacities and reliability**: as before.

- **Memoryless extension:**

$$\boldsymbol{x} = (x_1, \ldots, x_n) \to \boldsymbol{W}_{\boldsymbol{x}} = W_{x_1} \otimes \cdots \otimes W_{x_n}$$

- **Code**: M codewords $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\} \subset \mathcal{X}^n$
- **Decoder**: a POVM, collection of M positive operators $\{\Pi_1, \ldots, \Pi_M\}$ (positive semi-definite matrices) such that

$$I - \sum_{m=1}^{M} \Pi_m \geq 0$$

- Classical deterministic case: Π_m diagonal $\{0,1\}$-valued matrix, indicator function of $\boldsymbol{\mathcal{Y}}_m$
- **Probability of error**: $\mathsf{P}_{\mathrm{e}|m} = 1 - \mathrm{Tr}(\Pi_m \boldsymbol{W}_{\boldsymbol{x}_m})$
- **Capacities and reliability**: as before.

- **Memoryless extension:**

$$\boldsymbol{x} = (x_1, \ldots, x_n) \rightarrow \boldsymbol{W_x} = W_{x_1} \otimes \cdots \otimes W_{x_n}$$

- **Code**: M codewords $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\} \subset \mathcal{X}^n$
- **Decoder**: a POVM, collection of M positive operators $\{\Pi_1, \ldots, \Pi_M\}$ (positive semi-definite matrices) such that

$$I - \sum_{m=1}^{M} \Pi_m \geq 0$$

- Classical deterministic case: Π_m diagonal $\{0, 1\}$-valued matrix, indicator function of $\boldsymbol{\mathcal{Y}}_m$
- **Probability of error**: $\mathsf{P}_{\mathrm{e}|m} = 1 - \mathrm{Tr}(\Pi_m \boldsymbol{W}_{\boldsymbol{x}_m})$
- **Capacities and reliability**: as before.

Note on probabilities

- If $A = |a\rangle\langle a|$ and $B = |b\rangle\langle b|$ (pure states)

$$\operatorname{Tr} AB = |\langle a|b\rangle|^2$$

- If

$$A = \sum_i \alpha_i |a_i\rangle\langle a_i| \qquad B = \sum_j \beta_j |b_j\rangle\langle b_j|$$

  then

$$\operatorname{Tr} AB = \sum_{i,j} \alpha_i \beta_j |\langle a_i|b_j\rangle|^2$$

Results on C and $E(R)$?

Capacity

- Hausladen-Jozsa-Schumacher-Westmoreland-Wootters: (1996) pure states
- Holevo, Schumacher-Westmorelan (1998): general states

$E(R)$ - achievability

- Burnashev-Holevo (1998): random coding for pure states
- Holevo (2000): expurgate bound (general case)
- Hayashi (2006): best "random coding" bound for mixed states
- **Missing**: conjectured Gallager-like random coding exponent!

$E(R)$ - converse

- Dalai (2012): sphere-packing
- (using Berlekamp): zero-rate upper bound

Results on C and $E(R)$?

Capacity

- Hausladen-Jozsa-Schumacher-Westmoreland-Wootters: (1996) pure states
- Holevo, Schumacher-Westmorelan (1998): general states

$E(R)$ - achievability

- Burnashev-Holevo (1998): random coding for pure states
- Holevo (2000): expurgate bound (general case)
- Hayashi (2006): best "random coding" bound for mixed states
- **Missing**: conjectured Gallager-like random coding exponent!

$E(R)$ - converse

- Dalai (2012): sphere-packing
- (using Berlekamp): zero-rate upper bound

Results on C and $E(R)$?

Capacity

- Hausladen-Jozsa-Schumacher-Westmoreland-Wootters: (1996) pure states
- Holevo, Schumacher-Westmorelan (1998): general states

$E(R)$ - achievability

- Burnashev-Holevo (1998): random coding for pure states
- Holevo (2000): expurgate bound (general case)
- Hayashi (2006): best "random coding" bound for mixed states
- **Missing**: conjectured Gallager-like random coding exponent!

$E(R)$ - converse

- Dalai (2012): sphere-packing
- (using Berlekamp): zero-rate upper bound

Results on C and $E(R)$?

Capacity

- Hausladen-Jozsa-Schumacher-Westmoreland-Wootters: (1996) pure states
- Holevo, Schumacher-Westmorelan (1998): general states

$E(R)$ - achievability

- Burnashev-Holevo (1998): random coding for pure states
- Holevo (2000): expurgate bound (general case)
- Hayashi (2006): best "random coding" bound for mixed states
- **Missing**: conjectured Gallager-like random coding exponent!

$E(R)$ - converse

- Dalai (2012): sphere-packing
- (using Berlekamp): zero-rate upper bound

Capacity

- Hausladen-Jozsa-Schumacher-Westmoreland-Wootters: (1996) pure states
- Holevo, Schumacher-Westmorelan (1998): general states

$E(R)$ - achievability

- Burnashev-Holevo (1998): random coding for pure states
- Holevo (2000): expurgate bound (general case)
- Hayashi (2006): best “random coding” bound for mixed states
- **Missing**: conjectured Gallager-like random coding exponent!

$E(R)$ - converse

- Dalai (2012): sphere-packing
- (using Berlekamp): zero-rate upper bound

Capacity

- Hausladen-Jozsa-Schumacher-Westmoreland-Wootters: (1996) pure states
- Holevo, Schumacher-Westmorelan (1998): general states

$E(R)$ - achievability

- Burnashev-Holevo (1998): random coding for pure states
- Holevo (2000): expurgate bound (general case)
- Hayashi (2006): best "random coding" bound for mixed states
- **Missing**: conjectured Gallager-like random coding exponent!

$E(R)$ - converse

- Dalai (2012): sphere-packing
- (using Berlekamp): zero-rate upper bound

Results on C and $E(R)$?

Capacity

- Hausladen-Jozsa-Schumacher-Westmoreland-Wootters: (1996) pure states
- Holevo, Schumacher-Westmorelan (1998): general states

$E(R)$ - achievability

- Burnashev-Holevo (1998): random coding for pure states
- Holevo (2000): expurgate bound (general case)
- Hayashi (2006): best "random coding" bound for mixed states
- **Missing**: conjectured Gallager-like random coding exponent!

$E(R)$ - converse

- Dalai (2012): sphere-packing
- (using Berlekamp): zero-rate upper bound

Capacity

- Hausladen-Jozsa-Schumacher-Westmoreland-Wootters: (1996) pure states
- Holevo, Schumacher-Westmorelan (1998): general states

$E(R)$ - achievability

- Burnashev-Holevo (1998): random coding for pure states
- Holevo (2000): expurgate bound (general case)
- Hayashi (2006): best "random coding" bound for mixed states
- **Missing**: conjectured Gallager-like random coding exponent!

$E(R)$ - converse

- Dalai (2012): sphere-packing
- (using Berlekamp): zero-rate upper bound

Results on C and $E(R)$?

Capacity

- Hausladen-Jozsa-Schumacher-Westmoreland-Wootters: (1996) pure states
- Holevo, Schumacher-Westmorelan (1998): general states

$E(R)$ - achievability

- Burnashev-Holevo (1998): random coding for pure states
- Holevo (2000): expurgate bound (general case)
- Hayashi (2006): best "random coding" bound for mixed states
- **Missing**: conjectured Gallager-like random coding exponent!

$E(R)$ - converse

- Dalai (2012): sphere-packing
- (using Berlekamp): zero-rate upper bound

Results on C and $E(R)$?

Capacity

- Hausladen-Jozsa-Schumacher-Westmoreland-Wootters: (1996) pure states
- Holevo, Schumacher-Westmorelan (1998): general states

$E(R)$ - achievability

- Burnashev-Holevo (1998): random coding for pure states
- Holevo (2000): expurgate bound (general case)
- Hayashi (2006): best "random coding" bound for mixed states
- **Missing**: conjectured Gallager-like random coding exponent!

$E(R)$ - converse

- Dalai (2012): sphere-packing
- (using Berlekamp): zero-rate upper bound

- Consider again binary hypothesis testing
- Try both MIT approach and Harountunian's approach

Quantum Binary Hypothesis Testing

- Here σ_0, σ_1 are density operators, with

$$\mathsf{P}_{\mathrm{e}|\sigma_0} = \operatorname{Tr} \sigma_0^{\otimes n}(I - \Pi) \qquad \mathsf{P}_{\mathrm{e}|\sigma_1} = \operatorname{Tr} \sigma_1^{\otimes n}\Pi$$

- Error exponents:

$$-\frac{1}{n}\log \mathsf{P}_{\mathrm{e}|\sigma_0} = -\mu(s) + s\mu'(s) + o(1)$$
$$-\frac{1}{n}\log \mathsf{P}_{\mathrm{e}|\sigma_1} = -\mu(s) - (1-s)\mu'(s) + o(1)$$

where

$$\mu(s) = \log \operatorname{Tr} \sigma_0^{1-s}\sigma_1^{s}.$$
$$= -sD_{1-s}(\sigma_0\|\sigma_1)$$

and

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

Quantum Binary Hypothesis Testing

- Here σ_0, σ_1 are density operators, with

$$\mathsf{P}_{\mathrm{e}|\sigma_0} = \operatorname{Tr} \sigma_0^{\otimes n}(I - \Pi) \qquad \mathsf{P}_{\mathrm{e}|\sigma_1} = \operatorname{Tr} \sigma_1^{\otimes n}\Pi$$

- Error exponents:

$$-\frac{1}{n} \log \mathsf{P}_{\mathrm{e}|\sigma_0} = -\mu(s) + s\mu'(s) + o(1)$$
$$-\frac{1}{n} \log \mathsf{P}_{\mathrm{e}|\sigma_1} = -\mu(s) - (1-s)\mu'(s) + o(1)$$

where

$$\begin{aligned}\mu(s) &= \log \operatorname{Tr} \sigma_0^{1-s}\sigma_1^{s}.\\ &= -sD_{1-s}(\sigma_0\|\sigma_1)\end{aligned}$$

and

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

- Upon differentiation, one finds for example for $\mathsf{P}_{\mathrm{e}|\sigma_0}$

$$-\frac{1}{n}\log \mathsf{P}_{\mathrm{e}|\sigma_0} = -\log \mathrm{Tr}(\sigma_0^{1-s}\sigma_1^s)+\mathrm{Tr}\left[\frac{\sigma_0^{1-s}\sigma_1^s}{\mathrm{Tr}\,\sigma_0^{1-s}\sigma_1^s}\left(\log\sigma_1^s - \log\sigma_0^s\right)\right]+o$$

- When σ_0 and σ_1 commute, define

$$\sigma_s = \frac{\sigma_0^{1-s}\sigma_1^s}{\mathrm{Tr}\,\sigma_0^{1-s}\sigma_1^s}$$

and use $\log\sigma_1^s - \log\sigma_0^s = \log\sigma_0^{1-s}\sigma_1^s - \log\sigma_0$.

- This gives for example (same for $\mathsf{P}_{\mathrm{e}|\sigma_1}$)

$$\begin{aligned}-\frac{1}{n}\log \mathsf{P}_{\mathrm{e}|\sigma_0} &= \mathrm{Tr}\,\sigma_s(\log\sigma_s - \log\sigma_0) + o(1)\\ &= D(\sigma_s\|\sigma_0) + o(1).\end{aligned}$$

- But if σ_0, σ_1 do not commute, this form does not hold!

- Upon differentiation, one finds for example for $\mathsf{P}_{\mathrm{e}|\sigma_0}$

$$-\frac{1}{n}\log \mathsf{P}_{\mathrm{e}|\sigma_0} = -\log \mathrm{Tr}(\sigma_0^{1-s}\sigma_1^s) + \mathrm{Tr}\left[\frac{\sigma_0^{1-s}\sigma_1^s}{\mathrm{Tr}\,\sigma_0^{1-s}\sigma_1^s}\left(\log\sigma_1^s - \log\sigma_0^s\right)\right] + o$$

- When σ_0 and σ_1 commute, define

$$\sigma_s = \frac{\sigma_0^{1-s}\sigma_1^s}{\mathrm{Tr}\,\sigma_0^{1-s}\sigma_1^s}$$

  and use $\log\sigma_1^s - \log\sigma_0^s = \log\sigma_0^{1-s}\sigma_1^s - \log\sigma_0$.

- This gives for example (same for $\mathsf{P}_{\mathrm{e}|\sigma_1}$)

$$\begin{aligned}-\frac{1}{n}\log \mathsf{P}_{\mathrm{e}|\sigma_0} &= \mathrm{Tr}\,\sigma_s(\log\sigma_s - \log\sigma_0) + o(1)\\ &= D(\sigma_s\|\sigma_0) + o(1).\end{aligned}$$

- But if σ_0, σ_1 do not commute, this form does not hold!

Quantum Binary Hypothesis Testing

- Upon differentiation, one finds for example for $\mathsf{P}_{\mathrm{e}|\sigma_0}$

$$-\frac{1}{n}\log \mathsf{P}_{\mathrm{e}|\sigma_0} = -\log \mathrm{Tr}(\sigma_0^{1-s}\sigma_1^s) + \mathrm{Tr}\left[\frac{\sigma_0^{1-s}\sigma_1^s}{\mathrm{Tr}\,\sigma_0^{1-s}\sigma_1^s}\left(\log\sigma_1^s - \log\sigma_0^s\right)\right] +$$

- When σ_0 and σ_1 commute, define

$$\sigma_s = \frac{\sigma_0^{1-s}\sigma_1^s}{\mathrm{Tr}\,\sigma_0^{1-s}\sigma_1^s}$$

  and use $\log\sigma_1^s - \log\sigma_0^s = \log\sigma_0^{1-s}\sigma_1^s - \log\sigma_0$.

- This gives for example (same for $\mathsf{P}_{\mathrm{e}|\sigma_1}$)

$$\begin{aligned} -\frac{1}{n}\log \mathsf{P}_{\mathrm{e}|\sigma_0} &= \mathrm{Tr}\,\sigma_s(\log\sigma_s - \log\sigma_0) + o(1) \\ &= D(\sigma_s\|\sigma_0) + o(1). \end{aligned}$$

- But if σ_0, σ_1 do not commute, this form does not hold!

- Upon differentiation, one finds for example for $\mathsf{P}_{\mathrm{e}|\sigma_0}$

$$-\frac{1}{n}\log \mathsf{P}_{\mathrm{e}|\sigma_0} = -\log \mathrm{Tr}(\sigma_0^{1-s}\sigma_1^s) + \mathrm{Tr}\left[\frac{\sigma_0^{1-s}\sigma_1^s}{\mathrm{Tr}\,\sigma_0^{1-s}\sigma_1^s}(\log\sigma_1^s - \log\sigma_0^s)\right] + $$

- When σ_0 and σ_1 commute, define

$$\sigma_s = \frac{\sigma_0^{1-s}\sigma_1^s}{\mathrm{Tr}\,\sigma_0^{1-s}\sigma_1^s}$$

  and use $\log\sigma_1^s - \log\sigma_0^s = \log\sigma_0^{1-s}\sigma_1^s - \log\sigma_0$.
- This gives for example (same for $\mathsf{P}_{\mathrm{e}|\sigma_1}$)

$$\begin{aligned}-\frac{1}{n}\log \mathsf{P}_{\mathrm{e}|\sigma_0} &= \mathrm{Tr}\,\sigma_s(\log\sigma_s - \log\sigma_0) + o(1)\\ &= D(\sigma_s\|\sigma_0) + o(1).\end{aligned}$$

- But if σ_0, σ_1 do not commute, this form does not hold!

Example

- Non-orthogonal pure states $\sigma_0 = |\psi_0\rangle\langle\psi_0|$ and $\sigma_1 = |\psi_1\rangle\langle\psi_1|$
- Since $\sigma_0^{1-s} = \sigma_0$ and $\sigma_1^s = \sigma_1$

$$\mu(s) = \log |\langle\psi_0|\psi_1\rangle|^2$$

- At least one of the two error exponents is not larger than $-\log |\langle\psi_0|\psi_1\rangle|^2$.
- Thus, error exponents cannot be expressed as $D(\sigma_s\|\sigma_i)$

$$D(\rho\|\sigma_i) = \begin{cases} 0 & \rho = \sigma_i \\ +\infty & \rho \neq \sigma_i \end{cases} \quad , i = 0, 1 \,,$$

when σ_0 and σ_1 are pure.

Example

- Non-orthogonal pure states $\sigma_0 = |\psi_0\rangle\langle\psi_0|$ and $\sigma_1 = |\psi_1\rangle\langle\psi_1|$
- Since $\sigma_0^{1-s} = \sigma_0$ and $\sigma_1^s = \sigma_1$

$$\mu(s) = \log |\langle\psi_0|\psi_1\rangle|^2$$

- At least one of the two error exponents is not larger than $-\log |\langle\psi_0|\psi_1\rangle|^2$.
- Thus, error exponents cannot be expressed as $D(\sigma_s\|\sigma_i)$

$$D(\rho\|\sigma_i) = \begin{cases} 0 & \rho = \sigma_i \\ +\infty & \rho \neq \sigma_i \end{cases}, i = 0, 1,$$

when σ_0 and σ_1 are pure.

Example

- Non-orthogonal pure states $\sigma_0 = |\psi_0\rangle\langle\psi_0|$ and $\sigma_1 = |\psi_1\rangle\langle\psi_1|$
- Since $\sigma_0^{1-s} = \sigma_0$ and $\sigma_1^s = \sigma_1$

$$\mu(s) = \log |\langle\psi_0|\psi_1\rangle|^2$$

- At least one of the two error exponents is not larger than $-\log |\langle\psi_0|\psi_1\rangle|^2$.
- Thus, error exponents cannot be expressed as $D(\sigma_s\|\sigma_i)$

$$D(\rho\|\sigma_i) = \begin{cases} 0 & \rho = \sigma_i \\ +\infty & \rho \neq \sigma_i \end{cases} , i = 0, 1 \, ,$$

when σ_0 and σ_1 are pure.

Example

- Non-orthogonal pure states $\sigma_0 = |\psi_0\rangle\langle\psi_0|$ and $\sigma_1 = |\psi_1\rangle\langle\psi_1|$
- Since $\sigma_0^{1-s} = \sigma_0$ and $\sigma_1^s = \sigma_1$

$$\mu(s) = \log |\langle\psi_0|\psi_1\rangle|^2$$

- At least one of the two error exponents is not larger than $-\log |\langle\psi_0|\psi_1\rangle|^2$.
- Thus, error exponents cannot be expressed as $D(\sigma_s \| \sigma_i)$

$$D(\rho\|\sigma_i) = \begin{cases} 0 & \rho = \sigma_i \\ +\infty & \rho \neq \sigma_i \end{cases} \quad , i = 0, 1\,,$$

when σ_0 and σ_1 are pure.

Classical-quantum sphere packing

Channel and coding scheme

- W_x are density operators (classical case: diagonal)
- M codewords $\{\boldsymbol{x}_1, \dots \boldsymbol{x}_M\}$, where $\boldsymbol{x} \mapsto \boldsymbol{W}_{\boldsymbol{x}} = W_{x_1} \otimes \cdots \otimes W_{x_n}$
- Decoder: POVM $\{\boldsymbol{\Pi}_1, \dots, \boldsymbol{\Pi}_M\}$ and $P_{m'|m} = \operatorname{Tr} \boldsymbol{W}_{\boldsymbol{x}_m} \boldsymbol{\Pi}_{m'}$

MIT proof

- Extends using quantum Rényi divergence $D_\alpha(\rho \| \sigma)$
- Matches achievability at high rates for pure-state channels
- Auxiliary Q does *not* induce auxiliary channel V

Haroutunian's approach

- Extends using quantum KL divergence
- Trivial bound for pure-state channels:

$$E(R, P) \leq \infty, \quad R < I(P, W)$$

Channel and coding scheme

- W_x are density operators (classical case: diagonal)
- M codewords $\{\boldsymbol{x}_1, \ldots \boldsymbol{x}_M\}$, where $\boldsymbol{x} \mapsto \boldsymbol{W}_{\boldsymbol{x}} = W_{x_1} \otimes \cdots \otimes W_{x_n}$
- Decoder: POVM $\{\boldsymbol{\Pi}_1, \ldots, \boldsymbol{\Pi}_M\}$ and $P_{m'|m} = \operatorname{Tr} \boldsymbol{W}_{\boldsymbol{x}_m} \boldsymbol{\Pi}_{m'}$

MIT proof

- Extends using quantum Rényi divergence $D_\alpha(\rho\|\sigma)$
- Matches achievability at high rates for pure-state channels
- Auxiliary Q does *not* induce auxiliary channel V

Haroutunian's approach

- Extends using quantum KL divergence
- Trivial bound for pure-state channels:

$$E(R,P) \leq \infty, \quad R < I(P,W)$$

Classical-quantum sphere packing

Channel and coding scheme

- W_x are density operators (classical case: diagonal)
- M codewords $\{\boldsymbol{x}_1, \dots \boldsymbol{x}_M\}$, where $\boldsymbol{x} \mapsto \boldsymbol{W}_{\boldsymbol{x}} = W_{x_1} \otimes \cdots \otimes W_{x_n}$
- Decoder: POVM $\{\boldsymbol{\Pi}_1, \dots, \boldsymbol{\Pi}_M\}$ and $P_{m'|m} = \operatorname{Tr} \boldsymbol{W}_{\boldsymbol{x}_m} \boldsymbol{\Pi}_{m'}$

MIT proof

- Extends using quantum Rényi divergence $D_\alpha(\rho \| \sigma)$
- Matches achievability at high rates for pure-state channels
- Auxiliary Q does *not* induce auxiliary channel V

Haroutunian's approach

- Extends using quantum KL divergence
- Trivial bound for pure-state channels:

$$E(R, P) \leq \infty, \quad R < I(P, W)$$

Haroutunian's bound for pure-state channels

- The bound:

$$\frac{1}{n}\log\frac{1}{\mathsf{P}_{\mathrm{e}|\boldsymbol{W}_{\boldsymbol{x}_m}}} \leq \inf_{V:I(P,V)<R} D(V\|W|P)(1+o(1))$$

- Remember, for pure σ

$$D(\rho\|\sigma) = \begin{cases} 0 & \rho = \sigma \\ +\infty & \rho \neq \sigma \end{cases},$$

- If $R < I(P,W)$ then $I(P,V) < I(P,W)$
- Thus, we can only optimize over V such that $V_x \neq W_x$ for some "used" x
- Any such V gives $D(V\|W|P) = \infty$

Haroutunian's bound for pure-state channels

- The bound:

$$\frac{1}{n}\log\frac{1}{\mathsf{P}_{\mathrm{e}|\boldsymbol{W}_{\boldsymbol{x}_m}}} \leq \inf_{V:I(P,V)<R} D(V\|W|P)(1+o(1))$$

- Remember, for pure σ

$$D(\rho\|\sigma) = \begin{cases} 0 & \rho = \sigma \\ +\infty & \rho \neq \sigma \end{cases},$$

- If $R < I(P,W)$ then $I(P,V) < I(P,W)$
- Thus, we can only optimize over V such that $V_x \neq W_x$ for some "used" x
- Any such V gives $D(V\|W|P) = \infty$

Haroutunian's bound for pure-state channels

- The bound:

$$\frac{1}{n}\log\frac{1}{\mathsf{P}_{\mathrm{e}|\boldsymbol{W}_{\boldsymbol{x}_m}}} \leq \inf_{V:I(P,V)<R} D(V\|W|P)(1+o(1))$$

- Remember, for pure σ

$$D(\rho\|\sigma) = \begin{cases} 0 & \rho = \sigma \\ +\infty & \rho \neq \sigma \end{cases},$$

- If $R < I(P,W)$ then $I(P,V) < I(P,W)$
- Thus, we can only optimize over V such that $V_x \neq W_x$ for some "used" x
- Any such V gives $D(V\|W|P) = \infty$

Haroutunian's bound for pure-state channels

- The bound:

$$\frac{1}{n}\log\frac{1}{\mathsf{P}_{\mathrm{e}|\boldsymbol{W}_{\boldsymbol{x}_m}}} \leq \inf_{V:I(P,V)<R} D(V\|W|P)(1+o(1))$$

- Remember, for pure σ

$$D(\rho\|\sigma) = \begin{cases} 0 & \rho = \sigma \\ +\infty & \rho \neq \sigma \end{cases},$$

- If $R < I(P,W)$ then $I(P,V) < I(P,W)$
- Thus, we can only optimize over V such that $V_x \neq W_x$ for some "used" x
- Any such V gives $D(V\|W|P) = \infty$

Haroutunian's bound for pure-state channels

- The bound:

$$\frac{1}{n}\log\frac{1}{\mathsf{P}_{\mathrm{e}|\boldsymbol{W}_{\boldsymbol{x}_m}}} \leq \inf_{V:I(P,V)<R} D(V\|W|P)(1+o(1))$$

- Remember, for pure σ

$$D(\rho\|\sigma) = \begin{cases} 0 & \rho = \sigma \\ +\infty & \rho \neq \sigma \end{cases},$$

- If $R < I(P,W)$ then $I(P,V) < I(P,W)$
- Thus, we can only optimize over V such that $V_x \neq W_x$ for some "used" x
- Any such V gives $D(V\|W|P) = \infty$

What is the problem here?

What happened

- Using a constant Q we get a good bound
- Using an optimal channel V we don't
- Impossible... a constant Q is a "dummy channel" with $V_x = Q$

MIT Proof

- Dummy Q
- Converse for Q of the form $\operatorname{Tr} \boldsymbol{Q}\boldsymbol{\Pi}_m \leq e^{-nR}$
- Lower bound $\operatorname{Tr} \boldsymbol{W}_{\boldsymbol{x}_m}\boldsymbol{\Pi}_m$ using BHT between $\boldsymbol{Q}$ and $\boldsymbol{W}_{\boldsymbol{x}_m}$ in the regime where both error probabilities vanish exponentially

Haroutunian

- General channel V with $I(P, V) < R$
- Converse for V of the form $\operatorname{Tr} \boldsymbol{V}_{\boldsymbol{x}_m}\boldsymbol{\Pi}_m = o(1)$
- BHT between $\boldsymbol{V}_{\boldsymbol{x}_m}$ and $\boldsymbol{W}_{\boldsymbol{x}_m}$ in Stein's regime

What is the problem here?

What happened

- Using a constant Q we get a good bound
- Using an optimal channel V we don't
- Impossible... a constant Q is a "dummy channel" with $V_x = Q$

MIT Proof

- Dummy Q
- Converse for Q of the form $\operatorname{Tr} \boldsymbol{Q}\boldsymbol{\Pi}_m \leq e^{-nR}$
- Lower bound $\operatorname{Tr} \boldsymbol{W}_{\boldsymbol{x}_m}\boldsymbol{\Pi}_m$ using BHT between $\boldsymbol{Q}$ and $\boldsymbol{W}_{\boldsymbol{x}_m}$ in the regime where both error probabilities vanish exponentially

Haroutunian

- General channel V with $I(P,V) < R$
- Converse for V of the form $\operatorname{Tr} \boldsymbol{V}_{\boldsymbol{x}_m}\boldsymbol{\Pi}_m = o(1)$
- BHT between $\boldsymbol{V}_{\boldsymbol{x}_m}$ and $\boldsymbol{W}_{\boldsymbol{x}_m}$ in Stein's regime

What is the problem here?

What happened

- Using a constant Q we get a good bound
- Using an optimal channel V we don't
- Impossible... a constant Q is a "dummy channel" with $V_x = Q$

MIT Proof

- Dummy Q
- Converse for Q of the form $\operatorname{Tr} \boldsymbol{Q}\boldsymbol{\Pi}_m \leq e^{-nR}$
- Lower bound $\operatorname{Tr} \boldsymbol{W}_{\boldsymbol{x}_m}\boldsymbol{\Pi}_m$ using BHT between $\boldsymbol{Q}$ and $\boldsymbol{W}_{\boldsymbol{x}_m}$ in the regime where both error probabilities vanish exponentially

Haroutunian

- General channel V with $I(P,V) < R$
- Converse for V of the form $\operatorname{Tr} \boldsymbol{V}_{\boldsymbol{x}_m}\boldsymbol{\Pi}_m = o(1)$
- BHT between $\boldsymbol{V}_{\boldsymbol{x}_m}$ and $\boldsymbol{W}_{\boldsymbol{x}_m}$ in Stein's regime

What is the problem here?

What happened

- Using a constant Q we get a good bound
- Using an optimal channel V we don't
- Impossible... a constant Q is a "dummy channel" with $V_x = Q$

MIT Proof

- Dummy Q
- Converse for Q of the form $\mathrm{Tr}\,\boldsymbol{Q}\boldsymbol{\Pi}_m \leq e^{-nR}$
- Lower bound $\mathrm{Tr}\,\boldsymbol{W}_{\boldsymbol{x}_m}\boldsymbol{\Pi}_m$ using BHT between $\boldsymbol{Q}$ and $\boldsymbol{W}_{\boldsymbol{x}_m}$ in the regime where both error probabilities vanish exponentially

Haroutunian

- General channel V with $I(P, V) < R$
- Converse for V of the form $\mathrm{Tr}\,\boldsymbol{V}_{\boldsymbol{x}_m}\boldsymbol{\Pi}_m = o(1)$ ... too weak
- BHT between $\boldsymbol{V}_{\boldsymbol{x}_m}$ and $\boldsymbol{W}_{\boldsymbol{x}_m}$ in Stein's regime

Reconsidering a general V

What we should do

- Take an auxiliary V with $I(P,V) < R$
- Compute the correct strong converse $\operatorname{Tr} \boldsymbol{V}_{\boldsymbol{x}_m} \boldsymbol{\Pi}_m = e^{-nE_{sc}(R,P)}$
- BHT between $\boldsymbol{V}_{\boldsymbol{x}_m}$ and $\boldsymbol{W}_{\boldsymbol{x}_m}$ in the regime where both error probabilities vanish exponentially

Classical case

- Choosing $I(P,V) = 0$ (MIT) or $I(P,V) = R - \epsilon$ (Haroutunian) makes no difference
- No other choice can do better (I guess... list decoding)
- The strong converse exponent for V, and the BHT between $\boldsymbol{V}_{\boldsymbol{x}_m}$ and $\boldsymbol{W}_{\boldsymbol{x}_m}$ both involve Rény divergences

What we should do

- Take an auxiliary V with $I(P,V) < R$
- Compute the correct strong converse $\operatorname{Tr} \boldsymbol{V}_{\boldsymbol{x}_m} \boldsymbol{\Pi}_m = e^{-nE_{sc}(R,P)}$
- BHT between $\boldsymbol{V}_{\boldsymbol{x}_m}$ and $\boldsymbol{W}_{\boldsymbol{x}_m}$ in the regime where both error probabilities vanish exponentially

Classical case

- Choosing $I(P,V) = 0$ (MIT) or $I(P,V) = R - \epsilon$ (Haroutunian) makes no difference
- No other choice can do better (I guess... list decoding)
- The strong converse exponent for V, and the BHT between $\boldsymbol{V}_{\boldsymbol{x}_m}$ and $\boldsymbol{W}_{\boldsymbol{x}_m}$ both involve Rény divergences

What we should do

- Take an auxiliary V with $I(P, V) < R$
- Compute the correct strong converse $\operatorname{Tr} \boldsymbol{V}_{\boldsymbol{x}_m} \boldsymbol{\Pi}_m = e^{-nE_{sc}(R,P)}$
- BHT between $\boldsymbol{V}_{\boldsymbol{x}_m}$ and $\boldsymbol{W}_{\boldsymbol{x}_m}$ in the regime where both error probabilities vanish exponentially

Classical case

- Choosing $I(P, V) = 0$ (MIT) or $I(P, V) = R - \epsilon$ (Haroutunian) makes no difference
- No other choice can do better (I guess... list decoding)
- The strong converse exponent for V, and the BHT between $\boldsymbol{V}_{\boldsymbol{x}_m}$ and $\boldsymbol{W}_{\boldsymbol{x}_m}$ both involve Rény divergences

What we should do

- Take an auxiliary V with $I(P, V) < R$
- Compute the correct strong converse $\operatorname{Tr} \boldsymbol{V}_{\boldsymbol{x}_m} \boldsymbol{\Pi}_m = e^{-nE_{sc}(R,P)}$
- BHT between $\boldsymbol{V}_{\boldsymbol{x}_m}$ and $\boldsymbol{W}_{\boldsymbol{x}_m}$ in the regime where both error probabilities vanish exponentially

Classical case

- Choosing $I(P, V) = 0$ (MIT) or $I(P, V) = R - \epsilon$ (Haroutunian) makes no difference
- No other choice can do better (I guess... list decoding)
- The strong converse exponent for V, and the BHT between $\boldsymbol{V}_{\boldsymbol{x}_m}$ and $\boldsymbol{W}_{\boldsymbol{x}_m}$ both involve Rény divergences

What we should do

- Take an auxiliary V with $I(P, V) < R$
- Compute the correct strong converse $\operatorname{Tr} \boldsymbol{V}_{\boldsymbol{x}_m} \boldsymbol{\Pi}_m = e^{-nE_{sc}(R,P)}$
- BHT between $\boldsymbol{V}_{\boldsymbol{x}_m}$ and $\boldsymbol{W}_{\boldsymbol{x}_m}$ in the regime where both error probabilities vanish exponentially

Classical case

- Choosing $I(P, V) = 0$ (MIT) or $I(P, V) = R - \epsilon$ (Haroutunian) makes no difference
- No other choice can do better (I guess... list decoding)
- The strong converse exponent for V, and the BHT between $\boldsymbol{V}_{\boldsymbol{x}_m}$ and $\boldsymbol{W}_{\boldsymbol{x}_m}$ both involve Rény divergences

What we should do

- Take an auxiliary V with $I(P, V) < R$
- Compute the correct strong converse $\operatorname{Tr} \boldsymbol{V}_{\boldsymbol{x}_m} \boldsymbol{\Pi}_m = e^{-nE_{sc}(R,P)}$
- BHT between $\boldsymbol{V}_{\boldsymbol{x}_m}$ and $\boldsymbol{W}_{\boldsymbol{x}_m}$ in the regime where both error probabilities vanish exponentially

Classical case

- Choosing $I(P, V) = 0$ (MIT) or $I(P, V) = R - \epsilon$ (Haroutunian) makes no difference
- No other choice can do better (I guess... list decoding)
- The strong converse exponent for V, and the BHT between $\boldsymbol{V}_{\boldsymbol{x}_m}$ and $\boldsymbol{W}_{\boldsymbol{x}_m}$ both involve Rény divergences

Reconsidering a general V

What we should do

- Take an auxiliary V with $I(P,V) < R$
- Compute the correct strong converse $\operatorname{Tr} \boldsymbol{V}_{\boldsymbol{x}_m} \boldsymbol{\Pi}_m = e^{-nE_{sc}(R,P)}$
- BHT between $\boldsymbol{V}_{\boldsymbol{x}_m}$ and $\boldsymbol{W}_{\boldsymbol{x}_m}$ in the regime where both error probabilities vanish exponentially

Classical case

- Choosing $I(P,V) = 0$ (MIT) or $I(P,V) = R - \epsilon$ (Haroutunian) makes no difference
- No other choice can do better (I guess... list decoding)
- The strong converse exponent for V, and the BHT between $\boldsymbol{V}_{\boldsymbol{x}_m}$ and $\boldsymbol{W}_{\boldsymbol{x}_m}$ both involve Rény divergences

Classical-Quantum

- Choosing $I(P,V) = 0$ (MIT) or $I(P,V) = R - \epsilon$ (Haroutunian) does make a difference
- Is $I(P,V) = 0$ really optimal?
  $\rightarrow$ No matching achievability for mixed state channels.
- Strong converse exponent for c-q channels derived only very recently (Mosonyi and Ogawa 2014).
- Unlike the BHT between $\boldsymbol{V}_{\boldsymbol{x}_m}$ and $\boldsymbol{W}_{\boldsymbol{x}_m}$, strong converse involves so-called "sandwiched" Rényi divergence

$$\tilde{D}_\alpha(\rho,\sigma) = \frac{1}{\alpha-1}\log \operatorname{Tr}\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha.$$

Reconsidering a general V

Classical-Quantum

- Choosing $I(P,V)=0$ (MIT) or $I(P,V)=R-\epsilon$ (Haroutunian) does make a difference
- Is $I(P,V)=0$ really optimal?
  $\to$ No matching achievability for mixed state channels.
- Strong converse exponent for c-q channels derived only very recently (Mosonyi and Ogawa 2014).
- Unlike the BHT between $\boldsymbol{V}_{\boldsymbol{x}_m}$ and $\boldsymbol{W}_{\boldsymbol{x}_m}$, strong converse involves so-called “sandwiched” Rényi divergence

$$\tilde{D}_\alpha(\rho,\sigma)=\frac{1}{\alpha-1}\log \operatorname{Tr}\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha.$$

Classical-Quantum

- Choosing $I(P,V)=0$ (MIT) or $I(P,V)=R-\epsilon$ (Haroutunian) does make a difference
- Is $I(P,V)=0$ really optimal?
  $\rightarrow$ No matching achievability for mixed state channels.
- Strong converse exponent for c-q channels derived only very recently (Mosonyi and Ogawa 2014).
- Unlike the BHT between $\boldsymbol{V}_{\boldsymbol{x}_m}$ and $\boldsymbol{W}_{\boldsymbol{x}_m}$, strong converse involves so-called "sandwiched" Rényi divergence

$$\tilde{D}_\alpha(\rho,\sigma)=\frac{1}{\alpha-1}\log\mathrm{Tr}\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha.$$

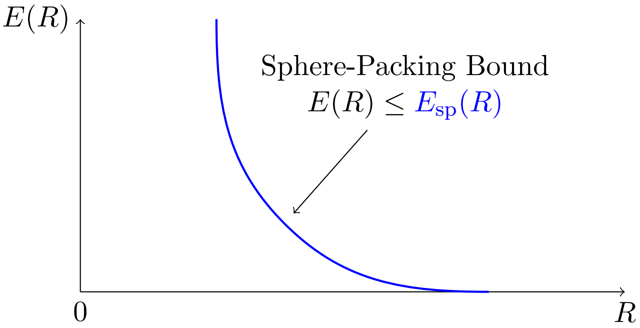
Reconsidering a general V

Classical-Quantum

- Choosing $I(P,V)=0$ (MIT) or $I(P,V)=R-\epsilon$ (Haroutunian) does make a difference
- Is $I(P,V)=0$ really optimal?
  $\rightarrow$ No matching achievability for mixed state channels.
- Strong converse exponent for c-q channels derived only very recently (Mosonyi and Ogawa 2014).
- Unlike the BHT between $\boldsymbol{V}_{\boldsymbol{x}_m}$ and $\boldsymbol{W}_{\boldsymbol{x}_m}$, strong converse involves so-called "sandwiched" Rényi divergence

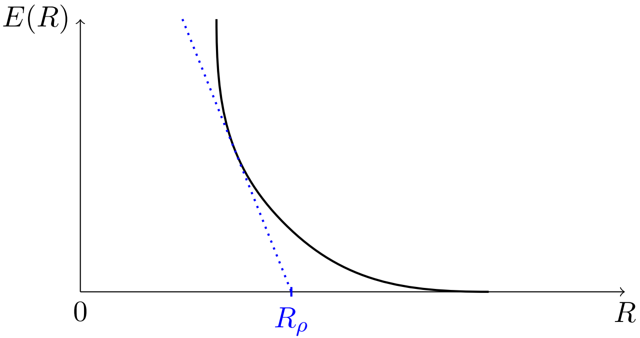
$$\tilde{D}_\alpha(\rho,\sigma)=\frac{1}{\alpha-1}\log\mathrm{Tr}\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right)^{\alpha}.$$


E(R)
0
R



Sphere packing

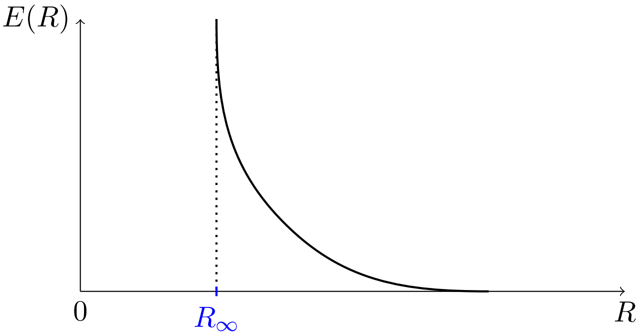
$$E_{\rm sp}(R) = \sup_{\rho \ge 0} \max_P \left[-\log \operatorname{Tr} \left(\sum_x P(x) W_x^{1/(1+\rho)} \right)^{1+\rho} - \rho R \right]$$



Minmax characterization

$$R_\rho = \min_F \max_x D_\alpha(W_x||F), \quad \alpha = 1/(1+\rho)$$

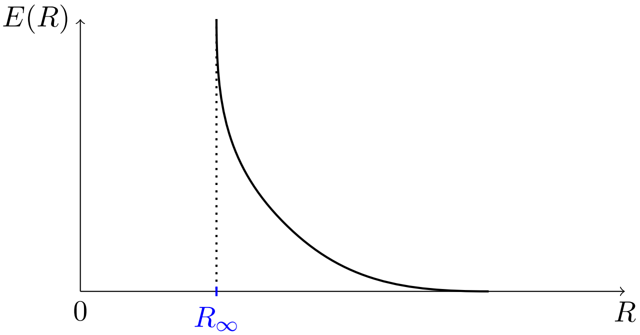
where F runs over density operators and
$D_\alpha(F_1||F_2) = \frac{1}{\alpha-1}\log \mathrm{Tr}(F_1^\alpha F_2^{1-\alpha})$ is the Rényi divergence



When $\rho \to \infty$

$$R_\infty = \min_F \max_x \log \frac{1}{\mathrm{Tr}\left(W_x^0 F\right)}$$

where W_x^0 is the projector onto the support of W_x



For pure-state channels $W_x = |u_x\rangle\langle u_x|$
Using pure-states $F = |f\rangle\langle f|$ we have $\mathrm{Tr}\left(W_x^0 F\right) = |\langle u_x|f\rangle|^2$.
So,

$$\begin{aligned} R_\infty &\leq \min_f \max_x \log \frac{1}{|\langle u_x|f\rangle|^2} \\ &= V(\{u_x\}) \end{aligned}$$

Orthonormal Representations and Auxiliary Channels

- For any representation $\{u_x\}$, the classical-quantum channel with pure-states $W_x = |u_x\rangle\langle u_x|$ satisfies $R_\infty \leq V(\{u_x\})$
- We can define
$$\vartheta_{sp} = \min_{\{W_x\}} R_\infty \quad (1)$$
where we minimize over all channels such that $\mathrm{Tr}\, W_x W_{x'} = 0$ if x and x' are not confusable
- Then
$$C_0 \leq \vartheta_{sp} \leq \vartheta$$
- Actually additional results in Lovász' paper imply $\vartheta \leq \vartheta_{sp}$ and hence $\vartheta_{sp} = \vartheta$.
- So, pure-state channels achieve the optimum in (1) and for some optimal channel some pure state $F = |f\rangle\langle f|$ achieves R_∞

Orthonormal Representations and Auxiliary Channels

- For any representation $\{u_x\}$, the classical-quantum channel with pure-states $W_x = |u_x\rangle\langle u_x|$ satisfies $R_\infty \leq V(\{u_x\})$
- We can define

$$\vartheta_{sp} = \min_{\{W_x\}} R_\infty \tag{1}$$

where we minimize over all channels such that $\operatorname{Tr} W_x W_{x'} = 0$ if x and x' are not confusable

- Then

$$C_0 \leq \vartheta_{sp} \leq \vartheta$$

- Actually additional results in Lovász' paper imply $\vartheta \leq \vartheta_{sp}$ and hence $\vartheta_{sp} = \vartheta$.
- So, pure-state channels achieve the optimum in (1) and for some optimal channel some pure state $F = |f\rangle\langle f|$ achieves R_∞

Orthonormal Representations and Auxiliary Channels

- For any representation $\{u_x\}$, the classical-quantum channel with pure-states $W_x = |u_x\rangle\langle u_x|$ satisfies $R_\infty \leq V(\{u_x\})$
- We can define
$$\vartheta_{sp} = \min_{\{W_x\}} R_\infty \quad (1)$$
where we minimize over all channels such that $\operatorname{Tr} W_x W_{x'} = 0$ if x and x' are not confusable
- Then
$$C_0 \leq \vartheta_{sp} \leq \vartheta$$
- Actually additional results in Lovász' paper imply $\vartheta \leq \vartheta_{sp}$ and hence $\vartheta_{sp} = \vartheta$.
- So, pure-state channels achieve the optimum in (1) and for some optimal channel some pure state $F = |f\rangle\langle f|$ achieves R_∞

Orthonormal Representations and Auxiliary Channels

- For any representation $\{u_x\}$, the classical-quantum channel with pure-states $W_x = |u_x\rangle\langle u_x|$ satisfies $R_\infty \leq V(\{u_x\})$
- We can define
$$\vartheta_{sp} = \min_{\{W_x\}} R_\infty \tag{1}$$
where we minimize over all channels such that $\operatorname{Tr} W_x W_{x'} = 0$ if x and x' are not confusable
- Then
$$C_0 \leq \vartheta_{sp} \leq \vartheta$$
- Actually additional results in Lovász' paper imply $\vartheta \leq \vartheta_{sp}$ and hence $\vartheta_{sp} = \vartheta$.
- So, pure-state channels achieve the optimum in (1) and for some optimal channel some pure state $F = |f\rangle\langle f|$ achieves R_∞

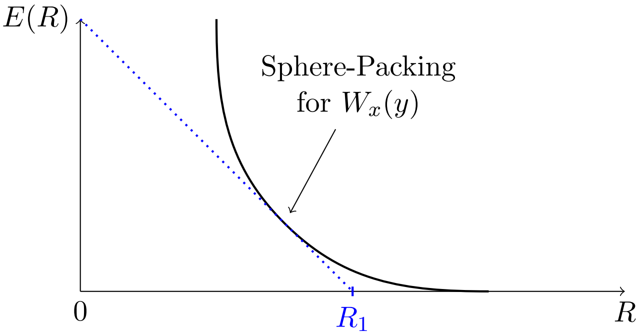
Orthonormal Representations and Auxiliary Channels

- For any representation $\{u_x\}$, the classical-quantum channel with pure-states $W_x = |u_x\rangle\langle u_x|$ satisfies $R_\infty \leq V(\{u_x\})$
- We can define
$$\vartheta_{sp} = \min_{\{W_x\}} R_\infty \tag{1}$$
where we minimize over all channels such that $\operatorname{Tr} W_x W_{x'} = 0$ if x and x' are not confusable
- Then
$$C_0 \leq \vartheta_{sp} \leq \vartheta$$
- Actually additional results in Lovász' paper imply $\vartheta \leq \vartheta_{sp}$ and hence $\vartheta_{sp} = \vartheta$.
- So, pure-state channels achieve the optimum in (1) and for some optimal channel some pure state $F = |f\rangle\langle f|$ achieves R_∞

Cutoff rates and R_∞

But... where are those cutoff rates?

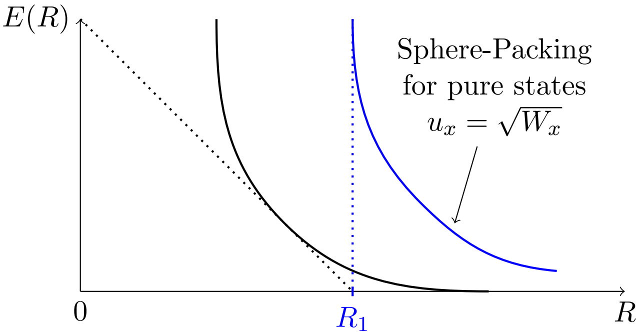
Cutoff rates and R_∞



But... where are those cutoff rates?

- We had previously identified R_1 with $V(\{\sqrt{W_x}\})$
- But then we ended up with a relation between ϑ and R_∞

Cutoff rates and R_∞



But... where are those cutoff rates?

- Mathematically, this is due to the fact that the cutoff rate of a channel W always equals the R_∞ rate of a pure-state channel with state vectors $u_x = \sqrt{W_x}$
- The true meaning of this... I do not know, but this sounds important

Suggested reading (to start with) I

- C. E. Shannon. "The Zero-Error Capacity of a Noisy Channel". In: *IRE Trans. Inform. Theory* IT-2 (1956), pp. 8–19.
- C. E. Shannon. "Certain results in coding theory for noisy channels". In: *Information and Control* 1 (1957), pp. 6–25.
- R. M. Fano. *Transmission of Information: A Statistical Theory of Communication.* Wiley, New York, 1961.
- R. G. Gallager. "A Simple Derivation of the Coding Theorem and Some Applications". In: *IEEE Trans. Inform. Theory* IT-11 (1965), pp. 3–18.
- C. E. Shannon, R. G. Gallager, and E. R. Berlekamp. "Lower Bounds to Error Probability for Coding in Discrete Memoryless Channels. I". In: *Information and Control* 10 (1967), pp. 65–103.

Suggested reading (to start with) II

- C. E. Shannon, R. G. Gallager, and E. R. Berlekamp. "Lower Bounds to Error Probability for Coding in Discrete Memoryless Channels. II". In: *Information and Control* 10 (1967), pp. 522–552.
- R. G. Gallager. *Information Theory and Reliable Communication*. Wiley, New York, 1968.
- E. A. Haroutunian. "Estimates of the Error Exponents for the semi-continuous memoryless channel". In: *(in Russian) Probl. Peredachi Inform.* 4.4 (1968), pp. 37–48.
- F. Jelinek. *Probabilistic Information Theory*. McGraw Hill, New York, 1968.
- R. E. Blahut. "Hypothesis testing and Information theory". In: *IEEE Trans. Inform. Theory* IT-20 (1974), pp. 405–417.

Suggested reading (to start with) III

- L. Lovász. "On the Shannon Capacity of a Graph". In: *IEEE Trans. Inform. Theory* 25.1 (1979), pp. 1–7.
- A. J. Viterbi and J. K. Omura. *Principles of Digital Communication and Coding*. McGraw-Hill, New York, 1979.
- I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1981.
- J. Korner and A. Orlitsky. "Zero-error information theory". In: *IEEE Trans. on Inform. Theory* 44.6 (Oct. 1998), pp. 2207–2229.
- A. S. Holevo. "Reliability Function of General Classical-Quantum Channel". In: *IEEE Trans. Inform. Theory* 46.6 (Sept. 2000), pp. 2256–2261.

Suggested reading (to start with) IV

- K. Audenaert, M. Nussbaum, A. Szkoła, and F. Verstraete. "Asymptotic Error Rates in Quantum Hypothesis Testing". In: *Communications in Mathematical Physics* 279 (1 2008), pp. 251–283.
- M. Nussbaum and A. Szkoła. "The Chernoff lower bound for symmetric quantum hypothesis testing". In: *Ann. Statist.* 37.2 (2009), pp. 1040–1057.
- H. Nagaoka. "The Converse Part of the Theorem for Quantum Hoeffding Bound". In: *arXiv:quant-ph/0611289v1* ().