Lovász's Theta Function, Rényi's Divergence and the Sphere-Packing Bound

Marco Dalai

Department of Information Engineering
University of Brescia - Italy

International Symposium on Information Theory, 2013

Classical DMCs

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

$$x \longrightarrow \boxed{W(y|x)} \longrightarrow y$$

- **Memoryless extension** $\mathcal{W}^{(n)} : \mathcal{X}^n \to \mathcal{Y}^n$

$$\mathbf{x} = (x_1, \ldots, x_n) \longrightarrow \boxed{W^{(n)}(\mathbf{y}|\mathbf{x})} \longrightarrow \mathbf{y} = (y_1, \ldots, y_n)$$

- **Code**: M codewords $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M\} \subset \mathcal{X}^n$
- **Decoder**: M disjoint decision regions $\{Y_1, \ldots, Y_M\} \subseteq \mathcal{Y}^n$
- **Probability of error**: $P_{e|m} = W^{(n)}(\overline{Y_m}|\mathbf{x}_m)$

Previous Results

Probability of error and reliability function

$$P_e \approx e^{-nE(R)}$$

Previous Results: Capacity C

Shannon, 1948

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

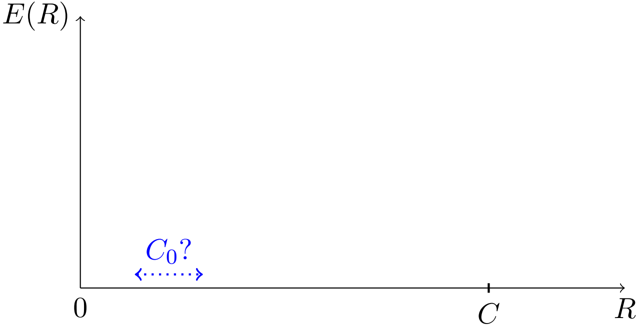
Previous Results: Capacity C

Shannon, 1948

$$\color{blue}{C} = \max_P \sum_{x,y} P(x)W(y|x) \log \frac{W(y|x)}{\sum_{x'} P(x')W(y|x')}$$

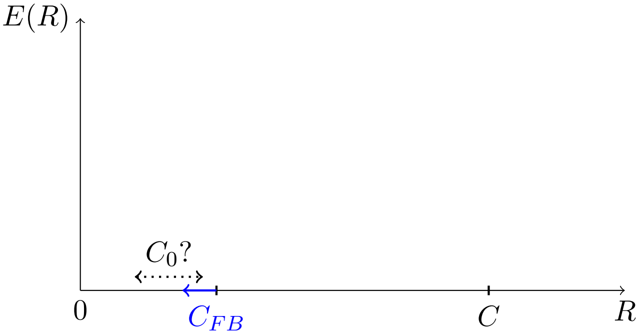
$$= \color{blue}{\min_Q \max_x} D(W(\cdot|x)||Q)$$

Previous Results: Zero-Error Capacity C_0



Shannon, 1956

Previous Results: Zero-Error Capacity C_0

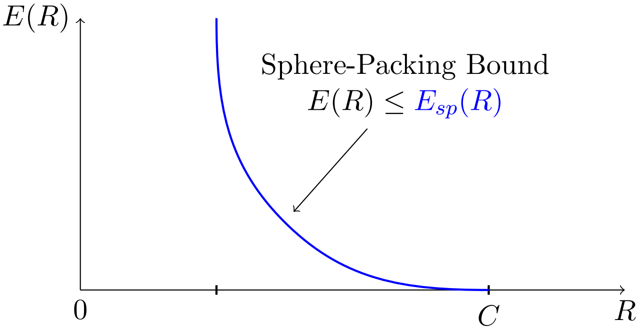


Shannon, 1956 (combinatorial)

Upper bounded by the zero-error capacity with feedback

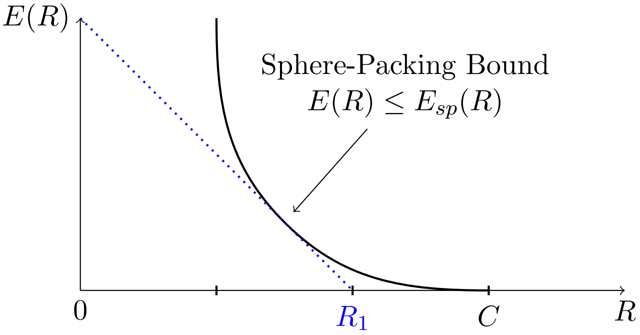
$$C_{FB} = \max_P \left[-\log \max_y \sum_{x:W(y|x)>0} P(x) \right]$$

Previous Results: Reliability Function $E(R)$



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

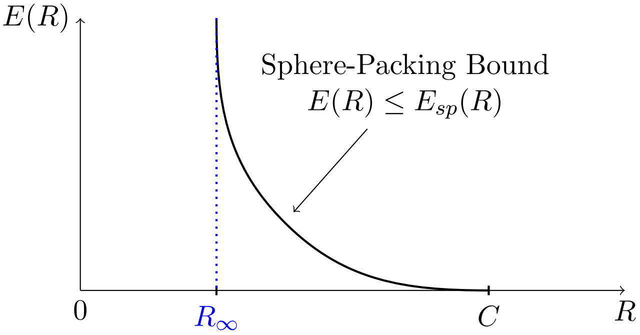
$$E_{sp}(R) = \sup_{\rho \geq 0} \max_{P} \left[-\log \sum_{y} \left(\sum_{x} P(x) W(y|x)^{1/(1+\rho)} \right)^{1+\rho} - \rho R \right]$$



Cutoff rate

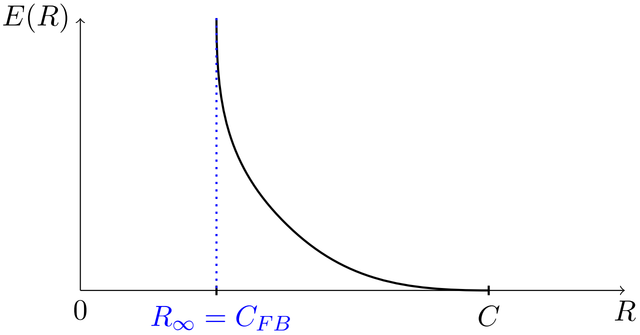
$$R_1 = \max_P \left[-\log \sum_y \left(\sum_x P(x) \sqrt{W(y|x)} \right)^2 \right]$$

Previous Results: Reliability Function $E(R)$



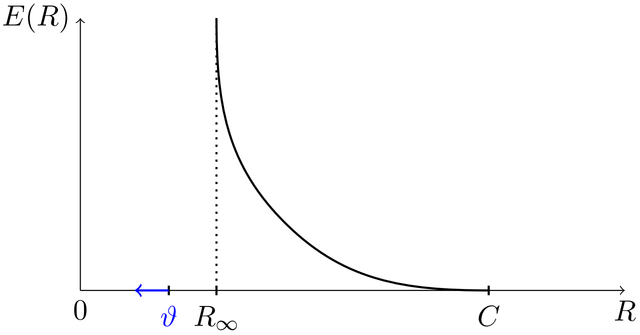
- $E_{sp}(R)$ is finite for $R > R_\infty$
- Hence, $C_0 \le R_\infty$

Previous Results: Reliability Function $E(R)$



- $E_{sp}(R)$ is finite for $R > R_\infty$
- Hence, $C_0 \leq R_\infty$
- It turns out that $R_\infty = C_{FB}$ (whenever $C_0 > 0$)
- Same bound for C_0 using combinatorial or probabilistic approaches
- We can then minimize R_∞ over auxiliary channels

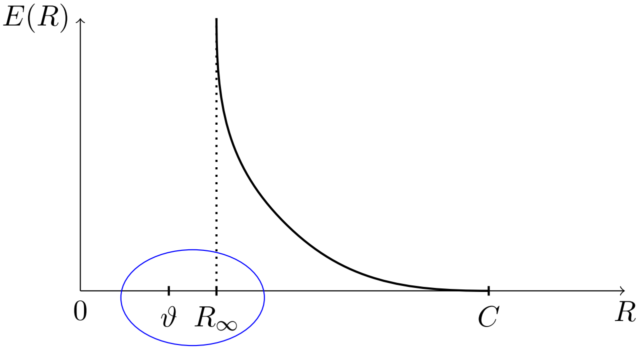
Previous Results: Lovász' upper bound to C_0



Lovász, 1979

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

This Paper

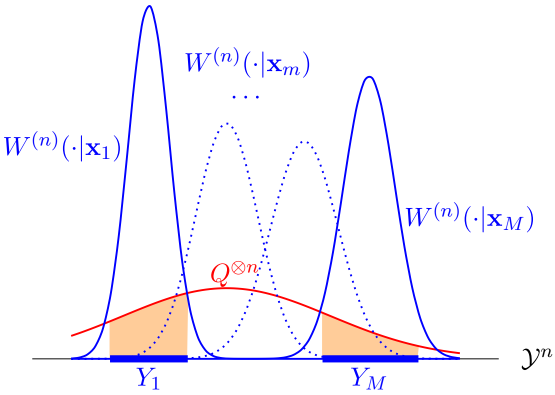


Lovász, 1979

- New bound: $C_0 \leq \vartheta$
- Using *geometric representations of graphs*
- Combinatorial, apparently no connection with probability
- **This paper**: $R_\infty = \vartheta$ for (optimal) classical-quantum channels

Sphere-Packing Bound: Sketch of Proof

Binary hypothesis testing: compare $Q^{\otimes n}$ with $W^{(n)}(\cdot|\mathbf{x}_m)$



- The decision regions $Y_1, \ldots, Y_M$ are disjoint
- $Q^{\otimes n}(Y_m) \leq 1/M$ for at least one m, since $\int Q^{\otimes n} = 1$
- $W^{(n)}(\overline{Y_m}|\mathbf{x}_m) \geq e^{-n(E_{sp}(R)+o(1))}$ using Neyman-Pearson/Chernoff

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

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

- **Example**: $u_x = \sqrt{W(\cdot|x)}$
- **Value**:

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

  The optimal c is the *handle* of the representation
- **The bound**:

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

- **Theta function**:

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

Lovász' Bound: Definitions

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

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

- **Example**: $u_x = \sqrt{W(\cdot|x)}$
- **Value**:

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

  The optimal c is the *handle* of the representation
- **The bound**:

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

- **Theta function**:

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

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

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

- **Example**: $u_x = \sqrt{W(\cdot|x)}$
- **Value**:

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

  The optimal c is the *handle* of the representation

- **The bound**:

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

- **Theta function**:

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

Lovász' Bound: Definitions

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

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

- **Example**: $u_x = \sqrt{W(\cdot|x)}$
- **Value**:

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

  The optimal c is the *handle* of the representation
- **The bound**:

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

- **Theta function**:

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

Lovász' Bound: Definitions

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

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

- **Example**: $u_x = \sqrt{W(\cdot|x)}$
- **Value**:

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

  The optimal c is the *handle* of the representation
- **The bound**:

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

- **Theta function**:

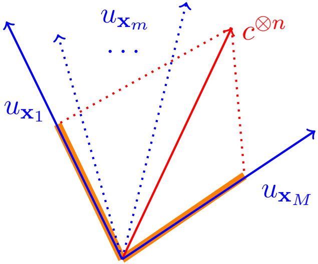
$$\vartheta = \min_{\{u_x\}} V(\{u_x\})$$

Lovász' Bound: Sketch of Proof

Representation for $W^{(n)}$

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

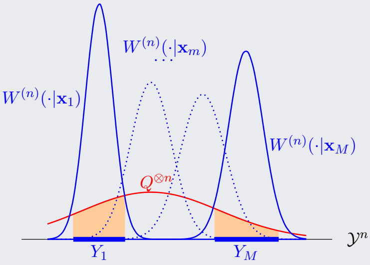
Handle ...... $c^{\otimes n} = c \otimes \cdots \otimes c$

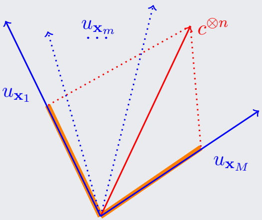


- For a zero-error code, the vectors $u_{\mathbf{x}_m}$ are pairwise orthogonal
- $|\langle u_{\mathbf{x}_m} | c^{\otimes n} \rangle|^2 \leq 1/M$ for at least one m, because $\|c^{\otimes n}\| = 1$
- $|\langle u_{\mathbf{x}_m} | c^{\otimes n} \rangle|^2 \geq e^{-nV(\{u_x\})}$ by construction. Hence $M \leq e^{nV(\{u_x\})}$

Analogies

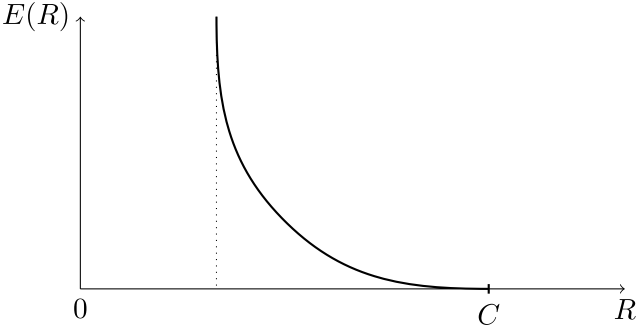
Sphere-Packing vs Lovász





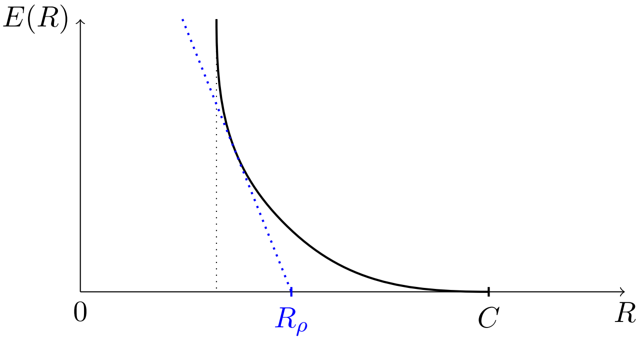
We note the following analogies

$$\sqrt{W(\cdot|\mathbf{x}_m)} \leftrightarrow u_{\mathbf{x}_m}$$
$$\sqrt{Q} \leftrightarrow c$$
$$Q^{\otimes n}(Y_m) \leftrightarrow |\langle u_{\mathbf{x}_m}|c^{\otimes n}\rangle|^2$$



What about min-max expressions?

Sphere-Packing Bound as an Information Radius

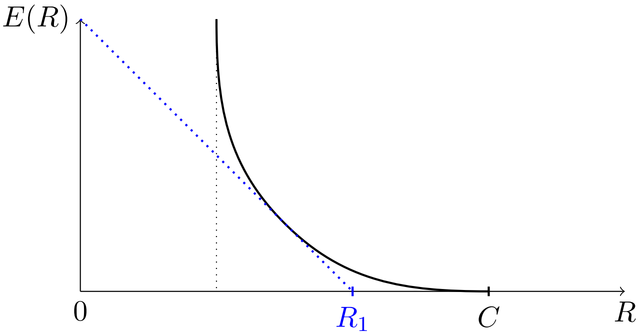


Csiszár, 1995 (implicit in S-G-B, 1967)

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

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

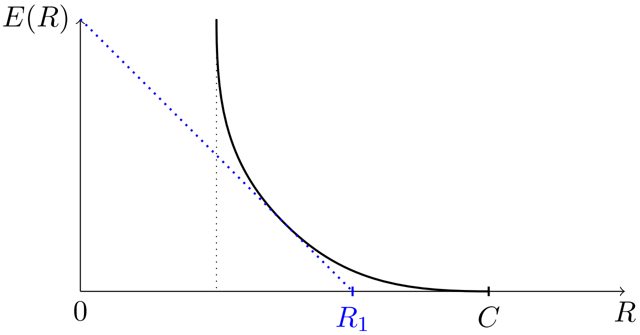
Sphere-Packing Bound as an Information Radius



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

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

Sphere-Packing Bound as an Information Radius



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

$$\textcolor{blue}{R_1} = \min_Q \max_x \log \frac{1}{\left(\sum_y \sqrt{W(y|x)Q(y)}\right)^2}$$

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

From Classical to Classical-Quantum

Representations, values and cutoff rates

- So,

$$u_x = \sqrt{W(\cdot|x)} \quad \Longrightarrow \quad V(\{u_x\}) = \text{cutoff rate}$$

- 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 $\sqrt{W(\cdot|x)}$

From Classical to Classical-Quantum

Representations, values and cutoff rates

- So,
$$u_x = \sqrt{W(\cdot|x)} \quad \Longrightarrow \quad V(\{u_x\}) = \text{cutoff rate}$$
- 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 $\sqrt{W(\cdot|x)}$

Representations, values and cutoff rates

- So,
$$u_x = \sqrt{W(\cdot|x)} \quad \Longrightarrow \quad V(\{u_x\}) = \text{cutoff rate}$$
- 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 $\sqrt{W(\cdot|x)}$

Representations, values and cutoff rates

- So,

$$u_x = \sqrt{W(\cdot|x)} \quad \Longrightarrow \quad V(\{u_x\}) = \text{cutoff rate}$$

- 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 $\sqrt{W(\cdot|x)}$

Definition

- **Basic Idea**

$$W(\cdot|x) \longrightarrow S_x \text{ density operator}$$

  S_x is a positive semi-definite matrix with unit trace

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

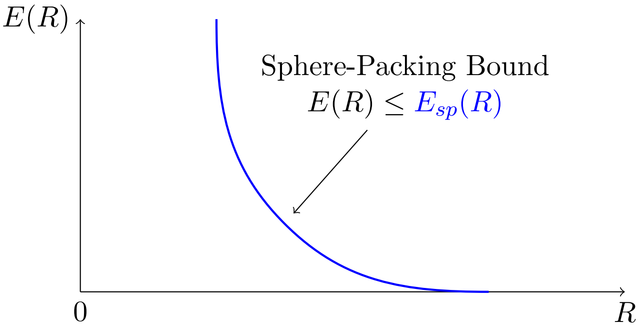
$$S_x = \begin{bmatrix} W(1|x) & 0 & \cdots & 0 \\ 0 & W(2|x) & \cdots & 0 \\ 0 & \cdots & \ddots & \end{bmatrix}$$

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

$$S_x = |u_x\rangle\langle u_x|$$

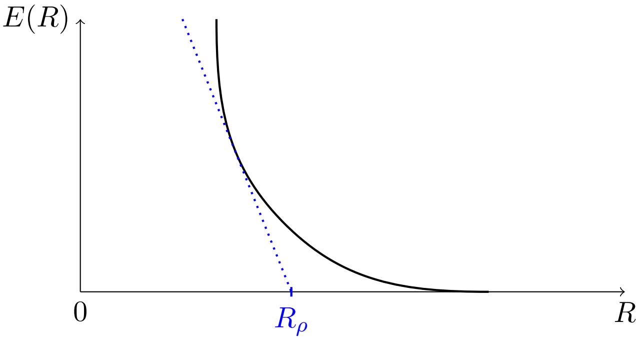
Classical-Quantum Channels: Reliability Function

Classical-Quantum Channels: Reliability Function



Dalai, ISIT 2012 (extending S-G-B 1967)

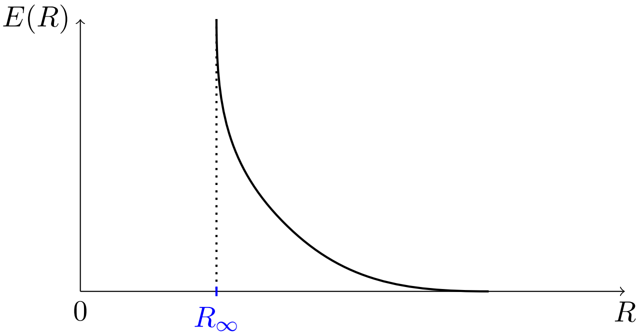
$$E_{sp}(R) = \sup_{\rho \geq 0} \max_{P} \left[-\log \operatorname{Tr} \left(\sum_{x} P(x) S_x^{1/(1+\rho)} \right)^{1+\rho} - \rho R \right]$$



This paper

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

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

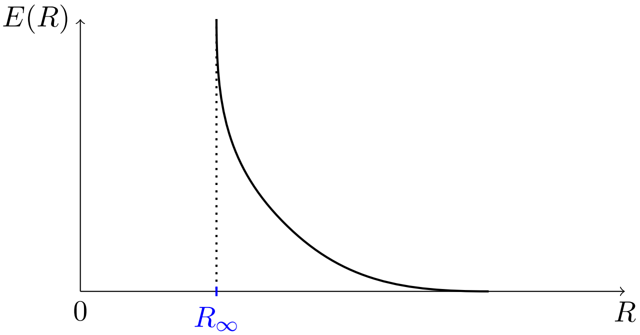


When $\rho \to \infty$

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

where S_x^0 is the projector into the support of S_x

Classical-Quantum Channels: Reliability Function



For pure-state channels $S_x = |u_x\rangle\langle u_x|$
Using pure-states $F = |f\rangle\langle f|$ we have $\mathrm{Tr}\left(S_x^0 F\right) = |\langle u_x|f\rangle|^2$.
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 $S_x = |u_x\rangle\langle u_x|$ satisfies $R_\infty \leq V(\{u_x\})$
- We can define
$$\vartheta_{sp} = \min_{\{S_x\}} R_\infty \qquad (1)$$
where we minimize over all channels such that $\operatorname{Tr} S_x S_{x'} = 0$ if x and x' are not confusable
- Then
$$C_0 \leq \vartheta_{sp} \leq \vartheta$$
- **Schrijver** observes that some additional results in Lovász' paper imply $\vartheta \leq \vartheta_{sp}$. Hence $\vartheta_{sp} = \vartheta$.
- So, pure-state channels achieve the optimum in (1) and for the 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 $S_x = |u_x\rangle\langle u_x|$ satisfies $R_\infty \leq V(\{u_x\})$
- We can define

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

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

- Then

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

- **Schrijver** observes that some additional results in Lovász' paper imply $\vartheta \leq \vartheta_{sp}$. Hence $\vartheta_{sp} = \vartheta$.
- So, pure-state channels achieve the optimum in (1) and for the 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 $S_x = |u_x\rangle\langle u_x|$ satisfies $R_\infty \leq V(\{u_x\})$
- We can define
$$\vartheta_{sp} = \min_{\{S_x\}} R_\infty \tag{1}$$
where we minimize over all channels such that $\operatorname{Tr} S_x S_{x'} = 0$ if x and x' are not confusable
- Then
$$C_0 \leq \vartheta_{sp} \leq \vartheta$$
- **Schrijver** observes that some additional results in Lovász' paper imply $\vartheta \leq \vartheta_{sp}$. Hence $\vartheta_{sp} = \vartheta$.
- So, pure-state channels achieve the optimum in (1) and for the 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 $S_x = |u_x\rangle\langle u_x|$ satisfies $R_\infty \leq V(\{u_x\})$
- We can define

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

  where we minimize over all channels such that $\operatorname{Tr} S_x S_{x'} = 0$ if x and x' are not confusable
- Then

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

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

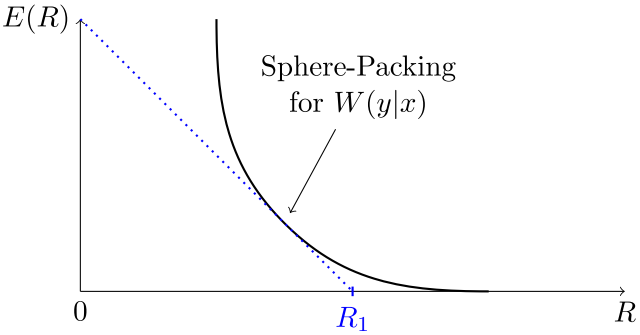
Orthonormal Representations and Auxiliary Channels

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

Cutoff rates and R_∞

But... where are those cutoff rates?

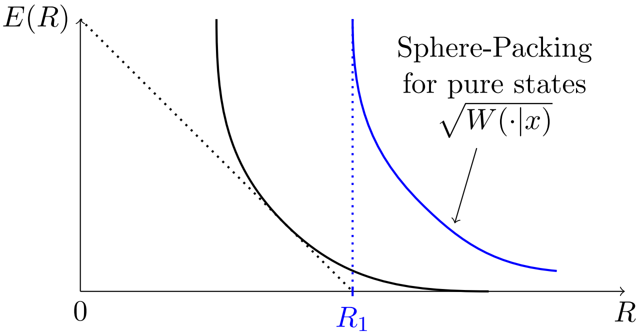
Cutoff rates and R_∞



But... where are those cutoff rates?

- We had previously identified R_1 with $V(\{\sqrt{W(\cdot|x)}\})$
- 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 $u_x = \sqrt{W(\cdot|x)}$
- The true meaning of this... I do not know, but this sounds important