On the Rényi divergence, the Joint Range of Relative Entropies, and a Channel Coding Theorem

Igal Sason

Department of Electrical Engineering
Technion, Haifa, Israel

ETH – Swiss Federal Institute of Technology
Zurich, Switzerland
August 25, 2015

Cast of Characters

- Probability measures P and Q defined on a measurable space $(\mathcal{A}, \mathscr{F})$.
- $X \sim P$.
- $Y \sim Q$.

Relative Information

$$i_{P\|Q}(x) = \log \frac{\mathrm{d}P}{\mathrm{d}Q}(x), \quad P \ll Q.$$

Total Variation (TV) Distance

$$|P-Q| = 2 \sup_{\mathcal{F} \in \mathscr{F}} |P(\mathcal{F}) - Q(\mathcal{F})|.$$

If $P \ll Q$

$$\begin{aligned} |P-Q| &= \mathbb{E}\left[\left|\frac{\mathrm{d}P}{\mathrm{d}Q}(Y) - 1\right|\right] \\ &= \mathbb{E}\Big[\Big|\exp\big(\imath_{P\|Q}(Y)\big) - 1\Big|\Big]. \end{aligned}$$

Simplification in the discrete case:

$$|P-Q| = \sum_{x \in \mathcal{A}} \big|P(x) - Q(x)\big| = |P-Q|_1.$$

Relative Entropy

$$D(P\|Q) = \mathbb{E}[\imath_{P\|Q}(X)] = \mathbb{E}[\imath_{P\|Q}(Y)\ \exp\big(\imath_{P\|Q}(Y)\big)].$$

Simplification in the discrete case:

$$D(P\|Q) = \sum_{x \in \mathcal{A}} P(x)\ \log \frac{P(x)}{Q(x)}.$$

The Rényi Divergence

Let

- $P \ll Q$.
- $X \sim P$ and $Y \sim Q$.
- $\alpha \in (0,1) \cup (1,\infty)$.

The Rényi divergence of order α is given by

$$\begin{aligned} D_\alpha(P\|Q) &= \frac{1}{\alpha-1} \log \mathbb{E}\Big[\exp\big(\alpha\, i_{P\|Q}(Y)\big)\Big] \\ &= \frac{1}{\alpha-1} \log \Big(\mathbb{E}\Big[\exp\big((\alpha-1) i_{P\|Q}(X)\big)\Big]\Big). \end{aligned}$$

Furthermore, $D_1(P\|Q) = D(P\|Q)$, and

$$\lim_{\alpha \to 1^-} D_\alpha(P\|Q) = D(P\|Q).$$

The Rényi Divergence

In the discrete case, we have for $\alpha \in (0,1) \cup (1,\infty)$

$$D_\alpha(P||Q) = \frac{1}{\alpha - 1} \, \log \left(\sum_{x \in A} P^\alpha(x) \, Q^{1-\alpha}(x) \right).$$

The Rényi Divergence

In the discrete case, we have for $\alpha \in (0,1) \cup (1,\infty)$

$$D_\alpha(P||Q) = \frac{1}{\alpha - 1} \log \left(\sum_{x \in A} P^\alpha(x)\, Q^{1-\alpha}(x) \right).$$

Extreme cases:

- If $\alpha = 0$ then $D_0(P||Q) = -\log Q(\text{Support}(P))$,
- If $\alpha = +\infty$ then $D_\infty(P||Q) = \log \left(\text{ess sup}\, \frac{P}{Q} \right)$.

The Rényi Divergence

In the discrete case, we have for $\alpha \in (0,1) \cup (1,\infty)$

$$D_\alpha(P||Q) = \frac{1}{\alpha - 1} \log \left(\sum_{x \in A} P^\alpha(x)\, Q^{1-\alpha}(x) \right).$$

Extreme cases:

- If $\alpha = 0$ then $D_0(P||Q) = -\log Q(\text{Support}(P))$,
- If $\alpha = +\infty$ then $D_\infty(P||Q) = \log \left(\text{ess sup } \frac{P}{Q} \right)$.

L'Hôpital's rule $\Rightarrow D(P||Q) = \lim_{\alpha \to 1^-} D_\alpha(P||Q)$.

Some Basic Properties of the Rényi Divergences

1. Non-negativity: $D_\alpha(P||Q) \geq 0$ with equality if and only if $P = Q$.
2. Monotonicity: $D_\alpha(P||Q)$ is monotonically increasing with α.
3. Convexity properties of $D_\alpha(P||Q)$:
   - jointly convex in (P, Q) for $\alpha \in [0, 1]$,
   - convex in Q for $\alpha \in [0, \infty]$, but not in P for $\alpha > 1$.
4. The Rényi divergence satisfies the data processing inequality (DPI).

Some Basic Properties of the Rényi Divergences

1. Non-negativity: $D_\alpha(P||Q) \geq 0$ with equality if and only if $P = Q$.
2. Monotonicity: $D_\alpha(P||Q)$ is monotonically increasing with α.
3. Convexity properties of $D_\alpha(P||Q)$:
   - jointly convex in (P, Q) for $\alpha \in [0, 1]$,
   - convex in Q for $\alpha \in [0, \infty]$, but not in P for $\alpha > 1$.
4. The Rényi divergence satisfies the data processing inequality (DPI).

Paper

T. van Erven and P. Harremoës, "Rényi divergence and Kullback-Leibler divergence," *IEEE Trans. on Information Theory*, vol. 60, no. 7, pp. 3797–3820, July 2014.

Information-Theoretic Applications of the Rényi divergence

- Channel coding error exponents (Gallager '65, Arimoto '73, Polyanskiy & Verdu '10).
- Generalized cutoff rates for hypothesis testing (Csiszár '95, Alajaji et al. '04).
- Multiple source adaptation (Mansour et al., '09).
- Generalized guessing moments (van Erven & Harremoes, '10).
- Two-sensor composite hypothesis testing (Shayevitz, '11).
- Bounds for joint source-channel coding (Tridenski & Zamir, '11)
- Strong data processing theorems for DMCs (Raginsky, '13).
- Strong converse theorems for networks (Fong & Tan, arXiv '14).
- IT applications of the logarithmic probability comparison bound (Atar & Merhav, arXiv '15).

Motivation: Characterization of the Joint Range of Relative Entropies

Question

Let

- $\varepsilon \in (0,2)$ *be fixed.*
- P_1, P_2 *be arbitrary PDs s.t.* $|P_1 - P_2| \geq \varepsilon$*.*
- Q *is an arbitrary PD s.t.* $Q \ll P_1, P_2$*.*

1. *What is the achievable region of* $\big(D(Q\|P_1),\, D(Q\|P_2)\big)$ *where none of these three distributions is fixed ?*
2. *Given an arbitrary point in this region, specify PDs* P_1, P_2, Q *that achieve this point.*

Approach for Solving the Problem

- Minimizing the Rényi divergence subject to a minimal TV distance: Finding the exact solution of the optimization problem

$$\inf_{P_1, P_2 \colon |P_1 - P_2| \geq \varepsilon} D_\alpha(P_1 \| P_2), \quad \varepsilon \in (0, 2).$$

Approach for Solving the Problem

- Minimizing the Rényi divergence subject to a minimal TV distance: Finding the exact solution of the optimization problem

$$\inf_{P_1, P_2 \colon |P_1 - P_2| \geq \varepsilon} D_\alpha(P_1 \| P_2), \quad \varepsilon \in (0, 2).$$

- Using the solution of this minimization problem for
  1. Providing an exact characterization of the joint region of relative entropies.

Approach for Solving the Problem

- Minimizing the Rényi divergence subject to a minimal TV distance: Finding the exact solution of the optimization problem

$$\inf_{P_1, P_2 \colon |P_1 - P_2| \geq \varepsilon} D_\alpha(P_1 \| P_2), \quad \varepsilon \in (0, 2).$$

- Using the solution of this minimization problem for
  1. Providing an exact characterization of the joint region of relative entropies.
  2. Proving that each point in this region and its boundary is achievable by a triple of 2-element PDs P_1, P_2 and Q.

Approach for Solving the Problem

- Minimizing the Rényi divergence subject to a minimal TV distance: Finding the exact solution of the optimization problem

$$\inf_{P_1, P_2 \colon |P_1 - P_2| \geq \varepsilon} D_\alpha(P_1 \| P_2), \quad \varepsilon \in (0, 2).$$

- Using the solution of this minimization problem for
  1. Providing an exact characterization of the joint region of relative entropies.
  2. Proving that each point in this region and its boundary is achievable by a triple of 2-element PDs P_1, P_2 and Q.
  3. Providing a geometric interpretation of the minimum of the Chernoff information under a constraint on the minimal TV distance.

Minimization of the Relative Entropy s.t. Minimal TV Distance

Consider the optimization problem

$$L(\varepsilon) \triangleq \inf_{P_1, P_2 \colon |P_1 - P_2| \geq \varepsilon} D(P_1 \| P_2), \quad \varepsilon \in [0, 2).$$

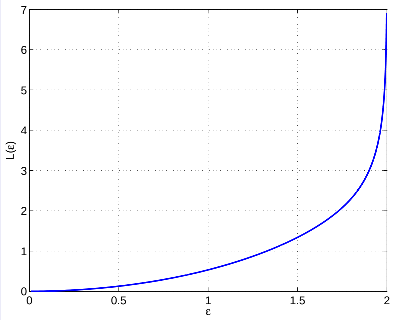
It was studied by Fedotov et al. (2003), Gilardoni (2006) and Reid and Williamson (2011), providing equivalent parameterizations of the solution.

Data Processing Inequality $\Rightarrow$ The minimization of the relative entropy subject to a minimal TV distance is attained by a pair of 2-element PDs.

Minimization of the Relative Entropy (Cont.)

$$\varepsilon(t) = t \left[1 - \left(\coth(t) - \frac{1}{t}\right)^2\right], \qquad t > 0$$

$$L\big(\varepsilon(t)\big) = \log\left(\frac{t}{\sinh(t)}\right) + t\coth(t) - \left(\frac{t}{\sinh(t)}\right)^2.$$



Minimization of the Rényi Divergence s.t. Minimal TV Distance

For $\alpha > 0$, let

$$g_\alpha(\varepsilon) \triangleq \inf_{P_1, P_2 \colon |P_1 - P_2| = \varepsilon} D_\alpha(P_1 || P_2), \quad \forall \, \varepsilon \in [0, 2).$$

Since $g_\alpha(\varepsilon)$ is monotonic non-decreasing in ε then

$$g_\alpha(\varepsilon) = \inf_{P_1, P_2 \colon |P_1 - P_2| \geq \varepsilon} D_\alpha(P_1 || P_2), \quad \forall \, \varepsilon \in [0, 2).$$

Minimization of the Rényi Divergence s.t. Minimal TV Distance

For $\alpha > 0$, let

$$g_\alpha(\varepsilon) \triangleq \inf_{P_1, P_2 :\, |P_1 - P_2| = \varepsilon} D_\alpha(P_1 || P_2), \quad \forall\, \varepsilon \in [0, 2).$$

Since $g_\alpha(\varepsilon)$ is monotonic non-decreasing in ε then

$$g_\alpha(\varepsilon) = \inf_{P_1, P_2 :\, |P_1 - P_2| \geq \varepsilon} D_\alpha(P_1 \| P_2), \quad \forall\, \varepsilon \in [0, 2).$$

For $\alpha \in [0, 1]$, since $D_\alpha(P \| Q)$ is jointly convex in (P, Q), it follows that g_α is convex, and the infimum above is a minimum.

Min. of the Rényi Divergence s.t. Minimal TV Distance (Cont.)

Claim: There is no loss of generality by restricting the minimization of $g_\alpha(\varepsilon)$, for $\varepsilon \in (0, 2)$, to pairs of 2-element PDs.

Min. of the Rényi Divergence s.t. Minimal TV Distance (Cont.)

Claim: There is no loss of generality by restricting the minimization of $g_\alpha(\varepsilon)$, for $\varepsilon \in (0,2)$, to pairs of 2-element PDs.

Proof

- Let P_1, P_2 be PDs defined on an arbitrary set $\mathcal{A}$ of $k \geq 2$ elements.
- Let $\phi \colon \mathcal{A} \to \{1,2\}$ be defined such that

$$\phi(x) = \begin{cases} 1, & \text{if } P_1(x) \geq P_2(x), \\ 2, & \text{if } P_1(x) < P_2(x) \end{cases}$$

  and define $\phi(P_i) = Q_i$ for $i \in \{1,2\}$ where

$$Q_i(j) \triangleq \sum_{x \in \mathcal{A} \colon \, \phi(x) = j} P_i(x), \quad \forall \, i, j \in \{1,2\}.$$

- It can be verified that $|P_1 - P_2| = |Q_1 - Q_2|$.
- Due to the DPI, $D_\alpha(P_1 \| P_2) \geq D_\alpha(Q_1 \| Q_2)$.

Min. of the Rényi Divergence s.t. Minimal TV Distance (Cont.)

Corollary

Let $\alpha \in (0,1) \cup (1, \infty)$, *and* $\varepsilon \in [0,2)$. *The function* g_α *satisfies*

$$g_\alpha(\varepsilon) = \min_{p,q \in [0,1] \, : \, |p-q| \geq \frac{\varepsilon}{2}} d_\alpha(p \| q)$$

where

$$d_\alpha(p \| q) \triangleq \frac{\log\Big(p^\alpha q^{1-\alpha} + (1-p)^\alpha (1-q)^{1-\alpha}\Big)}{\alpha - 1}$$

is the binary Rényi divergence of order α.
The minimizing probability distributions: $P_1 = (p, 1-p)$, $P_2 = (q, 1-q)$.

Min. of the Rényi Divergence s.t. Minimal TV Distance (Cont.)

Corollary

For $\alpha \in (0,1)$ *and* $\varepsilon \in [0,2)$

$$g_\alpha(\varepsilon) = \left(\frac{\alpha}{1-\alpha}\right) g_{1-\alpha}(\varepsilon),$$

and

$$g_\alpha(\varepsilon) \geq c_1(\alpha) \, \log\left(\frac{1}{1-\frac{\varepsilon}{2}}\right) + c_2(\alpha),$$

where $c_1(\alpha) \triangleq \min\left\{1, \frac{\alpha}{1-\alpha}\right\}$, *and* $c_2(\alpha) \triangleq -\frac{\log(2)}{1-\alpha}$.

Min. of the Rényi Divergence s.t. Minimal TV Distance (Cont.)

Corollary

For $\alpha \in (0,1)$ *and* $\varepsilon \in [0,2)$

$$g_\alpha(\varepsilon) = \left(\frac{\alpha}{1-\alpha}\right) g_{1-\alpha}(\varepsilon),$$

and

$$g_\alpha(\varepsilon) \geq c_1(\alpha) \, \log\left(\frac{1}{1-\frac{\varepsilon}{2}}\right) + c_2(\alpha),$$

where $c_1(\alpha) \triangleq \min\left\{1, \frac{\alpha}{1-\alpha}\right\}$, *and* $c_2(\alpha) \triangleq -\frac{\log(2)}{1-\alpha}$.

Remark

Note that the above corollary yields that $\lim_{\varepsilon \to 2^-} g_\alpha(\varepsilon) = +\infty$, as opposed to Pinsker-type inequalities.

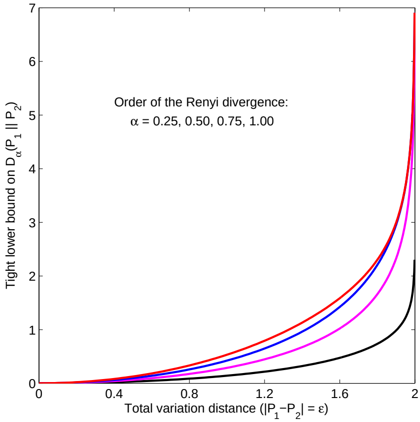


Figure: A plot of the minimum of the Rényi divergence $D_\alpha(P_1||P_2)$ of order $\alpha = 0.25, 0.50, 0.75, 1.00$ ($\alpha = 1$ gives the relative entropy) as a function of the total variation distance $|P_1 - P_2| = \varepsilon \in [0, 2)$.

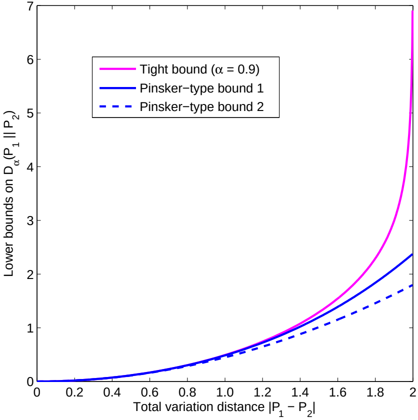


Figure: A plot of the minimum of the Rényi divergence $D_\alpha(P_1 \| P_2)$ of order $\alpha = 0.90$ subject to a fixed total variation distance $|P_1 - P_2| = \varepsilon \in [0, 2)$. This tight lower bound is compared with the Pinsker-type lower bounds by Gilardoni.

Exact Characterization of the Joint Range of the Relative Entropies

Question

Let

- $\varepsilon \in (0, 2)$ *be fixed.*
- P_1, P_2 *be arbitrary PDs s.t.* $|P_1 - P_2| \geq \varepsilon$.
- Q *is an arbitrary PD s.t.* $Q \ll P_1, P_2$.

1. *What is the achievable region of* $\big(D(Q\|P_1),\, D(Q\|P_2)\big)$ *where none of these three distributions is fixed ?*
2. *Given an arbitrary point in this region, specify PDs* P_1, P_2, Q *that achieve this point.*

Exact Characterization of the Joint Range of the Relative Entropies

Question

Let

- $\varepsilon \in (0, 2)$ *be fixed.*
- P_1, P_2 *be arbitrary PDs s.t.* $|P_1 - P_2| \geq \varepsilon$.
- Q *is an arbitrary PD s.t.* $Q \ll P_1, P_2$.

1. *What is the achievable region of* $\big(D(Q\|P_1),\, D(Q\|P_2)\big)$ *where none of these three distributions is fixed ?*
2. *Given an arbitrary point in this region, specify PDs* P_1, P_2, Q *that achieve this point.*

We find this region, and show that each point in this region is attained by a certain triple of 2-element PDs P_1, P_2 and Q.

An identity for the Rényi divergence

For $\alpha \in (0,1) \cup (1,\infty) \setminus \{1\}$

$$D_\alpha(P_1 \| P_2) = D(Q \| P_2) + \frac{\alpha}{1-\alpha} \cdot D(Q \| P_1) + \frac{1}{\alpha - 1} \cdot D(Q \| Q_\alpha)$$

where Q_α is given by

$$Q_\alpha(x) \triangleq \frac{P_1^\alpha(x)\, P_2^{1-\alpha}(x)}{\sum_u P_1^\alpha(u)\, P_2^{1-\alpha}(u)}, \quad \forall\, x \in \text{Supp}(P_1).$$

An identity for the Rényi divergence

For $\alpha \in (0,1) \cup (1,\infty) \setminus \{1\}$

$$D_\alpha(P_1 \| P_2) = D(Q \| P_2) + \frac{\alpha}{1-\alpha} \cdot D(Q \| P_1) + \frac{1}{\alpha - 1} \cdot D(Q \| Q_\alpha)$$

where Q_α is given by

$$Q_\alpha(x) \triangleq \frac{P_1^\alpha(x)\, P_2^{1-\alpha}(x)}{\sum_u P_1^\alpha(u)\, P_2^{1-\alpha}(u)}, \quad \forall\, x \in \mathrm{Supp}(P_1).$$

This comes as a direct calculation, following a result by Shayevitz (ISIT '11) where for $\alpha > 1$

$$D_\alpha(P_1 \| P_2) = \max_{Q \ll P_1} \left\{ D(Q \| P_2) + \frac{\alpha}{\alpha - 1} \cdot D(Q \| P_1) \right\}$$

and the $\max$ is replaced by $\min$ for $\alpha \in (0,1)$.

Proof of the identity

$$
\begin{aligned}
& D(Q\|P_2) + \frac{\alpha}{1-\alpha} \cdot D(Q\|P_1) + \frac{1}{\alpha-1} \, D(Q\|Q_\alpha) \\
&= \sum_x Q(x) \, \log \frac{Q(x)}{P_2(x)} + \frac{\alpha}{1-\alpha} \sum_x Q(x) \log \frac{Q(x)}{P_1(x)} + \frac{1}{\alpha-1} \sum_x Q(x) \log \frac{Q(x)}{Q_\alpha(x)} \\
&= \frac{1}{\alpha-1} \sum_x Q(x) \log \left(\frac{P_1^\alpha(x) \, P_2^{1-\alpha}(x)}{Q_\alpha(x)} \right) \\
&= \frac{1}{\alpha-1} \sum_x Q(x) \log \left(\sum_u P_1^\alpha(u) P_2^{1-\alpha}(u) \right) \\
&= \frac{1}{\alpha-1} \, \log \left(\sum_u P_1^\alpha(u) P_2^{1-\alpha}(u) \right) \\
&= D_\alpha(P_1\|P_2)
\end{aligned}
$$

An Exact Characterization of the Region

The boundary is determined by letting α increase continuously in (0,1), and drawing the straight lines in the plane of $\big(D(Q\|P_1), D(Q\|P_2)\big)$:

$$D(Q\|P_2) + \frac{\alpha}{1-\alpha} \cdot D(Q\|P_1) = g_\alpha(\varepsilon), \quad \forall\, \alpha \in (0,1).$$

Every point on the boundary is a tangent point to one of the straight lines.

An Exact Characterization of the Region (Cont.)

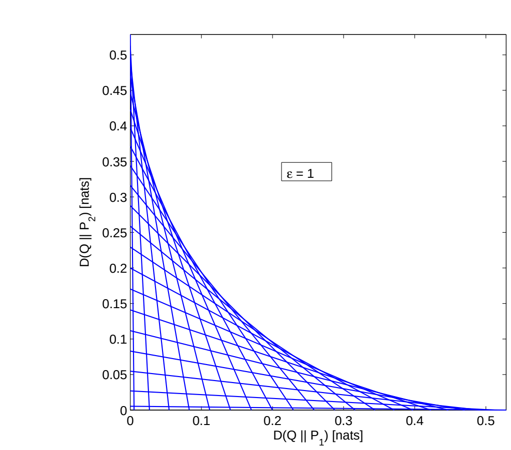


Figure: The achievable region of $(D(Q\|P_1), D(Q\|P_2))$ where $|P_1 - P_2| \geq 1$ is the upper envelope of the straight lines.

An Exact Characterization of the Region (Cont.)

The boundary is determined by letting α increase continuously in (0,1), and drawing the straight lines in the plane of $\big(D(Q\|P_1), D(Q\|P_2)\big)$:

$$D(Q\|P_2) + \frac{\alpha}{1-\alpha} \cdot D(Q\|P_1) = g_\alpha(\varepsilon), \quad \forall\, \alpha \in (0,1).$$

Every point on the boundary is a tangent point to one of the straight lines.

The triple of 2-element PDs P_1, P_2 and Q that achieves an arbitrary point on the boundary of this region is determined as follows:

- Find the slope s of the tangent line $(s < 0)$, and determine $\alpha \in (0,1)$ such that $-\frac{\alpha}{1-\alpha} = s \;\Rightarrow\; \alpha = -\frac{s}{1-s}$.

An Exact Characterization of the Region (Cont.)

The boundary is determined by letting α increase continuously in (0,1), and drawing the straight lines in the plane of $\big(D(Q\|P_1), D(Q\|P_2)\big)$:

$$D(Q\|P_2) + \frac{\alpha}{1-\alpha} \cdot D(Q\|P_1) = g_\alpha(\varepsilon), \quad \forall\, \alpha \in (0,1).$$

Every point on the boundary is a tangent point to one of the straight lines.

The triple of 2-element PDs P_1, P_2 and Q that achieves an arbitrary point on the boundary of this region is determined as follows:

- Find the slope s of the tangent line $(s < 0)$, and determine $\alpha \in (0,1)$ such that $-\frac{\alpha}{1-\alpha} = s \;\Rightarrow\; \alpha = -\frac{s}{1-s}$.
- Determine the 2-element PDs $P_1 = (p, 1-p)$, $P_2 = (q, 1-q)$ such that $d_\alpha(p\|q) = g_\alpha(\varepsilon)$.

An Exact Characterization of the Region (Cont.)

The boundary is determined by letting α increase continuously in (0,1), and drawing the straight lines in the plane of $\big(D(Q\|P_1), D(Q\|P_2)\big)$:

$$D(Q\|P_2) + \frac{\alpha}{1-\alpha} \cdot D(Q\|P_1) = g_\alpha(\varepsilon), \quad \forall\, \alpha \in (0,1).$$

Every point on the boundary is a tangent point to one of the straight lines.

The triple of 2-element PDs P_1, P_2 and Q that achieves an arbitrary point on the boundary of this region is determined as follows:

- Find the slope s of the tangent line $(s < 0)$, and determine $\alpha \in (0,1)$ such that $-\frac{\alpha}{1-\alpha} = s \;\Rightarrow\; \alpha = -\frac{s}{1-s}$.
- Determine the 2-element PDs $P_1 = (p, 1-p)$, $P_2 = (q, 1-q)$ such that $d_\alpha(p\|q) = g_\alpha(\varepsilon)$.
- Calculate the 2-element PD $Q = Q_\alpha$ (as above) for the calculated α, p and q.

An Exact Characterization of the Region (Cont.)

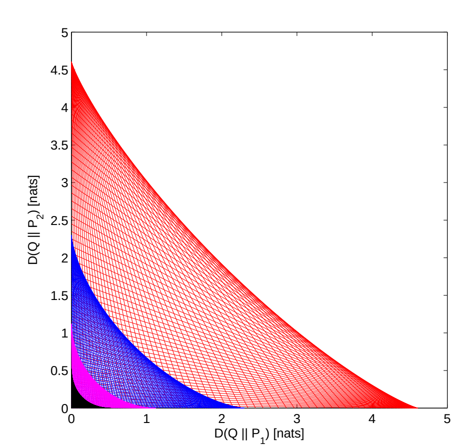


Figure: The boundary of the achievable region of $(D(Q\|P_1), D(Q\|P_2))$ where the TV distance $|P_1 - P_2|$ is at least $\varepsilon = 1.00, 1.40, 1.80, 1.98$.

An Exact Characterization of the Region (Cont.)

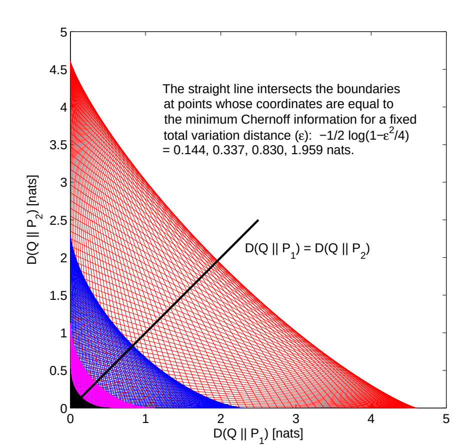


Figure: The boundary of the achievable region of $(D(Q\|P_1), D(Q\|P_2))$ where the TV distance $|P_1 - P_2|$ is at least $\varepsilon = 1.00, 1.40, 1.80, 1.98$.

Motivation for Part II of the Talk

- Performance analysis of linear codes under ML decoding is of interest for the study of the potential performance of these codes under optimal decoding.
- It is also of interest for the evaluation of the degradation in performance that is incurred by the use of sub-optimal and practical decoding algorithms.
- Similarly to the Shulman-Feder bound and related studies, the upper bound in the following theorem quantifies the degradation in the performance of block codes under ML decoding in terms of the deviation of their distance spectra from the binomial distribution.
- The latter distribution characterizes the average distance spectrum of the ensemble of fully random binary block codes, achieving the capacity of any memoryless binary-input output-symmetric channel.

Theorem: A New Upper Bound on the ML Decoding Error Probability

- Consider a binary linear block code of length N and rate $R = \frac{\log(M)}{N}$ where M designates the number of codewords.

Theorem: A New Upper Bound on the ML Decoding Error Probability

- Consider a binary linear block code of length N and rate $R = \frac{\log(M)}{N}$ where M designates the number of codewords.
- Let $S_0 = 0$ and, for $l \in \{1, \ldots, N\}$, let S_l be the number of non-zero codewords of Hamming weight l.

Theorem: A New Upper Bound on the ML Decoding Error Probability

- Consider a binary linear block code of length N and rate $R = \frac{\log(M)}{N}$ where M designates the number of codewords.
- Let $S_0 = 0$ and, for $l \in \{1, \ldots, N\}$, let S_l be the number of non-zero codewords of Hamming weight l.
- Assume that the transmission of the code takes place over a memoryless, binary-input and output-symmetric channel.

Theorem: A New Upper Bound on the ML Decoding Error Probability

- Consider a binary linear block code of length N and rate $R = \frac{\log(M)}{N}$ where M designates the number of codewords.
- Let $S_0 = 0$ and, for $l \in \{1, \dots, N\}$, let S_l be the number of non-zero codewords of Hamming weight l.
- Assume that the transmission of the code takes place over a memoryless, binary-input and output-symmetric channel.
- Assume that the code is maximum-likelihood (ML) decoded.

Theorem: A New Upper Bound (Cont.)

The block error probability satisfies

$$P_{\mathsf{e}} = P_{\mathsf{e}|0} \leq \exp\left(-N \sup_{r \geq 1} \max_{0 \leq \rho' \leq \frac{1}{r}} \left[E_0\left(\rho', \underline{q} = \left(\frac{1}{2}, \frac{1}{2}\right)\right) - \rho'\left(rR + \frac{D_s(P_N \| Q_N)}{N}\right)\right]\right)$$

where

- $s \triangleq s(r) = \frac{r}{r-1}$ for $r \geq 1$ (with the convention that $s = \infty$ for $r = 1$),
- Q_N is the binomial distribution with parameter $\frac{1}{2}$ and N i.i.d. trials,
- P_N is the PMF defined by $P_N(l) = \frac{S_l}{M-1}$ for $l \in \{0, \ldots, N\}$,
- $D_s(\cdot \| \cdot)$ is the Rényi divergence of order s,
- $E_0(\rho, \underline{q})$ is the Gallager random coding error exponent.

Special Case: The Shulman-Feder Bound

Loosening the bound by taking $r = 1 \;\Rightarrow s = \infty$ gives

$$\begin{aligned} P_{\mathsf{e}} &= P_{\mathsf{e}|0} \\ &\leq \exp\left(-N\, E_{\mathsf{r}}\left(R + \frac{D_\infty(P_N \| Q_N)}{N}\right)\right) \\ &= \exp\left(-N\, E_{\mathsf{r}}\left(R + \frac{1}{N}\, \log \max_{0 \leq l \leq N} \frac{P_N(l)}{Q_N(l)}\right)\right) \\ &= \exp\left(-N\, E_{\mathsf{r}}\left(R + \frac{1}{N}\, \log \max_{0 \leq l \leq N} \frac{S_l}{e^{-N(\log 2 - R)}\binom{N}{l}}\right)\right) \end{aligned}$$

which coincides with the Shulman-Feder bound.

Related Papers on Variations of the Gallager Bounds

1. S. Shamai and I. Sason, "Variations on the Gallager bounds, connections, and applications," *IEEE Trans. on Information Theory*, vol. 48, no. 12, pp. 3029–3051, December 2002.
2. I. Sason and S. Shamai, *Performance Analysis of Linear Codes under Maximum-Likelihood Decoding: A Tutorial, Foundations and Trends in Communications and Information Theory*, vol. 3, no. 1–2, pp. 1–222, NOW Publishers, Delft, the Netherlands, July 2006.

Novelty of the Bound & Proof

- The proof of this theorem has an overlap with Appendix A in the paper by Shamai and Sason (2002).
- The novelty here is in working with the Rényi divergence of order $s \geq 1$, instead of the Kullback-Leibler divergence as a lower bound, reveals a need for an optimization of the error exponent:
  1. If $r \geq 1$ is increased, $s = \frac{r}{r-1} \geq 1$ is decreased, and $D_s(P_N \| Q_N)$ is decreased (unless it is 0; note that P_N, Q_N do not depend on r, s).
  2. The maximization of the error exponent in the theorem aims to find a proper balance between the two summands rR and $\frac{D_s(P_N \| Q_N)}{N}$ in the exponent of the new bound, while also optimizing $\rho' \in \left[0, \frac{1}{r}\right]$.

Applicability of the New Bound to Code Ensembles

The bound can be shown to be applicable to code ensembles of binary linear block codes:

- In the probability distribution P_N, the distance spectrum is replaced by the average distance spectrum of the ensemble.

Combination of the New Bound with an Existing Approach

- We borrow a concept of bounding by Miller and Burshtein, and propose to combine it with the new bound.
- In order to utilize the Shulman-Feder bound for binary linear block codes in a clever way, they partitioned the binary linear block code $\mathcal{C}$ into two subcodes $\mathcal{C}_1$ and $\mathcal{C}_2$ where

$$\mathcal{C}_1 \cup \mathcal{C}_2 = \mathcal{C}, \quad \mathcal{C}_1 \cap \mathcal{C}_2 = \{0\}.$$

- The subcode $\mathcal{C}_1$ contains the all-zero codeword and all the codewords of $\mathcal{C}$ whose Hamming weights l belong to a subset $\mathcal{L} \subseteq \{1, 2, ..., N\}$.
- The subcode $\mathcal{C}_2$ contains the other codewords of $\mathcal{C}$ (with Hamming weights of $l \in \mathcal{L}^{\mathrm{c}} \triangleq \{1, 2, ..., N\} \setminus \mathcal{L}$), and the all-zero codeword.

Idea in Selecting $\mathcal{C}_1$

Select $\mathcal{C}_1$ such that it includes the codewords whose hamming weights correspond to the portion of the distance spectrum which is close to the binomial distribution:

$$P_N(l) \approx Q_N(l), \quad \forall\, l \in \mathcal{L}.$$

This selection implies that the normalized Rényi divergence $\frac{D_s(P_N \| Q_N)}{N}$ in the exponent of the new bound has a marginal effect on the conditional ML decoding error probability of the subcode $\mathcal{C}_1$.

Combination of the New Bound with an Existing Approach (Cont.)

- From the symmetry of the channel,

$$P_{\mathsf{e}} = P_{\mathsf{e}|0} \leq P_{\mathsf{e}|0}(\mathcal{C}_1) + P_{\mathsf{e}|0}(\mathcal{C}_2)$$

  where $P_{\mathsf{e}|0}(\mathcal{C}_1)$ and $P_{\mathsf{e}|0}(\mathcal{C}_2)$ are the conditional ML decoding error probabilities of $\mathcal{C}_1$ and $\mathcal{C}_2$ given that the zero codeword is transmitted.
- One can rely on different upper bounds on the conditional error probabilities $P_{\mathsf{e}|0}(\mathcal{C}_1)$ and $P_{\mathsf{e}|0}(\mathcal{C}_2)$:
  1. Bound $P_{\mathsf{e}|0}(\mathcal{C}_1)$ by invoking the new bound, due to the closeness of its distance spectrum to the binomial distribution.
  2. Rely on an alternative approach for bounding $P_{\mathsf{e}|0}(\mathcal{C}_2)$ (e.g., using the union bound w.r.t. the fixed composition codes of the subcode $\mathcal{C}_2$).

Example: Performance Bounds for an Ensemble of Turbo-Block Bodes

Consider

- An ensemble of uniformly interleaved turbo codes whose two component codes are chosen uniformly at random from the ensemble of (1072, 1000) binary systematic linear block codes.
- The overall code rate is 0.8741 bits per channel use.
- The transmission of these codes takes place over an additive white Gaussian noise (AWGN) channel.
- The codes are BPSK modulated and coherently detected.

Example: Turbo-block codes (Cont.)

The following upper bounds under ML decoding are compared:

- The tangential-sphere bound (TSB) of Herzberg and Poltyrev.
- The suggested combination of the union bound (UB) and the new bound (NB). An optimal partitioning is performed to obtain the tightest bound within this form.

Example: Turbo-block codes (Cont.)

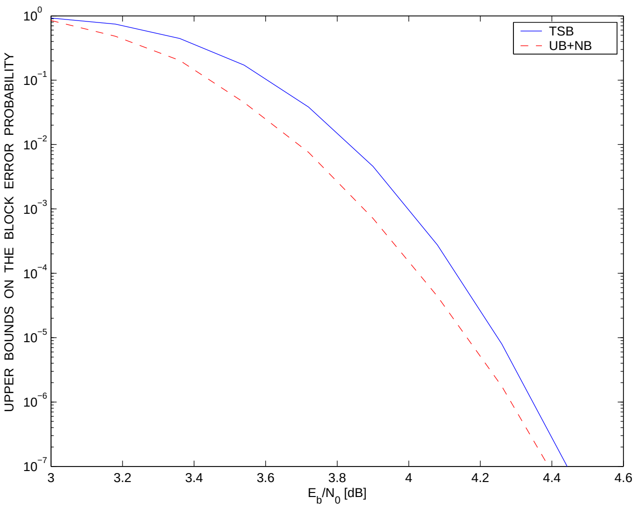


Figure: Comparison between upper bounds on the block error probability.

Summary (Part I)

- The problem of the minimization of the Rényi divergence subject to a fixed (or minimal) total variation distance is solved.
- We determine the joint range of $(D(Q\|P_1), D(Q\|P_2))$ where $|P_1 - P_2| \geq \varepsilon$ for a fixed ε, and Q is arbitrary.
- All the points $(D(Q\|P_1), D(Q\|P_2))$ of this convex region are characterized, and all these points are achieved by PDs (P_1, P_2, Q) that are defined on 2-element sets.
- A geometric interpretation of the minimum of the Chernoff information subject to a fixed total variation distance is provided.

Summary (Part II)

- A new bound on the ML decoding error probability has been derived, involving the Rényi divergence.
- The derivation of this bound relies on variations of the Gallager bounds (the Duman and Salehi bound).
- It reproduces the 1965 random coding Gallager bound, and the Shulman-Feder bound for binary linear block codes (or ensembles). Furthermore, it has in general an additional parameter that is subject to optimization.
- The bound has been applied to an ensemble of uniformly interleaved turbo-block codes with systematic random component codes, and its superiority has been exemplified.

Full Paper Version

http://arxiv.org/abs/1501.03616.
Submitted to the *IEEE Trans. on Information Theory*, February 2015.
Presented in part at ISIT 2015.