Euclidean Information Theory

S. Borade and L. Zheng

STATUS QUO

Network information theory with optimization of K-L divergence

- *rarely analytical solution*
- *global vs. local optimum*
- *structure of optimal solution*

K-L divergence too complicated

NEW INSIGHTS

Local approximation to K-L divergence

- **K-L ---- quadratic**
- **Information geometry ---- Euclidean geometry**
- **Large Deviation ---- CLT**

$$D(P||Q) \approx \sum_i \frac{1}{2Q_i}(Q_i - P_i)^2$$

ACHIEVEMENT DESCRIPTION

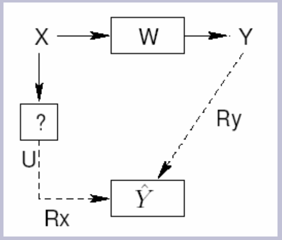
MAIN RESULT:

- **Single letter solution for very noisy BC channel and slope of capacity region for general cases**
- **Relation between rate tradeoff and SVD of divergence translation matrix**

$$B = \left[\sqrt{P_Y}\right]^{-1} W \left[\sqrt{P_X}\right]$$

HOW IT WORKS:

Divergence translation over channel, joint distribution, and linear mapping



ASSUMPTIONS AND LIMITATIONS:

- **Very noisy, local approximation**
- **Canonical example of variation analysis of distributions**
- **Layered coding over networks**

END-OF-PHASE GOAL

Our approach provides new opportunities for traditional network information theory problems, security, multi-terminal source coding.

COMMUNITY CHALLENGE

Geometric view of distributions, related to robust Hypo-testing, compress sensing

Simplifying Information Geometry to Allow Analytical Solutions for Network Problems

What is Euclidean?

- Motivation:
  - Most information theoretic problems are optimization of KL divergence;
  - Finding the structure of the optimal solutions is important: eg. multi-letter $\to$ single letter;
  - The geometry of distributions is in general difficult;
  - Local approximation = Euclidean geometry is the simplest special case
- Local approximation of divergence:

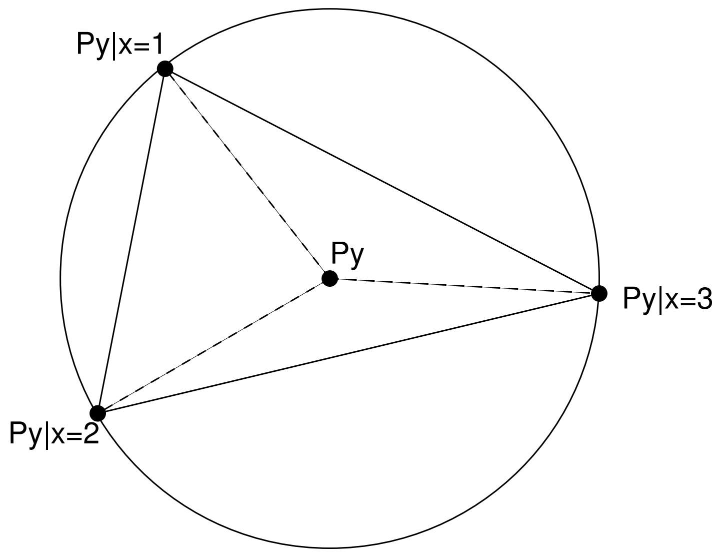
$$
\begin{aligned}
D(P||Q) &= -\sum_i Pi \log \frac{Q_i}{P_i} = -\sum_i P_i \log \left(1 + \frac{Q_i - P_i}{P_i}\right) \\
&\leq\approx -\sum_i P_i \left[\left(\frac{Q_i - P_i}{P_i}\right) - \frac{1}{2}\left(\frac{Q_i - P_i}{P_i}\right)^2\right] \\
&= \sum_i \frac{1}{2P_i}(Q_i - P_i)^2 \quad \approx \sum_i \frac{1}{2Q_i}(Q_i - P_i)^2 \\
&\triangleq \|P - Q\|_P^2 \quad \approx \|P - Q\|_Q^2
\end{aligned}
$$

Mutual Information as Variance

- Consider a DMC $W : \mathcal{X} \to \mathcal{Y}$

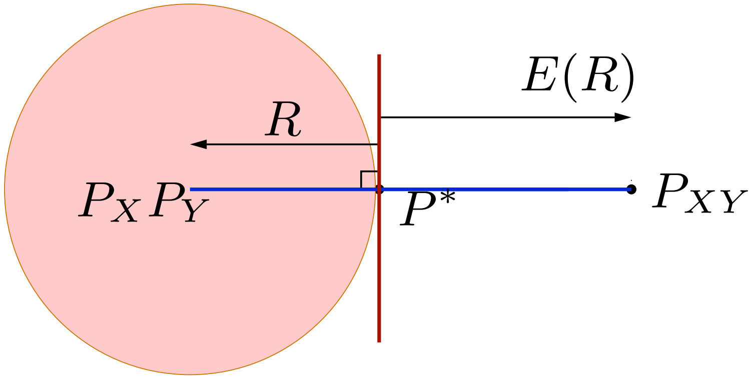
$$I(X;Y) = I(P_X;W) \quad = \quad E_{P_X}[D(W(\cdot|x)||P_Y(\cdot))]$$
$$\approx \quad E_{P_X}\left[||W(\cdot|x) - P_Y(\cdot)||^2_{P_0}\right]$$

- Local approximation valid for very noisy channel: $W(\cdot|x) \sim P_0, \forall x$.
- Mutual information as a variance.
- Optimal output P_Y as circular center.



A Picture of Error Exponent

- Assume random code P_x over channel W
  - Correct codeword $\mathbf{x}_0$ and received string $\mathbf{y} \sim$ i.i.d. P_{XY}
  - e^{nR} incorrect codewords $\mathbf{x}_i, \mathbf{y} \sim$ i.i.d. $(P_X P_Y)$.
  - Independence vs. conditional independence

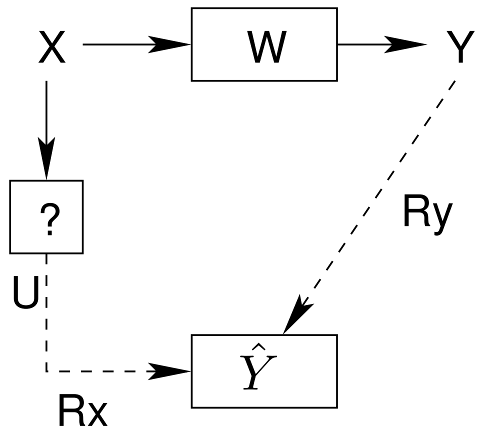


- With a little bit "cheating"

$$\left.\begin{aligned} R = \quad & D(P^* \| P_X P_Y) & \approx k \cdot t^2 \\ E_r = \quad & D(P^* \| P_{XY}) & \approx k \cdot (1-t)^2 \\ I(X;Y) = \quad & D(P_{XY} \| P_X P_Y) & \approx k \end{aligned}\right\} \Longrightarrow \boxed{\sqrt{R} + \sqrt{E_r} = \sqrt{C}}$$

Source Coding Helper Problem

Csiszar & Korner:



- X, Y are joint DMS with $P_{XY} = P_X \cdot W$.
- Goal: lossless reconstruction $\hat{Y}$
- Helper observes X, summarize into U, send rate

$$R_X = I(U; X)$$

- Source coding Y with side information U,

$$R_Y = H(Y|U)$$

$$\boxed{\max_{U \leftrightarrow X \leftrightarrow Y : I(U;X) \leq R_X} I(U;Y)}$$

- Not convex optimization;
- Multi-letter problem in general.

Euclidean Version of the Problem

$$\boxed{\max_{U \leftrightarrow X \leftrightarrow Y : I(U;X) \leq R_X} I(U;Y)} \Longleftrightarrow \boxed{\begin{array}{ll} \max & E_{P_U}\left[\|P_{Y|U}(\cdot|u) - P_Y(\cdot)\|^2_{P_Y}\right] \\ E_{P_U}\left[\|P_{X|U}(\cdot|u) - P_X(\cdot)\|^2_{P_X}\right] < R_x & \end{array}}$$

- Linear relation, for any u, $P_{Y|U}(\cdot|u) = W_{|\mathcal{Y}|\times|\mathcal{X}|} P_{X|U}(\cdot|u)$
- Solution by SVD of W.
- Compensate for the scaling: divergence translation matrix

$$B \triangleq \quad = \quad \left[\sqrt{P_Y}\right]^{-1} W \left[\sqrt{P_X}\right]$$

- New problem

$$\boxed{\max E_{P_U}[\|B \cdot \underline{\mathbf{v}}_{X|u}\|^2] \qquad \text{subject to} \qquad E_{P_U}[\|\underline{\mathbf{v}}_{X|u}\|^2] \leq R_x}$$

SVD of Divergence Translation Matrix

Lemma 1: B has a singular value $\sigma_1 = 1$, corresponding to the input/ output singular vectors $[\dots\sqrt{P_X(x)}\dots]^T, [\dots\sqrt{P_Y(y)}\dots]^T$.

Lemma 2: All other singular values of B have $\sigma_i \leq 1$

Proposition

$$\boxed{I(U;Y) \leq \sigma_2^2 \cdot I(U;X)}$$

with equality achieved by choosing $\underline{\mathbf{v}}_{X|u}$ along the singular vector of B, corresponding to the second largest singular value σ_2.

$$P_{X|U}(\cdot|u) = P_X(\cdot) + \epsilon[\sqrt{P_X}] \cdot \underline{\mathbf{v}}_2$$

Multi-Letter Problem

$$\max I(U;\underline{Y}) \qquad \text{subject to } I(U;\underline{X}) \le nR_X$$

- $P_{\underline{X}}$ can be written as $P_X \otimes P_X \otimes \ldots \otimes P_X$.
- Probability transition matrix $P_{\underline{Y}|\underline{X}} = W \otimes W \otimes \ldots \otimes W$
- Divergence translation from $P_{\underline{X}}$ to $P_{\underline{Y}}$: $B^{(n)} = B \otimes B \otimes \ldots \otimes B$.

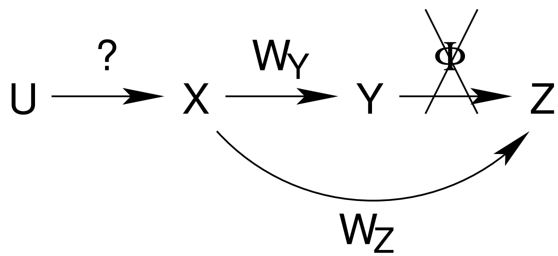
  Lemma: If $\underline{\mathbf{v}}_i, \underline{\mathbf{v}}_j$ are singular vectors of B, corresponding to singular values σ_i, σ_j, then $\underline{\mathbf{v}}_i \otimes \underline{\mathbf{v}}_j$ is a singular vector of $B \otimes B$, with singular value $\sigma_i \cdot \sigma_j$.
- Optimal choice: for all u

$$\begin{aligned} P_{X_1,X_2|U=u} &= P_X \otimes P_X + [\sqrt{P_X \otimes P_X}] \cdot (\epsilon_1 \cdot \underline{\mathbf{v}}_1 \otimes \underline{\mathbf{v}}_2 + \epsilon_2 \cdot \underline{\mathbf{v}}_2 \otimes \underline{\mathbf{v}}_1) \\ &= (P_X + \epsilon_1 \cdot [\sqrt{P_X}] \cdot \underline{\mathbf{v}}_2) \otimes (P_X + \epsilon_2 \cdot [\sqrt{P_X}] \cdot \underline{\mathbf{v}}_2) \end{aligned}$$

- Single letter solution optimal.

Euclidean Approach for Broadcasting Channels

- General very noisy BC channel

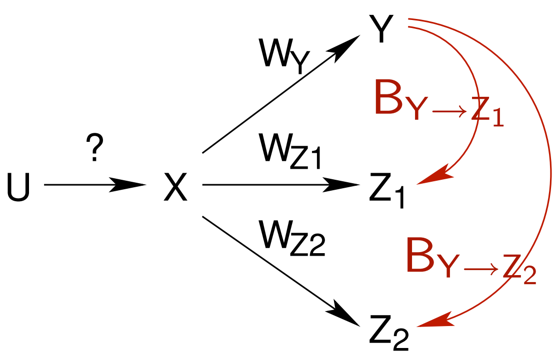


- Divergence translation matrix from P_Y to P_Z:

$$B_{Y\to Z} = [\sqrt{P_Z}]^{-1} \cdot W_Z \cdot W_Y^{-1}[\sqrt{P_Y}]$$

- SV might > 1, related to the shape of capacity region.

- 3-user broadcasting with degraded message sets



- All channels are very noisy.
- Two levels of degraded messages, private message to Y only.
- Single letter "order optimal".

- Single letter solution optimal.
- Two different divergence translation matrices, from P_Y space, $B_{Y\to Z_1}$ $B_{Y\to Z_2}$ can have different singular vectors.