Common Information and Non-Interactive Correlation Distillation

Vincent Y. F. Tan

Department of ECE and Maths, National University of Singapore

Special thanks to **Lei Yu** (Nankai University)

2021 East Asian School on Information Theory

Style of Tutorial

- Based on an upcoming monograph by myself and Lei Yu

Style of Tutorial

- Based on an upcoming monograph by myself and Lei Yu
- Will cover classical stuff and more recent advances based on the speaker's knowledge and preferences

Style of Tutorial

- Based on an upcoming monograph by myself and Lei Yu
- Will cover classical stuff and more recent advances based on the speaker's knowledge and preferences
- Will not be able to touch all bases, e.g., everything I will talk about is discrete

Style of Tutorial

- Based on an upcoming monograph by myself and Lei Yu
- Will cover classical stuff and more recent advances based on the speaker's knowledge and preferences
- Will not be able to touch all bases, e.g., everything I will talk about is discrete
- Will do some proof sketches (since this is a tutorial)

Style of Tutorial

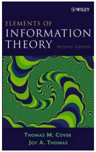
- Based on an upcoming monograph by myself and Lei Yu
- Will cover classical stuff and more recent advances based on the speaker's knowledge and preferences
- Will not be able to touch all bases, e.g., everything I will talk about is discrete
- Will do some proof sketches (since this is a tutorial)
- May get a bit technical (no apologies for that)

Style of Tutorial

- Based on an upcoming monograph by myself and Lei Yu
- Will cover classical stuff and more recent advances based on the speaker's knowledge and preferences
- Will not be able to touch all bases, e.g., everything I will talk about is discrete
- Will do some proof sketches (since this is a tutorial)
- May get a bit technical (no apologies for that)
- But will try to provide as much intuition as possible

Style of Tutorial

- Based on an upcoming monograph by myself and Lei Yu
- Will cover classical stuff and more recent advances based on the speaker's knowledge and preferences
- Will not be able to touch all bases, e.g., everything I will talk about is discrete
- Will do some proof sketches (since this is a tutorial)
- May get a bit technical (no apologies for that)
- But will try to provide as much intuition as possible
- Prerequisite: Information theory at the level of [Cover and Thomas, 2006]



Outline

Measures of Information Among Random Variables

- Given two random variables $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ with joint distribution π_{XY}, how common are they?

Measures of Information Among Random Variables

- Given two random variables $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ with joint distribution π_{XY}, how common are they?
- One may conceive of the following measures of "common information".

Measures of Information Among Random Variables

- Given two random variables $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ with joint distribution π_{XY}, how common are they?
- One may conceive of the following measures of "common information".
- Pearson correlation coefficient

$$\rho(X;Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \in [-1,1].$$

Measures of Information Among Random Variables

- Given two random variables $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ with joint distribution π_{XY}, how common are they?
- One may conceive of the following measures of "common information".
- Pearson correlation coefficient

$$\rho(X;Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \in [-1,1].$$

- Mutual Information

$$I_\pi(X;Y) = \mathsf{E}\left[\log \frac{\pi_{XY}(X,Y)}{\pi_X(X)\pi_Y(Y)}\right] = D(\pi_{XY}\|\pi_X\pi_Y).$$

Measures of Information Among Random Variables

- Given two random variables $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ with joint distribution π_{XY}, how common are they?
- One may conceive of the following measures of "common information".
- Pearson correlation coefficient

$$\rho(X;Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \in [-1,1].$$

- Mutual Information

$$I_\pi(X;Y) = \mathsf{E}\left[\log \frac{\pi_{XY}(X,Y)}{\pi_X(X)\pi_Y(Y)}\right] = D(\pi_{XY}\|\pi_X\pi_Y).$$

- As information theorists, we like operational interpretations

Measures of Information Among Random Variables

- Given two random variables $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ with joint distribution π_{XY}, how common are they?
- One may conceive of the following measures of "common information".
- Pearson correlation coefficient

$$\rho(X;Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \in [-1,1].$$

- Mutual Information

$$I_\pi(X;Y) = \mathsf{E}\left[\log \frac{\pi_{XY}(X,Y)}{\pi_X(X)\pi_Y(Y)}\right] = D(\pi_{XY} \| \pi_X \pi_Y).$$

- As information theorists, we like operational interpretations
- Wyner's CI and Gäcs–Körner–Witsenhausen's CI are the two archetypal notions of information among RVs that admit operational interpretations.

Outline

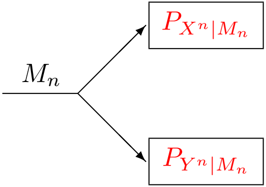
Wyner's Common Information [Wyner, 1975]

Wyner's Common Information [Wyner, 1975]

M_n

- M_n is uniformly distributed over $\mathcal{M}_n = [2^{nR}] := \{1, \ldots, 2^{nR}\}$

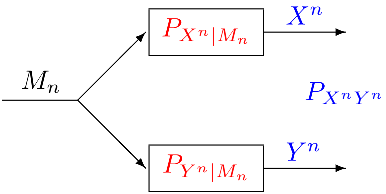
Wyner's Common Information [Wyner, 1975]



- M_n is uniformly distributed over $\mathcal{M}_n = [2^{nR}] := \{1, \ldots, 2^{nR}\}$
- An (n, R)-synthesis code consists of

$$P_{X^n|M_n} : \mathcal{M}_n \to \mathcal{X}^n \quad \text{and} \quad P_{Y^n|M_n} : \mathcal{M}_n \to \mathcal{Y}^n.$$

Wyner's Common Information [Wyner, 1975]



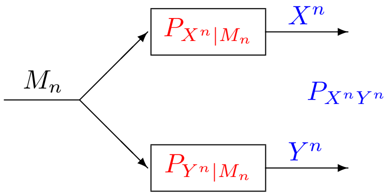
- M_n is uniformly distributed over $\mathcal{M}_n = [2^{nR}] := \{1, \ldots, 2^{nR}\}$
- An (n, R)-synthesis code consists of

$$P_{X^n|M_n} : \mathcal{M}_n \to \mathcal{X}^n \quad \text{and} \quad P_{Y^n|M_n} : \mathcal{M}_n \to \mathcal{Y}^n.$$

- The distribution induced by the code $(P_{X^n|M_n}, P_{Y^n|M_n})$ is

$$P_{X^nY^n}(x^n, y^n) := \frac{1}{|\mathcal{M}_n|} \sum_{m \in \mathcal{M}_n} P_{X^n|M_n}(x^n|m) P_{Y^n|M_n}(y^n|m)$$

Wyner's Common Information [Wyner, 1975]



- M_n is uniformly distributed over $\mathcal{M}_n = [2^{nR}] := \{1, \ldots, 2^{nR}\}$
- An (n, R)-synthesis code consists of

$$P_{X^n|M_n} : \mathcal{M}_n \to \mathcal{X}^n \quad \text{and} \quad P_{Y^n|M_n} : \mathcal{M}_n \to \mathcal{Y}^n.$$

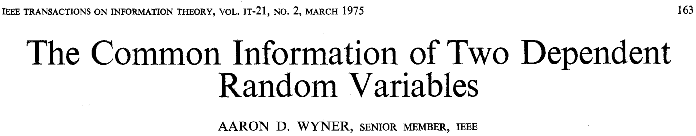
- The distribution induced by the code $(P_{X^n|M_n}, P_{Y^n|M_n})$ is

$$P_{X^nY^n}(x^n, y^n) := \frac{1}{|\mathcal{M}_n|} \sum_{m \in \mathcal{M}_n} P_{X^n|M_n}(x^n|m) P_{Y^n|M_n}(y^n|m)$$

- Desideratum:

$$P_{X^nY^n} \approx \pi_{XY}^n \qquad \text{(target distribution)}$$

Wyner's Common Information [Wyner, 1975]


IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. IT-21, NO. 2, MARCH 1975 163

The Common Information of Two Dependent Random Variables

AARON D. WYNER, SENIOR MEMBER, IEEE

Wyner's Common Information [Wyner, 1975]

IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. IT-21, NO. 2, MARCH 1975 163

The Common Information of Two Dependent Random Variables

AARON D. WYNER, SENIOR MEMBER, IEEE

Normalized relative entropy to measure the "distance" between $P_{X^nY^n}$ and π_{XY}^n

Wyner's Common Information [Wyner, 1975]

IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. IT-21, NO. 2, MARCH 1975 163

The Common Information of Two Dependent Random Variables

AARON D. WYNER, SENIOR MEMBER, IEEE

Normalized relative entropy to measure the "distance" between $P_{X^nY^n}$ and π_{XY}^n

Theorem ([Wyner, 1975])

$$\inf \left\{ R : \frac{1}{n} D(P_{X^nY^n} \| \pi_{XY}^n) \to 0 \right\}$$

Wyner's Common Information [Wyner, 1975]

IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. IT-21, NO. 2, MARCH 1975 163

The Common Information of Two Dependent Random Variables

AARON D. WYNER, SENIOR MEMBER, IEEE

Normalized relative entropy to measure the "distance" between $P_{X^nY^n}$ and π_{XY}^n

Theorem ([Wyner, 1975])

$$\inf\left\{R : \frac{1}{n} D(P_{X^nY^n} \| \pi_{XY}^n) \to 0\right\} = \min_{P_W P_{X|W} P_{Y|W} : P_{XY} = \pi_{XY}} I(XY;W)$$

Wyner's Common Information [Wyner, 1975]

IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. IT-21, NO. 2, MARCH 1975 163

The Common Information of Two Dependent Random Variables

AARON D. WYNER, SENIOR MEMBER, IEEE

Normalized relative entropy to measure the "distance" between $P_{X^nY^n}$ and π_{XY}^n

Theorem ([Wyner, 1975])

$$\begin{aligned}
&\inf\left\{R : \frac{1}{n} D(P_{X^nY^n} \| \pi_{XY}^n) \to 0\right\} \\
&= \min_{P_W P_{X|W} P_{Y|W} : P_{XY} = \pi_{XY}} I(XY;W) \\
&=: C_{\mathrm{W}}(\pi_{XY})
\end{aligned}$$

Wyner's Common Information [Wyner, 1975]

IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. IT-21, NO. 2, MARCH 1975 163

The Common Information of Two Dependent Random Variables

AARON D. WYNER, SENIOR MEMBER, IEEE

Normalized relative entropy to measure the "distance" between $P_{X^nY^n}$ and π^n_{XY}

Theorem ([Wyner, 1975])

$$\inf\left\{R : \frac{1}{n} D(P_{X^nY^n} \| \pi^n_{XY}) \to 0\right\}$$

$$= \min_{P_W P_{X|W} P_{Y|W} : P_{XY} = \pi_{XY}} I(XY;W)$$

$$=: C_{\mathrm{W}}(\pi_{XY})$$

Wyner's Common Information [Wyner, 1975]

IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. IT-21, NO. 2, MARCH 1975 163

The Common Information of Two Dependent Random Variables

AARON D. WYNER, SENIOR MEMBER, IEEE

Normalized relative entropy to measure the "distance" between $P_{X^nY^n}$ and π_{XY}^n

Theorem ([Wyner, 1975])

$$\begin{aligned} &\inf\left\{R : \frac{1}{n} D(P_{X^nY^n} \| \pi_{XY}^n) \to 0\right\} \\ &= \min_{P_W P_{X|W} P_{Y|W} : P_{XY} = \pi_{XY}} I(XY;W) \\ &=: C_{\mathrm{W}}(\pi_{XY}) \end{aligned}$$

where $C_{\mathrm{W}}(\pi_{XY})$ is named Wyner's Common Information.

Sanity Check I

- So Wyner said that a reasonable notion of common information is

$$C_{\mathrm{W}}(\pi_{XY}) = \min_{P_W P_{X|W} P_{Y|W} : P_{XY} = \pi_{XY}} I(XY;W).$$

Sanity Check I

- So Wyner said that a reasonable notion of common information is

$$C_{\mathrm{W}}(\pi_{XY}) = \min_{P_W P_{X|W} P_{Y|W} : P_{XY} = \pi_{XY}} I(XY; W).$$

- Let's test this on $X = (\tilde{X}, V)$ and $Y = (\tilde{Y}, V)$ with $\tilde{X}, \tilde{Y}, V$ independent.

Sanity Check I

- So Wyner said that a reasonable notion of common information is

$$C_{\mathrm{W}}(\pi_{XY}) = \min_{P_W P_{X|W} P_{Y|W}:\, P_{XY}=\pi_{XY}} I(XY;W).$$

- Let's test this on $X = (\tilde{X}, V)$ and $Y = (\tilde{Y}, V)$ with $\tilde{X}, \tilde{Y}, V$ independent.
- Intuitively, we should get $H(V)$ as the common information. Do we?

Sanity Check I

- So Wyner said that a reasonable notion of common information is

$$C_{\mathrm{W}}(\pi_{XY}) = \min_{P_W P_{X|W} P_{Y|W} :\, P_{XY} = \pi_{XY}} I(XY;W).$$

- Let's test this on $X = (\tilde{X}, V)$ and $Y = (\tilde{Y}, V)$ with $\tilde{X}, \tilde{Y}, V$ independent.
- Intuitively, we should get $H(V)$ as the common information. Do we?
- Take $W = V$, satisfies $X - W - Y$. Then

$$I(XY;W) = I(XY;V) \leq H(V) \qquad \text{so far so good...}$$

Sanity Check II

- Now comes the other part, i.e., to show $C_{\mathrm{W}}(\pi_{XY}) \geq H(V)$.

Sanity Check II

- Now comes the other part, i.e., to show $C_{\mathrm{W}}(\pi_{XY}) \geq H(V)$.
- Obviously $X = (\tilde{X}, V)$ and $Y = (\tilde{Y}, V)$ and so

$$V - X - W - Y - V.$$

Sanity Check II

- Now comes the other part, i.e., to show $C_{\mathrm{W}}(\pi_{XY}) \geq H(V)$.
- Obviously $X = (\tilde{X}, V)$ and $Y = (\tilde{Y}, V)$ and so

$$V - X - W - Y - V.$$

- So V is a function of W and

$$I(X, Y; W) = I(\tilde{X}, \tilde{Y}, V; W, V) \geq H(V)$$

Sanity Check II

- Now comes the other part, i.e., to show $C_{\mathrm{W}}(\pi_{XY}) \geq H(V)$.
- Obviously $X = (\tilde{X}, V)$ and $Y = (\tilde{Y}, V)$ and so

$$V - X - W - Y - V.$$

- So V is a function of W and

$$I(X, Y; W) = I(\tilde{X}, \tilde{Y}, V; W, V) \geq H(V)$$

- Minimize over $X - W - Y$ so

$$C_{\mathrm{W}}(\pi_{XY}) \geq H(V)$$

Proof Idea of the Achievability Part

Lemma (Soft-covering lemma [Wyner, 1975] [Cuff, 2012])

Let $(U, W) \sim P_{UW}$ *have mutual information* $I(U;W)$. *For any*

$$R > I(U;W),$$

there exists a sequence of codebooks $\mathcal{C}_n = \{w^n(m) : m \in [2^{nR}]\}$ *such that the synthesized distribution*

$$P_{U^n}(u^n) = \frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} P^n_{U|W}(u^n | w^n(m)) \qquad \forall\, n \in \mathbb{N}$$

Proof Idea of the Achievability Part

Lemma (Soft-covering lemma [Wyner, 1975] [Cuff, 2012])

Let $(U, W) \sim P_{UW}$ *have mutual information* $I(U;W)$*. For any*

$$R > I(U;W),$$

there exists a sequence of codebooks $\mathcal{C}_n = \{w^n(m) : m \in [2^{nR}]\}$ *such that the synthesized distribution*

$$P_{U^n}(u^n) = \frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} P_{U|W}^n(u^n | w^n(m)) \qquad \forall\, n \in \mathbb{N}$$

satisfies

$$\lim_{n\to\infty} \frac{1}{n} D(P_{U^n} \| P_U^n) = 0 \quad \textit{and} \quad \lim_{n\to\infty} |P_{U^n} - P_U^n| = 0 \quad \textit{(TV dist).}$$

Proof Idea of the Achievability Part

Lemma (Soft-covering lemma [Wyner, 1975] [Cuff, 2012])

Let $(U, W) \sim P_{UW}$ *have mutual information* $I(U;W)$*. For any*

$$R > I(U;W),$$

there exists a sequence of codebooks $\mathcal{C}_n = \{w^n(m) : m \in [2^{nR}]\}$ *such that the synthesized distribution*

$$P_{U^n}(u^n) = \frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} P_{U|W}^n(u^n|w^n(m)) \qquad \forall\, n \in \mathbb{N}$$

satisfies

$$\lim_{n\to\infty} \frac{1}{n} D(P_{U^n} \| P_U^n) = 0 \quad \textit{and} \quad \lim_{n\to\infty} |P_{U^n} - P_U^n| = 0 \quad \textit{(TV dist).}$$

Also known as resolvability [Han and Verdú, 1993], [Hayashi, 2006], [Hayashi, 2011] and [Yu and Tan, 2019c].

Proof Idea of the Achievability Part

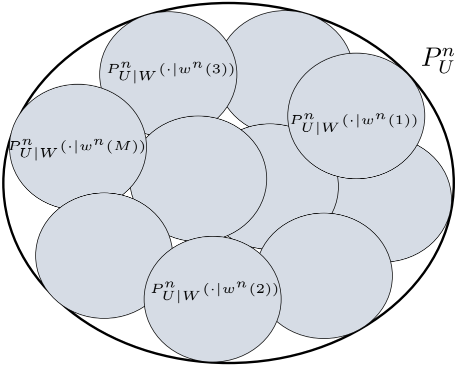


Figure: If $M = 2^{nR}$ and $R > I(U;W)$, then $\frac{1}{n} D(P_{U^n} \| P_U^n) \to 0$.

Proof Idea of the Achievability Part

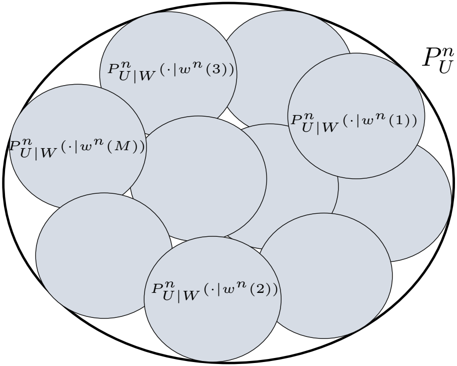
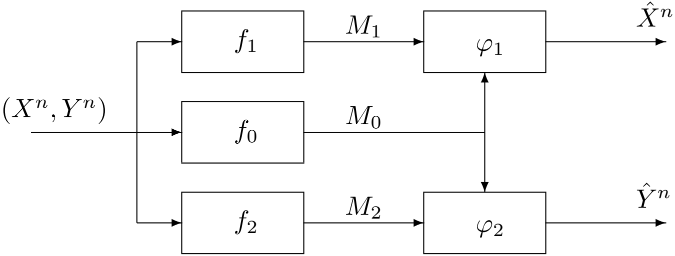


Figure: If $M = 2^{nR}$ and $R > I(U;W)$, then $\frac{1}{n}D(P_{U^n}\|P_U^n) \to 0$.

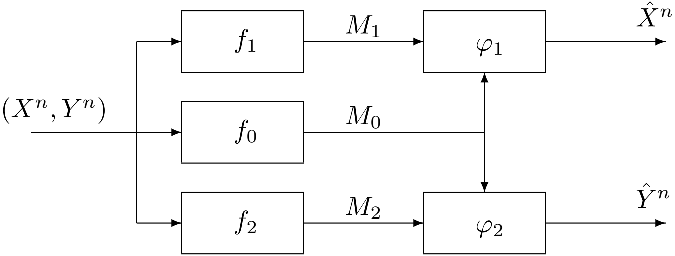
Now take $U = (X,Y) \sim \pi_{XY}$ and note by Markovity $X - W - Y$ that

$$P_{X^n|M_n}(x^n|m)P_{Y^n|M_n}(y^n|m) = P_{U^n|W^n}(u^n|w^n(m)) \text{ and } I(W;U) = I(W;XY).$$

Alternative Interpretation of Wyner's Common Information

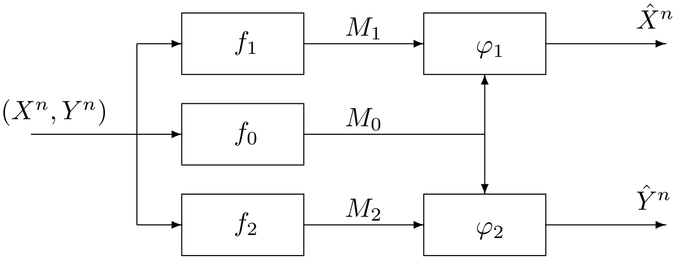


Alternative Interpretation of Wyner's Common Information



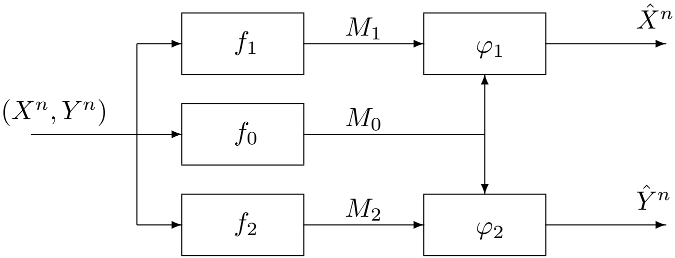
- An (n, R_0, R_1, R_2)-Gray-Wyner code [Gray and Wyner, 1974] consists of

Alternative Interpretation of Wyner's Common Information



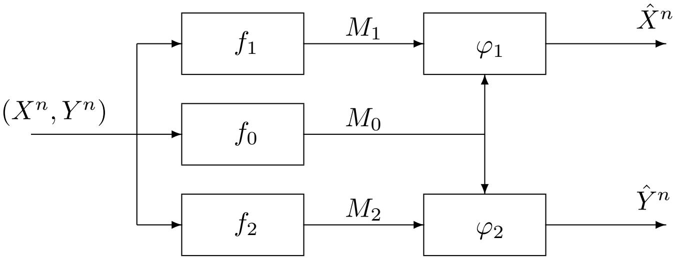
- An (n, R_0, R_1, R_2)-Gray-Wyner code [Gray and Wyner, 1974] consists of
  - Three encoders $f_i : \mathcal{X}^n \times \mathcal{Y}^n \to [2^{nR_i}]$ where $i = 0, 1, 2$;

Alternative Interpretation of Wyner's Common Information



- An (n, R_0, R_1, R_2)-Gray-Wyner code [Gray and Wyner, 1974] consists of
  - Three encoders $f_i : \mathcal{X}^n \times \mathcal{Y}^n \to [2^{nR_i}]$ where $i = 0, 1, 2$;
  - Two decoders $\varphi_1 : [2^{nR_0}] \times [2^{nR_1}] \to \mathcal{X}^n$ and $\varphi_2 : [2^{nR_0}] \times [2^{nR_2}] \to \mathcal{Y}^n$.

Alternative Interpretation of Wyner's Common Information



- An (n, R_0, R_1, R_2)-Gray-Wyner code [Gray and Wyner, 1974] consists of
  - Three encoders $f_i : \mathcal{X}^n \times \mathcal{Y}^n \to [2^{nR_i}]$ where $i = 0, 1, 2$;
  - Two decoders $\varphi_1 : [2^{nR_0}] \times [2^{nR_1}] \to \mathcal{X}^n$ and $\varphi_2 : [2^{nR_0}] \times [2^{nR_2}] \to \mathcal{Y}^n$.
- The probability of error of the code is

$$\Pr\big((\varphi_1(M_0, M_1), \varphi_2(M_0, M_2)) \neq (X^n, Y^n)\big).$$

where $M_i = f_i(X^n, Y^n)$ for $i = 0, 1, 2$.

Alternative Interpretation of Wyner's Common Information

Common information based on the Gray-Wyner system $T_{\mathrm{GW}}(\pi_{XY})$ for $(X,Y) \sim \pi_{XY}$

$\Longleftrightarrow$

Smallest common rate R_0 such that for all $\epsilon > 0$, there exists sequence of (n, R_0, R_1, R_2) Gray-Wyner codes $\{(f_{0,n}, f_{1,n}, f_{2,n}, \varphi_{1,n}, \varphi_{2,n})\}_{n=1}^{\infty}$ such that

$$R_0 + R_1 + R_2 \leq H(XY) + \epsilon$$

and the probability of error vanishes.

Alternative Interpretation of Wyner's Common Information

Common information based on the Gray-Wyner system $T_{\mathrm{GW}}(\pi_{XY})$ for $(X,Y) \sim \pi_{XY}$

$\Longleftrightarrow$

Smallest common rate R_0 such that for all $\epsilon > 0$, there exists sequence of (n, R_0, R_1, R_2) Gray-Wyner codes $\{(f_{0,n}, f_{1,n}, f_{2,n}, \varphi_{1,n}, \varphi_{2,n})\}_{n=1}^{\infty}$ such that

$$R_0 + R_1 + R_2 \leq H(XY) + \epsilon$$

and the probability of error vanishes.

Theorem ([Wyner, 1975])

$$T_{\mathrm{GW}}(\pi_{XY}) = C_{\mathrm{W}}(\pi_{XY})$$

Example: Doubly Symmetric Binary Source (DSBS)

- Consider a DSBS $(X, Y) \in \{0, 1\}^2$ which is defined for $p \in (0, 1/2)$ by

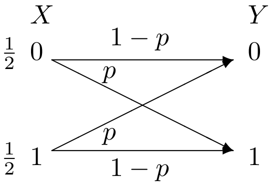
$$\pi_{XY} = \begin{bmatrix} (1-p)/2 & p/2 \\ p/2 & (1-p)/2 \end{bmatrix}$$

Example: Doubly Symmetric Binary Source (DSBS)

- Consider a DSBS $(X, Y) \in \{0, 1\}^2$ which is defined for $p \in (0, 1/2)$ by

$$\pi_{XY} = \begin{bmatrix} (1-p)/2 & p/2 \\ p/2 & (1-p)/2 \end{bmatrix}$$

- Interpretation in terms of $X - W - Y$

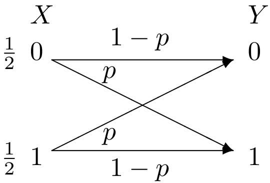


Example: Doubly Symmetric Binary Source (DSBS)

- Consider a DSBS $(X, Y) \in \{0,1\}^2$ which is defined for $p \in (0, 1/2)$ by

$$\pi_{XY} = \begin{bmatrix} (1-p)/2 & p/2 \\ p/2 & (1-p)/2 \end{bmatrix}$$

- Interpretation in terms of $X - W - Y$

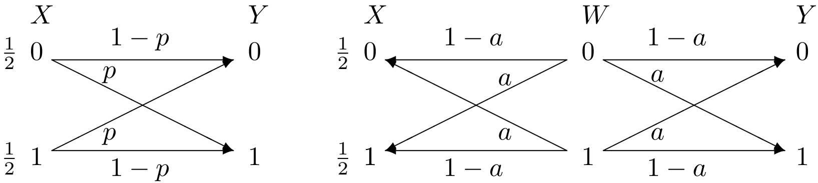


Example: Doubly Symmetric Binary Source (DSBS)

- Consider a DSBS $(X, Y) \in \{0,1\}^2$ which is defined for $p \in (0, 1/2)$ by

$$\pi_{XY} = \begin{bmatrix} (1-p)/2 & p/2 \\ p/2 & (1-p)/2 \end{bmatrix}$$

- Interpretation in terms of $X - W - Y$

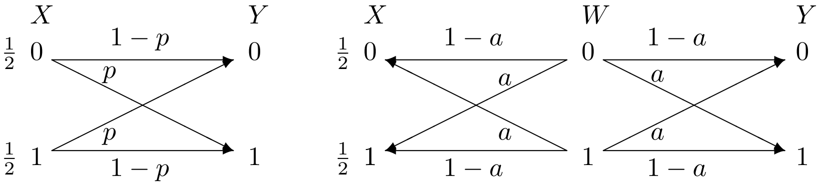


Example: Doubly Symmetric Binary Source (DSBS)

- Consider a DSBS $(X, Y) \in \{0,1\}^2$ which is defined for $p \in (0, 1/2)$ by

$$\pi_{XY} = \begin{bmatrix} (1-p)/2 & p/2 \\ p/2 & (1-p)/2 \end{bmatrix}$$

- Interpretation in terms of $X - W - Y$



- Here, $a * a = p$ and

$$a = \frac{1 - \sqrt{1-2p}}{2} \in (0, 1/2).$$

Example: DSBS

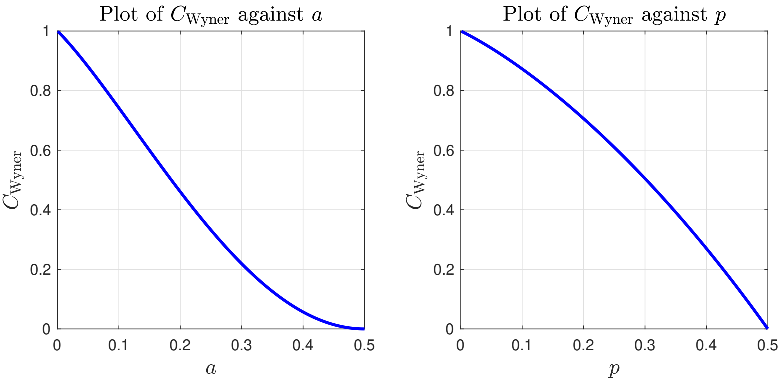


Figure: Plots of Wyner's common information for the DSBS in terms of p and a

Outline

Motivation for Alternative Measures

- Wyner used the normalized relative entropy, i.e.,

$$\inf\left\{R : \lim_{n\to\infty} \frac{D(P_{X^nY^n}\|\pi_{XY}^n)}{n} = 0\right\} = C_{\mathrm{W}}(\pi_{XY}) = \min_{X-W-Y} I(W;XY).$$

Motivation for Alternative Measures

- Wyner used the normalized relative entropy, i.e.,

$$\inf\left\{R : \lim_{n\to\infty} \frac{D(P_{X^nY^n}\|\pi_{XY}^n)}{n} = 0\right\} = C_{\mathrm{W}}(\pi_{XY}) = \min_{X-W-Y} I(W;XY).$$

- What if we do not normalize?

$$\tilde{T}(\pi_{XY}) := \inf\left\{R : \lim_{n\to\infty} D(P_{X^nY^n}\|\pi_{XY}^n) = 0\right\} \geq C_{\mathrm{W}}(\pi_{XY}).$$

We get a stronger measure of dependence.

Motivation for Alternative Measures

- Wyner used the normalized relative entropy, i.e.,

$$\inf\left\{R : \lim_{n\to\infty} \frac{D(P_{X^nY^n}\|\pi_{XY}^n)}{n} = 0\right\} = C_{\mathrm{W}}(\pi_{XY}) = \min_{X-W-Y} I(W;XY).$$

- What if we do not normalize?

$$\tilde{T}(\pi_{XY}) := \inf\left\{R : \lim_{n\to\infty} D(P_{X^nY^n}\|\pi_{XY}^n) = 0\right\} \geq C_{\mathrm{W}}(\pi_{XY}).$$

  We get a stronger measure of dependence.
- What if we want an even stronger measure of dependence?

Motivation for Alternative Measures

- Wyner used the normalized relative entropy, i.e.,

$$\inf\left\{R : \lim_{n\to\infty} \frac{D(P_{X^nY^n}\|\pi_{XY}^n)}{n} = 0\right\} = C_{\mathrm{W}}(\pi_{XY}) = \min_{X-W-Y} I(W;XY).$$

- What if we do not normalize?

$$\tilde{T}(\pi_{XY}) := \inf\left\{R : \lim_{n\to\infty} D(P_{X^nY^n}\|\pi_{XY}^n) = 0\right\} \ge C_{\mathrm{W}}(\pi_{XY}).$$

  We get a stronger measure of dependence.

- What if we want an even stronger measure of dependence?
- Rényi common information for orders ≥ 1 [Yu and Tan, 2018]!

$$T_{1+s}(\pi_{XY}) := \inf\left\{R : \lim_{n\to\infty} \frac{D_{1+s}(P_{X^nY^n}\|\pi_{XY}^n)}{n} = 0\right\}$$

$$\tilde{T}_{1+s}(\pi_{XY}) := \inf\left\{R : \lim_{n\to\infty} D_{1+s}(P_{X^nY^n}\|\pi_{XY}^n) = 0\right\}$$

Rényi Common Information

- Rényi divergence

$$D_{1+s}(P\|Q) := \frac{1}{s}\log\sum_{x\in\mathrm{supp}(P)} P(x)\left(\frac{P(x)}{Q(x)}\right)^s \quad s\in[-1,\infty)$$

$$D_\infty(P\|Q) := \log\max_{x\in\mathrm{supp}(P)}\frac{P(x)}{Q(x)}.$$

Rényi Common Information

- Rényi divergence

$$D_{1+s}(P\|Q) := \frac{1}{s}\log\sum_{x\in\mathrm{supp}(P)} P(x)\left(\frac{P(x)}{Q(x)}\right)^s \qquad s\in[-1,\infty)$$

$$D_\infty(P\|Q) := \log\max_{x\in\mathrm{supp}(P)}\frac{P(x)}{Q(x)}.$$

- The Rényi divergence if monotonically non-decreasing, i.e.,

$$D_{1+s}(P\|Q)\le D_{1+t}(P\|Q) \qquad s\le t.$$

Rényi Common Information

- Rényi divergence

$$D_{1+s}(P\|Q) := \frac{1}{s} \log \sum_{x \in \mathrm{supp}(P)} P(x) \left(\frac{P(x)}{Q(x)} \right)^s \qquad s \in [-1, \infty)$$

$$D_\infty(P\|Q) := \log \max_{x \in \mathrm{supp}(P)} \frac{P(x)}{Q(x)}.$$

- The Rényi divergence if monotonically non-decreasing, i.e.,

$$D_{1+s}(P\|Q) \le D_{1+t}(P\|Q) \qquad s \le t.$$

- Hence, the Rényi common information is also non-decreasing, i.e.,

(normalized) $$T_{1+s}(\pi_{XY}) \le T_{1+t}(\pi_{XY}) \qquad s \le t.$$

and

(unnormalized) $$\tilde{T}_{1+s}(\pi_{XY}) \le \tilde{T}_{1+t}(\pi_{XY}) \qquad s \le t.$$

Rényi Common Information

- Rényi divergence

$$D_{1+s}(P\|Q) := \frac{1}{s} \log \sum_{x \in \mathrm{supp}(P)} P(x) \left(\frac{P(x)}{Q(x)} \right)^s \qquad s \in [-1, \infty)$$

$$D_\infty(P\|Q) := \log \max_{x \in \mathrm{supp}(P)} \frac{P(x)}{Q(x)}.$$

- The Rényi divergence if monotonically non-decreasing, i.e.,

$$D_{1+s}(P\|Q) \le D_{1+t}(P\|Q) \qquad s \le t.$$

- Hence, the Rényi common information is also non-decreasing, i.e.,

(normalized) $$T_{1+s}(\pi_{XY}) \le T_{1+t}(\pi_{XY}) \qquad s \le t.$$

and

(unnormalized) $$\tilde{T}_{1+s}(\pi_{XY}) \le \tilde{T}_{1+t}(\pi_{XY}) \qquad s \le t.$$

- And for a fixed order $1 + s \in [0, \infty]$,

$$T_{1+s}(\pi_{XY}) \le \tilde{T}_{1+s}(\pi_{XY}).$$

Are we doing math for the sake of doing math?

Are we doing math for the sake of doing math?

- The sceptic in you might wonder whether we are just doing math.

Are we doing math for the sake of doing math?

- The sceptic in you might wonder whether we are just doing math.
- In fact not! We show in the sequel that

$$\tilde{T}_\infty(\pi_{XY}) = \text{Exact Common Information of } \pi_{XY}.$$

Exact Common Information was introduced by [Kumar et al., 2014].

Are we doing math for the sake of doing math?

- The sceptic in you might wonder whether we are just doing math.
- In fact not! We show in the sequel that

$$\tilde{T}_\infty(\pi_{XY}) = \text{Exact Common Information of } \pi_{XY}.$$

  Exact Common Information was introduced by [Kumar et al., 2014].
- And it is through this unexpected connection that we show that

$$\text{Exact Common Information of } \pi_{XY} > C_{\mathrm{W}}(\pi_{XY})$$

  for some joint sources π_{XY}.

Are we doing math for the sake of doing math?

- The sceptic in you might wonder whether we are just doing math.
- In fact not! We show in the sequel that

$$\tilde{T}_\infty(\pi_{XY}) = \text{Exact Common Information of } \pi_{XY}.$$

  Exact Common Information was introduced by [Kumar et al., 2014].
- And it is through this unexpected connection that we show that

$$\text{Exact Common Information of } \pi_{XY} > C_{\mathrm{W}}(\pi_{XY})$$

  for some joint sources π_{XY}.
- But let's soldier on and tackle the Rényi common information for now.

Rényi Common Information: The Weaker Case

Rényi Common Information: The Weaker Case

Let's start with a simple exercise. Consider the case $s \in (-1, 0]$ in which

$$T_{1+s}(\pi_{XY}) \le C_{\mathrm{W}}(\pi_{XY})$$

Rényi Common Information: The Weaker Case

Let's start with a simple exercise. Consider the case $s \in (-1, 0]$ in which

$$T_{1+s}(\pi_{XY}) \le C_{\mathrm{W}}(\pi_{XY})$$

Theorem ([Yu and Tan, 2018] [Yu and Tan, 2020a])

For Rényi orders in $(0, 1]$ (i.e., $s \in (-1, 0]$),

$$T_{1+s}(\pi_{XY}) = \tilde{T}_{1+s}(\pi_{XY}) = C_{\mathrm{W}}(\pi_{XY}).$$

Rényi Common Information: The Weaker Case

Let's start with a simple exercise. Consider the case $s \in (-1, 0]$ in which

$$T_{1+s}(\pi_{XY}) \le C_{\mathrm{W}}(\pi_{XY})$$

Theorem ([Yu and Tan, 2018] [Yu and Tan, 2020a])

For Rényi orders in $(0, 1]$ *(i.e.,* $s \in (-1, 0]$*),*

$$T_{1+s}(\pi_{XY}) = \tilde{T}_{1+s}(\pi_{XY}) = C_{\mathrm{W}}(\pi_{XY}).$$

Our stepping stone...

Rényi Common Information: The Weaker Case

Let's start with a simple exercise. Consider the case $s \in (-1, 0]$ in which

$$T_{1+s}(\pi_{XY}) \leq C_{\mathrm{W}}(\pi_{XY})$$

Theorem ([Yu and Tan, 2018] [Yu and Tan, 2020a])

For Rényi orders in $(0, 1]$ *(i.e.,* $s \in (-1, 0]$*),*

$$T_{1+s}(\pi_{XY}) = \tilde{T}_{1+s}(\pi_{XY}) = C_{\mathrm{W}}(\pi_{XY}).$$

Our stepping stone... Total variation distance $|P - Q| := \frac{1}{2} \sum_x |P(x) - Q(x)|$.

Rényi Common Information: The Weaker Case

Let's start with a simple exercise. Consider the case $s \in (-1, 0]$ in which

$$T_{1+s}(\pi_{XY}) \leq C_{\mathrm{W}}(\pi_{XY})$$

Theorem ([Yu and Tan, 2018] [Yu and Tan, 2020a])

For Rényi orders in $(0, 1]$ *(i.e.,* $s \in (-1, 0]$*),*

$$T_{1+s}(\pi_{XY}) = \tilde{T}_{1+s}(\pi_{XY}) = C_{\mathrm{W}}(\pi_{XY}).$$

Our stepping stone... Total variation distance $|P - Q| := \frac{1}{2} \sum_x |P(x) - Q(x)|$.

Theorem ([Yu and Tan, 2018])

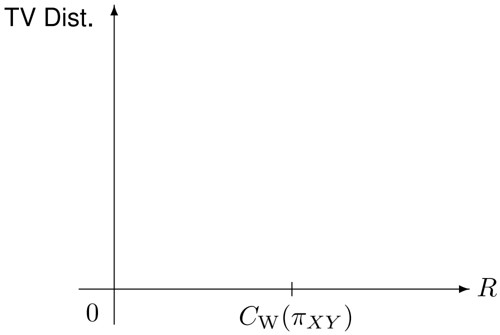
For any $\varepsilon \in [0, 1)$,

$$T_{\varepsilon}^{\mathrm{TV}}(\pi_{XY}) = C_{\mathrm{W}}(\pi_{XY}), \qquad \textit{(Strong converse)}$$

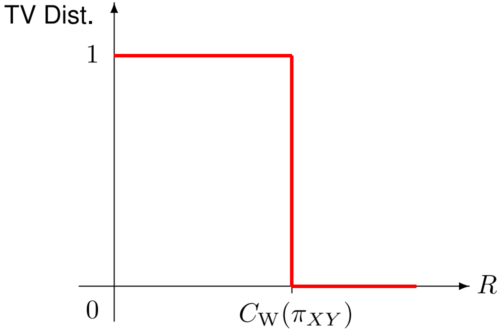
where $T_{\varepsilon}^{\mathrm{TV}}(\pi_{XY})$ *is the minimum simulation rate required to ensure*

$$\limsup_{n \to \infty} |P_{X^n Y^n} - \pi_{XY}^n| \leq \varepsilon.$$

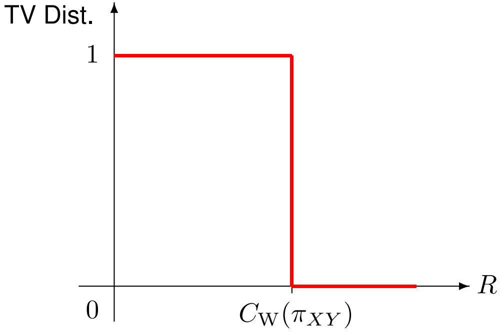
Total Variation Common Information



Total Variation Common Information



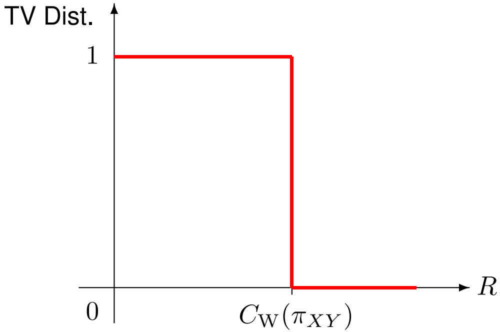
Total Variation Common Information



In fact, we have an exponential strong converse, i.e., if $R < C_{\mathrm{W}}(\pi_{XY})$,

$$|P_{X^nY^n} - \pi_{XY}^n| \geq 1 - 2^{-nE} \quad \text{for some} \quad E > 0.$$

Total Variation Common Information



In fact, we have an exponential strong converse, i.e., if $R < C_{\mathrm{W}}(\pi_{XY})$,

$$|P_{X^nY^n} - \pi_{XY}^n| \geq 1 - 2^{-nE} \quad \text{for some} \quad E > 0.$$

Amenable to second-order?

Total Variation Common Information

- Achievability part follows from the soft-covering lemma.

$$\text{If} \quad R > I(XY;W) \quad \text{then} \quad \lim_{n\to\infty} |P_{X^nY^n} - \pi_{XY}^n| = 0.$$

Total Variation Common Information

- Achievability part follows from the soft-covering lemma.

$$\text{If} \quad R > I(XY;W) \quad \text{then} \quad \lim_{n\to\infty} |P_{X^nY^n} - \pi_{XY}^n| = 0.$$

- Converse requires a very cool information spectrum, single-letterization idea from [Oohama, 2018].

entropy MDPI

Article

Exponential Strong Converse for Source Coding with Side Information at the Decoder †

Yasutada Oohama

Department of Communication Engineering and Informatics, University of Electro-Communications, Tokyo 182-8585, Japan; oohama@uec.ac.jp; Tel.: +81-42-443-5358

† This paper is an extended version of our paper published in 2016 International Symposium on Information Theory and Its Applications, Monterey, CA, USA, 6–9 November 2016; pp. 171–175.

Received: 31 January 2018; Accepted: 20 April 2018; Published: 8 May 2018

check for updates

Going Back to Rényi CI: The Weaker Case $s \in (-1, 0]$

- Because $T_{1+s}(\pi_{XY}) \le C_{\mathrm{W}}(\pi_{XY})$, only have to prove the converse.

Going Back to Rényi CI: The Weaker Case $s \in (-1, 0]$

- Because $T_{1+s}(\pi_{XY}) \leq C_{\mathrm{W}}(\pi_{XY})$, only have to prove the converse.
- Main idea is a Pinsker-type inequality due to [Sason, 2016].

IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 62, NO. 1, JANUARY 2016 23

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

Igal Sason, *Senior Member, IEEE*

Going Back to Rényi CI: The Weaker Case $s \in (-1, 0]$

- Because $T_{1+s}(\pi_{XY}) \le C_{\mathrm{W}}(\pi_{XY})$, only have to prove the converse.
- Main idea is a Pinsker-type inequality due to [Sason, 2016].

IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 62, NO. 1, JANUARY 2016 23

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

Igal Sason, *Senior Member, IEEE*

Lemma

For any $s \in (-1, 0]$,

$$\inf_{P_X, Q_X : |P_X - Q_X| \ge \epsilon} D_{1+s}(P_X \| Q_X) = \inf_{q \in [0, 1-\epsilon]} d_{1+s}(q + \epsilon \| q)$$

and

$$\inf_{q \in [0, 1-\epsilon]} d_{1+s}(q + \epsilon \| q) \ge \left[\min\left\{ 1, \frac{1+s}{s} \right\} \log \frac{1}{1-\epsilon} + \frac{1}{s} \log 2 \right]^{+}$$

Going Back to Rényi CI: The Weaker Case $s \in (-1, 0]$

- From [Sason, 2016], we have

$$\inf_{P_X, Q_X : |P_X - Q_X| \geq \epsilon} D_{1+s}(P_X \| Q_X) \geq \left[\min\left\{1, \frac{1+s}{s}\right\} \log \frac{1}{1-\epsilon} + \frac{1}{s} \log 2 \right]^+$$

Going Back to Rényi CI: The Weaker Case $s \in (-1, 0]$

- From [Sason, 2016], we have

$$\inf_{P_X, Q_X : |P_X - Q_X| \geq \epsilon} D_{1+s}(P_X \| Q_X) \geq \left[\min\left\{1, \frac{1+s}{s}\right\} \log \frac{1}{1-\epsilon} + \frac{1}{s} \log 2 \right]^+$$

- If $R < C_{\mathrm{W}}(\pi_{XY})$, exponential strong converse to TV CI says

$$|P_{X^n Y^n} - \pi_{XY}^n| \geq 1 - 2^{-nE} \quad \text{for some} \quad E > 0.$$

Going Back to Rényi CI: The Weaker Case $s \in (-1, 0]$

- From [Sason, 2016], we have

$$\inf_{P_X, Q_X : |P_X - Q_X| \ge \epsilon} D_{1+s}(P_X \| Q_X) \ge \left[\min\left\{1, \frac{1+s}{s}\right\} \log \frac{1}{1-\epsilon} + \frac{1}{s} \log 2 \right]^+$$

- If $R < C_{\mathrm{W}}(\pi_{XY})$, exponential strong converse to TV CI says

$$|P_{X^n Y^n} - \pi_{XY}^n| \ge 1 - 2^{-nE} \quad \text{for some} \quad E > 0.$$

- Thus, if $R < C_{\mathrm{W}}(\pi_{XY})$

$$\frac{1}{n} \inf_{P_X, Q_X : |P_X - Q_X| \ge \epsilon} D_{1+s}(P_X \| Q_X) \ge \frac{1}{n} \left[\min\left\{1, \frac{1+s}{s}\right\} nE + \frac{1}{s} \log 2 \right]^+$$

  and the normalized Rényi divergence cannot vanish.

Rényi CI: The Stronger Case $s \in (0,1] \cup \{\infty\}$

- For $s \in (0,1] \cup \{\infty\}$,
$$C_{\mathrm{W}}(\pi_{XY}) \le T_{1+s}(\pi_{XY}).$$

Rényi CI: The Stronger Case $s \in (0, 1] \cup \{\infty\}$

- For $s \in (0, 1] \cup \{\infty\}$,
$$C_{\mathrm{W}}(\pi_{XY}) \le T_{1+s}(\pi_{XY}).$$
- We only discuss the case $s = \infty$ in this tutorial.

Rényi CI: The Stronger Case $s \in (0, 1] \cup \{\infty\}$

- For $s \in (0, 1] \cup \{\infty\}$,
$$C_{\mathrm{W}}(\pi_{XY}) \le T_{1+s}(\pi_{XY}).$$
- We only discuss the case $s = \infty$ in this tutorial.
- For the other cases (i.e., $s \ge 1$ finite), see our upcoming monograph.

Rényi CI: The Stronger Case $s \in (0,1] \cup \{\infty\}$

- For $s \in (0,1] \cup \{\infty\}$,
$$C_{\mathrm{W}}(\pi_{XY}) \le T_{1+s}(\pi_{XY}).$$
- We only discuss the case $s = \infty$ in this tutorial.
- For the other cases (i.e., $s \ge 1$ finite), see our upcoming monograph.

Definition

The maximal cross entropy w.r.t. $(X,Y) \sim \pi_{XY}$ over couplings of (P_X, P_Y) is

$$\mathsf{H}_\infty(P_X, P_Y \| \pi_{XY}) := \max_{Q_{XY} \in \mathcal{C}(P_X, P_Y)} \sum_{x,y} Q_{XY}(x,y) \log \frac{1}{\pi_{XY}(x,y)},$$

where

$$\mathcal{C}(P_X, P_Y) := \{Q_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) : Q_X = P_X, Q_Y = P_Y\}.$$

- For $s \in (0,1] \cup \{\infty\}$,
$$C_{\mathrm{W}}(\pi_{XY}) \le T_{1+s}(\pi_{XY}).$$
- We only discuss the case $s = \infty$ in this tutorial.
- For the other cases (i.e., $s \ge 1$ finite), see our upcoming monograph.

Definition

The maximal cross entropy w.r.t. $(X,Y) \sim \pi_{XY}$ over couplings of (P_X, P_Y) is

$$\mathrm{H}_\infty(P_X, P_Y \| \pi_{XY}) := \max_{Q_{XY} \in \mathcal{C}(P_X, P_Y)} \sum_{x,y} Q_{XY}(x,y) \log \frac{1}{\pi_{XY}(x,y)},$$

where

$$\mathcal{C}(P_X, P_Y) := \{Q_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) : Q_X = P_X, Q_Y = P_Y\}.$$

- $\mathrm{H}_\infty(\pi_X, \pi_Y \| \pi_{XY}) \ge H_\pi(X;Y)$ with equality iff $\pi_{XY} = \pi_X \pi_Y$.

Intuition for the Maximal Cross Entropy

- Take a sequence of n-types $T_X^{(n)} \in \mathcal{P}_n(\mathcal{X})$ and $T_Y^{(n)} \in \mathcal{P}_n(\mathcal{Y})$.

Intuition for the Maximal Cross Entropy

- Take a sequence of n-types $T_X^{(n)} \in \mathcal{P}_n(\mathcal{X})$ and $T_Y^{(n)} \in \mathcal{P}_n(\mathcal{Y})$.
- Let them converge as

$$T_X^{(n)} \to P_X \quad \text{and} \quad T_Y^{(n)} \to P_Y .$$

Intuition for the Maximal Cross Entropy

- Take a sequence of n-types $T_X^{(n)} \in \mathcal{P}_n(\mathcal{X})$ and $T_Y^{(n)} \in \mathcal{P}_n(\mathcal{Y})$.
- Let them converge as

$$T_X^{(n)} \to P_X \quad \text{and} \quad T_Y^{(n)} \to P_Y .$$

- What's the minimum π_{XY}^n-probability of (x^n, y^n) where x^n has type $T_X^{(n)}$ and y^n has type $T_Y^{(n)}$, i.e.,

$$\min_{T_{x^n}=T_X^{(n)},T_{y^n}=T_Y^{(n)}} \pi_{XY}^n(x^n, y^n)?$$

Intuition for the Maximal Cross Entropy

- Take a sequence of n-types $T_X^{(n)} \in \mathcal{P}_n(\mathcal{X})$ and $T_Y^{(n)} \in \mathcal{P}_n(\mathcal{Y})$.
- Let them converge as

$$T_X^{(n)} \to P_X \quad \text{and} \quad T_Y^{(n)} \to P_Y .$$

- What's the minimum π_{XY}^n-probability of (x^n, y^n) where x^n has type $T_X^{(n)}$ and y^n has type $T_Y^{(n)}$, i.e.,

$$\min_{T_{x^n}=T_X^{(n)}, T_{y^n}=T_Y^{(n)}} \pi_{XY}^n(x^n, y^n)?$$

- By type gymnastics,

$$\min_{T_{x^n}=T_X^{(n)}, T_{y^n}=T_Y^{(n)}} \pi_{XY}^n(x^n, y^n) \doteq \exp\big(- n\mathsf{H}_\infty(P_X, P_Y \| \pi_{XY})\big).$$

Intuition for the Maximal Cross Entropy

- Take a sequence of n-types $T_X^{(n)} \in \mathcal{P}_n(\mathcal{X})$ and $T_Y^{(n)} \in \mathcal{P}_n(\mathcal{Y})$.
- Let them converge as

$$T_X^{(n)} \to P_X \quad \text{and} \quad T_Y^{(n)} \to P_Y .$$

- What's the minimum π_{XY}^n-probability of (x^n, y^n) where x^n has type $T_X^{(n)}$ and y^n has type $T_Y^{(n)}$, i.e.,

$$\min_{T_{x^n} = T_X^{(n)}, T_{y^n} = T_Y^{(n)}} \pi_{XY}^n(x^n, y^n)?$$

- By type gymnastics,

$$\min_{T_{x^n} = T_X^{(n)}, T_{y^n} = T_Y^{(n)}} \pi_{XY}^n(x^n, y^n) \doteq \exp\big(- n\mathsf{H}_\infty(P_X, P_Y \| \pi_{XY})\big).$$

- So $\mathsf{H}_\infty(P_X, P_Y \| \pi_{XY})$ is the exponential decay rate of this probability.

Upper and Lower Pseudo Common Informations

Definition

The upper pseudo-common information is

$$\overline{\Gamma}_\infty(\pi_{XY}) := \min_{\substack{P_W P_{X|W} P_{Y|W}:\\ P_{XY}=\pi_{XY}}} -H(XY|W) + \mathsf{E}_{P_W}\left[\mathsf{H}_\infty(P_{X|W}, P_{Y|W} \| \pi_{XY})\right]$$

Upper and Lower Pseudo Common Informations

Definition

The upper pseudo-common information is

$$\overline{\Gamma}_\infty(\pi_{XY}) := \min_{\substack{P_W P_{X|W} P_{Y|W}: \\ P_{XY}=\pi_{XY}}} -H(XY|W) + \mathsf{E}_{P_W}\left[\mathsf{H}_\infty(P_{X|W}, P_{Y|W}\|\pi_{XY})\right]$$

Contrast to Wyner's common information

$$C_{\mathrm{W}}(\pi_{XY}) = \min_{\substack{P_W P_{X|W} P_{Y|W}: \\ P_{XY}=\pi_{XY}}} -H(XY|W) + H(XY).$$

Upper and Lower Pseudo Common Informations

Definition

The upper pseudo-common information is

$$\overline{\Gamma}_\infty(\pi_{XY}) := \min_{\substack{P_W P_{X|W} P_{Y|W}: \\ P_{XY}=\pi_{XY}}} -H(XY|W) + \mathsf{E}_{P_W}\left[\mathsf{H}_\infty(P_{X|W}, P_{Y|W} \| \pi_{XY})\right]$$

Contrast to Wyner's common information

$$C_{\mathrm{W}}(\pi_{XY}) = \min_{\substack{P_W P_{X|W} P_{Y|W}: \\ P_{XY}=\pi_{XY}}} -H(XY|W) + H(XY).$$

Definition

The lower pseudo-common information is

$$\underline{\Gamma}_\infty(\pi_{XY}) := \inf_{\substack{P_W P_{X|W} P_{Y|W}: \\ P_{XY}=\pi_{XY}}} -H(XY|W) + \inf_{Q_{WW'} \in \mathcal{C}(P_W, P_W)} \mathsf{E}_{Q_{WW'}}\left[\mathsf{H}_\infty(P_{X|W}, P_{Y|W'} \| \pi_{XY})\right].$$

Upper and Lower Pseudo Common Informations

Definition

The upper pseudo-common information is

$$\overline{\Gamma}_\infty(\pi_{XY}) := \min_{\substack{P_W P_{X|W} P_{Y|W}: \\ P_{XY}=\pi_{XY}}} -H(XY|W) + \mathsf{E}_{P_W}\left[\mathsf{H}_\infty(P_{X|W}, P_{Y|W} \| \pi_{XY})\right]$$

Contrast to Wyner's common information

$$C_{\mathrm{W}}(\pi_{XY}) = \min_{\substack{P_W P_{X|W} P_{Y|W}: \\ P_{XY}=\pi_{XY}}} -H(XY|W) + H(XY).$$

Definition

The lower pseudo-common information is

$$\underline{\Gamma}_\infty(\pi_{XY}) := \inf_{\substack{P_W P_{X|W} P_{Y|W}: \\ P_{XY}=\pi_{XY}}} -H(XY|W) + \inf_{Q_{WW'} \in \mathcal{C}(P_W, P_W)} \mathsf{E}_{Q_{WW'}}\left[\mathsf{H}_\infty(P_{X|W}, P_{Y|W'} \| \pi_{XY})\right].$$

Rényi Common Information of order ∞

Theorem ([Yu and Tan, 2020a] [Yu and Tan, 2020c])

The order-∞ Rényi common information admits the following single-letter bounds

$$\tilde{T}_\infty(\pi_{XY}) \geq T_\infty(\pi_{XY}) \geq \max\{\underline{\Gamma}_\infty(\pi_{XY}), C_{\mathrm{W}}(\pi_{XY})\}$$

and

$$T_\infty(\pi_{XY}) \leq \tilde{T}_\infty(\pi_{XY}) \leq \overline{\Gamma}_\infty(\pi_{XY}).$$

Rényi Common Information of order ∞

Theorem ([Yu and Tan, 2020a] [Yu and Tan, 2020c])

The order-∞ Rényi common information admits the following single-letter bounds

$$\tilde{T}_\infty(\pi_{XY}) \geq T_\infty(\pi_{XY}) \geq \max\{\underline{\Gamma}_\infty(\pi_{XY}), C_{\mathrm{W}}(\pi_{XY})\}$$

and

$$T_\infty(\pi_{XY}) \leq \tilde{T}_\infty(\pi_{XY}) \leq \overline{\Gamma}_\infty(\pi_{XY}).$$

Achievability: Rényi soft-covering [Yu and Tan, 2019d] and truncated product distributions.

Rényi Common Information of order ∞

Theorem ([Yu and Tan, 2020a] [Yu and Tan, 2020c])

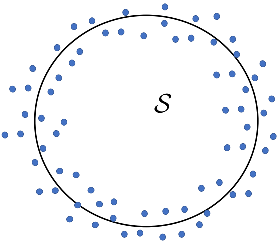
The order-∞ Rényi common information admits the following single-letter bounds

$$\tilde{T}_\infty(\pi_{XY}) \geq T_\infty(\pi_{XY}) \geq \max\{\underline{\Gamma}_\infty(\pi_{XY}), C_{\mathrm{W}}(\pi_{XY})\}$$

and

$$T_\infty(\pi_{XY}) \leq \tilde{T}_\infty(\pi_{XY}) \leq \overline{\Gamma}_\infty(\pi_{XY}).$$

Achievability: Rényi soft-covering [Yu and Tan, 2019d] and truncated product distributions.



Product distribution

$$P_W^n(w^n) = \prod_{i=1}^n P_W(w_i)$$

Rényi Common Information of order ∞

Theorem ([Yu and Tan, 2020a] [Yu and Tan, 2020c])

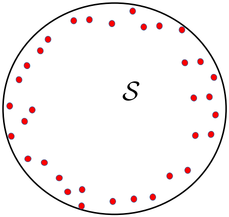
The order-∞ Rényi common information admits the following single-letter bounds

$$\tilde{T}_\infty(\pi_{XY}) \geq T_\infty(\pi_{XY}) \geq \max\{\underline{\Gamma}_\infty(\pi_{XY}), C_{\mathrm{W}}(\pi_{XY})\}$$

and

$$T_\infty(\pi_{XY}) \leq \tilde{T}_\infty(\pi_{XY}) \leq \overline{\Gamma}_\infty(\pi_{XY}).$$

Achievability: Rényi soft-covering [Yu and Tan, 2019d] and truncated product distributions.



Truncated product distribution

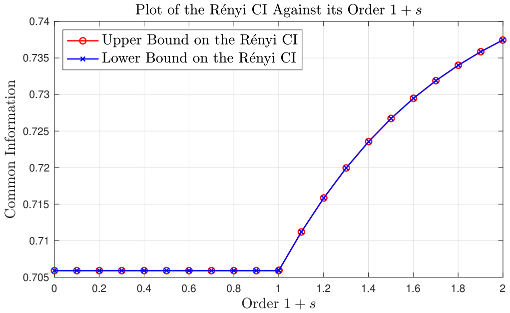
$$P_{W^n}(w^n) \propto \left(\prod_{i=1}^{n} P_W(w_i)\right) \mathbb{1}\{w^n \in \mathcal{S}\}$$

Rényi Common Information of other orders $\in (1, \infty)$?

- Can obtain similar bounds [Yu and Tan, 2020a]

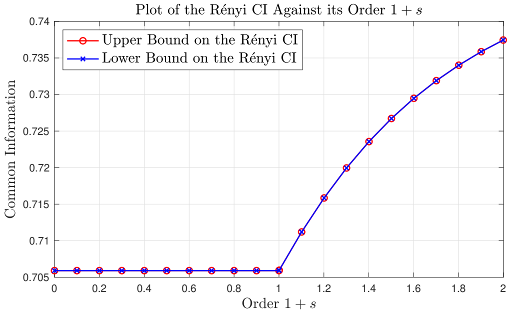
Rényi Common Information of other orders $\in (1,\infty)$?

- Can obtain similar bounds [Yu and Tan, 2020a]
- For the DSBS, for $1+s \in [0,2]$, after some calculations, we get



Rényi Common Information of other orders $\in (1, \infty)$?

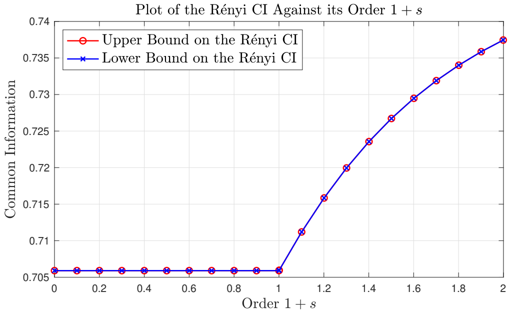
- Can obtain similar bounds [Yu and Tan, 2020a]
- For the DSBS, for $1 + s \in [0, 2]$, after some calculations, we get



- Rényi common information for the DSBS increases with $1 + s \in [1, 2]$!

Rényi Common Information of other orders $\in (1, \infty)$?

- Can obtain similar bounds [Yu and Tan, 2020a]
- For the DSBS, for $1 + s \in [0, 2]$, after some calculations, we get



- Rényi common information for the DSBS increases with $1 + s \in [1, 2]$!

Does this have more profound implications?

Outline

Exact Common Information?

- In the distributed source simulation problem à la Wyner, we mandated that

$$\frac{1}{n} D(P_{X^n Y^n} \| \pi_{XY}^n) \to 0.$$

Exact Common Information?

- In the distributed source simulation problem à la Wyner, we mandated that

$$\frac{1}{n}D(P_{X^nY^n}\|\pi_{XY}^n) \to 0.$$

- What if we require

$$P_{X^nY^n} = \pi_{XY}^n \quad \text{for some} \quad n \in \mathbb{N}?$$

Exact Common Information?

- In the distributed source simulation problem à la Wyner, we mandated that

$$\frac{1}{n} D(P_{X^n Y^n} \| \pi_{XY}^n) \to 0.$$

- What if we require

$$P_{X^n Y^n} = \pi_{XY}^n \quad \text{for some} \quad n \in \mathbb{N}?$$

- Using fixed-length block codes, we need rate $\lim_{n\to\infty} \frac{1}{n} \log |\mathcal{W}_n|$ over $W \in \mathcal{W}_n$ such that $X^n - W - Y^n$! Potentially up to $\min\{\log |\mathcal{X}|, \log |\mathcal{Y}|\}$.

Exact Common Information?

- In the distributed source simulation problem à la Wyner, we mandated that

$$\frac{1}{n} D(P_{X^n Y^n} \| \pi_{XY}^n) \to 0.$$

- What if we require

$$P_{X^n Y^n} = \pi_{XY}^n \quad \text{for some} \quad n \in \mathbb{N}?$$

- Using fixed-length block codes, we need rate $\lim_{n\to\infty} \frac{1}{n} \log |\mathcal{W}_n|$ over $W \in \mathcal{W}_n$ such that $X^n - W - Y^n$! Potentially up to $\min\{\log |\mathcal{X}|, \log |\mathcal{Y}|\}$.

- In come [Kumar et al., 2014], who introduced

2014 IEEE International Symposium on Information Theory

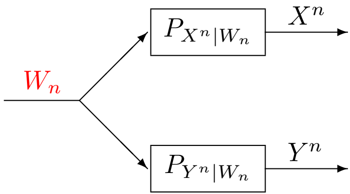
Exact Common Information

Gowtham Ramani Kumar
Electrical Engineering
Stanford University
Email: gowthamr@stanford.edu

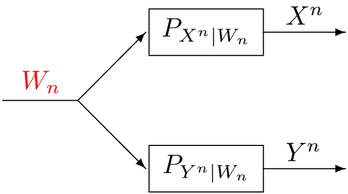
Cheuk Ting Li
Electrical Engineering
Stanford University
Email: ctli@stanford.edu

Abbas El Gamal
Electrical Engineering
Stanford University
Email: abbas@stanford.edu

Exact Common Information

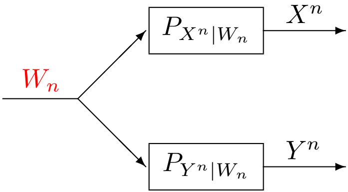


Exact Common Information



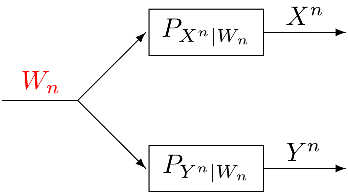
- A synthesis code $(P_{W_n}, P_{X^n|W_n}, P_{Y^n|W_n})$

Exact Common Information



- A synthesis code $(P_{W_n}, P_{X^n|W_n}, P_{Y^n|W_n})$
- W_n can be any (not necessarily uniform) discrete random variable

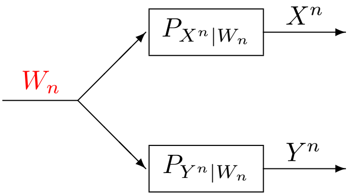
Exact Common Information



- A synthesis code $(P_{W_n}, P_{X^n|W_n}, P_{Y^n|W_n})$
- W_n can be any (not necessarily uniform) discrete random variable
- Distribution induced by the code is

$$P_{X^nY^n}(x^n, y^n) := \sum_w P_{W_n}(w)P_{X^n|W_n}(x^n|w)P_{Y^n|W_n}(y^n|w).$$

Exact Common Information



- A synthesis code $(P_{W_n}, P_{X^n|W_n}, P_{Y^n|W_n})$
- W_n can be any (not necessarily uniform) discrete random variable
- Distribution induced by the code is

$$P_{X^nY^n}(x^n, y^n) := \sum_{w} P_{W_n}(w) P_{X^n|W_n}(x^n|w) P_{Y^n|W_n}(y^n|w).$$

- Require

$$P_{X^nY^n} = \pi_{XY}^n \quad \text{for some} \quad n \in \mathbb{N}.$$

Exact Common Information

Asymptotic rate induced by the code is

$$\lim_{n\to\infty} \frac{H(W_n)}{n}$$

Exact Common Information

Asymptotic rate induced by the code is

$$\lim_{n\to\infty} \frac{H(W_n)}{n}$$

- Compress W_n by a prefix-free, zero-error variable-length code (e.g., Shannon-Fano or Huffman code)

$$f : \mathcal{W}_n \to \{0,1\}^* := \bigcup_{n\geq 1} \{0,1\}^n$$

- Let the length of W_n be $\ell(W_n)$.

Exact Common Information

Asymptotic rate induced by the code is

$$\lim_{n\to\infty} \frac{H(W_n)}{n}$$

- Compress W_n by a prefix-free, zero-error variable-length code (e.g., Shannon-Fano or Huffman code)

$$f : \mathcal{W}_n \to \{0,1\}^* := \bigcup_{n\geq 1} \{0,1\}^n$$

- Let the length of W_n be $\ell(W_n)$.
- Then, by Shannon's zero-error compression theorem, the optimal expected codeword length $L(W_n) = \mathbb{E}[\ell(W_n)]$ satisfies

$$H(W_n) \leq L(W_n) < H(W_n) + 1$$

which implies that

$$\lim_{n\to\infty} \frac{L(W_n)}{n} = \lim_{n\to\infty} \frac{H(W_n)}{n}.$$

Exact Common Information

Definition

The exact common information is defined as

$$T_{\mathrm{Ex}}(\pi_{XY}) := \inf\left\{\lim_{n\to\infty}\frac{L(W_n)}{n} : P_{X^nY^n} = \pi_{XY}^n \text{ for some } n \geq 1\right\}$$

Exact Common Information

Definition

The exact common information is defined as

$$T_{\mathrm{Ex}}(\pi_{XY}) := \inf \left\{ \lim_{n\to\infty} \frac{L(W_n)}{n} \; : \; P_{X^nY^n} = \pi_{XY}^n \text{ for some } n \ge 1 \right\}$$

Theorem ([Kumar et al., 2014])

$$T_{\mathrm{Ex}}(\pi_{XY}) = \lim_{n\to\infty} \frac{1}{n} \min_{\substack{P_{W_n} P_{X^n|W_n} P_{Y^n|W_n} : \\ P_{X^nY^n} = \pi_{XY}^n}} H(W_n).$$

Exact Common Information

Definition

The exact common information is defined as

$$T_{\mathrm{Ex}}(\pi_{XY}) := \inf\left\{ \lim_{n\to\infty} \frac{L(W_n)}{n} \;:\; P_{X^nY^n} = \pi_{XY}^n \text{ for some } n \ge 1 \right\}$$

Theorem ([Kumar et al., 2014])

$$T_{\mathrm{Ex}}(\pi_{XY}) = \lim_{n\to\infty} \frac{1}{n} \min_{\substack{P_{W_n}P_{X^n|W_n}P_{Y^n|W_n}: \\ P_{X^nY^n}=\pi_{XY}^n}} H(W_n).$$

- Multi-letter characterization!

Definition

The exact common information is defined as

$$T_{\mathrm{Ex}}(\pi_{XY}) := \inf\left\{ \lim_{n\to\infty} \frac{L(W_n)}{n} : P_{X^nY^n} = \pi_{XY}^n \text{ for some } n \geq 1 \right\}$$

Theorem ([Kumar et al., 2014])

$$T_{\mathrm{Ex}}(\pi_{XY}) = \lim_{n\to\infty} \frac{1}{n} \min_{\substack{P_{W_n}P_{X^n|W_n}P_{Y^n|W_n}: \\ P_{X^nY^n}=\pi_{XY}^n}} H(W_n).$$

- Multi-letter characterization!
- Exact CI $\geq$ Wyner's CI

Exact Common Information

Definition

The exact common information is defined as

$$T_{\mathrm{Ex}}(\pi_{XY}) := \inf\left\{ \lim_{n\to\infty} \frac{L(W_n)}{n} : P_{X^nY^n} = \pi_{XY}^n \text{ for some } n \geq 1 \right\}$$

Theorem ([Kumar et al., 2014])

$$T_{\mathrm{Ex}}(\pi_{XY}) = \lim_{n\to\infty} \frac{1}{n} \min_{\substack{P_{W_n}P_{X^n|W_n}P_{Y^n|W_n}: \\ P_{X^nY^n}=\pi_{XY}^n}} H(W_n).$$

- Multi-letter characterization!
- Exact CI $\geq$ Wyner's CI
- Exact CI $>$ Wyner's CI?
- Open problem posed by [Kumar et al., 2014]

Exact Common Information

Definition

The exact common information is defined as

$$T_{\mathrm{Ex}}(\pi_{XY}) := \inf\left\{ \lim_{n\to\infty} \frac{L(W_n)}{n} : P_{X^nY^n} = \pi_{XY}^n \text{ for some } n \geq 1 \right\}$$

Theorem ([Kumar et al., 2014])

$$T_{\mathrm{Ex}}(\pi_{XY}) = \lim_{n\to\infty} \frac{1}{n} \min_{\substack{P_{W_n} P_{X^n|W_n} P_{Y^n|W_n} : \\ P_{X^nY^n} = \pi_{XY}^n}} H(W_n).$$

- Multi-letter characterization!
- Exact CI $\geq$ Wyner's CI
- Exact CI $>$ Wyner's CI?
- Open problem posed by [Kumar et al., 2014]

As expected the exact common information rate is greater than or equal to the Wyner common information.

Proposition 3.

$$\overline{G}(X;Y) \geq J(X;Y).$$

In the following section, we show that they are equal for the SBES in Example 1. We do not know if this is the case in general, however.

From [Kumar et al., 2014]

Exact Common Information

Definition

The exact common information is defined as

$$T_{\mathrm{Ex}}(\pi_{XY}) := \inf\left\{\lim_{n\to\infty}\frac{L(W_n)}{n} : P_{X^nY^n} = \pi_{XY}^n \text{ for some } n \geq 1\right\}$$

Theorem ([Kumar et al., 2014])

$$T_{\mathrm{Ex}}(\pi_{XY}) = \lim_{n\to\infty}\frac{1}{n}\min_{\substack{P_{W_n}P_{X^n|W_n}P_{Y^n|W_n}:\\ P_{X^nY^n}=\pi_{XY}^n}} H(W_n).$$

- Multi-letter characterization!
- Exact CI $\geq$ Wyner's CI
- Exact CI $>$ Wyner's CI?
- Open problem posed by [Kumar et al., 2014]

the exact common information rate. While this multiletter characterization is in general greater than or equal to the Wyner common information, we showed that they are equal for the SBES. The main open question is whether the exact common information rate has a single letter characterization in general. Is it always equal to the Wyner common information? Is there an example 2-DMS for which the exact common information rate is strictly larger than the Wyner common information? It would also be interesting to further explore the application to machine learning.

From [Kumar et al., 2014]

Surprising Equivalence: ∞-Rényi CI and Exact CI

Theorem ([Yu and Tan, 2020c])

For a bivariate source π_{XY} on a finite alphabet,

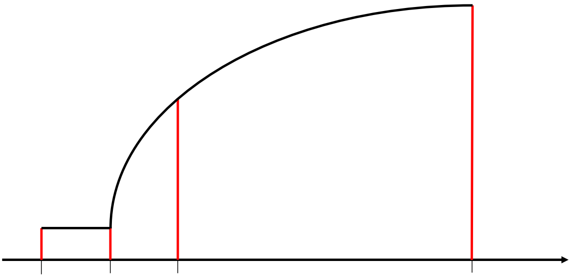
$$T_{\mathrm{Ex}}(\pi_{XY}) = \tilde{T}_{\infty}(\pi_{XY}).$$

Surprising Equivalence: ∞-Rényi CI and Exact CI

Theorem ([Yu and Tan, 2020c])

For a bivariate source π_{XY} on a finite alphabet,

$$T_{\mathrm{Ex}}(\pi_{XY}) = \tilde{T}_{\infty}(\pi_{XY}).$$

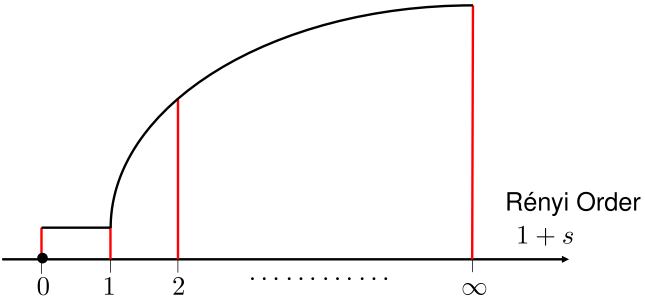


Surprising Equivalence: ∞-Rényi CI and Exact CI

Theorem ([Yu and Tan, 2020c])

For a bivariate source π_{XY} *on a finite alphabet,*

$$T_{\mathrm{Ex}}(\pi_{XY}) = \tilde{T}_{\infty}(\pi_{XY}).$$

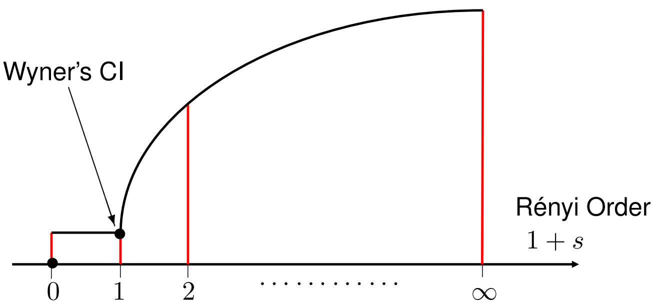


Surprising Equivalence: ∞-Rényi CI and Exact CI

Theorem ([Yu and Tan, 2020c])

For a bivariate source π_{XY} on a finite alphabet,

$$T_{\mathrm{Ex}}(\pi_{XY}) = \tilde{T}_{\infty}(\pi_{XY}).$$

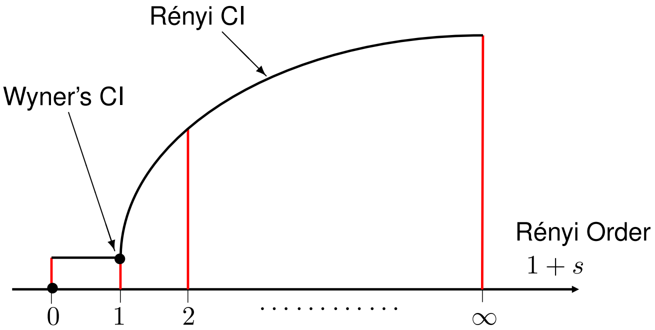


Surprising Equivalence: ∞-Rényi CI and Exact CI

Theorem ([Yu and Tan, 2020c])

For a bivariate source π_{XY} on a finite alphabet,

$$T_{\mathrm{Ex}}(\pi_{XY}) = \tilde{T}_{\infty}(\pi_{XY}).$$

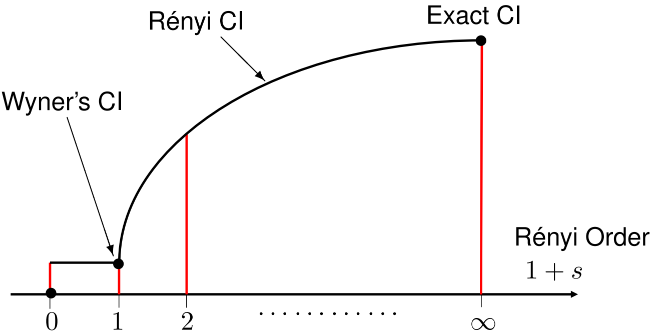


Surprising Equivalence: ∞-Rényi CI and Exact CI

Theorem ([Yu and Tan, 2020c])

For a bivariate source π_{XY} on a finite alphabet,

$$T_{\mathrm{Ex}}(\pi_{XY}) = \tilde{T}_{\infty}(\pi_{XY}).$$



Proof of $\Longrightarrow$ Part of Equivalence Theorem

Lemma ([Kumar et al., 2014], [Vellambi and Kliewer, 2016])

$\exists$ *rate-*R ∞*-Rényi CI code* $\Longrightarrow$ $\exists$ *rate-*R *Exact CI code*

Proof of $\Longrightarrow$ Part of Equivalence Theorem

Lemma ([Kumar et al., 2014], [Vellambi and Kliewer, 2016])

$\exists$ *rate-*R ∞*-Rényi CI code* $\Longrightarrow$ $\exists$ *rate-*R *Exact CI code*

- $\exists$ rate-R ∞-Rényi CI code

$$D_\infty(P_{X^nY^n}\|\pi_{XY}^n) < \epsilon \quad \Longrightarrow \quad P_{X^nY^n}(x^n, y^n) < 2^\epsilon \pi_{XY}^n(x^n, y^n)$$

Proof of $\Longrightarrow$ Part of Equivalence Theorem

Lemma ([Kumar et al., 2014], [Vellambi and Kliewer, 2016])

$\exists$ *rate-R ∞-Rényi CI code* $\Longrightarrow$ $\exists$ *rate-R Exact CI code*

- $\exists$ rate-R ∞-Rényi CI code

$$D_\infty(P_{X^nY^n}\|\pi_{XY}^n) < \epsilon \quad \Longrightarrow \quad P_{X^nY^n}(x^n, y^n) < 2^\epsilon \pi_{XY}^n(x^n, y^n)$$

- Define

$$\widehat{P}_{X^nY^n}(x^n, y^n) := \frac{2^\epsilon \pi_{XY}^n(x^n, y^n) - P_{X^nY^n}(x^n, y^n)}{2^\epsilon - 1},$$

then obviously, $\widehat{P}_{X^nY^n}(x^n, y^n)$ is a valid distribution.

Proof of $\Longrightarrow$ Part of Equivalence Theorem

Lemma ([Kumar et al., 2014], [Vellambi and Kliewer, 2016])

$\exists$ *rate-R ∞-Rényi CI code* $\Longrightarrow$ $\exists$ *rate-R Exact CI code*

- $\exists$ rate-R ∞-Rényi CI code

$$D_\infty(P_{X^nY^n} \| \pi_{XY}^n) < \epsilon \quad \Longrightarrow \quad P_{X^nY^n}(x^n, y^n) < 2^\epsilon \pi_{XY}^n(x^n, y^n)$$

- Define

$$\widehat{P}_{X^nY^n}(x^n, y^n) := \frac{2^\epsilon \pi_{XY}^n(x^n, y^n) - P_{X^nY^n}(x^n, y^n)}{2^\epsilon - 1},$$

then obviously, $\widehat{P}_{X^nY^n}(x^n, y^n)$ is a valid distribution.

- Hence π_{XY}^n can be written as a mixture distribution

$$\pi_{XY}^n(x^n, y^n) = 2^{-\epsilon} P_{X^nY^n}(x^n, y^n) + \left(1 - 2^{-\epsilon}\right) \widehat{P}_{X^nY^n}(x^n, y^n)$$

Proof of $\Longrightarrow$ Part of Equivalence Theorem

$$\pi_{XY}^{n}(x^n, y^n) = 2^{-\epsilon} P_{X^nY^n}(x^n, y^n) + \left(1 - 2^{-\epsilon}\right) \widehat{P}_{X^nY^n}(x^n, y^n)$$

- A time-sharing variable-length scheme:
  - The encoder first generates $U \sim \mathrm{Bern}(2^{-\epsilon})$, and transmits it to two generators using 1 bit
  - If $U = 1$, then the encoder and two generators use the rate-R ∞-Rényi CI code to generate $P_{X^nY^n}$
  - If $U = 0$, then the encoder generates $(X^n, Y^n) \sim \widehat{P}_{X^nY^n}$, and compresses it with rate $\log(|\mathcal{X}||\mathcal{Y}|)$ to generate $\widehat{P}_{X^nY^n}$

Proof of $\Longrightarrow$ Part of Equivalence Theorem

$$\pi_{XY}^n(x^n, y^n) = 2^{-\epsilon} P_{X^nY^n}(x^n, y^n) + \left(1 - 2^{-\epsilon}\right) \widehat{P}_{X^nY^n}(x^n, y^n)$$

- A time-sharing variable-length scheme:
  - The encoder first generates $U \sim \mathrm{Bern}(2^{-\epsilon})$, and transmits it to two generators using 1 bit
  - If $U = 1$, then the encoder and two generators use the rate-R ∞-Rényi CI code to generate $P_{X^nY^n}$
  - If $U = 0$, then the encoder generates $(X^n, Y^n) \sim \widehat{P}_{X^nY^n}$, and compresses it with rate $\log(|\mathcal{X}||\mathcal{Y}|)$ to generate $\widehat{P}_{X^nY^n}$
- The induced distribution is π_{XY}^n exactly

Proof of $\Longrightarrow$ Part of Equivalence Theorem

$$\pi_{XY}^n(x^n, y^n) = 2^{-\epsilon} P_{X^nY^n}(x^n, y^n) + \left(1 - 2^{-\epsilon}\right) \widehat{P}_{X^nY^n}(x^n, y^n)$$

- A time-sharing variable-length scheme:
  - The encoder first generates $U \sim \mathrm{Bern}(2^{-\epsilon})$, and transmits it to two generators using 1 bit
  - If $U = 1$, then the encoder and two generators use the rate-R ∞-Rényi CI code to generate $P_{X^nY^n}$
  - If $U = 0$, then the encoder generates $(X^n, Y^n) \sim \widehat{P}_{X^nY^n}$, and compresses it with rate $\log(|\mathcal{X}||\mathcal{Y}|)$ to generate $\widehat{P}_{X^nY^n}$
- The induced distribution is π_{XY}^n exactly
- The total code rate

$$\leq \frac{1}{n} + 2^{-\epsilon} R + \left(1 - 2^{-\epsilon}\right) \log(|\mathcal{X}||\mathcal{Y}|) \to R$$

as $n \to \infty, \epsilon \to 0$

Proof of $\Longleftarrow$ Part of Equivalence Theorem

Lemma

$\exists$ *rate-*R ∞*-Rényi CI code* $\Longleftarrow$ $\exists$ *rate-*R *Exact CI code*

Proof of $\Longleftarrow$ Part of Equivalence Theorem

Lemma

$\exists$ rate-R ∞-Rényi CI code $\Longleftarrow$ $\exists$ rate-R Exact CI code

- Let $\{(P_{W_k}, P_{X^k|W_k}, P_{Y^k|W_k})\}_{k\in\mathbb{N}}$ be rate-R exact CI codes such that

$$\lim_{k\to\infty} \frac{1}{k} H(P_{W_k}) = R$$

but W_k is not uniform. 

Proof of $\Longleftarrow$ Part of Equivalence Theorem

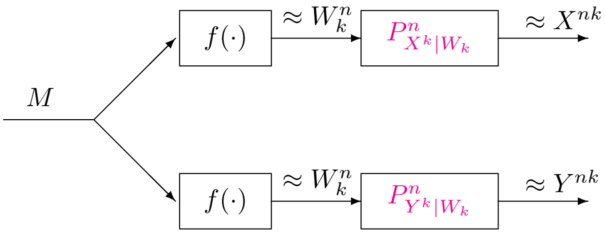
Lemma

$\exists$ rate-R ∞-Rényi CI code $\Longleftarrow$ *$\exists$ rate-R Exact CI code*

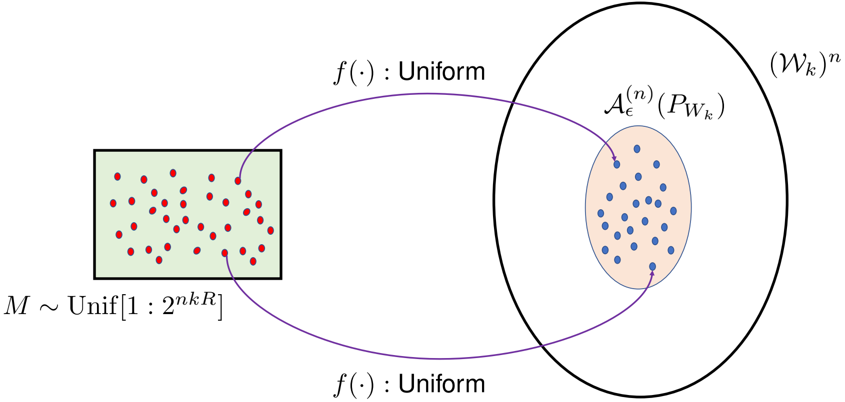
- Let $\{(P_{W_k}, P_{X^k|W_k}, P_{Y^k|W_k})\}_{k\in\mathbb{N}}$ be rate-R exact CI codes such that

$$\lim_{k\to\infty} \frac{1}{k} H(P_{W_k}) = R$$

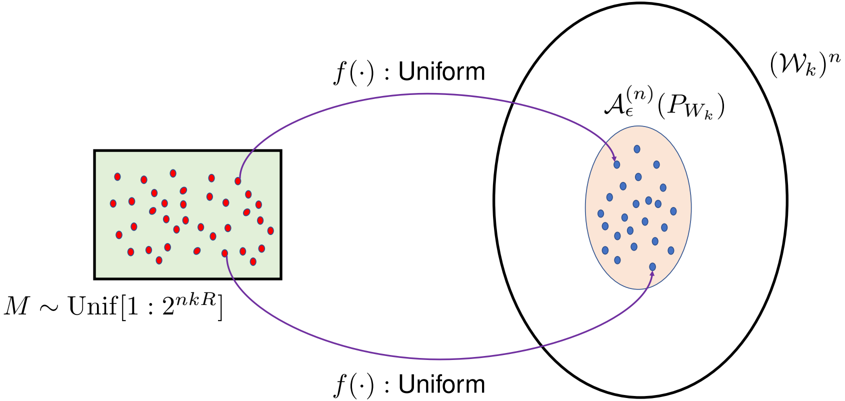
  but W_k is not uniform.
- Simulate W_k^n using two Rényi source resolvability codes!



Proof of $\Longleftarrow$ Part of Equivalence Theorem



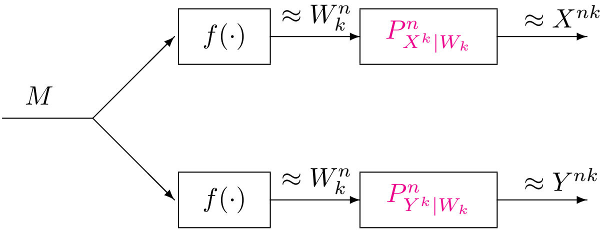
Proof of $\Longleftarrow$ Part of Equivalence Theorem



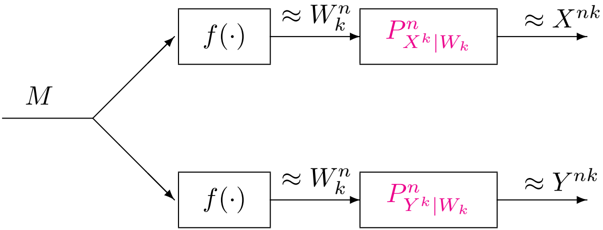
Succeed in the sense of $D_\infty(P_{f(M)} \| P_{W_k}^n) \to 0$ if [Yu and Tan, 2019d]

$$R > \frac{1}{k} H(P_{W_k})$$

Proof of $\Longleftarrow$ Part of Equivalence Theorem



Proof of $\Longleftarrow$ Part of Equivalence Theorem

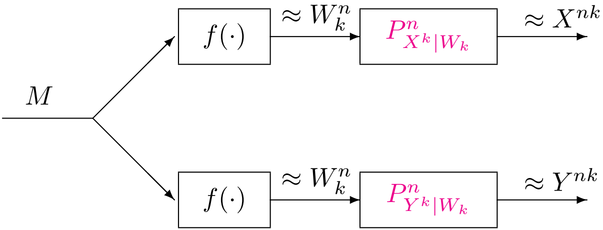


- For the given stochastic kernel (channel) $P^n_{X^k|W_k} P^n_{Y^k|W_k}$,

$$P_W^n \longrightarrow P^n_{X^k|W_k} P^n_{Y^k|W_k} \longrightarrow \pi^{kn}_{XY}$$

$$P_{f(M)} \longrightarrow P^n_{X^k|W_k} P^n_{Y^k|W_k} \longrightarrow P_{X^{kn}Y^{kn}}$$

Proof of $\Longleftarrow$ Part of Equivalence Theorem



- For the given stochastic kernel (channel) $P^n_{X^k|W_k} P^n_{Y^k|W_k}$,

$$P_W^n \longrightarrow P^n_{X^k|W_k} P^n_{Y^k|W_k} \longrightarrow \pi^{kn}_{XY}$$

$$P_{f(M)} \longrightarrow P^n_{X^k|W_k} P^n_{Y^k|W_k} \longrightarrow P_{X^{kn} Y^{kn}}$$

- By the data processing inequality (DPI) for Rényi divergence,

$$D_\infty(P_{X^{kn} Y^{kn}} \| \pi^{kn}_{XY}) \le D_\infty(P_{f(M)} \| P^n_{W_k}) \overset{n\to\infty}{\longrightarrow} 0$$

Combining with Single-Letter Bounds from Rényi CI

Theorem ([Yu and Tan, 2020c])

For $(X, Y) \sim \pi_{XY}$ *on a finite alphabet,*

$$\underline{\Gamma}_\infty(\pi_{XY}) \le T_{\mathrm{Ex}}(\pi_{XY}) = \tilde{T}_\infty(\pi_{XY}) \le \overline{\Gamma}_\infty(\pi_{XY}).$$

Combining with Single-Letter Bounds from Rényi CI

Theorem ([Yu and Tan, 2020c])

For $(X, Y) \sim \pi_{XY}$ *on a finite alphabet,*

$$\underline{\Gamma}_\infty(\pi_{XY}) \le T_{\mathrm{Ex}}(\pi_{XY}) = \tilde{T}_\infty(\pi_{XY}) \le \overline{\Gamma}_\infty(\pi_{XY}).$$

- Gone from a multi-letter expression by [Kumar et al., 2014]

$$\lim_{n\to\infty} \frac{1}{n} \min_{\substack{P_{W_n} P_{X^n|W_n} P_{Y^n|W_n} : \\ P_{X^n Y^n} = \pi_{XY}^n}} H(W_n)$$

to single-letter bounds.

Combining with Single-Letter Bounds from Rényi CI

Theorem ([Yu and Tan, 2020c])

For $(X,Y) \sim \pi_{XY}$ on a finite alphabet,

$$\underline{\Gamma}_\infty(\pi_{XY}) \le T_{\mathrm{Ex}}(\pi_{XY}) = \tilde{T}_\infty(\pi_{XY}) \le \overline{\Gamma}_\infty(\pi_{XY}).$$

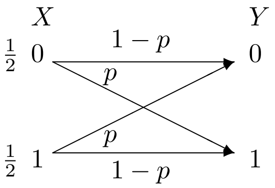
- Gone from a multi-letter expression by [Kumar et al., 2014]

$$\lim_{n\to\infty} \frac{1}{n} \min_{\substack{P_{W_n} P_{X^n|W_n} P_{Y^n|W_n}: \\ P_{X^nY^n}=\pi_{XY}^n}} H(W_n)$$

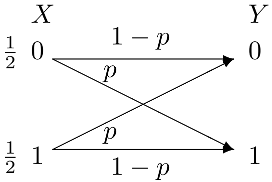
  to single-letter bounds.

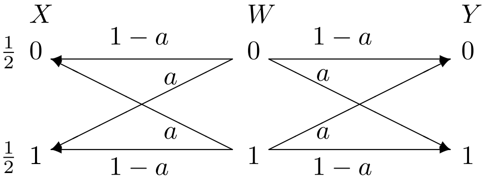
- Presumably the bounds are more amenable to numerical evaluation?

Revisiting the DBSS

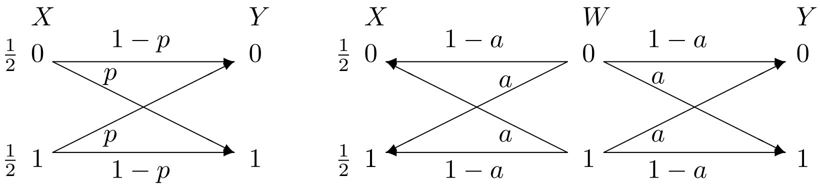


Revisiting the DBSS





Revisiting the DBSS



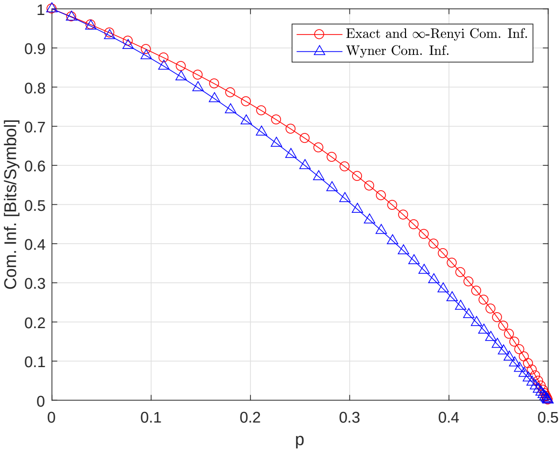
Theorem (Evaluation of Upper and Lower Bounds for DSBS(p))

For a DSBS $(X, Y) \sim \mathrm{DSBS}(p)$ *with crossover probability* $p \in (0, 1/2)$*,*

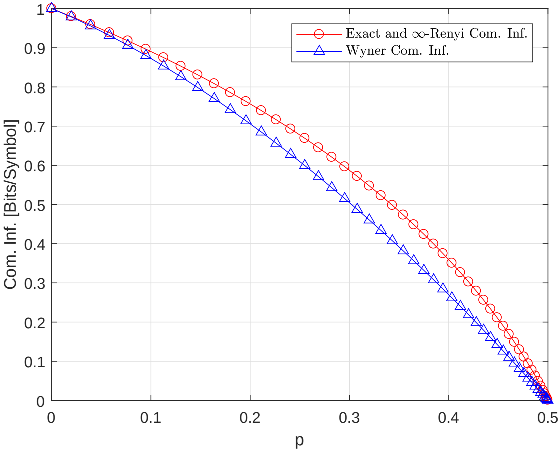
$$\begin{aligned}\tilde{T}_\infty(\pi_{XY}) &= T_{\mathrm{Ex}}(\pi_{XY}) \\ &= -2h(a) - (1-2a)\log\left[\frac{1}{2}\left(a^2 + (1-a)^2\right)\right] - 2a\log\left[a(1-a)\right],\end{aligned}$$

where $a := \frac{1-\sqrt{1-2p}}{2} \in (0, \frac{1}{2})$ *and* $h(a) := -a\log a - (1-a)\log(1-a)$*.*

Numerical Results — DSBS



Numerical Results — DSBS

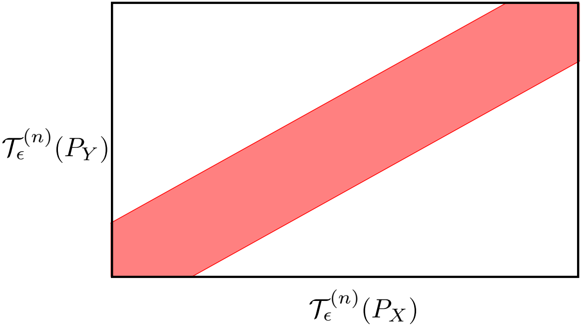


$$T_{\text{Ex}}(\text{DSBS}(p)) > C_{\text{W}}(\text{DSBS}(p)) \quad \forall\, p \in (0, 1/2).$$

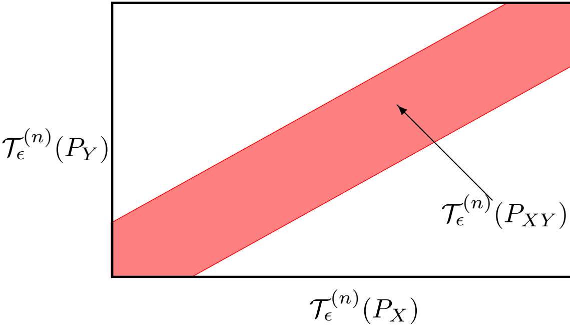
Answers the open question in [Kumar et al., 2014].

Why is Exact CI (or ∞-Rényi CI) $>$ Wyner's CI?

Why is Exact CI (or ∞-Rényi CI) $>$ Wyner's CI?



Why is Exact CI (or ∞-Rényi CI) $>$ Wyner's CI?

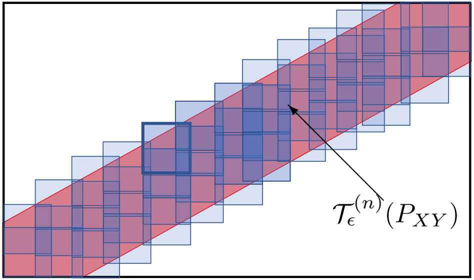


Wyner's common information requires

$$\frac{P_{X^nY^n}(x^n, y^n)}{\pi_{XY}^n(x^n, y^n)} = 1 + o(1) \quad \text{for almost all} \quad (x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}(\pi_{XY})$$

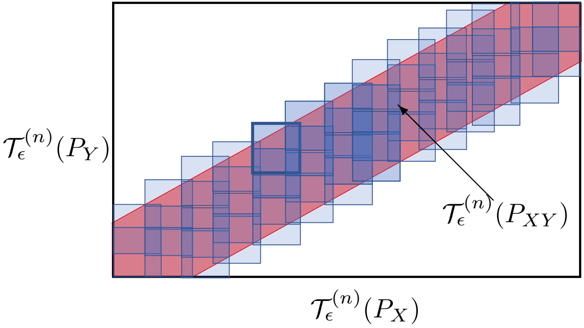
Why is Exact CI (or ∞-Rényi CI) $>$ Wyner's CI?

$\mathcal{T}_\epsilon^{(n)}(P_Y)$



$\mathcal{T}_\epsilon^{(n)}(P_X)$

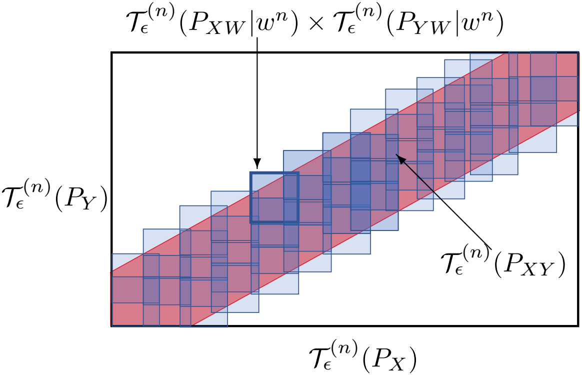
Why is Exact CI (or ∞-Rényi CI) $>$ Wyner's CI?



Rényi CI of order ∞ or Exact CI requires

$$\frac{P_{X^nY^n}(x^n,y^n)}{\pi^n_{XY}(x^n,y^n)} = 1 + o(1) \quad \text{for all} \quad (x^n,y^n) \in \mathrm{supp}(P_{X^nY^n})$$

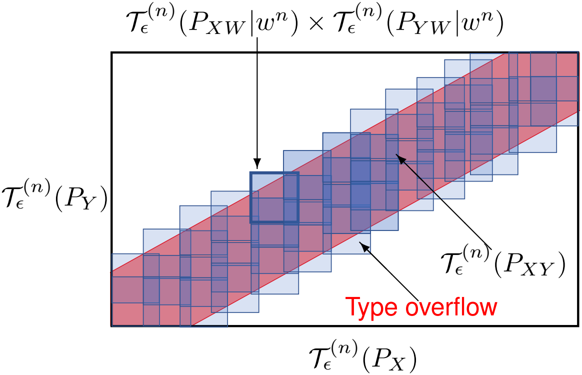
Why is Exact CI (or ∞-Rényi CI) $>$ Wyner's CI?



Rényi CI of order ∞ or Exact CI requires

$$\frac{P_{X^nY^n}(x^n,y^n)}{\pi_{XY}^n(x^n,y^n)} = 1+o(1) \quad \text{for all} \quad (x^n,y^n) \in \bigcup_w \mathrm{supp}\Big(P_{X^n|W_n}(\cdot|w)P_{Y^n|W_n}(\cdot|w)\Big)$$

Why is Exact CI (or ∞-Rényi CI) $>$ Wyner's CI?

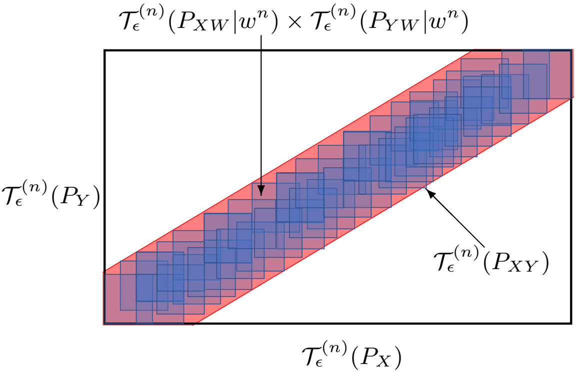


Rényi CI of order ∞ or Exact CI requires

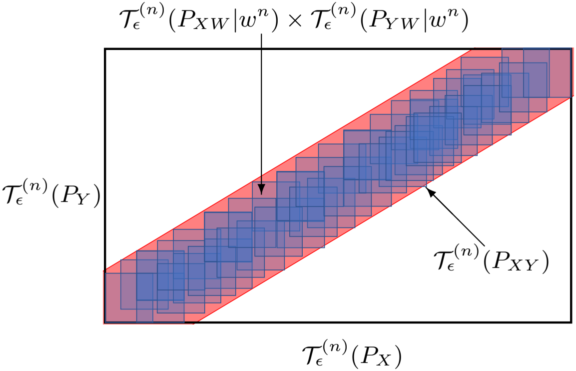
$$\frac{P_{X^nY^n}(x^n,y^n)}{\pi_{XY}^n(x^n,y^n)} = 1+o(1) \quad \text{for all} \quad (x^n,y^n) \in \bigcup_w \mathrm{supp}\Big(P_{X^n|W_n}(\cdot|w)P_{Y^n|W_n}(\cdot|w)\Big)$$

When is Exact CI (or ∞-Rényi CI) $=$ Wyner's CI?

When is Exact CI (or ∞-Rényi CI) = Wyner's CI?

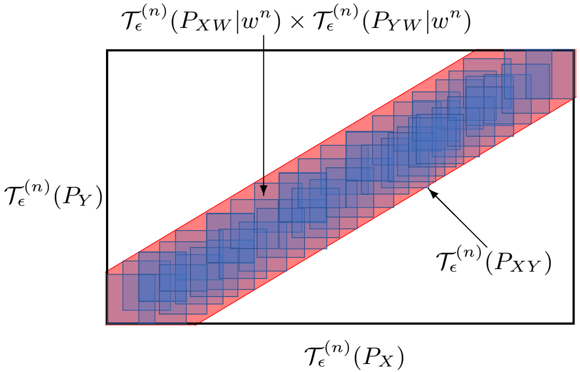


When is Exact CI (or ∞-Rényi CI) $=$ Wyner's CI?



Sufficient Condition [Vellambi and Kliewer, 2016]

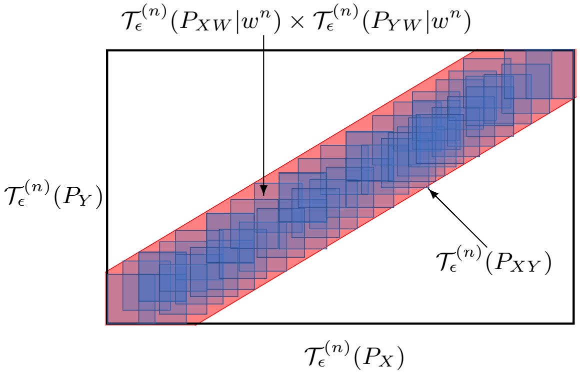
When is Exact CI (or ∞-Rényi CI) $=$ Wyner's CI?



Sufficient Condition [Vellambi and Kliewer, 2016]

$$H(X|W=w)H(Y|W=w)=0 \quad \text{for each} \quad w$$

When is Exact CI (or ∞-Rényi CI) $=$ Wyner's CI?

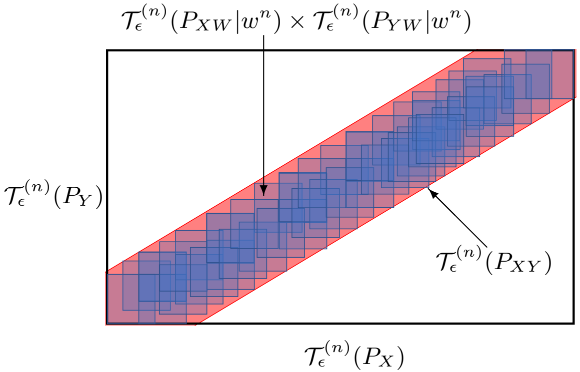


Sufficient Condition [Vellambi and Kliewer, 2016]

$$H(X|W=w)H(Y|W=w)=0 \quad \text{for each} \quad w$$

$$\iff \mathcal{C}(P_{X|W}, P_{Y|W}) = \{P_{X|W}P_{Y|W}\}$$

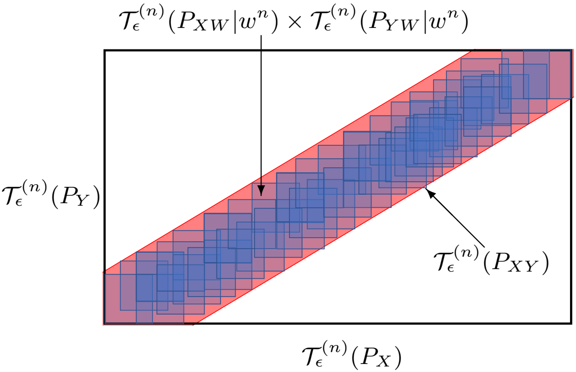
When is Exact CI (or ∞-Rényi CI) $=$ Wyner's CI?



Sufficient Condition [Vellambi and Kliewer, 2016]

$$H(X|W=w)H(Y|W=w)=0 \quad \text{for each} \quad w$$
$$\iff \mathcal{C}(P_{X|W},P_{Y|W}) = \{P_{X|W}P_{Y|W}\}$$
$$\iff \mathcal{T}_\epsilon^{(n)}(P_{XY}) \approx \bigcup_{w^n\in\mathcal{C}} \Big(\mathcal{T}_\epsilon^{(n)}(P_{XW}|w^n) \times \mathcal{T}_\epsilon^{(n)}(P_{YW}|w^n)\Big)$$

When is Exact CI (or ∞-Rényi CI) $=$ Wyner's CI?



Sufficient Condition [Vellambi and Kliewer, 2016]

$$H(X|W=w)H(Y|W=w)=0 \quad \text{for each} \quad w$$
$$\Longleftrightarrow \mathcal{C}(P_{X|W},P_{Y|W})=\{P_{X|W}P_{Y|W}\}$$
$$\Longleftrightarrow \mathcal{T}_\epsilon^{(n)}(P_{XY}) \approx \bigcup_{w^n\in\mathcal{C}} \Big(\mathcal{T}_\epsilon^{(n)}(P_{XW}|w^n) \times \mathcal{T}_\epsilon^{(n)}(P_{YW}|w^n)\Big)$$
$$\Longleftrightarrow \mathcal{T}_\epsilon^{(n)}(P_{XY}) \approx \mathrm{supp}(P_{X^nY^n}) \qquad \text{(No type overflow)}$$

When Exact CI (or ∞-Rényi CI) $=$ Wyner's CI

Example for Sufficient Condition:

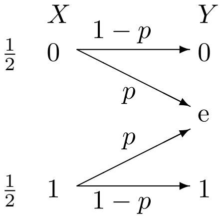
$$H(X|W = w)H(Y|W = w) = 0 \qquad \forall\, w$$

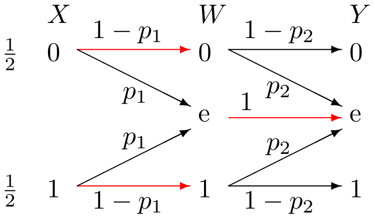
When Exact CI (or ∞-Rényi CI) $=$ Wyner's CI

Example for Sufficient Condition:

$$H(X|W=w)H(Y|W=w)=0 \qquad \forall\, w$$

- Symmetric Binary Erasure Source (SBES)



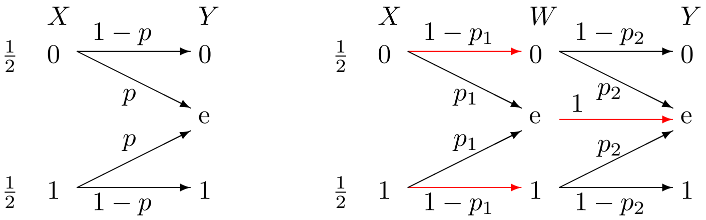


When Exact CI (or ∞-Rényi CI) $=$ Wyner's CI

Example for Sufficient Condition:

$$H(X|W=w)H(Y|W=w)=0 \qquad \forall w$$

- Symmetric Binary Erasure Source (SBES)



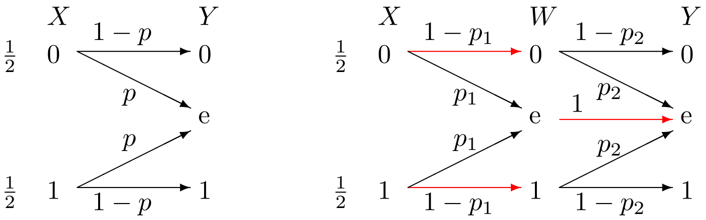
- $(1-p_1)(1-p_2)=1-p$.

When Exact CI (or ∞-Rényi CI) = Wyner's CI

Example for Sufficient Condition:

$$H(X|W=w)H(Y|W=w)=0 \qquad \forall\, w$$

- Symmetric Binary Erasure Source (SBES)



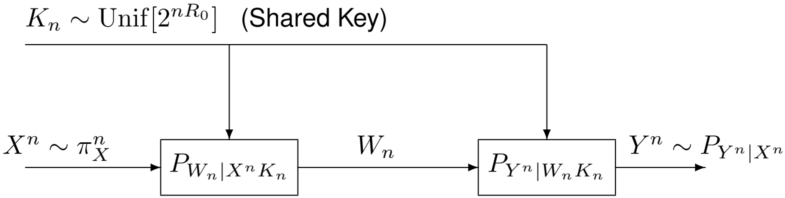
- $(1-p_1)(1-p_2)=1-p$.
- The Exact CI is equal to Wyner's CI and

$$\tilde{T}_\infty(\pi_{XY}) = T_{\text{Exact}}(\pi_{XY}) = C_{\text{Wyner}}(\pi_{XY}) = \begin{cases} 1 & p \le \frac{1}{2} \\ H(p) & p > \frac{1}{2} \end{cases}.$$

Outline

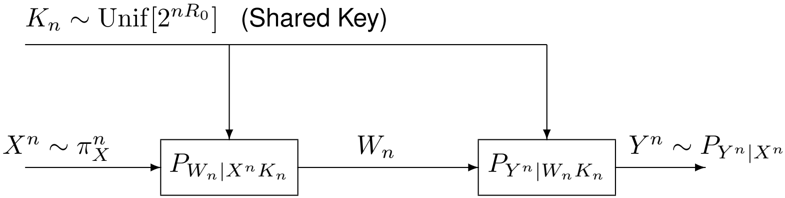
Channel Synthesis

- Given $\pi_{XY} = \pi_X \pi_{Y|X}$ consider the following task:



Channel Synthesis

- Given $\pi_{XY} = \pi_X \pi_{Y|X}$ consider the following task:

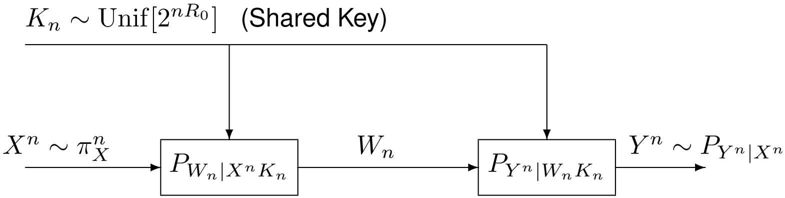


- Goal: Ensure that

$$P_{X^nY^n} \approx \pi_{XY}^n \text{ (Approximate)} \quad \text{or} \quad P_{X^nY^n} = \pi_{XY}^n \text{ (Exact)}.$$

Channel Synthesis

- Given $\pi_{XY} = \pi_X \pi_{Y|X}$ consider the following task:



- Goal: Ensure that

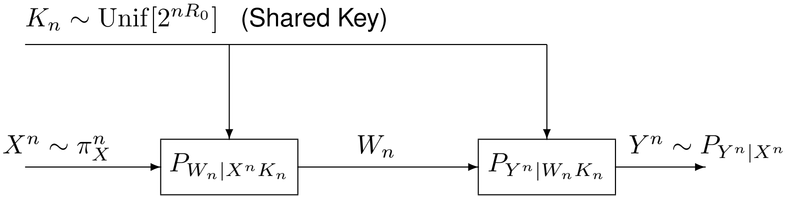
$$P_{X^nY^n} \approx \pi_{XY}^n \text{ (Approximate)} \quad \text{or} \quad P_{X^nY^n} = \pi_{XY}^n \text{ (Exact).}$$

- Equivalently,

$$P_{Y^n|X^n} \approx \pi_{Y|X}^n \text{ (Approximate)} \quad \text{or} \quad P_{Y^n|X^n} = \pi_{Y|X}^n \text{ (Exact).}$$

Channel Synthesis

- Given $\pi_{XY} = \pi_X \pi_{Y|X}$ consider the following task:



- Goal: Ensure that

$$P_{X^nY^n} \approx \pi_{XY}^n \text{ (Approximate)} \quad \text{or} \quad P_{X^nY^n} = \pi_{XY}^n \text{ (Exact).}$$

- Equivalently,

$$P_{Y^n|X^n} \approx \pi_{Y|X}^n \text{ (Approximate)} \quad \text{or} \quad P_{Y^n|X^n} = \pi_{Y|X}^n \text{ (Exact).}$$

- Known as channel synthesis [Cuff, 2012].

Approximate Channel Synthesis

- Consider approximate channel synthesis under TV criterion, i.e.,

$$\lim_{n\to\infty} |P_{X^nY^n} - \pi_{XY}^n| = 0.$$

Approximate Channel Synthesis

- Consider approximate channel synthesis under TV criterion, i.e.,

$$\lim_{n\to\infty} |P_{X^nY^n} - \pi_{XY}^n| = 0.$$

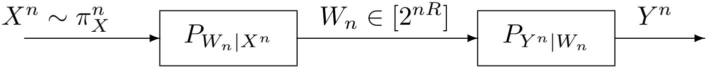
- Let the region of achievable rate pairs (R, R_0) be $\mathcal{R}_{\mathrm{W}}(\pi_{XY})$.

Approximate Channel Synthesis

- Consider approximate channel synthesis under TV criterion, i.e.,

$$\lim_{n\to\infty} |P_{X^nY^n} - \pi_{XY}^n| = 0.$$

- Let the region of achievable rate pairs (R, R_0) be $\mathcal{R}_{\mathrm{W}}(\pi_{XY})$.
- When $R_0 = 0$, problem reduces to approximate distributed source simulation



so the minimum compression rate is Wyner's common information

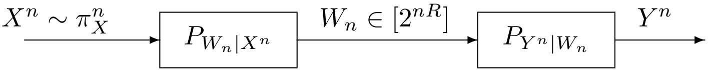
$$R^*(R_0 = 0|\pi_{XY}) = C_{\mathrm{W}}(\pi_{XY})$$

Approximate Channel Synthesis

- Consider approximate channel synthesis under TV criterion, i.e.,

$$\lim_{n\to\infty} |P_{X^nY^n} - \pi^n_{XY}| = 0.$$

- Let the region of achievable rate pairs (R, R_0) be $\mathcal{R}_{\mathrm{W}}(\pi_{XY})$.
- When $R_0 = 0$, problem reduces to approximate distributed source simulation



so the minimum compression rate is Wyner's common information

$$R^*(R_0 = 0|\pi_{XY}) = C_{\mathrm{W}}(\pi_{XY})$$

- When $R_0 = \infty$,

$$R^*(R_0 = \infty|\pi_{XY}) = I_\pi(X;Y)$$

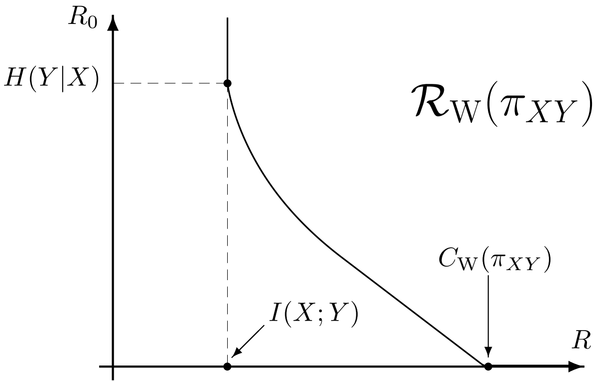
Approximate Channel Synthesis

It was shown in [Cuff, 2012] that

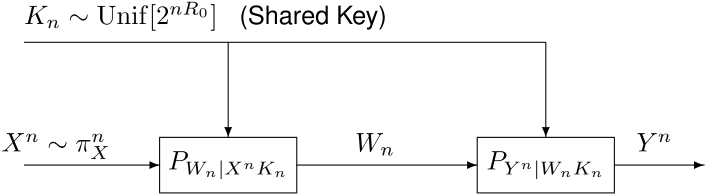
$$\mathcal{R}_{\mathrm{W}}(\pi_{XY}) := \bigcup_{\substack{P_W P_{X|W} P_{Y|W}: \\ P_{XY}=\pi_{XY}}} \left\{ (R, R_0) \, : \, \begin{array}{r} R \geq I(X;W) \\ R + R_0 \geq I(XY;W) \end{array} \right\}.$$

It was shown in [Cuff, 2012] that

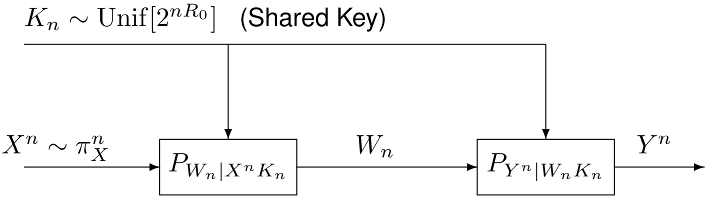
$$\mathcal{R}_{\mathrm{W}}(\pi_{XY}) := \bigcup_{\substack{P_W P_{X|W} P_{Y|W}: \\ P_{XY}=\pi_{XY}}} \left\{ (R, R_0) \ : \begin{array}{c} R \geq I(X;W) \\ R + R_0 \geq I(XY;W) \end{array} \right\}.$$



Exact Channel Synthesis



Exact Channel Synthesis

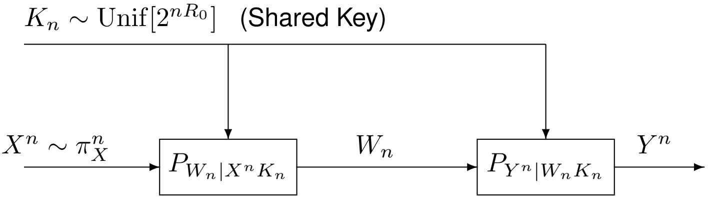


- Now, similarly to exact common information, we demand that

$$P_{X^nY^n} = \pi_{XY}^n \quad \text{for some large enough} \quad n \in \mathbb{N}$$

but just like exact CI, we allow variable-length codes for W_n.

Exact Channel Synthesis

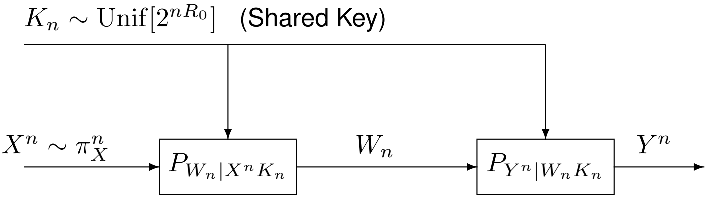


- Now, similarly to exact common information, we demand that

$$P_{X^nY^n} = \pi_{XY}^n \quad \text{for some large enough} \quad n \in \mathbb{N}$$

  but just like exact CI, we allow variable-length codes for W_n.
- If $R_0 = \infty$, [Bennett et al., 2002] showed that the minimum R is $I(X;Y)$.

Exact Channel Synthesis

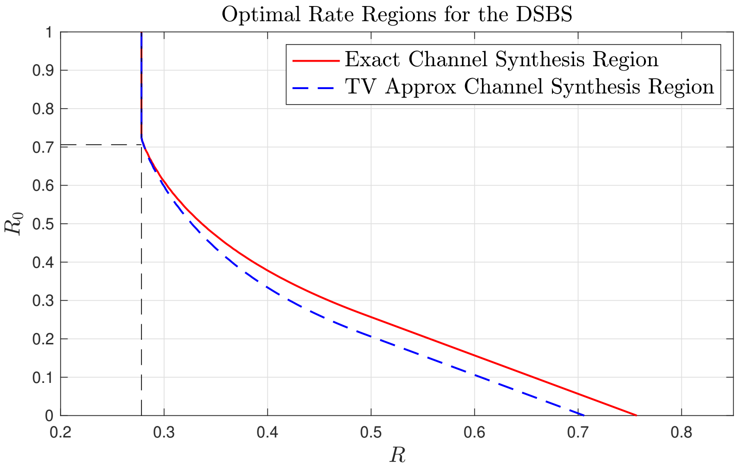


- Now, similarly to exact common information, we demand that

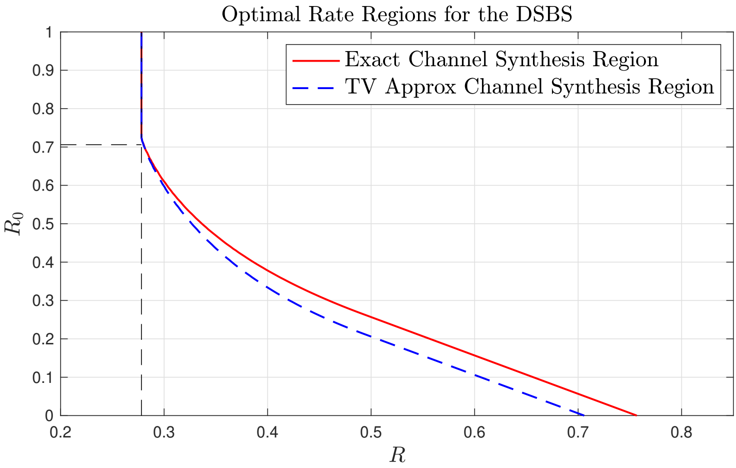
$$P_{X^n Y^n} = \pi_{XY}^n \quad \text{for some large enough} \quad n \in \mathbb{N}$$

  but just like exact CI, we allow variable-length codes for W_n.
- If $R_0 = \infty$, [Bennett et al., 2002] showed that the minimum R is $I(X;Y)$.
- Best tradeoff between R and R_0 in the non-extremal cases considered by [Yu and Tan, 2020b].

Doubly Binary Symmetric Sources

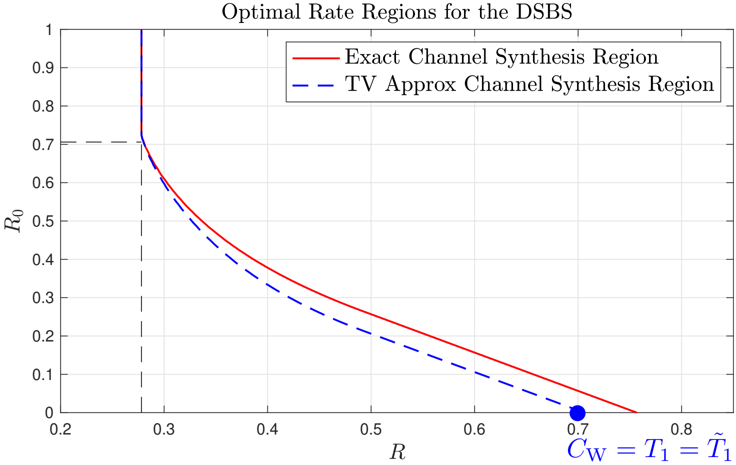


Doubly Binary Symmetric Sources



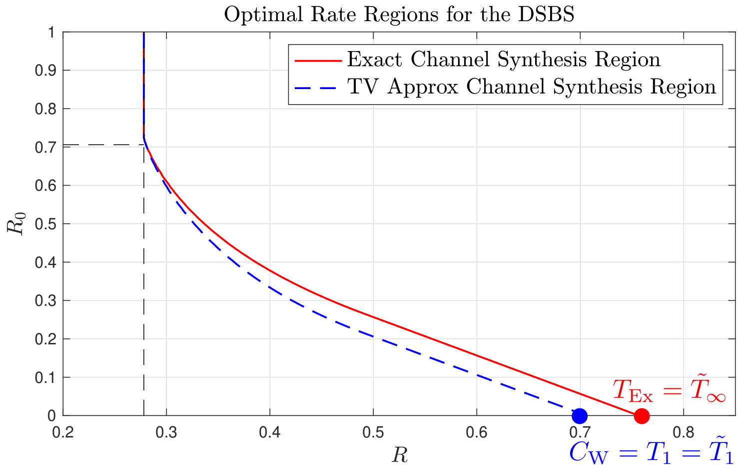
Exact channel synthesis region is strictly smaller than $\mathcal{R}_{\mathrm{W}}(\pi_{XY})$

Doubly Binary Symmetric Sources



Exact channel synthesis region is strictly smaller than $\mathcal{R}_{\mathrm{W}}(\pi_{XY})$

Doubly Binary Symmetric Sources



Exact channel synthesis region is strictly smaller than $\mathcal{R}_{\mathrm{W}}(\pi_{XY})$

Outline

Nonnegative Matrix Factorization

- Given a matrix $\mathbf{M} \in \mathbb{R}_+^{m \times k}$, find $\mathbf{U} \in \mathbb{R}_+^{m \times r}$ and $\mathbf{V} \in \mathbb{R}_+^{r \times k}$ such that

$$\mathbf{M} \approx \mathbf{U}\mathbf{V} \qquad \text{or} \qquad \mathbf{M} = \mathbf{U}\mathbf{V}.$$

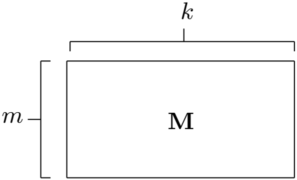
Many applications. See [Cichocki et al., 2009] or [Gillis, 2020].

Nonnegative Matrix Factorization

- Given a matrix $\mathbf{M} \in \mathbb{R}_+^{m \times k}$, find $\mathbf{U} \in \mathbb{R}_+^{m \times r}$ and $\mathbf{V} \in \mathbb{R}_+^{r \times k}$ such that

$$\mathbf{M} \approx \mathbf{UV} \qquad \text{or} \qquad \mathbf{M} = \mathbf{UV}.$$

  Many applications. See [Cichocki et al., 2009] or [Gillis, 2020].
- Dimensionality reduction:

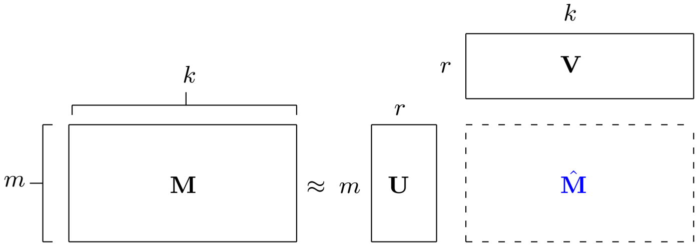


Nonnegative Matrix Factorization

- Given a matrix $\mathbf{M} \in \mathbb{R}_+^{m \times k}$, find $\mathbf{U} \in \mathbb{R}_+^{m \times r}$ and $\mathbf{V} \in \mathbb{R}_+^{r \times k}$ such that

$$\mathbf{M} \approx \mathbf{U}\mathbf{V} \qquad \text{or} \qquad \mathbf{M} = \mathbf{U}\mathbf{V}.$$

  Many applications. See [Cichocki et al., 2009] or [Gillis, 2020].
- Dimensionality reduction:

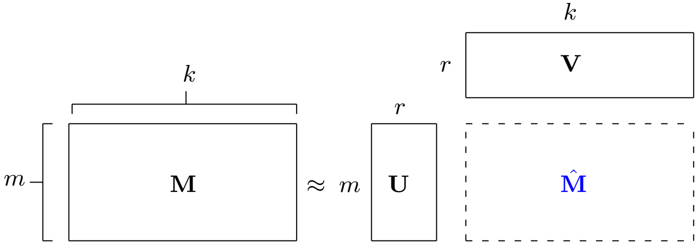


Nonnegative Matrix Factorization

- Given a matrix $\mathbf{M} \in \mathbb{R}_+^{m \times k}$, find $\mathbf{U} \in \mathbb{R}_+^{m \times r}$ and $\mathbf{V} \in \mathbb{R}_+^{r \times k}$ such that

$$\mathbf{M} \approx \mathbf{U}\mathbf{V} \qquad \text{or} \qquad \mathbf{M} = \mathbf{U}\mathbf{V}.$$

  Many applications. See [Cichocki et al., 2009] or [Gillis, 2020].
- Dimensionality reduction:



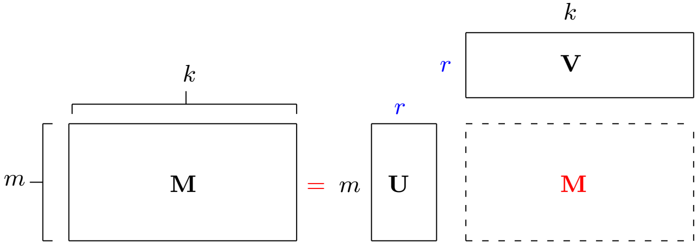
- Only interested in exact factorization.

Nonnegative Matrix Factorization

- Given a matrix $\mathbf{M} \in \mathbb{R}_+^{m \times k}$, find $\mathbf{U} \in \mathbb{R}_+^{m \times r}$ and $\mathbf{V} \in \mathbb{R}_+^{r \times k}$ such that

$$\mathbf{M} \approx \mathbf{UV} \qquad \text{or} \qquad \mathbf{M} = \mathbf{UV}.$$

  Many applications. See [Cichocki et al., 2009] or [Gillis, 2020].
- Dimensionality reduction:



- Only interested in exact factorization.
- What is the minimum r to achieve exact factorization? Is this connected to information theory?

Nonnegative Rank

Definition

The nonnegative rank of $\mathbf{M} \in \mathbb{R}_+^{m\times k}$, denoted as $\mathrm{rank}_+(\mathbf{M})$, is the smallest integer r such that

$$\mathbf{M} = \sum_{w=1}^{r} \mathbf{u}_w \mathbf{v}_w^\top$$

for some nonnegative vectors $\mathbf{u}_w \in \mathbb{R}_+^m$ and $\mathbf{v}_w \in \mathbb{R}_+^k$.

Nonnegative Rank

Definition

The nonnegative rank of $\mathbf{M} \in \mathbb{R}_+^{m \times k}$, denoted as $\mathrm{rank}_+(\mathbf{M})$, is the smallest integer r such that

$$\mathbf{M} = \sum_{w=1}^{r} \mathbf{u}_w \mathbf{v}_w^\top$$

for some nonnegative vectors $\mathbf{u}_w \in \mathbb{R}_+^m$ and $\mathbf{v}_w \in \mathbb{R}_+^k$.

- Obviously, $\mathrm{rank}(\mathbf{M}) \leq \mathrm{rank}_+(\mathbf{M})$

Nonnegative Rank

Definition

The nonnegative rank of $\mathbf{M} \in \mathbb{R}_+^{m\times k}$, denoted as $\mathrm{rank}_+(\mathbf{M})$, is the smallest integer r such that

$$\mathbf{M} = \sum_{w=1}^{r} \mathbf{u}_w \mathbf{v}_w^\top$$

for some nonnegative vectors $\mathbf{u}_w \in \mathbb{R}_+^m$ and $\mathbf{v}_w \in \mathbb{R}_+^k$.

- Obviously, $\mathrm{rank}(\mathbf{M}) \le \mathrm{rank}_+(\mathbf{M})$
- Gap can be large. Fix $\{a_1, \ldots, a_m\} \subset \mathbb{R}$ and consider distance matrix

$$\mathbf{M} = \begin{bmatrix} 0 & (a_1-a_2)^2 & (a_1-a_3)^2 & \ldots & (a_1-a_m)^2 \\ (a_2-a_1)^2 & 0 & (a_2-a_3)^2 & \ldots & (a_2-a_m)^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (a_m-a_1)^2 & (a_m-a_2)^2 & (a_m-a_3)^2 & \ldots & 0 \end{bmatrix}.$$

Nonnegative Rank

Definition

The nonnegative rank of $\mathbf{M} \in \mathbb{R}_+^{m\times k}$, denoted as $\mathrm{rank}_+(\mathbf{M})$, is the smallest integer r such that

$$\mathbf{M} = \sum_{w=1}^{r} \mathbf{u}_w \mathbf{v}_w^\top$$

for some nonnegative vectors $\mathbf{u}_w \in \mathbb{R}_+^m$ and $\mathbf{v}_w \in \mathbb{R}_+^k$.

- Obviously, $\mathrm{rank}(\mathbf{M}) \le \mathrm{rank}_+(\mathbf{M})$
- Gap can be large. Fix $\{a_1, \ldots, a_m\} \subset \mathbb{R}$ and consider distance matrix

$$\mathbf{M} = \begin{bmatrix} a_1^2 & 1 & -2a_1 \\ a_2^2 & 1 & -2a_2 \\ \vdots & \ddots & \vdots \\ a_m^2 & 1 & -2a_m \end{bmatrix} \begin{bmatrix} 1 & 1 & \ldots & 1 \\ a_1^2 & a_2^2 & \ldots & a_m^2 \\ a_1 & a_2 & \ldots & a_m \end{bmatrix}$$

- $\mathrm{rank}(\mathbf{M}) \le 3$. [Beasley and Laffey, 2009] showed $\mathrm{rank}_+(\mathbf{M}) = \Omega(\log m)$.

Wyner's CI for Normalized Nonnegative Matrices

- Let $\mathbf{M} \in \mathbb{R}_+^{r \times k}$ be a nonnegative matrix.

Wyner's CI for Normalized Nonnegative Matrices

- Let $\mathbf{M} \in \mathbb{R}_+^{r\times k}$ be a nonnegative matrix.
- We may define

$$\pi_{XY}(x,y) := \frac{M_{x,y}}{\|\mathbf{M}\|_1} \qquad (x,y) \in [m] \times [k] = \mathcal{X} \times \mathcal{Y}.$$

Wyner's CI for Normalized Nonnegative Matrices

- Let $\mathbf{M} \in \mathbb{R}_+^{r \times k}$ be a nonnegative matrix.
- We may define

$$\pi_{XY}(x, y) := \frac{M_{x,y}}{\|\mathbf{M}\|_1} \qquad (x, y) \in [m] \times [k] = \mathcal{X} \times \mathcal{Y}.$$

- A discrete r.v. $W \in \mathcal{W}$ is a seed for π_{XY}, or equivalently $\mathbf{M}$, if

$$X - W - Y.$$

Wyner's CI for Normalized Nonnegative Matrices

- Let $\mathbf{M} \in \mathbb{R}_+^{r\times k}$ be a nonnegative matrix.
- We may define

$$\pi_{XY}(x,y) := \frac{M_{x,y}}{\|\mathbf{M}\|_1} \qquad (x,y) \in [m]\times[k] = \mathcal{X}\times\mathcal{Y}.$$

- A discrete r.v. $W \in \mathcal{W}$ is a seed for π_{XY}, or equivalently $\mathbf{M}$, if

$$X - W - Y.$$

- Every NMF of

$$\mathbf{M} = \sum_w \mathbf{u}_w \mathbf{v}_w^\top$$

induces a seed W for $\mathbf{M}$.

Wyner's CI for Normalized Nonnegative Matrices

- Let $\mathbf{M} \in \mathbb{R}_+^{r\times k}$ be a nonnegative matrix.
- We may define

$$\pi_{XY}(x,y) := \frac{M_{x,y}}{\|\mathbf{M}\|_1} \qquad (x,y) \in [m] \times [k] = \mathcal{X} \times \mathcal{Y}.$$

- A discrete r.v. $W \in \mathcal{W}$ is a seed for π_{XY}, or equivalently $\mathbf{M}$, if

$$X - W - Y.$$

- Every NMF of

$$\mathbf{M} = \sum_w \mathbf{u}_w \mathbf{v}_w^\top$$

induces a seed W for $\mathbf{M}$.
- Wyner's common information for $\mathbf{M}$ is

$$C_{\mathrm{W}}(\mathbf{M}) := C_{\mathrm{W}}(\pi_{XY}).$$

Playing With Definitions

Theorem ([Jain et al., 2013], [Braun and Pokutta, 2013])

$$C_{\mathrm{W}}(\mathbf{M}) \leq \log \mathrm{rank}_{+}(\mathbf{M}).$$

Playing With Definitions

Theorem ([Jain et al., 2013], [Braun and Pokutta, 2013])

$$C_{\mathrm{W}}(\mathbf{M}) \le \log \operatorname{rank}_+(\mathbf{M}).$$

Proof.

Let $\mathbf{M}$ have an optimal NMF $\mathbf{M} = \sum_w \mathbf{u}_w \mathbf{v}_w^\top$. Define seed W as

$$P_{W|XY}(w|x,y) = \begin{cases} \dfrac{[\mathbf{u}_w]_x [\mathbf{v}_w]_y}{M_{x,y}} & M_{x,y} > 0 \\ \text{arbitrary} & M_{x,y} = 0 \end{cases}.$$

Playing With Definitions

Theorem ([Jain et al., 2013], [Braun and Pokutta, 2013])

$$C_{\mathrm{W}}(\mathbf{M}) \le \log \mathrm{rank}_+(\mathbf{M}).$$

Proof.

Let $\mathbf{M}$ have an optimal NMF $\mathbf{M} = \sum_w \mathbf{u}_w \mathbf{v}_w^\top$. Define seed W as

$$P_{W|XY}(w|x,y) = \begin{cases} \dfrac{[\mathbf{u}_w]_x [\mathbf{v}_w]_y}{M_{x,y}} & M_{x,y} > 0 \\ \text{arbitrary} & M_{x,y} = 0 \end{cases}.$$

By Bayes rule,

$$P_{XY|W}(x,y|w) = \frac{[\mathbf{u}_w]_x [\mathbf{v}_w]_y}{\sum_{x',y'} [\mathbf{u}_w]_{x'} [\mathbf{v}_w]_{y'}} \qquad (x,y) \in \mathcal{X} \times \mathcal{Y}.$$

Playing With Definitions

Theorem ([Jain et al., 2013], [Braun and Pokutta, 2013])

$$C_{\mathrm{W}}(\mathbf{M}) \le \log \mathrm{rank}_+(\mathbf{M}).$$

Proof.

Let $\mathbf{M}$ have an optimal NMF $\mathbf{M} = \sum_w \mathbf{u}_w \mathbf{v}_w^\top$. Define seed W as

$$P_{W|XY}(w|x,y) = \begin{cases} \dfrac{[\mathbf{u}_w]_x [\mathbf{v}_w]_y}{M_{x,y}} & M_{x,y} > 0 \\ \text{arbitrary} & M_{x,y} = 0 \end{cases}.$$

By Bayes rule,

$$P_{XY|W}(x,y|w) = \frac{[\mathbf{u}_w]_x [\mathbf{v}_w]_y}{\sum_{x',y'} [\mathbf{u}_w]_{x'} [\mathbf{v}_w]_{y'}} \qquad (x,y) \in \mathcal{X} \times \mathcal{Y}.$$

So, $X - W - Y$ and

$$C_{\mathrm{W}}(\mathbf{M}) \le I_P(XY;W) \le H(W) \le \log |\mathcal{W}| = \log \mathrm{rank}_+(\mathbf{M}).$$

□

Gap Between $C_{\mathrm{W}}(\mathbf{M})$ and $\log \mathrm{rank}_{+}(\mathbf{M})$?

- Consider the diagonal matrix

$$\mathbf{M} = \frac{1}{\sum_{j=1}^{m} 2^j} \begin{bmatrix} 2^1 & 0 & 0 & \ldots & 0 \\ 0 & 2^2 & 0 & \ldots & 0 \\ 0 & 0 & 2^3 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 2^m \end{bmatrix}.$$

Gap Between $C_{\mathrm{W}}(\mathbf{M})$ and $\log \mathrm{rank}_+(\mathbf{M})$?

- Consider the diagonal matrix

$$\mathbf{M} = \frac{1}{\sum_{j=1}^m 2^j} \begin{bmatrix} 2^1 & 0 & 0 & \dots & 0 \\ 0 & 2^2 & 0 & \dots & 0 \\ 0 & 0 & 2^3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 2^m \end{bmatrix}.$$

- $\mathrm{rank}_+(\mathbf{M}) = m$

Gap Between $C_{\mathrm{W}}(\mathbf{M})$ and $\log \operatorname{rank}_+(\mathbf{M})$?

- Consider the diagonal matrix

$$\mathbf{M} = \frac{1}{\sum_{j=1}^m 2^j} \begin{bmatrix} 2^1 & 0 & 0 & \dots & 0 \\ 0 & 2^2 & 0 & \dots & 0 \\ 0 & 0 & 2^3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 2^m \end{bmatrix}.$$

- $\operatorname{rank}_+(\mathbf{M}) = m$
- But

$$\begin{aligned} C_{\mathrm{W}}(\mathbf{M}) &\le H_\pi(XY) = H(\pi_X) \\ &= H\left(\frac{2}{\sum_{j\in[m]} 2^j}, \frac{2^2}{\sum_{j\in[m]} 2^j}, \dots, \frac{2^m}{\sum_{j\in[m]} 2^j} \right) \\ &= -\sum_{i\in[m]} \frac{2^i}{\sum_{j\in[m]} 2^j} \log\left(\frac{2^i}{\sum_{j\in[m]} 2^j} \right) \le 2 \quad \forall\, m \in \mathbb{N}. \end{aligned}$$

Gap Between $C_{\mathrm{W}}(\mathbf{M})$ and $\log \mathrm{rank}_+(\mathbf{M})$?

- Consider the diagonal matrix

$$\mathbf{M} = \frac{1}{\sum_{j=1}^m 2^j} \begin{bmatrix} 2^1 & 0 & 0 & \dots & 0 \\ 0 & 2^2 & 0 & \dots & 0 \\ 0 & 0 & 2^3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 2^m \end{bmatrix}.$$

- $\mathrm{rank}_+(\mathbf{M}) = m$
- But

$$\begin{aligned} C_{\mathrm{W}}(\mathbf{M}) &\le H_\pi(XY) = H(\pi_X) \\ &= H\left(\frac{2}{\sum_{j\in[m]} 2^j}, \frac{2^2}{\sum_{j\in[m]} 2^j}, \dots, \frac{2^m}{\sum_{j\in[m]} 2^j} \right) \\ &= -\sum_{i\in[m]} \frac{2^i}{\sum_{j\in[m]} 2^j} \log\left(\frac{2^i}{\sum_{j\in[m]} 2^j} \right) \le 2 \quad \forall\, m \in \mathbb{N}. \end{aligned}$$

- Gap can be arbitrarily large.

Gap Between $C_{\mathrm{W}}(\mathbf{M})$ and $\log \mathrm{rank}_+(\mathbf{M})$?

- Consider the diagonal matrix

$$\mathbf{M} = \frac{1}{\sum_{j=1}^m 2^j} \begin{bmatrix} 2^1 & 0 & 0 & \ldots & 0 \\ 0 & 2^2 & 0 & \ldots & 0 \\ 0 & 0 & 2^3 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 2^m \end{bmatrix}.$$

- $\mathrm{rank}_+(\mathbf{M}) = m$
- But

$$\begin{aligned} C_{\mathrm{W}}(\mathbf{M}) &\le H_\pi(XY) = H(\pi_X) \\ &= H\left(\frac{2}{\sum_{j\in[m]} 2^j}, \frac{2^2}{\sum_{j\in[m]} 2^j}, \ldots, \frac{2^m}{\sum_{j\in[m]} 2^j}\right) \\ &= -\sum_{i\in[m]} \frac{2^i}{\sum_{j\in[m]} 2^j} \log\left(\frac{2^i}{\sum_{j\in[m]} 2^j}\right) \le 2 \quad \forall\, m \in \mathbb{N}. \end{aligned}$$

- Gap can be arbitrarily large.
- Is the relation between $C_{\mathrm{W}}(\mathbf{M})$ and $\log \mathrm{rank}_+(\mathbf{M})$ fundamental?

Amortization Comes to the Rescue

Theorem ([Braun et al., 2017])

Let $\mathbf{M} \in \mathbb{R}_+^{m \times k}$ *be such that* $\|\mathbf{M}\|_1 = \sum_{x,y} M_{x,y} = 1$. *For any* $\epsilon, \delta > 0$, *if* $n \geq n_0(\epsilon, \delta, m, k, C_{\mathrm{W}}(\mathbf{M}))$ *is sufficiently large, there exists* $\mathbf{M}_{\epsilon,\delta,n} \in \mathbb{R}_+^{m^n \times k^n}$ *with*

$$\left\| \mathbf{M}^{\otimes n} - \mathbf{M}_{\epsilon,\delta,n} \right\|_1 \leq \delta.$$

and

$$\lim_{\epsilon \downarrow 0} \lim_{n \to \infty} \frac{1}{n} \log \operatorname{rank}_+(\mathbf{M}_{\epsilon,\delta,n}) = C_{\mathrm{W}}(\mathbf{M}).$$

Amortization Comes to the Rescue

Theorem ([Braun et al., 2017])

Let $\mathbf{M} \in \mathbb{R}_+^{m\times k}$ *be such that* $\|\mathbf{M}\|_1 = \sum_{x,y} M_{x,y} = 1$. *For any* $\epsilon, \delta > 0$, *if* $n \geq n_0(\epsilon, \delta, m, k, C_{\mathrm{W}}(\mathbf{M}))$ *is sufficiently large, there exists* $\mathbf{M}_{\epsilon,\delta,n} \in \mathbb{R}_+^{m^n \times k^n}$ *with*

$$\left\| \mathbf{M}^{\otimes n} - \mathbf{M}_{\epsilon,\delta,n} \right\|_1 \leq \delta.$$

and

$$\lim_{\epsilon \downarrow 0} \lim_{n\to\infty} \frac{1}{n} \log \mathrm{rank}_+(\mathbf{M}_{\epsilon,\delta,n}) = C_{\mathrm{W}}(\mathbf{M}).$$

- Normalized logarithm of the nonnegative rank of an ℓ_1-perturbed version of $\mathbf{M}^{\otimes n}$ for large enough n.

Amortization Comes to the Rescue

Theorem ([Braun et al., 2017])

Let $\mathbf{M} \in \mathbb{R}_+^{m\times k}$ *be such that* $\|\mathbf{M}\|_1 = \sum_{x,y} M_{x,y} = 1$. *For any* $\epsilon, \delta > 0$, *if* $n \geq n_0(\epsilon, \delta, m, k, C_{\mathrm{W}}(\mathbf{M}))$ *is sufficiently large, there exists* $\mathbf{M}_{\epsilon,\delta,n} \in \mathbb{R}_+^{m^n \times k^n}$ *with*

$$\left\|\mathbf{M}^{\otimes n} - \mathbf{M}_{\epsilon,\delta,n}\right\|_1 \leq \delta.$$

and

$$\lim_{\epsilon\downarrow 0} \lim_{n\to\infty} \frac{1}{n} \log \mathrm{rank}_+(\mathbf{M}_{\epsilon,\delta,n}) = C_{\mathrm{W}}(\mathbf{M}).$$

- Normalized logarithm of the nonnegative rank of an ℓ_1-perturbed version of $\mathbf{M}^{\otimes n}$ for large enough n.
- TV common information $=$ Wyner's common information [Cuff, 2012].

Amortization Comes to the Rescue

Theorem ([Braun et al., 2017])

Let $\mathbf{M} \in \mathbb{R}_+^{m \times k}$ *be such that* $\|\mathbf{M}\|_1 = \sum_{x,y} M_{x,y} = 1$. *For any* $\epsilon, \delta > 0$, *if* $n \geq n_0(\epsilon, \delta, m, k, C_{\mathrm{W}}(\mathbf{M}))$ *is sufficiently large, there exists* $\mathbf{M}_{\epsilon,\delta,n} \in \mathbb{R}_+^{m^n \times k^n}$ *with*

$$\left\| \mathbf{M}^{\otimes n} - \mathbf{M}_{\epsilon,\delta,n} \right\|_1 \leq \delta.$$

and

$$\lim_{\epsilon \downarrow 0} \lim_{n \to \infty} \frac{1}{n} \log \mathrm{rank}_+(\mathbf{M}_{\epsilon,\delta,n}) = C_{\mathrm{W}}(\mathbf{M}).$$

- Normalized logarithm of the nonnegative rank of an ℓ_1-perturbed version of $\mathbf{M}^{\otimes n}$ for large enough n.
- TV common information $=$ Wyner's common information [Cuff, 2012].

Amortization Comes to the Rescue

Theorem ([Braun et al., 2017])

Let $\mathbf{M} \in \mathbb{R}_+^{m \times k}$ *be such that* $\|\mathbf{M}\|_1 = \sum_{x,y} M_{x,y} = 1$. *For any* $\epsilon, \delta > 0$, *if* $n \geq n_0(\epsilon, \delta, m, k, C_{\mathrm{W}}(\mathbf{M}))$ *is sufficiently large, there exists* $\mathbf{M}_{\epsilon,\delta,n} \in \mathbb{R}_+^{m^n \times k^n}$ *with*

$$\left\| \mathbf{M}^{\otimes n} - \mathbf{M}_{\epsilon,\delta,n} \right\|_1 \leq \delta.$$

and

$$\lim_{\epsilon \downarrow 0} \lim_{n \to \infty} \frac{1}{n} \log \mathrm{rank}_+(\mathbf{M}_{\epsilon,\delta,n}) = C_{\mathrm{W}}(\mathbf{M}).$$

- Normalized logarithm of the nonnegative rank of an ℓ_1-perturbed version of $\mathbf{M}^{\otimes n}$ for large enough n.
- TV common information $=$ Wyner's common information [Cuff, 2012].

Amortization Comes to the Rescue

Theorem ([Braun et al., 2017])

Let $\mathbf{M} \in \mathbb{R}_+^{m \times k}$ *be such that* $\|\mathbf{M}\|_1 = \sum_{x,y} M_{x,y} = 1$. *For any* $\epsilon, \delta > 0$, *if* $n \geq n_0(\epsilon, \delta, m, k, C_{\mathrm{W}}(\mathbf{M}))$ *is sufficiently large, there exists* $\mathbf{M}_{\epsilon,\delta,n} \in \mathbb{R}_+^{m^n \times k^n}$ *with*

$$\left\| \mathbf{M}^{\otimes n} - \mathbf{M}_{\epsilon,\delta,n} \right\|_1 \leq \delta.$$

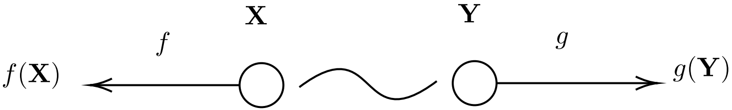
and

$$\lim_{\epsilon \downarrow 0} \lim_{n \to \infty} \frac{1}{n} \log \mathrm{rank}_+(\mathbf{M}_{\epsilon,\delta,n}) = C_{\mathrm{W}}(\mathbf{M}).$$

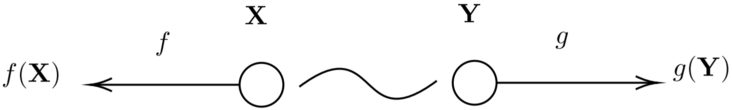
- Normalized logarithm of the nonnegative rank of an ℓ_1-perturbed version of $\mathbf{M}^{\otimes n}$ for large enough n.
- TV common information $=$ Wyner's common information [Cuff, 2012].

Outline

Gács–Körner–Witsenhausen's System

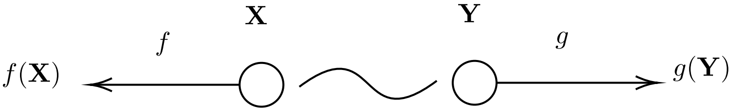


Gács–Körner–Witsenhausen's System



- $(\mathbf{X}, \mathbf{Y}) \sim P_{XY}^n$: a pair of correlated sources

Gács–Körner–Witsenhausen's System



- $(\mathbf{X}, \mathbf{Y}) \sim P_{XY}^n$: a pair of correlated sources
- Define one-sided ϵ-GKW common information:

$$T_X(\epsilon) := \liminf_{n\to\infty} \max_{f,g:\mathbb{P}[f(\mathbf{X})\neq g(\mathbf{Y})]\leq\epsilon} \frac{1}{n} H(f(\mathbf{X}))$$

$$T_Y(\epsilon) := \liminf_{n\to\infty} \max_{f,g:\mathbb{P}[f(\mathbf{X})\neq g(\mathbf{Y})]\leq\epsilon} \frac{1}{n} H(g(\mathbf{Y}))$$

Gács–Körner–Witsenhausen's CI

Problems of Control and Information Theory, Vol 2 (2), pp 119–162 (1973)

COMMON INFORMATION IS FAR LESS THAN MUTUAL INFORMATION

P. GÁCS and J KÖRNER

(Budapest)

(Received February 5, 1972)

Gács–Körner–Witsenhausen's CI

Problems of Control and Information Theory, Vol 2 (2), pp 149–162 (1973)

COMMON INFORMATION IS FAR LESS THAN MUTUAL INFORMATION

P. GÁCS and J KÖRNER
(Budapest)

(Received February 5, 1972)

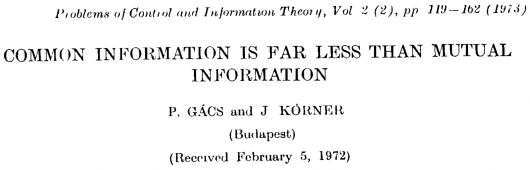
Theorem ([Gács and Körner, 1973])

$$\lim_{\epsilon \downarrow 0} T_X(\epsilon) = \lim_{\epsilon \downarrow 0} T_Y(\epsilon) = C_{\mathrm{GKW}}(X;Y),$$

where

$$C_{\mathrm{GKW}}(X;Y) := \max_{f,g: f(X)=g(Y)} H(f(X))$$

Gács–Körner–Witsenhausen's CI


Problems of Control and Information Theory, Vol 2 (2), pp 149–162 (1973)

COMMON INFORMATION IS FAR LESS THAN MUTUAL INFORMATION

P. GÁCS and J KÖRNER

(Budapest)

(Received February 5, 1972)

Theorem ([Gács and Körner, 1973])

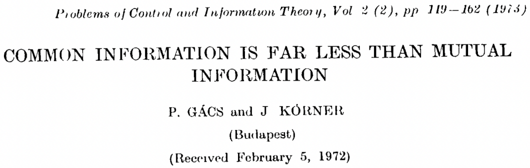
$$\lim_{\epsilon\downarrow 0} T_X(\epsilon) = \lim_{\epsilon\downarrow 0} T_Y(\epsilon) = C_{\mathrm{GKW}}(X;Y),$$

where

$$C_{\mathrm{GKW}}(X;Y) := \max_{f,g: f(X)=g(Y)} H(f(X))$$

- $C_{\mathrm{GKW}}(X;Y)$ called Gács–Körner–Witsenhausen's (GKW's) CI

Gács–Körner–Witsenhausen's CI


Problems of Control and Information Theory, Vol 2 (2), pp 149–162 (1973)

COMMON INFORMATION IS FAR LESS THAN MUTUAL INFORMATION

P. GÁCS and J KÖRNER

(Budapest)

(Received February 5, 1972)

Theorem ([Gács and Körner, 1973])

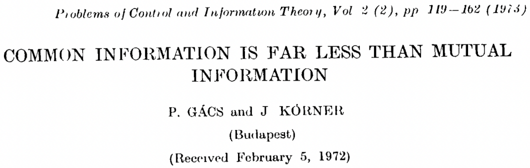
$$\lim_{\epsilon \downarrow 0} T_X(\epsilon) = \lim_{\epsilon \downarrow 0} T_Y(\epsilon) = C_{\mathrm{GKW}}(X;Y),$$

where

$$C_{\mathrm{GKW}}(X;Y) := \max_{f,g:f(X)=g(Y)} H(f(X))$$

- $C_{\mathrm{GKW}}(X;Y)$ called Gács–Körner–Witsenhausen's (GKW's) CI
- Abridged version of GKW's system as in [Csiszár and Narayan, 2000]

Gács–Körner–Witsenhausen's CI


Problems of Control and Information Theory, Vol 2 (2), pp 149–162 (1973)

COMMON INFORMATION IS FAR LESS THAN MUTUAL INFORMATION

P. GÁCS and J KÖRNER

(Budapest)

(Received February 5, 1972)

Theorem ([Gács and Körner, 1973])

$$\lim_{\epsilon \downarrow 0} T_X(\epsilon) = \lim_{\epsilon \downarrow 0} T_Y(\epsilon) = C_{\mathrm{GKW}}(X;Y),$$

where

$$C_{\mathrm{GKW}}(X;Y) := \max_{f,g: f(X)=g(Y)} H(f(X))$$

- $C_{\mathrm{GKW}}(X;Y)$ called Gács–Körner–Witsenhausen's (GKW's) CI
- Abridged version of GKW's system as in [Csiszár and Narayan, 2000]
- Other interesting operational interpretations in [Yu and Tan, 2019a]

Undesirable Properties of GKW's CI

- Fact: Gács–Körner–Witsenhausen's CI $= 0$ for Gaussian sources and doubly symmetric binary sources (DSBSes)
- More unfortunately, we cannot extract even one pair of identical bits from $(\mathbf{X}, \mathbf{Y})$, if $(\mathbf{X}, \mathbf{Y})$ is jointly Gaussian or if $(\mathbf{X}, \mathbf{Y})$ is a DSBS.
- How to measure "common information" for this case?
- Literally, "common information" $\iff$ "correlated bits"

Undesirable Properties of GKW's CI

- Fact: Gács–Körner–Witsenhausen's CI $= 0$ for Gaussian sources and doubly symmetric binary sources (DSBSes)
- More unfortunately, we cannot extract even one pair of identical bits from $(\mathbf{X}, \mathbf{Y})$, if $(\mathbf{X}, \mathbf{Y})$ is jointly Gaussian or if $(\mathbf{X}, \mathbf{Y})$ is a DSBS.
- How to measure "common information" for this case?
- Literally, "common information" $\Longleftrightarrow$ "correlated bits"

- A Variant of CI: What is the maximal possible **correlation of a pair of bits that can be extracted from $\mathbf{X}, \mathbf{Y}$ individually?**

Undesirable Properties of GKW's CI

- Fact: Gács–Körner–Witsenhausen's CI $= 0$ for Gaussian sources and doubly symmetric binary sources (DSBSes)
- More unfortunately, we cannot extract even one pair of identical bits from $(\mathbf{X}, \mathbf{Y})$, if $(\mathbf{X}, \mathbf{Y})$ is jointly Gaussian or if $(\mathbf{X}, \mathbf{Y})$ is a DSBS.
- How to measure "common information" for this case?
- Literally, "common information" $\Longleftrightarrow$ "correlated bits"

- A Variant of CI: What is the maximal possible **correlation of a pair of bits that can be extracted from $\mathbf{X}, \mathbf{Y}$ individually?**
- Coined the binary decision problem [Witsenhausen, 1975], the noninteractive correlation distillation (NICD) problem [Mossel et al., 2006], the noninteractive binary simulation problem [Kamath and Anantharam, 2016]

Outline

Doubly Symmetric Binary Source (DSBS)

- In this section, we only consider the DSBS

$$P_{XY} = \begin{bmatrix} \frac{1+\rho}{4} & \frac{1-\rho}{4} \\ \frac{1-\rho}{4} & \frac{1+\rho}{4} \end{bmatrix}$$

with correlation $\rho \in (0,1)$, and

$$(\mathbf{X}, \mathbf{Y}) \sim P_{XY}^n$$

Doubly Symmetric Binary Source (DSBS)

- In this section, we only consider the DSBS

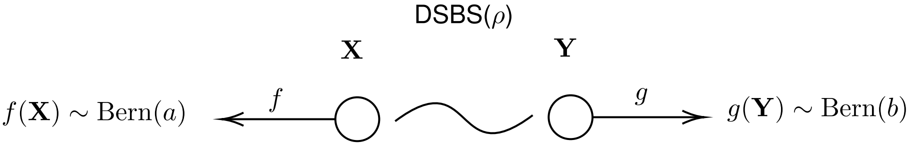
$$P_{XY} = \begin{bmatrix} \frac{1+\rho}{4} & \frac{1-\rho}{4} \\ \frac{1-\rho}{4} & \frac{1+\rho}{4} \end{bmatrix}$$

with correlation $\rho \in (0,1)$, and

$$(\mathbf{X}, \mathbf{Y}) \sim P_{XY}^n$$

- If you are interested in other sources, please refer to [Ahlswede and Gács, 1976, Borell, 1985, Carlen and Cordero-Erausquin, 2009, Mossel and Neeman, 2015, Beigi and Nair, 2016, Yu et al., 2021, Yu, 2021b]...

Non-Interactive Correlation Distillation



$\max \mathbb{P}(f(\mathbf{X}) = g(\mathbf{Y}))$ or equivalently, $\max \mathbb{P}(f(\mathbf{X}) = g(\mathbf{Y}) = 1)$

Non-Interactive Correlation Distillation

- Formally, for $a, b \in [0,1]$, define the Forward Joint Probability as

$$\overline{\Gamma}^{(n)}(a,b) := \max_{\substack{f,g:\{0,1\}^n \to \{0,1\}: \mathbb{P}(f(\mathbf{X})=1) \le a, \\ \mathbb{P}(g(\mathbf{Y})=1) \le b}} \mathbb{P}\left(f(\mathbf{X}) = g(\mathbf{Y}) = 1\right)$$

$$= \max_{\substack{A,B \subseteq \{0,1\}^n : P_X^n(A) \le a, \\ P_Y^n(B) \le b}} P_{XY}^n(A \times B), \qquad (f = 1_A, g = 1_B)$$

Non-Interactive Correlation Distillation

- Formally, for $a, b \in [0,1]$, define the Forward Joint Probability as

$$\overline{\Gamma}^{(n)}(a,b) := \max_{\substack{f,g:\{0,1\}^n \to \{0,1\}:\mathbb{P}(f(\mathbf{X})=1)\le a,\\ \mathbb{P}(g(\mathbf{Y})=1)\le b}} \mathbb{P}\left(f(\mathbf{X}) = g(\mathbf{Y}) = 1\right)$$

$$= \max_{\substack{A,B\subseteq\{0,1\}^n:P_X^n(A)\le a,\\ P_Y^n(B)\le b}} P_{XY}^n(A\times B), \qquad (f = 1_A, g = 1_B)$$

- Define the Reverse Joint Probability as

$$\underline{\Gamma}^{(n)}(a,b) := \min_{\substack{A,B\subseteq\{0,1\}^n:P_X^n(A)\ge a,\\ P_Y^n(B)\ge b}} P_{XY}^n(A\times B)$$

Non-Interactive Correlation Distillation

- Formally, for $a, b \in [0,1]$, define the Forward Joint Probability as

$$\overline{\Gamma}^{(n)}(a,b) := \max_{\substack{f,g:\{0,1\}^n \to \{0,1\}:\mathbb{P}(f(\mathbf{X})=1)\le a,\\ \mathbb{P}(g(\mathbf{Y})=1)\le b}} \mathbb{P}\left(f(\mathbf{X}) = g(\mathbf{Y}) = 1\right)$$

$$= \max_{\substack{A,B\subseteq\{0,1\}^n : P_X^n(A)\le a,\\ P_Y^n(B)\le b}} P_{XY}^n(A\times B), \qquad (f = 1_A, g = 1_B)$$

- Define the Reverse Joint Probability as

$$\underline{\Gamma}^{(n)}(a,b) := \min_{\substack{A,B\subseteq\{0,1\}^n : P_X^n(A)\ge a,\\ P_Y^n(B)\ge b}} P_{XY}^n(A\times B)$$

- For $a = \frac{M}{2^n}, b = \frac{N}{2^n}$ (with integers M, N), the “inequalities” in the constraints can be replaced by “equalities”

Non-Interactive Correlation Distillation

- Formally, for $a, b \in [0,1]$, define the Forward Joint Probability as

$$\overline{\Gamma}^{(n)}(a,b) := \max_{\substack{f,g:\{0,1\}^n \to \{0,1\}: \mathbb{P}(f(\mathbf{X})=1)\le a,\\ \mathbb{P}(g(\mathbf{Y})=1)\le b}} \mathbb{P}(f(\mathbf{X}) = g(\mathbf{Y}) = 1)$$

$$= \max_{\substack{A,B\subseteq\{0,1\}^n: P_X^n(A)\le a,\\ P_Y^n(B)\le b}} P_{XY}^n(A\times B), \qquad (f = 1_A, g = 1_B)$$

- Define the Reverse Joint Probability as

$$\underline{\Gamma}^{(n)}(a,b) := \min_{\substack{A,B\subseteq\{0,1\}^n: P_X^n(A)\ge a,\\ P_Y^n(B)\ge b}} P_{XY}^n(A\times B)$$

- For $a = \frac{M}{2^n}, b = \frac{N}{2^n}$ (with integers M, N), the "inequalities" in the constraints can be replaced by "equalities"
- Equivalence:

$$\overline{\Gamma}^{(\infty)}(1-a,b) = b - \underline{\Gamma}^{(\infty)}(a,b),$$

where $\overline{\Gamma}^{(\infty)}$, $\underline{\Gamma}^{(\infty)}$ denote the pointwise limits of $\overline{\Gamma}^{(n)}$, $\underline{\Gamma}^{(n)}$ as $n \to \infty$.

Asymptotic Regimes and Exponents

Asymptotic cases as $n \to \infty$

Asymptotic Regimes and Exponents

Asymptotic cases as $n \to \infty$

- Central Limit (CL) regime: $a = 2^{-\alpha}, b = 2^{-\beta}$ are fixed

  (Forward and Reverse) CL Exponents: For $\alpha, \beta \in (0, \infty)$,

$$\underline{\Theta}_{\mathrm{CL}}^{(n)}(\alpha, \beta) := -\log \overline{\Gamma}^{(n)}\left(2^{-\alpha}, 2^{-\beta}\right) \qquad \overline{\Theta}_{\mathrm{CL}}^{(n)}(\alpha, \beta) := -\log \underline{\Gamma}^{(n)}\left(2^{-\alpha}, 2^{-\beta}\right)$$

Asymptotic Regimes and Exponents

Asymptotic cases as $n \to \infty$

- Central Limit (CL) regime: $a = 2^{-\alpha}, b = 2^{-\beta}$ are fixed

  (Forward and Reverse) CL Exponents: For $\alpha, \beta \in (0, \infty)$,

$$\underline{\Theta}_{\mathrm{CL}}^{(n)}(\alpha, \beta) := -\log \overline{\Gamma}^{(n)}\left(2^{-\alpha}, 2^{-\beta}\right) \qquad \overline{\Theta}_{\mathrm{CL}}^{(n)}(\alpha, \beta) := -\log \underline{\Gamma}^{(n)}\left(2^{-\alpha}, 2^{-\beta}\right)$$

- Large Deviation (LD) regime: $a = 2^{-n\alpha}, b = 2^{-n\beta}$ are exponentially small

  (Forward and Reverse) LD Exponents: For $\alpha, \beta \in (0, 1)$,

$$\underline{\Theta}_{\mathrm{LD}}^{(n)}(\alpha, \beta) := -\frac{1}{n} \log \overline{\Gamma}^{(n)}\left(2^{-n\alpha}, 2^{-n\beta}\right) \qquad \overline{\Theta}_{\mathrm{LD}}^{(n)}(\alpha, \beta) := -\frac{1}{n} \log \underline{\Gamma}^{(n)}\left(2^{-n\alpha}, 2^{-n\beta}\right)$$

Asymptotic Regimes and Exponents

Asymptotic cases as $n \to \infty$

- Central Limit (CL) regime: $a = 2^{-\alpha}, b = 2^{-\beta}$ are fixed

  (Forward and Reverse) CL Exponents: For $\alpha, \beta \in (0, \infty)$,

$$\underline{\Theta}_{\mathrm{CL}}^{(n)}(\alpha, \beta) := -\log \overline{\Gamma}^{(n)}\left(2^{-\alpha}, 2^{-\beta}\right) \qquad \overline{\Theta}_{\mathrm{CL}}^{(n)}(\alpha, \beta) := -\log \underline{\Gamma}^{(n)}\left(2^{-\alpha}, 2^{-\beta}\right)$$

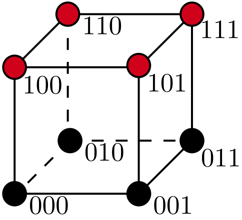
- Large Deviation (LD) regime: $a = 2^{-n\alpha}, b = 2^{-n\beta}$ are exponentially small

  (Forward and Reverse) LD Exponents: For $\alpha, \beta \in (0, 1)$,

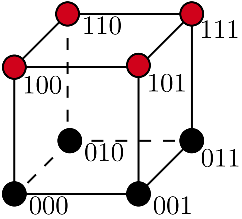
$$\underline{\Theta}_{\mathrm{LD}}^{(n)}(\alpha, \beta) := -\frac{1}{n} \log \overline{\Gamma}^{(n)}\left(2^{-n\alpha}, 2^{-n\beta}\right) \quad \overline{\Theta}_{\mathrm{LD}}^{(n)}(\alpha, \beta) := -\frac{1}{n} \log \underline{\Gamma}^{(n)}\left(2^{-n\alpha}, 2^{-n\beta}\right)$$

- Denote $\underline{\Theta}_{\mathrm{CL}}^{(\infty)}$, $\overline{\Theta}_{\mathrm{CL}}^{(\infty)}$, $\underline{\Theta}_{\mathrm{LD}}^{(\infty)}$, $\overline{\Theta}_{\mathrm{LD}}^{(\infty)}$, as the pointwise limits as $n \to \infty$.

Achievability: Hamming Subcubes

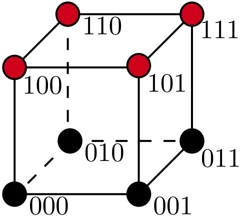


Achievability: Hamming Subcubes



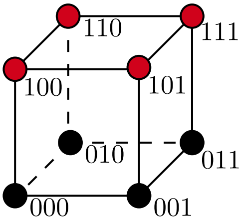
- An $(n-k)$-subcube $\mathcal{C}_{n-k}$ is a set of $\mathbf{x}$ with k components fixed

Achievability: Hamming Subcubes



- An $(n-k)$-subcube $\mathcal{C}_{n-k}$ is a set of $\mathbf{x}$ with k components fixed
  - Special case $\mathcal{C}_{n-1}$: e.g., $\{1\} \times \{0,1\}^{n-1}$ (Indicator $\mathbf{x} \mapsto x_1$ called a dictator function)

Achievability: Hamming Subcubes



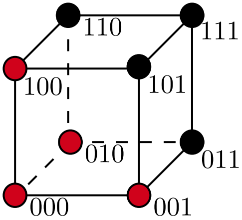
- An $(n-k)$-subcube $\mathcal{C}_{n-k}$ is a set of $\mathbf{x}$ with k components fixed
  - Special case $\mathcal{C}_{n-1}$: e.g., $\{1\} \times \{0,1\}^{n-1}$ (Indicator $\mathbf{x} \mapsto x_1$ called a dictator function)
- Case of $a = b = 2^{-k}$: $A = B = \mathcal{C}_{n-k}$ (identical) $\Longrightarrow$

$$P_{XY}^n(A \times B) = P_{XY}(1,1)^k = \left(\frac{1+\rho}{4}\right)^k$$

$A = \mathbf{1} - B = \mathcal{C}_{n-k}$ (anti-symmetric) $\Longrightarrow$

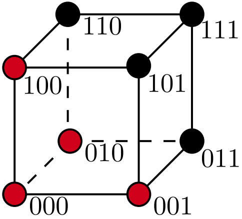
$$P_{XY}^n(A \times B) = P_{XY}(1,0)^k = \left(\frac{1-\rho}{4}\right)^k$$

Achievability: Hamming Balls (CL Regime)



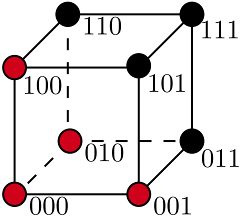
- Hamming Ball: $\mathbb{B}_r(\mathbf{0}) := \{\mathbf{x} : d_{\mathrm{H}}(\mathbf{x}, \mathbf{0}) \le r\} \iff \{\mathbf{x} : \sum_{i=1}^{n} x_i \le r\}$

Achievability: Hamming Balls (CL Regime)



- Hamming Ball: $\mathbb{B}_r(\mathbf{0}) := \{\mathbf{x} : d_{\mathrm{H}}(\mathbf{x}, \mathbf{0}) \le r\} \iff \{\mathbf{x} : \sum_{i=1}^n x_i \le r\}$
- CL regime: Choose $A = \mathbb{B}_{r_n}(\mathbf{0}), B = \mathbb{B}_{s_n}(\mathbf{0})$ with $r_n = \frac{n}{2} + \frac{\lambda\sqrt{n}}{2}, s_n = \frac{n}{2} + \frac{\mu\sqrt{n}}{2}$ where $\lambda, \mu \in \mathbb{R}$

Achievability: Hamming Balls (CL Regime)



- Hamming Ball: $\mathbb{B}_r(\mathbf{0}) := \{\mathbf{x} : d_{\mathrm{H}}(\mathbf{x}, \mathbf{0}) \le r\} \Longleftrightarrow \{\mathbf{x} : \sum_{i=1}^n x_i \le r\}$
- CL regime: Choose $A = \mathbb{B}_{r_n}(\mathbf{0}), B = \mathbb{B}_{s_n}(\mathbf{0})$ with $r_n = \frac{n}{2} + \frac{\lambda\sqrt{n}}{2}, s_n = \frac{n}{2} + \frac{\mu\sqrt{n}}{2}$ where $\lambda, \mu \in \mathbb{R}$
- By the univariate and multivariate CL theorems,

$$P_X^n(A) \to \Phi(\lambda), \qquad P_Y^n(B) \to \Phi(\mu), \qquad P_{XY}^n(A \times B) \to \Phi_\rho(\lambda, \mu)$$

where Φ is the CDF of the standard Gaussian, and $\Phi_\rho(\cdot,\cdot)$ is the CDF of the zero-mean bivariate Gaussian with covariance matrix $\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$.

Achievability: Hamming Balls (CL Regime)

- Achievable CL probabilities:

$$\overline{\Gamma}^{(\infty)}(a,b) \geq \Lambda_{\rho}(a,b) \qquad \text{(by concentric balls)}$$

  - Bivariate normal copula (or Gaussian quadrant probability function):

$$\Lambda_{\rho}(a,b) := \Phi_{\rho}\left(\Phi^{-1}(a), \Phi^{-1}(b)\right)$$

Achievability: Hamming Balls (CL Regime)

- Achievable CL probabilities:

$$\overline{\Gamma}^{(\infty)}(a,b) \ge \Lambda_{\rho}(a,b) \qquad \text{(by concentric balls)}$$

  - Bivariate normal copula (or Gaussian quadrant probability function):

$$\Lambda_{\rho}(a,b) := \Phi_{\rho}\left(\Phi^{-1}(a), \Phi^{-1}(b)\right)$$

- By equivalence of forward and reverse joint probabilities,

$$\underline{\Gamma}^{(\infty)}(a,b) \le \Lambda_{-\rho}(a,b) \qquad \text{(by anti-concentric balls)}$$

  - $\Lambda_{-\rho}(a,b)$ is attained by anti-concentric balls $A = \mathbb{B}_{r_n}(\mathbf{0}), B = \mathbb{B}_{s_n}(\mathbf{1})$

Achievability: Hamming Balls (CL Regime)

- Achievable CL probabilities:

$$\overline{\Gamma}^{(\infty)}(a,b) \geq \Lambda_{\rho}(a,b) \qquad \text{(by concentric balls)}$$

  - Bivariate normal copula (or Gaussian quadrant probability function):

$$\Lambda_{\rho}(a,b) := \Phi_{\rho}\left(\Phi^{-1}(a), \Phi^{-1}(b)\right)$$

- By equivalence of forward and reverse joint probabilities,

$$\underline{\Gamma}^{(\infty)}(a,b) \leq \Lambda_{-\rho}(a,b) \qquad \text{(by anti-concentric balls)}$$

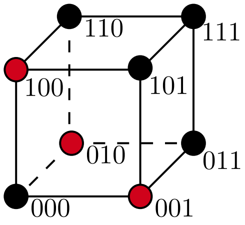
  - $\Lambda_{-\rho}(a,b)$ is attained by anti-concentric balls $A = \mathbb{B}_{r_n}(\mathbf{0}), B = \mathbb{B}_{s_n}(\mathbf{1})$
- Considering exponents,

$$\underline{\Theta}_{\mathrm{CL}}^{(\infty)}(\alpha,\beta) \leq \underline{\Theta}_{\mathrm{CL}}(\alpha,\beta) \qquad \overline{\Theta}_{\mathrm{CL}}^{(\infty)}(\alpha,\beta) \geq \overline{\Theta}_{\mathrm{CL}}(\alpha,\beta).$$

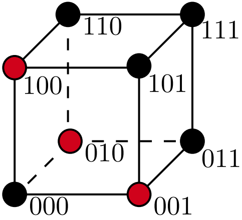
  - Exponents of Λ_{ρ} and $\Lambda_{-\rho}$:

$$\underline{\Theta}_{\mathrm{CL}}(\alpha,\beta) := -\log \Lambda_{\rho}\left(e^{-\alpha}, e^{-\beta}\right), \qquad \overline{\Theta}_{\mathrm{CL}}(\alpha,\beta) := -\log \Lambda_{-\rho}\left(e^{-\alpha}, e^{-\beta}\right)$$

Achievability: Hamming Spheres (LD Regime)

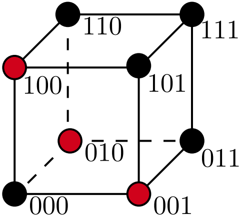


Achievability: Hamming Spheres (LD Regime)



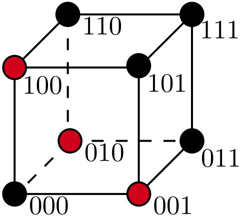
- Hamming Sphere: For $r \in [0:n]$, $\mathbb{S}_r(\mathbf{0}) := \{\mathbf{x} : d_{\mathrm{H}}(\mathbf{x}, \mathbf{0}) = r\} \Longleftrightarrow \{\mathbf{x} : \sum_{i=1}^{n} x_i = r\}$

Achievability: Hamming Spheres (LD Regime)



- Hamming Sphere: For $r \in [0:n]$, $\mathbb{S}_r(\mathbf{0}) := \{\mathbf{x} : d_{\mathrm{H}}(\mathbf{x}, \mathbf{0}) = r\} \iff \{\mathbf{x} : \sum_{i=1}^{n} x_i = r\}$
- It can be regarded as a type class with type $(\lambda, \bar{\lambda})$ in Hamming space, where $\lambda := \frac{r}{n}$ and $\bar{\lambda} := 1 - \lambda$

Achievability: Hamming Spheres (LD Regime)



- Hamming Sphere: For $r \in [0:n]$, $\mathbb{S}_r(\mathbf{0}) := \{\mathbf{x} : d_{\mathrm{H}}(\mathbf{x}, \mathbf{0}) = r\} \iff \{\mathbf{x} : \sum_{i=1}^n x_i = r\}$
- It can be regarded as a type class with type $(\lambda, \bar{\lambda})$ in Hamming space, where $\lambda := \frac{r}{n}$ and $\bar{\lambda} := 1 - \lambda$
- LD regime: Choose $A = \mathbb{S}_{r_n}(\mathbf{0}), B = \mathbb{S}_{s_n}(\mathbf{0})$ with $r_n = \lambda n, s_n = \mu n$ where $\lambda, \mu \in [0, 1]$

Achievability: Hamming Spheres (LD Regime)

By LD theory (or Sanov's theorem),

$$-\frac{1}{n}\log P_X^n(A) \to D\left((\lambda,\bar{\lambda})\,\|P_X\right) = 1 - H_2(\lambda)$$
$$-\frac{1}{n}\log P_Y^n(B) \to D\left((\mu,\bar{\mu})\,\|P_Y\right) = 1 - H_2(\mu)$$
$$-\frac{1}{n}\log P_{XY}^n(A\times B) \to \mathbb{D}\left((\lambda,\bar{\lambda}),(\mu,\bar{\mu})\,\|P_{XY}\right),$$

where the minimum-relative-entropy over couplings of (Q_X, Q_Y) is

$$\mathbb{D}(Q_X, Q_Y\|P_{XY}) := \min_{Q_{XY}\in\mathcal{C}(Q_X,Q_Y)} D(Q_{XY}\|P_{XY})$$

with $\mathcal{C}(Q_X, Q_Y) := \{Q_{XY} \text{ with marginals } Q_X, Q_Y\}$ denoting the coupling set of Q_X and Q_Y.

Achievability: Hamming Spheres (LD Regime)

[Ordentlich et al., 2020] proved...

- Optimizing $\mathbb{D}(Q_X, Q_Y \| P_{XY})$ over feasible $Q_X := \left(\lambda, \bar{\lambda}\right), Q_Y := (\mu, \bar{\mu}) \Longrightarrow$

$$\underline{\Theta}_{\mathrm{LD}}^{(\infty)}(\alpha, \beta) \le \underline{\Theta}_{\mathrm{LD}}(\alpha, \beta) := \min_{\substack{Q_X, Q_Y : D(Q_X \| P_X) \ge \alpha, \\ D(Q_Y \| P_Y) \ge \beta}} \mathbb{D}(Q_X, Q_Y \| P_{XY}),$$

$$\overline{\Theta}_{\mathrm{LD}}^{(\infty)}(\alpha, \beta) \ge \overline{\Theta}_{\mathrm{LD}}(\alpha, \beta) := \max_{\substack{Q_X, Q_Y : D(Q_X \| P_X) \le \alpha, \\ D(Q_Y \| P_Y) \le \beta}} \mathbb{D}(Q_X, Q_Y \| P_{XY}).$$

- Attained by concentric and anti-concentric Hamming spheres or balls

Achievability: Hamming Spheres (LD Regime)

[Ordentlich et al., 2020] proved...

- Optimizing $\mathbb{D}\left(Q_X, Q_Y \| P_{XY}\right)$ over feasible $Q_X := \left(\lambda, \bar{\lambda}\right), Q_Y := (\mu, \bar{\mu}) \Longrightarrow$

$$\underline{\Theta}_{\mathrm{LD}}^{(\infty)}(\alpha, \beta) \le \underline{\Theta}_{\mathrm{LD}}(\alpha, \beta) := \min_{\substack{Q_X, Q_Y : D(Q_X \| P_X) \ge \alpha, \\ D(Q_Y \| P_Y) \ge \beta}} \mathbb{D}\left(Q_X, Q_Y \| P_{XY}\right),$$

$$\overline{\Theta}_{\mathrm{LD}}^{(\infty)}(\alpha, \beta) \ge \overline{\Theta}_{\mathrm{LD}}(\alpha, \beta) := \max_{\substack{Q_X, Q_Y : D(Q_X \| P_X) \le \alpha, \\ D(Q_Y \| P_Y) \le \beta}} \mathbb{D}\left(Q_X, Q_Y \| P_{XY}\right).$$

- Attained by concentric and anti-concentric Hamming spheres or balls

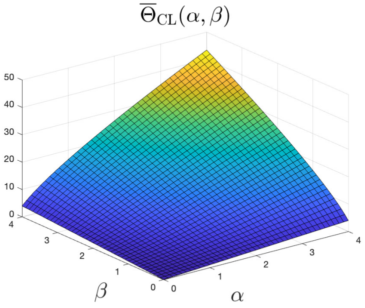
[Ordentlich et al., 2020] conjectured...

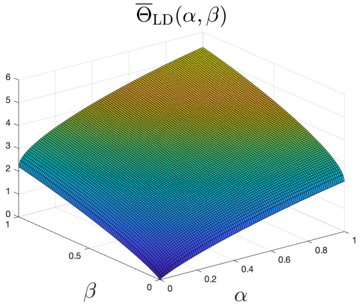
Conjecture (Ordentlich–Polyanskiy–Shayevitz (2020))

For the DSBS and $\alpha, \beta \in (0, 1)$,

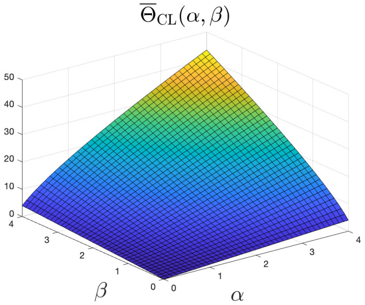
$$\underline{\Theta}_{\mathrm{LD}}^{(\infty)}(\alpha, \beta) \stackrel{?}{=} \underline{\Theta}_{\mathrm{LD}}(\alpha, \beta), \qquad \overline{\Theta}_{\mathrm{LD}}^{(\infty)}(\alpha, \beta) \stackrel{?}{=} \overline{\Theta}_{\mathrm{LD}}(\alpha, \beta).$$

Exponents induced by Hamming Spheres for $\rho = 0.9$



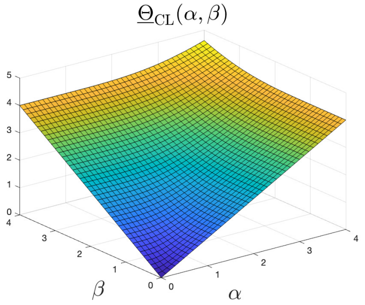


Exponents induced by Hamming Spheres for $\rho = 0.9$

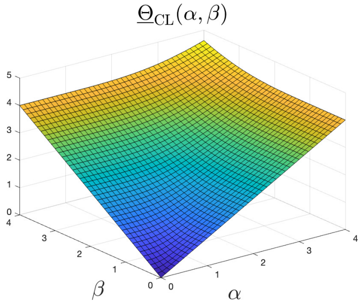


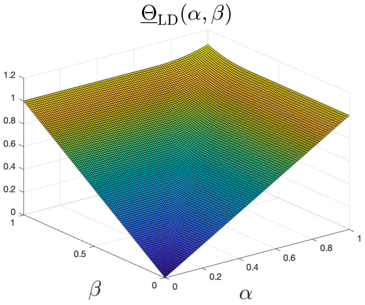
Remark that $\overline{\Theta}_{\mathrm{LD}}$ looks concave! Has implications for OPS' conjecture.

Exponents induced by Hamming Spheres for $\rho = 0.9$



Exponents induced by Hamming Spheres for $\rho = 0.9$





Remark that $\underline{\Theta}_{\mathrm{LD}}$ looks convex! Has implications for OPS' conjecture.

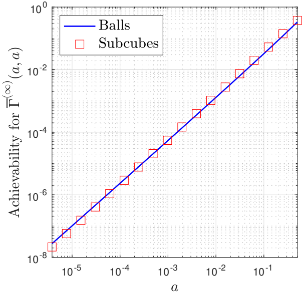
Comparison: Hamming Subcubes vs. Hamming Balls

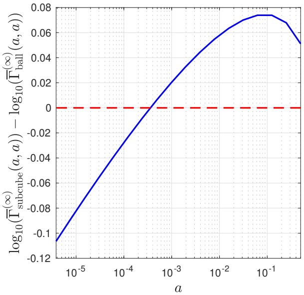
Regime	Central Limit		Large Deviation
a, b	fixed and large a, b	fixed but small a, b	exp. small a, b
Subcubes	Better	Worse	Worse
Balls	Worse	Better	Better

Comparison: Hamming Subcubes vs. Hamming Balls

Regime	Central Limit		Large Deviation
a, b	fixed and large a, b	fixed but small a, b	exp. small a, b
Subcubes	Better	Worse	Worse
Balls	Worse	Better	Better

- For large a, b, subcubes are better; for small a, b, balls are better



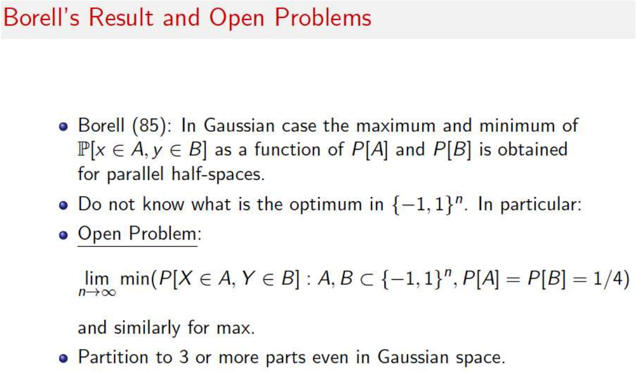


Natural Questions on Optimality I

- Question: Are Hamming subcubes optimal for large a, b (CL regime)?
- Are subcubes optimal for $a = b \in \left\{\frac{1}{2}, \frac{1}{4}\right\}$?
- Mossel's mean-1/4 stability problem

Natural Questions on Optimality I

- Question: Are Hamming subcubes optimal for large a, b (CL regime)?
- Are subcubes optimal for $a = b \in \left\{\frac{1}{2}, \frac{1}{4}\right\}$?
- Mossel's mean-1/4 stability problem



Natural Questions on Optimality II

- Question: Are Hamming balls optimal for exp. small a, b (LD regime)?
- Ordentlich–Polyanskiy–Shayevitz's conjecture
- Excerpt from [Ordentlich et al., 2020]...

Natural Questions on Optimality II

- Question: Are Hamming balls optimal for exp. small a, b (LD regime)?
- Ordentlich–Polyanskiy–Shayevitz's conjecture
- Excerpt from [Ordentlich et al., 2020]...

Our interest is in the greatest and smallest exponential decay rate of $P_{XY}(A \times B)$ among all possible sets A, B of sizes $2^{n\alpha}$ and $2^{n\beta}$, respectively. To that end, for fixed $0 < \alpha, \beta < 1$ we define

$$\overline{E}(\alpha, \beta, \rho) \triangleq -\limsup_{n\to\infty} \max_{\{A\},\{B\}} \frac{1}{n} \log P_{XY}(A \times B), \quad (8)$$

$$\underline{E}(\alpha, \beta, \rho) \triangleq -\liminf_{n\to\infty} \min_{\{A\},\{B\}} \frac{1}{n} \log P_{XY}(A \times B), \quad (9)$$

where $\max_{\{A\},\{B\}}$ and $\min_{\{A\},\{B\}}$ denote optimizations over the sequences of sets $A_n \subset \{0,1\}^n$, $B_n \subset \{0,1\}^n$, $n \in \mathbb{Z}_+$ such that

$$|A_n| = 2^{n\alpha + o(n)}, \qquad |B_n| = 2^{n\beta + o(n)}.$$

Our **main conjecture** is that *both* $\overline{E}(\alpha, \beta, \rho)$ *and* $\underline{E}(\alpha, \beta, \rho)$ *are optimized by concentric (resp., anti-concentric) Hamming balls*. In this work we show partial progress towards establishing this conjecture. Our conjecture is in line with the well-known facts that among all pairs of sets $A, B \subset \{0,1\}^n$ of given sizes, the maximal distance $d_{\max}(A, B) = \max_{a\in A, b\in B} d(a, b)$ is minimized by concentric Hamming (quasi) balls [19], [20], whereas the minimum distance $d_{\min}(A, B) = \min_{a\in A, b\in B} d(a, b)$ is maximized by anti-concentric Hamming (quasi) balls [21].



Converse for $a = b = \frac{1}{2}$: Subcubes/dictators optimal?

Converse for $a = b = \frac{1}{2}$: Subcubes/dictators optimal?

- Confirmed positively by Witsenhausen (1975) using maximal correlation

Converse for $a = b = \frac{1}{2}$: Subcubes/dictators optimal?

- Confirmed positively by Witsenhausen (1975) using maximal correlation
- The (Hirschfeld–Gebelein–Rényi) maximal correlation

$$\rho_{\mathrm{m}}(X;Y) := \sup_{f,g} \rho(f(X);g(Y)),$$

  - $\rho(U;V) := \frac{\mathbb{E}[UV]}{\sqrt{\mathrm{var}[U]\mathrm{var}[V]}}$ is the Pearson correlation coefficient
  - the supremum is taken over all real-valued functions with finite variances

Converse for $a = b = \frac{1}{2}$: Subcubes/dictators optimal?

- Confirmed positively by Witsenhausen (1975) using maximal correlation
- The (Hirschfeld–Gebelein–Rényi) maximal correlation

$$\rho_{\mathrm{m}}(X;Y) := \sup_{f,g} \rho(f(X); g(Y)),$$

  - $\rho(U;V) := \frac{\mathbb{E}[UV]}{\sqrt{\mathrm{var}[U]\mathrm{var}[V]}}$ is the Pearson correlation coefficient
  - the supremum is taken over all real-valued functions with finite variances
- Tensorization: For $(\mathbf{X}, \mathbf{Y}) = \{(X_i, Y_i)\}_{i=1}^n$ i.i.d.,

$$\rho_{\mathrm{m}}(\mathbf{X};\mathbf{Y}) = \rho_{\mathrm{m}}(X;Y).$$

Converse for $a = b = \frac{1}{2}$: Subcubes/dictators optimal?

- Confirmed positively by Witsenhausen (1975) using maximal correlation
- The (Hirschfeld–Gebelein–Rényi) maximal correlation

$$\rho_{\mathrm{m}}(X;Y) := \sup_{f,g} \rho(f(X);g(Y)),$$

  - $\rho(U;V) := \frac{\mathbb{E}[UV]}{\sqrt{\mathrm{var}[U]\mathrm{var}[V]}}$ is the Pearson correlation coefficient
  - the supremum is taken over all real-valued functions with finite variances
- Tensorization: For $(\mathbf{X},\mathbf{Y}) = \{(X_i,Y_i)\}_{i=1}^n$ i.i.d.,

$$\rho_{\mathrm{m}}(\mathbf{X};\mathbf{Y}) = \rho_{\mathrm{m}}(X;Y).$$

- Data Processing Inequality (DPI): For a Markov chain $U - X - Y - V$,

$$\rho_{\mathrm{m}}(U;V) \le \rho_{\mathrm{m}}(X;Y).$$

Converse for $a = b = \frac{1}{2}$: Subcubes/dictators optimal?

- Confirmed positively by Witsenhausen (1975) using maximal correlation
- The (Hirschfeld–Gebelein–Rényi) maximal correlation

$$\rho_{\mathrm{m}}(X;Y) := \sup_{f,g} \rho(f(X);g(Y)),$$

  - $\rho(U;V) := \frac{\mathbb{E}[UV]}{\sqrt{\mathrm{var}[U]\mathrm{var}[V]}}$ is the Pearson correlation coefficient
  - the supremum is taken over all real-valued functions with finite variances
- Tensorization: For $(\mathbf{X}, \mathbf{Y}) = \{(X_i, Y_i)\}_{i=1}^n$ i.i.d.,

$$\rho_{\mathrm{m}}(\mathbf{X};\mathbf{Y}) = \rho_{\mathrm{m}}(X;Y).$$

- Data Processing Inequality (DPI): For a Markov chain $U - X - Y - V$,

$$\rho_{\mathrm{m}}(U;V) \leq \rho_{\mathrm{m}}(X;Y).$$

- For binary X, Y, $\rho_{\mathrm{m}}(X;Y) = |\rho(X;Y)|$.

Converse for $a = b = \frac{1}{2}$: Subcubes/dictators optimal?

Theorem ([Witsenhausen, 1975])

Let $\bar{a} = 1 - a$*. For any* A, B *with* $P_X^n(A) = a, P_Y^n(B) = b$*,*

$$ab - \rho\sqrt{a\bar{a}b\bar{b}} \le P_{XY}^n(A \times B) \le ab + \rho\sqrt{a\bar{a}b\bar{b}}.$$

Converse for $a = b = \frac{1}{2}$: Subcubes/dictators optimal?

Theorem ([Witsenhausen, 1975])

Let $\bar{a} = 1 - a$. For any A, B with $P_X^n(A) = a, P_Y^n(B) = b$,

$$ab - \rho\sqrt{a\bar{a}b\bar{b}} \le P_{XY}^n(A \times B) \le ab + \rho\sqrt{a\bar{a}b\bar{b}}.$$

Proof: Setting $U = 1_A(\mathbf{X}), V = 1_B(\mathbf{Y})$, we have $U - \mathbf{X} - \mathbf{Y} - V$

$$\frac{|P_{XY}^n(A \times B) - ab|}{\sqrt{a\bar{a}}\sqrt{b\bar{b}}} = |\rho(U;V)|$$

.

Converse for $a = b = \frac{1}{2}$: Subcubes/dictators optimal?

Theorem ([Witsenhausen, 1975])

Let $\bar{a} = 1 - a$. For any A, B with $P_X^n(A) = a, P_Y^n(B) = b$,

$$ab - \rho\sqrt{a\bar{a}b\bar{b}} \le P_{XY}^n(A \times B) \le ab + \rho\sqrt{a\bar{a}b\bar{b}}.$$

Proof: Setting $U = 1_A(\mathbf{X}), V = 1_B(\mathbf{Y})$, we have $U - \mathbf{X} - \mathbf{Y} - V$

$$\frac{|P_{XY}^n(A \times B) - ab|}{\sqrt{a\bar{a}}\sqrt{b\bar{b}}} = |\rho(U;V)|$$

$$= \rho_{\mathrm{m}}(U;V) \qquad \text{[Binary]}$$

.

Converse for $a = b = \frac{1}{2}$: Subcubes/dictators optimal?

Theorem ([Witsenhausen, 1975])

Let $\bar{a} = 1 - a$. For any A, B with $P_X^n(A) = a, P_Y^n(B) = b$,

$$ab - \rho\sqrt{a\bar{a}b\bar{b}} \le P_{XY}^n(A \times B) \le ab + \rho\sqrt{a\bar{a}b\bar{b}}.$$

Proof: Setting $U = 1_A(\mathbf{X}), V = 1_B(\mathbf{Y})$, we have $U - \mathbf{X} - \mathbf{Y} - V$

$$\begin{aligned}
\frac{|P_{XY}^n(A \times B) - ab|}{\sqrt{a\bar{a}}\sqrt{b\bar{b}}} &= |\rho(U;V)| \\
&= \rho_{\mathrm{m}}(U;V) && \text{[Binary]} \\
&\le \rho_{\mathrm{m}}(\mathbf{X};\mathbf{Y}) && \text{[DPI]}
\end{aligned}$$

.

Converse for $a = b = \frac{1}{2}$: Subcubes/dictators optimal?

Theorem ([Witsenhausen, 1975])

Let $\bar{a} = 1 - a$. *For any* A, B *with* $P_X^n(A) = a, P_Y^n(B) = b$,

$$ab - \rho\sqrt{a\bar{a}b\bar{b}} \le P_{XY}^n(A \times B) \le ab + \rho\sqrt{a\bar{a}b\bar{b}}.$$

Proof: Setting $U = 1_A(\mathbf{X}), V = 1_B(\mathbf{Y})$, we have $U - \mathbf{X} - \mathbf{Y} - V$

$$\begin{aligned}
\frac{|P_{XY}^n(A \times B) - ab|}{\sqrt{a\bar{a}}\sqrt{b\bar{b}}} &= |\rho(U;V)| \\
&= \rho_{\mathrm{m}}(U;V) && \text{[Binary]} \\
&\le \rho_{\mathrm{m}}(\mathbf{X};\mathbf{Y}) && \text{[DPI]} \\
&= \rho_{\mathrm{m}}(X;Y) && \text{[Tensorization]}
\end{aligned}$$

.

Converse for $a = b = \frac{1}{2}$: Subcubes/dictators optimal?

Theorem ([Witsenhausen, 1975])

Let $\bar{a} = 1 - a$. *For any* A, B *with* $P_X^n(A) = a, P_Y^n(B) = b$,

$$ab - \rho\sqrt{a\bar{a}b\bar{b}} \le P_{XY}^n(A \times B) \le ab + \rho\sqrt{a\bar{a}b\bar{b}}.$$

Proof: Setting $U = 1_A(\mathbf{X}), V = 1_B(\mathbf{Y})$, we have $U - \mathbf{X} - \mathbf{Y} - V$

$$\begin{aligned}
\frac{|P_{XY}^n(A \times B) - ab|}{\sqrt{a\bar{a}}\sqrt{b\bar{b}}} &= |\rho(U;V)| \\
&= \rho_{\mathrm{m}}(U;V) && \text{[Binary]} \\
&\le \rho_{\mathrm{m}}(\mathbf{X};\mathbf{Y}) && \text{[DPI]} \\
&= \rho_{\mathrm{m}}(X;Y) && \text{[Tensorization]} \\
&= \rho.
\end{aligned}$$

Converse for $a = b = \frac{1}{2}$: Subcubes/dictators optimal?

Important Consequence:

- For $a = b = 1/2$,

$$\frac{1-\rho}{4} \le P_{XY}^n(A \times B) \le \frac{1+\rho}{4}.$$

Converse for $a = b = \frac{1}{2}$: Subcubes/dictators optimal?

Important Consequence:

- For $a = b = 1/2$,
$$\frac{1-\rho}{4} \le P_{XY}^n(A \times B) \le \frac{1+\rho}{4}.$$

- Upper bound is attained by
$$f(\mathbf{x}) = g(\mathbf{x}) = x_i$$
and lower bound by
$$f(\mathbf{x}) = -g(\mathbf{x}) = x_i.$$

Converse for $a = b = \frac{1}{2}$: Subcubes/dictators optimal?

Important Consequence:

- For $a = b = 1/2$,

$$\frac{1-\rho}{4} \le P_{XY}^n(A \times B) \le \frac{1+\rho}{4}.$$

- Upper bound is attained by

$$f(\mathbf{x}) = g(\mathbf{x}) = x_i$$

  and lower bound by

$$f(\mathbf{x}) = -g(\mathbf{x}) = x_i.$$

- Dictators (subcubes) are optimal for $a = b = 1/2$, i.e.,

$$\overline{\Gamma}^{(n)}\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{1+\rho}{4} \qquad \underline{\Gamma}^{(n)}\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{1-\rho}{4}$$

Converse for $a = b = \frac{1}{4}$: Are subcubes optimal? —— Mossel's mean-1/4 stability problem

Converse for $a = b = \frac{1}{4}$: Are subcubes optimal? —— Mossel's mean-1/4 stability problem

- Confirmed positively by [Yu and Tan, 2021] using Fourier analysis

Converse for $a = b = \frac{1}{4}$: Are subcubes optimal? —— Mossel's mean-1/4 stability problem

- Confirmed positively by [Yu and Tan, 2021] using Fourier analysis
- Fourier coefficients of $f : \{0,1\}^n \to \{0,1\}$ are

$$\hat{f}(\mathbf{y}) := \frac{1}{2^n} \sum_{\mathbf{x}} f(\mathbf{x}) (-1)^{\langle \mathbf{x}, \mathbf{y} \rangle}$$

Converse for $a = b = \frac{1}{4}$: Are subcubes optimal? —— Mossel's mean-1/4 stability problem

- Confirmed positively by [Yu and Tan, 2021] using Fourier analysis
- Fourier coefficients of $f : \{0,1\}^n \to \{0,1\}$ are

$$\hat{f}(\mathbf{y}) := \frac{1}{2^n} \sum_{\mathbf{x}} f(\mathbf{x}) (-1)^{\langle \mathbf{x}, \mathbf{y} \rangle}$$

- Fourier expansion of f is

$$f(\mathbf{x}) = \sum_{\mathbf{y}} \hat{f}(\mathbf{y}) (-1)^{\langle \mathbf{x}, \mathbf{y} \rangle}$$

Converse for $a = b = \frac{1}{4}$: Are subcubes optimal? —— Mossel's mean-1/4 stability problem

- Confirmed positively by [Yu and Tan, 2021] using Fourier analysis
- Fourier coefficients of $f : \{0,1\}^n \to \{0,1\}$ are

$$\hat{f}(\mathbf{y}) := \frac{1}{2^n} \sum_{\mathbf{x}} f(\mathbf{x}) (-1)^{\langle \mathbf{x}, \mathbf{y} \rangle}$$

- Fourier expansion of f is

$$f(\mathbf{x}) = \sum_{\mathbf{y}} \hat{f}(\mathbf{y}) (-1)^{\langle \mathbf{x}, \mathbf{y} \rangle}$$

- Define the k-degree Fourier weight as

$$\mathbf{W}_k[f] := \sum_{|\mathbf{y}|=k} \hat{f}(\mathbf{y})^2$$

where $|\mathbf{y}|$ denotes the Hamming weight of $\mathbf{y}$.

Converse for $a = b = \frac{1}{4}$: Are subcubes optimal?

- Properties: For a Boolean f with mean a,

$$\mathbf{W}_0[f] = a^2 \qquad \sum_{k=0}^{n} \mathbf{W}_k[f] = a$$

and

$$\mathbb{P}\left(f(\mathbf{X}) = f(\mathbf{Y}) = 1\right) = \sum_{k=0}^{n} \mathbf{W}_k[f] \rho^k.$$

Converse for $a = b = \frac{1}{4}$: Are subcubes optimal?

- Properties: For a Boolean f with mean a,

$$\mathbf{W}_0[f] = a^2 \qquad \sum_{k=0}^{n} \mathbf{W}_k[f] = a$$

and

$$\mathbb{P}\left(f(\mathbf{X}) = f(\mathbf{Y}) = 1\right) = \sum_{k=0}^{n} \mathbf{W}_k[f]\rho^k.$$

- Linear Programming bound on $\mathbf{W}_1[f]$ [Fu et al., 2001, Yu and Tan, 2019b]:

$$\mathbf{W}_1[f] \le \varphi(a) := \begin{cases} 2a\left(\sqrt{a} - a\right) & 0 \le a \le 1/4 \\ a/2 & 1/4 < a \le 1/2 \end{cases}$$

Converse for $a = b = \frac{1}{4}$: Are subcubes optimal?

- Properties: For a Boolean f with mean a,

$$\mathbf{W}_0[f] = a^2 \qquad \sum_{k=0}^{n} \mathbf{W}_k[f] = a$$

and

$$\mathbb{P}\left(f(\mathbf{X}) = f(\mathbf{Y}) = 1\right) = \sum_{k=0}^{n} \mathbf{W}_k[f]\rho^k.$$

- Linear Programming bound on $\mathbf{W}_1[f]$ [Fu et al., 2001, Yu and Tan, 2019b]:

$$\mathbf{W}_1[f] \le \varphi(a) := \begin{cases} 2a\left(\sqrt{a} - a\right) & 0 \le a \le 1/4 \\ a/2 & 1/4 < a \le 1/2 \end{cases}$$

- Fact (Cauchy–Schwarz inequality):

$$\mathbb{P}\left(f(\mathbf{X}) = g(\mathbf{Y}) = 1\right) \le \max\left\{\mathbb{P}\left(f(\mathbf{X}) = f(\mathbf{Y}) = 1\right), \mathbb{P}\left(g(\mathbf{X}) = g(\mathbf{Y}) = 1\right)\right\}$$

Suffices to consider identical Boolean functions for $\overline{\Gamma}^{(n)}(a, a)$.

Converse for $a = b = \frac{1}{4}$: Are subcubes optimal?

Theorem ([Yu and Tan, 2021])

$$\overline{\Gamma}^{(n)}(a,a) \le a^2 + \rho\varphi(a) + \rho^2\left(a - a^2 - \varphi(a)\right).$$

Converse for $a = b = \frac{1}{4}$: Are subcubes optimal?

Theorem ([Yu and Tan, 2021])

$$\overline{\Gamma}^{(n)}(a, a) \leq a^2 + \rho\varphi(a) + \rho^2\left(a - a^2 - \varphi(a)\right).$$

- Consequence: For $a = 1/4$, the upper bound reduces to $\left(\frac{1+\rho}{4}\right)^2 \implies$

$$\overline{\Gamma}^{(n)}\left(\frac{1}{4}, \frac{1}{4}\right) = \left(\frac{1+\rho}{4}\right)^2$$

for $n \geq 2$, attained by $(n-2)$-subcubes!

Converse for $a = b = \frac{1}{4}$: Are subcubes optimal?

Theorem ([Yu and Tan, 2021])

$$\overline{\Gamma}^{(n)}(a,a) \le a^2 + \rho\varphi(a) + \rho^2\left(a - a^2 - \varphi(a)\right).$$

- Consequence: For $a = 1/4$, the upper bound reduces to $\left(\frac{1+\rho}{4}\right)^2 \implies$

$$\overline{\Gamma}^{(n)}\left(\frac{1}{4},\frac{1}{4}\right) = \left(\frac{1+\rho}{4}\right)^2$$

for $n \ge 2$, attained by $(n-2)$-subcubes!

- Resolution of forward part of Mossel's mean-$1/4$ stability problem!

Converse for $a = b = \frac{1}{4}$: Are subcubes optimal?

Theorem ([Yu and Tan, 2021])

$$\overline{\Gamma}^{(n)}(a,a) \leq a^2 + \rho\varphi(a) + \rho^2\left(a - a^2 - \varphi(a)\right).$$

- Consequence: For $a = 1/4$, the upper bound reduces to $\left(\frac{1+\rho}{4}\right)^2 \implies$

$$\overline{\Gamma}^{(n)}\left(\frac{1}{4},\frac{1}{4}\right) = \left(\frac{1+\rho}{4}\right)^2$$

  for $n \geq 2$, attained by $(n-2)$-subcubes!

- Resolution of forward part of Mossel's mean-$1/4$ stability problem!

- However, $\underline{\Gamma}^{(n)}\left(\frac{1}{4},\frac{1}{4}\right)$ is still open!

Converse for LD: Strong Small-Set Expansion Theorem

Theorem (Strong Small-Set Expansion [Yu et al., 2021, Yu, 2021b])

For any $n \geq 1$ and $\alpha, \beta \in (0, 1]$,

$$\underline{\Theta}_{\mathrm{LD}}^{(n)}(\alpha, \beta) \geq \mathbb{L}\left[\underline{\Theta}_{\mathrm{LD}}\right](\alpha, \beta) \quad \text{and}$$
$$\overline{\Theta}_{\mathrm{LD}}^{(n)}(\alpha, \beta) \leq \mathbb{U}\left[\overline{\Theta}_{\mathrm{LD}}\right](\alpha, \beta),$$

where $\mathbb{L}[f]$ and $\mathbb{U}[f]$ respectively denote the lower convex and upper concave envelopes of a function f.

Converse for LD: Strong Small-Set Expansion Theorem

Theorem (Strong Small-Set Expansion [Yu et al., 2021, Yu, 2021b])

For any $n \geq 1$ and $\alpha, \beta \in (0,1]$,

$$\underline{\Theta}_{\mathrm{LD}}^{(n)}(\alpha,\beta) \geq \mathbb{L}\left[\underline{\Theta}_{\mathrm{LD}}\right](\alpha,\beta) \quad \text{and}$$
$$\overline{\Theta}_{\mathrm{LD}}^{(n)}(\alpha,\beta) \leq \mathbb{U}\left[\overline{\Theta}_{\mathrm{LD}}\right](\alpha,\beta),$$

where $\mathbb{L}[f]$ and $\mathbb{U}[f]$ respectively denote the lower convex and upper concave envelopes of a function f.

- Recall: $\underline{\Theta}_{\mathrm{LD}}(\alpha,\beta), \overline{\Theta}_{\mathrm{LD}}(\alpha,\beta)$ are achieved by spheres/balls

Converse for LD: Strong Small-Set Expansion Theorem

Theorem (Strong Small-Set Expansion [Yu et al., 2021, Yu, 2021b])

For any $n \geq 1$ and $\alpha, \beta \in (0,1]$,

$$\underline{\Theta}_{\mathrm{LD}}^{(n)}(\alpha,\beta) \geq \mathbb{L}\left[\underline{\Theta}_{\mathrm{LD}}\right](\alpha,\beta) \quad \text{and}$$
$$\overline{\Theta}_{\mathrm{LD}}^{(n)}(\alpha,\beta) \leq \mathbb{U}\left[\overline{\Theta}_{\mathrm{LD}}\right](\alpha,\beta),$$

where $\mathbb{L}[f]$ and $\mathbb{U}[f]$ respectively denote the lower convex and upper concave envelopes of a function f.

- Recall: $\underline{\Theta}_{\mathrm{LD}}(\alpha,\beta), \overline{\Theta}_{\mathrm{LD}}(\alpha,\beta)$ are achieved by spheres/balls
- Consequence: Time-sharing certain Hamming spheres/balls is optimal in LD regime! —— A weaker version of OPS's conjecture

Converse for LD: Strong Small-Set Expansion Theorem

Converse for LD: Strong Small-Set Expansion Theorem

Lemma ([Yu, 2021a])

$\underline{\Theta}_{\mathrm{LD}}$ *is convex, and* $\overline{\Theta}_{\mathrm{LD}}$ *is concave.*

Converse for LD: Strong Small-Set Expansion Theorem

Lemma ([Yu, 2021a])

$\underline{\Theta}_{\mathrm{LD}}$ is convex, and $\overline{\Theta}_{\mathrm{LD}}$ is concave.

- $\Longrightarrow \mathbb{L}\left[\underline{\Theta}_{\mathrm{LD}}\right] = \underline{\Theta}_{\mathrm{LD}}$ and $\mathbb{U}\left[\overline{\Theta}_{\mathrm{LD}}\right] = \overline{\Theta}_{\mathrm{LD}}$.

Converse for LD: Strong Small-Set Expansion Theorem

Lemma ([Yu, 2021a])

$\underline{\Theta}_{\mathrm{LD}}$ is convex, and $\overline{\Theta}_{\mathrm{LD}}$ is concave.

- $\Longrightarrow \mathbb{L}\left[\underline{\Theta}_{\mathrm{LD}}\right] = \underline{\Theta}_{\mathrm{LD}}$ and $\mathbb{U}\left[\overline{\Theta}_{\mathrm{LD}}\right] = \overline{\Theta}_{\mathrm{LD}}$.
- Substituting these to Strong SSE Theorem $\Longrightarrow$

Converse for LD: Strong Small-Set Expansion Theorem

Lemma ([Yu, 2021a])

$\underline{\Theta}_{\mathrm{LD}}$ *is convex, and* $\overline{\Theta}_{\mathrm{LD}}$ *is concave.*

- $\Longrightarrow \mathbb{L}\left[\underline{\Theta}_{\mathrm{LD}}\right] = \underline{\Theta}_{\mathrm{LD}}$ and $\mathbb{U}\left[\overline{\Theta}_{\mathrm{LD}}\right] = \overline{\Theta}_{\mathrm{LD}}$.
- Substituting these to Strong SSE Theorem $\Longrightarrow$

OPS's conjecture is true:
Balls/spheres are optimal in LD regime!

Converse for LD: Strong Small-Set Expansion Theorem

Lemma ([Yu, 2021a])

$\underline{\Theta}_{\mathrm{LD}}$ is convex, and $\overline{\Theta}_{\mathrm{LD}}$ is concave.

- $\Longrightarrow \mathbb{L}\left[\underline{\Theta}_{\mathrm{LD}}\right] = \underline{\Theta}_{\mathrm{LD}}$ and $\mathbb{U}\left[\overline{\Theta}_{\mathrm{LD}}\right] = \overline{\Theta}_{\mathrm{LD}}$.
- Substituting these to Strong SSE Theorem $\Longrightarrow$

OPS's conjecture is true:
Balls/spheres are optimal in LD regime!

- Note:
  - The limiting cases as $\rho \to 0$ or 1 were previously proven in [Ordentlich et al., 2020].
  - The special case with $\alpha = \beta$ was previously proven in [Kirshner and Samorodnitsky, 2021].

References I

Ahlswede, R. and Gács, P. (1976).
Spreading of sets in product spaces and hypercontraction of the markov operator.
The Annals of Probability, pages 925–939.

Beasley, L. B. and Laffey, T. J. (2009).
Real rank versus nonnegative rank.
Linear Algebra and its Applications, 431(12):2330–2335.

Beigi, S. and Nair, C. (2016).
Equivalent characterization of reverse Brascamp-Lieb-type inequalities using information measures.
In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 1038–1042. IEEE.

Bennett, C. H., Shor, P. W., Smolin, J. A., and Thapliyal, A. V. (2002).
Entanglement-assisted capacity of a quantum channel and the reverse shannon theorem.
IEEE Trans. on Inform. Th., 48(10):2637–2655.

Borell, C. (1985).
Geometric bounds on the Ornstein-Uhlenbeck velocity process.
Probability Theory and Related Fields, 70(1):1–13.

Braun, G., Jain, R., Lee, T., and Pokutta, S. (2017).
Information-theoretic approximations of the nonnegative rank.
Computational Complexity, 26:147–197.

Braun, G. and Pokutta, S. (2013).
Common information and unique disjointness.
In *Proc. of the 54th IEEE Symposium on Foundations of Computer Science*, pages 688–697.

References II

Carlen, E. A. and Cordero-Erausquin, D. (2009).
Subadditivity of the entropy and its relation to Brascamp–Lieb type inequalities.
Geometric and Functional Analysis, 19(2):373–405.

Cichocki, A., Zdunek, R., Phan, A. H., and i. Amari, S. (2009).
Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation.
Wiley.

Cover, T. M. and Thomas, J. A. (2006).
Elements of Information Theory.
Wiley-Interscience, 2nd edition.

Csiszár, I. and Narayan, P. (2000).
Common randomness and secret key generation with a helper.
IEEE Trans. Inf. Theory, 46(2):344–366.

Cuff, P. (2012).
Distributed channel synthesis.
IEEE Trans. on Inform. Th., 59(11):7071–7096.

Fu, F.-W., Wei, V. K., and Yeung, R. W. (2001).
On the minimum average distance of binary codes: Linear programming approach.
Discrete Applied Mathematics, 111(3):263–281.

Gács, P. and Körner, J. (1973).
Common information is far less than mutual information.
Problems of Control and Information Theory, 2(2):149–162.

References III

Gillis, N. (2020).
Nonnegative Matrix Factorization.
Society for Industrial & Applied Mathematics.

Gray, R. M. and Wyner, A. D. (1974).
Source coding for a simple network.
The Bell Systems Technical Journal, 53:1681–1721.

Han, T. S. and Verdú, S. (1993).
Approximation theory of output statistics.
IEEE Trans. on Inform. Th., 39(3):752–772.

Hayashi, M. (2006).
General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel.
IEEE Trans. on Inform. Th., 52(4):1562–1575.

Hayashi, M. (2011).
Exponential decreasing rate of leaked information in universal random privacy amplification.
IEEE Trans. on Inform. Th., 57(6):3989–4001.

Jain, R., Shi, Y., Wei, Z., and Zhang, S. (2013).
Efficient protocols for generating bipartite classical distributions and quantum states.
In *Proc. of SODA*.

Kamath, S. and Anantharam, V. (2016).
On non-interactive simulation of joint distributions.
IEEE Trans. on Inform. Th., 62(6):3419–3435.

References IV

Kirshner, N. and Samorodnitsky, A. (2021).
A moment ratio bound for polynomials and some extremal properties of Krawchouk polynomials and Hamming spheres.
IEEE Trans. Inf. Theory, 67(6):3509–3541.

Kumar, G. R., Li, C. T., and El Gamal, A. (2014).
Exact common information.
In *Proc. IEEE Int. Symp. Inform. Theory*, pages 161–165, Honolulu, Hawaii.

Mossel, E. and Neeman, J. (2015).
Robust optimality of Gaussian noise stability.
Journal of the European Mathematical Society, 17(2):433–482.

Mossel, E., O'Donnell, R., Regev, O., Steif, J. E., and Sudakov, B. (2006).
Non-interactive correlation distillation, inhomogeneous Markov chains, and the reverse Bonami-Beckner inequality.
Israel Journal of Mathematics, 154(1):299–336.

Oohama, Y. (2018).
Exponential strong converse for source coding with side information at the decoder.
Entropy, 20(5):352.

Ordentlich, O., Polyanskiy, Y., and Shayevitz, O. (2020).
A note on the probability of rectangles for correlated binary strings.
IEEE Trans. on Inform. Th., 66(11):7878–7886.

References V

Sason, I. (2016).
On the Rényi divergence, joint range of relative entropies, and a channel coding theorem.
IEEE Trans. on Inform. Th., 62(1):23–34.

Vellambi, B. N. and Kliewer, J. (2016).
Sufficient conditions for the equality of exact and Wyner common information.
In *Proceedings of Allerton Conference on Communication, Control, and Computing*, Monticello, IL.

Witsenhausen, H. S. (1975).
On sequences of pairs of dependent random variables.
SIAM Journal on Applied Mathematics, 28(1):100–113.

Wyner, A. D. (1975).
The common information of two dependent random variables.
IEEE Trans. on Inform. Th., 21(2):163–179.

Yu, L. (2021a).
The convexity and concavity of envelopes of the minimum-relative-entropy region for the DSBS.
arXiv preprint arXiv:2106.03654.

Yu, L. (2021b).
Strong Brascamp–Lieb inequalities.
arXiv preprint arXiv:2102.06935.

Yu, L., Anantharam, V., and Chen, J. (2021).
Graphs of joint types, noninteractive simulation, and stronger hypercontractivity.
arXiv preprint arXiv:2102.00668.

References VI

Yu, L. and Tan, V. Y. F. (2018).
Wyner's common information under Rényi divergence measures.
IEEE Trans. on Inform. Th., 64(5):3616–3623.

Yu, L. and Tan, V. Y. F. (2019a).
Asymptotic coupling and its applications in information theory.
IEEE Trans. on Inform. Th., 65(3):1321–1344.

Yu, L. and Tan, V. Y. F. (2019b).
An improved linear programming bound on the average distance of a binary code.
arXiv preprint arXiv:1910.09416.

Yu, L. and Tan, V. Y. F. (2019c).
Rényi resolvability and its applications to the wiretap channel.
IEEE Trans. on Inform. Th., 65(3):1862–1897.

Yu, L. and Tan, V. Y. F. (2019d).
Simulation of random variables under Rényi divergence measures of all orders.
IEEE Trans. on Inform. Th., 65(6):3349–3383.

Yu, L. and Tan, V. Y. F. (2020a).
Corrections to "Wyner's common information under Rényi divergence measures".
IEEE Trans. on Inform. Th., 66(4):2599–2608.

Yu, L. and Tan, V. Y. F. (2020b).
Exact channel synthesis.
IEEE Trans. on Inform. Th., 66(5):2299–2818.

References VII

Yu, L. and Tan, V. Y. F. (2020c).
On exact and ∞-Rényi common information.
IEEE Trans. on Inform. Th., 66(6):3366–3406.

Yu, L. and Tan, V. Y. F. (2021).
On non-interactive simulation of binary random variables.
IEEE Trans. on Inform. Th., 67(4):2528–2538.