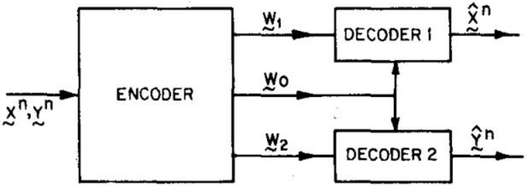
Variations on Common Information

Robert Graczyk

TELECOM ParisTech

January 25, 2023

Wyner's Common Information

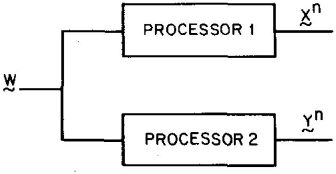


Definition 1 (Wyner 1975)

The common information $\mathrm{C}_1(X;Y)$ between X and Y is the least common rate R_0 under which almost-lossless compression of (X^n, Y^n) is possible subject to the no-excess-rate constraint

$$R_0 + R_1 + R_2 = \mathsf{H}(X, Y)$$

Wyner's Common Information



Definition 2 (Wyner 1975)

The common information $\mathrm{C}_2(X;Y)$ between X and Y is the least common rate R_W under which simulation of (X^n, Y^n) is possible in the sense that

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

Wyner's Common Information

Theorem (Wyner 1975)

$$\mathrm{C}_1(X;Y) = \mathrm{C}_2(X;Y) = \mathrm{C}_{\mathrm{W}}(X;Y)$$

where

$$\mathrm{C}_{\mathrm{W}}(X;Y) \triangleq \min_{\substack{P_{W|XY}: \\ X \to W \to Y}} \mathrm{I}(W;X,Y)$$

Variation 1

Recall the Rényi entropy of $X \sim P_X$

$$\mathsf{H}_\alpha(X) = \frac{1}{1-\alpha} \log\left(\sum_x P_X(x)^\alpha \right), \quad \alpha \in \mathbb{R} \setminus \{1\}$$

Variation 1

Recall the Rényi entropy of $X \sim P_X$

$$\mathsf{H}_\alpha(X) = \frac{1}{1-\alpha} \log\left(\sum_x P_X(x)^\alpha\right), \quad \alpha \in \mathbb{R} \setminus \{1\}$$

The Rényi entropy generalizes Shannon's

$$\mathsf{H}(X) = \lim_{\alpha \to 1} \mathsf{H}_\alpha(X)$$

Variation 1

Recall the Rényi entropy of $X \sim P_X$

$$\mathrm{H}_\alpha(X) = \frac{1}{1-\alpha} \log\left(\sum_x P_X(x)^\alpha\right), \quad \alpha \in \mathbb{R} \setminus \{1\}$$

The Rényi entropy generalizes Shannon's

$$\mathrm{H}(X) = \lim_{\alpha \to 1} \mathrm{H}_\alpha(X)$$

Task: Find a Rényi-type common information measure $\mathrm{C}_\alpha(X;Y)$

$$\mathrm{C}_\mathrm{W}(X;Y) = \lim_{\alpha \to 1} \mathrm{C}_\alpha(X;Y)$$

Variation 2

When $X = (U, V)$ and $Y = (V, W)$ with $U \perp\!\!\!\perp V \perp\!\!\!\perp W$

$$\mathrm{C_W}(X;Y) = \mathsf{H}(V)$$

Variation 2

When $X = (U, V)$ and $Y = (V, W)$ with $U \perp\!\!\!\perp V \perp\!\!\!\perp W$

$$\mathrm{C_W}(X; Y) = \mathrm{H}(V)$$

Assume now that Z is independent of U and W but not of V
The common information between X and Y relevant to Z is

$$\mathrm{C}(X; Y \to Z) = \mathrm{I}(V; Z)$$

Variation 2

When $X = (U, V)$ and $Y = (V, W)$ with $U \perp\!\!\!\perp V \perp\!\!\!\perp W$

$$\mathrm{C_W}(X;Y) = \mathrm{H}(V)$$

Assume now that Z is independent of U and W but not of V
The common information between X and Y relevant to Z is

$$\mathrm{C}(X;Y \to Z) = \mathrm{I}(V;Z)$$

Task: Generalize this notion of relevant common information

An Operational Approach Towards $\mathrm{C}_\alpha(X;Y)$

Yu and Tan suggested to replace the criterion

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

in Wyner's definition of $\mathrm{C}_2(X;Y)$ with

$$\frac{1}{n}\,\mathscr{D}_{\alpha}(P_{X^nY^n}\,\|\,P_{XY}^n) \to 0, \quad \alpha \in \mathbb{R}$$

where

$$\mathscr{D}_\alpha(P\,\|\,Q) = \frac{1}{\alpha-1}\log\left(\sum_x \frac{P(x)^\alpha}{Q(x)^{\alpha-1}}\right), \quad \alpha \in \mathbb{R}\setminus\{1\}$$

An Operational Approach Towards $\mathrm{C}_\alpha(X;Y)$

Theorem (Yu and Tan, 2018)

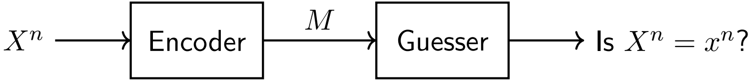
$$\mathrm{C}_\alpha^{\mathsf{YT}}(X;Y) = \begin{cases} 0 & \text{when } \alpha = 0 \\ \mathrm{C}_{\mathrm{W}}(X;Y) & \text{when } \alpha \in (0,1] \end{cases}$$

Yu and Tan also obtained bounds on $\mathrm{C}_\alpha^{\mathsf{YT}}(X;Y)$ for $\alpha \in (1,2]$

A Different Operational Approach Towards $\mathrm{C}_\alpha(X;Y)$

We propose a different approach based on

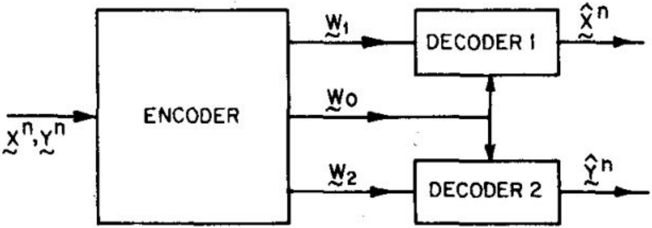
A Simple Observation



The least description rate R of M that is required to drive $\frac{1}{n}\log \mathsf{E}[G(X^n \,|\, M)^\rho]$ to zero is $\mathsf{H}_{\tilde{\rho}}(X)$ where $\tilde{\rho} = 1/(1+\rho)$

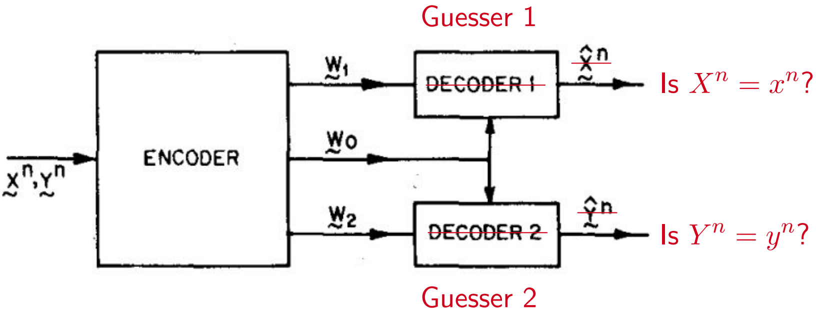
A Different Operational Approach Towards $\mathrm{C}_\alpha(X;Y)$

Instead of ...



A Different Operational Approach Towards $\mathrm{C}_\alpha(X;Y)$

... consider the setup



Guessing on the Gray-Wyner Network

The rate triple (R_0, R_1, R_2) is (E_X, E_Y)-achievable if there exists a joint encoder and guessing strategies for X^n and Y^n that satisfy

$$\lim_{n\to\infty} \frac{1}{n} \log \mathsf{E}[G(X^n \mid M_0, M_1)^\rho] \le E_X$$
$$\lim_{n\to\infty} \frac{1}{n} \log \mathsf{E}[G(Y^n \mid M_0, M_2)^\rho] \le E_Y$$

Guessing on the Gray-Wyner Network

Theorem (Graczyk and Lapidoth, 2021)

The set of all (E_X, E_Y)-achievable rate triples (R_0, R_1, R_2) is given by

$$\bigcap_{Q_{XY}} \bigcup_{Q_{W|XY}} \Big\{ (R_0, R_1, R_2) \colon \\ R_0 \geq \mathrm{I}_Q(W; X, Y) \\ R_1 \geq \mathsf{H}_Q(X \mid W) - \frac{1}{\rho}\big(E_X + \mathscr{D}(Q_{XY} \| P_{XY})\big) \\ R_2 \geq \mathsf{H}_Q(Y \mid W) - \frac{1}{\rho}\big(E_Y + \mathscr{D}(Q_{XY} \| P_{XY})\big) \Big\}$$

Guessing on the Gray-Wyner Network

Theorem (Graczyk and Lapidoth, 2021)

The set of all (E_X, E_Y)-achievable rate triples (R_0, R_1, R_2) is given by

$$\bigcap_{Q_{XY}} \bigcup_{Q_{W|XY}} \Big\{(R_0, R_1, R_2) \colon \; R_0 \geq \mathrm{I}_Q(W; X, Y), \; R_1 \geq \mathrm{H}_Q(X \mid W) - \frac{1}{\rho}\big(E_X + \mathscr{D}(Q_{XY} \| P_{XY})\big), \; R_2 \geq \mathrm{H}_Q(Y \mid W) - \frac{1}{\rho}\big(E_Y + \mathscr{D}(Q_{XY} \| P_{XY})\big)\Big\}$$

Red: Gray-Wyner Source Coding Region

Proof Outline

1. Assume (X^n, Y^n) are equiprobable over a type class $\mathcal{T}^{(n)}(Q_{XY})$ (rather than IID according to P_{XY})

2. Find the (E_X, E_Y)-achievable rate triples under that assumption (direct part via type covering, converse mostly standard)

3. Account for the assumption via the factor $2^{-n\mathscr{D}(Q_{XY} \,\|\, P_{XY})}$

An Operational Approach Towards $\mathrm{C}_{\tilde{\rho}}(X;Y)$

Define the Rényi common information $\mathrm{C}_{\tilde{\rho}}(X;Y)$ of order $\tilde{\rho} = 1/(1+\rho)$ between X and Y as the least common rate R_0 under which

$$\lim_{n\to\infty} \frac{1}{n} \log \mathsf{E}[G(X^n \mid M_0, M_1)^\rho] = 0$$

and

$$\lim_{n\to\infty} \frac{1}{n} \log \mathsf{E}[G(Y^n \mid M_0, M_2)^\rho] = 0$$

with (R_0, R_1, R_2) obeying the no-excess-rate constraint

An Operational Approach Towards $\mathrm{C}_{\tilde{\rho}}(X;Y)$

Define the Rényi common information $\mathrm{C}_{\tilde{\rho}}(X;Y)$ of order $\tilde{\rho} = 1/(1+\rho)$ between X and Y as the least common rate R_0 under which

$$\lim_{n\to\infty} \frac{1}{n} \log \mathsf{E}[G(X^n \mid M_0, M_1)^\rho] = 0$$

and

$$\lim_{n\to\infty} \frac{1}{n} \log \mathsf{E}[G(Y^n \mid M_0, M_2)^\rho] = 0$$

with (R_0, R_1, R_2) obeying the no-excess-rate constraint

We shall focus on the $(0,0)$-achievable rate triples (R_0, R_1, R_2)

Least Sum Rate in the Gray-Wyner Guessing Problem

To incorporate the no-excess-rate constraint
we need the least $(0,0)$-achievable sum rate

Least Sum Rate in the Gray-Wyner Guessing Problem

To incorporate the no-excess-rate constraint
we need the least $(0, 0)$-achievable sum rate

Because

$$(R_0, R_1, R_2) \text{ is achievable} \implies (R_0 + R_1 + R_2,\, 0,\, 0) \text{ is achievable}$$

Least Sum Rate in the Gray-Wyner Guessing Problem

To incorporate the no-excess-rate constraint
we need the least $(0,0)$-achievable sum rate

Because

$$(R_0, R_1, R_2) \text{ is achievable} \implies (R_0 + R_1 + R_2, 0, 0) \text{ is achievable}$$

the least sum rate equals the least R_0 in

$$\bigcap_{Q_{XY}} \bigcup_{Q_{W|XY}} \left\{ R_0 : \begin{aligned} R_0 &\geq \mathrm{I}_Q(W; X, Y) \\ 0 &\geq \mathrm{H}_Q(X \mid W) - \frac{1}{\rho}\big(0 + \mathscr{D}(Q_{XY} \| P_{XY})\big) \\ 0 &\geq \mathrm{H}_Q(Y \mid W) - \frac{1}{\rho}\big(0 + \mathscr{D}(Q_{XY} \| P_{XY})\big) \end{aligned} \right\}$$

Least Sum Rate in the Gray-Wyner Guessing Problem

Lemma (Graczyk and Lapidoth, 2021)

The least $(0,0)$-achievable sum rate R_{Σ}^{*} equals

$$\sup_{Q_{XY}} \inf_{\substack{Q_{W|XY}: \\ \mathrm{H}_Q(X|W) \le \mathscr{D}(Q_{XY} \| P_{XY})/\rho \\ \mathrm{H}_Q(Y|W) \le \mathscr{D}(Q_{XY} \| P_{XY})/\rho}} \mathrm{I}_Q(W; X, Y)$$

Least Sum Rate in the Gray-Wyner Guessing Problem

Lemma (Graczyk and Lapidoth, 2021)

The least $(0,0)$-achievable sum rate R_Σ^* equals

$$\sup_{Q_{XY}} \inf_{\substack{Q_{W|XY}: \; \mathsf{H}_Q(X \mid W) \le \mathscr{D}(Q_{XY} \| P_{XY})/\rho \\ \mathsf{H}_Q(Y \mid W) \le \mathscr{D}(Q_{XY} \| P_{XY})/\rho}} \mathrm{I}_Q(W; X, Y)$$

Two observations:

1. In general $R_\Sigma^* < \mathsf{H}_{\tilde{\rho}}(X, Y)$
2. Because $\operatorname{argmax} Q_{XY} \to P_{XY}$ as $\rho \to 0$

$$\lim_{\rho \to 0} R_\Sigma^* = \mathsf{H}_P(X, Y)$$

An Operational Approach Towards $\mathrm{C}_{\tilde{\rho}}(X;Y)$

Theorem (Graczyk and Lapidoth, 2021)

The Rényi common information $\mathrm{C}_{\tilde{\rho}}(X;Y)$ of order $\tilde{\rho}$ between X and Y equals

$$\sup_{Q_{XY}} \inf_{\substack{Q_{W|XY}: \; (\mathrm{H}_Q(X|W) - \mathscr{D}(Q_{XY} \| P_{XY})/\rho)^+ \\ + (\mathrm{H}_Q(Y|W) - \mathscr{D}(Q_{XY} \| P_{XY})/\rho)^+ \\ + \mathrm{I}_Q(W;X,Y) \leq R_\Sigma^*}} \mathrm{I}_Q(W;X,Y)$$

where $(\cdot)^+ = \max(\cdot, 0)$

Does $\mathrm{C}_{\tilde{\rho}}(X;Y)$ Generalize $\mathrm{C}_{\mathrm{W}}(X;Y)$?

A final observation about $\mathrm{C}_{\tilde{\rho}}(X;Y)$:

Because $\operatorname{argmax} Q_{XY} \to P_{XY}$ and $R^*_\Sigma \to \mathsf{H}_P(X,Y)$ as $\rho \to 0$

$$\begin{aligned}
&(\mathsf{H}_Q(X \mid W) - \mathscr{D}(Q_{XY} \,\|\, P_{XY})/\rho)^+ \\
&+ (\mathsf{H}_Q(Y \mid W) - \mathscr{D}(Q_{XY} \,\|\, P_{XY})/\rho)^+ \\
&+ \mathrm{I}_Q(W;X,Y) \le R^*_\Sigma
\end{aligned}$$

is about equivalent to

$$\mathsf{H}_P(X \mid W) + \mathsf{H}_P(Y \mid W) + \mathrm{I}_P(W;X,Y) \le \mathsf{H}_P(X,Y)$$

for ρ sufficiently small

Does $\mathrm{C}_{\tilde{\rho}}(X;Y)$ Generalize $\mathrm{C}_{\mathrm{W}}(X;Y)$?

A final observation about $\mathrm{C}_{\tilde{\rho}}(X;Y)$:

Because $\operatorname{argmax} Q_{XY} \to P_{XY}$ and $R_\Sigma^* \to \mathsf{H}_P(X,Y)$ as $\rho \to 0$

$$
\begin{aligned}
&(\mathsf{H}_Q(X \mid W) - \mathscr{D}(Q_{XY} \| P_{XY})/\rho)^+ \\
&+ (\mathsf{H}_Q(Y \mid W) - \mathscr{D}(Q_{XY} \| P_{XY})/\rho)^+ \\
&+ \mathrm{I}_Q(W;X,Y) \le R_\Sigma^*
\end{aligned}
$$

is about equivalent to

$$\mathsf{H}_P(X \mid W) + \mathsf{H}_P(Y \mid W) + \mathrm{I}_P(W;X,Y) \le \mathsf{H}_P(X,Y)$$

for ρ sufficiently small

Does $\mathrm{C}_{\tilde{\rho}}(X;Y)$ Generalize $\mathrm{C}_{\mathrm{W}}(X;Y)$?

A final observation about $\mathrm{C}_{\tilde{\rho}}(X;Y)$:

Because $\arg\max Q_{XY} \to P_{XY}$ and $R^*_{\Sigma} \to \mathsf{H}_P(X,Y)$ as $\rho \to 0$

$$\begin{aligned} &(\mathsf{H}_Q(X \mid W) - \mathscr{D}(Q_{XY} \| P_{XY})/\rho)^+ \\ &+ (\mathsf{H}_Q(Y \mid W) - \mathscr{D}(Q_{XY} \| P_{XY})/\rho)^+ \\ &+ \mathrm{I}_Q(W;X,Y) \le R^*_{\Sigma} \end{aligned}$$

is about equivalent to

$$\mathsf{H}_P(X \mid W) + \mathsf{H}_P(Y \mid W) + \mathrm{I}_P(W;X,Y) \le \mathsf{H}_P(X,Y)$$

for ρ sufficiently small

Does $\mathrm{C}_{\tilde{\rho}}(X;Y)$ Generalize $\mathrm{C}_{\mathrm{W}}(X;Y)$?

A final observation about $\mathrm{C}_{\tilde{\rho}}(X;Y)$:

Because $\operatorname{argmax} Q_{XY} \to P_{XY}$ and $R_{\Sigma}^{*} \to \mathsf{H}_P(X,Y)$ as $\rho \to 0$

$$\begin{aligned} &(\mathsf{H}_Q(X \mid W) - \mathscr{D}(Q_{XY} \,\|\, P_{XY})/\rho)^{+} \\ &+ (\mathsf{H}_Q(Y \mid W) - \mathscr{D}(Q_{XY} \,\|\, P_{XY})/\rho)^{+} \\ &+ \mathrm{I}_Q(W;X,Y) \le R_{\Sigma}^{*} \end{aligned}$$

is about equivalent to

$$\mathsf{H}_P(X \mid W) + \mathsf{H}_P(Y \mid W) + \mathrm{I}_P(W;X,Y) \le \mathsf{H}_P(X,Y)$$

for ρ sufficiently small

Does $\mathrm{C}_{\tilde{\rho}}(X;Y)$ Generalize $\mathrm{C}_{\mathrm{W}}(X;Y)$?

A final observation about $\mathrm{C}_{\tilde{\rho}}(X;Y)$:

Because $\operatorname{argmax} Q_{XY} \to P_{XY}$ and $R^*_\Sigma \to \mathsf{H}_P(X,Y)$ as $\rho \to 0$

$$\begin{aligned} &(\mathsf{H}_Q(X \mid W) - \mathscr{D}(Q_{XY} \| P_{XY})/\rho)^+ \\ &+ (\mathsf{H}_Q(Y \mid W) - \mathscr{D}(Q_{XY} \| P_{XY})/\rho)^+ \\ &+ \mathrm{I}_Q(W;X,Y) \leq R^*_\Sigma \end{aligned}$$

is about equivalent to

$$\mathsf{H}_P(X \mid W) + \mathsf{H}_P(Y \mid W) + \mathrm{I}_P(W;X,Y) \leq \mathsf{H}_P(X,Y)$$

for ρ sufficiently small

Does $\mathrm{C}_{\tilde{\rho}}(X;Y)$ **Generalize** $\mathrm{C}_{\mathrm{W}}(X;Y)$**?**

Because

$$\begin{aligned} &\mathsf{H}_P(X \mid W) + \mathsf{H}_P(Y \mid W) + \mathrm{I}_P(W;X,Y) \leq \mathsf{H}_P(W;X,Y) \\ &\iff \mathsf{H}_P(X \mid W) + \mathsf{H}_P(Y \mid W) \leq \mathsf{H}_P(X,Y \mid W) \\ &\iff X \to W \to Y \text{ under } P \end{aligned}$$

we indeed have

$$\lim_{\rho \to 0} \mathrm{C}_{\tilde{\rho}}(X;Y) = \min_{\substack{P_{W|XY}: \\ X \to W \to Y}} \mathrm{I}_P(W;X,Y) = \mathrm{C}_{\mathrm{W}}(X;Y)$$

Three Follow-Up Questions

1. Can $\mathrm{C}_{\tilde{\rho}}(X;Y)$ be expressed in terms of $\mathrm{H}_{\tilde{\rho}}(\cdot)$, $\mathscr{D}_{\tilde{\rho}}(\cdot\|\cdot)$, and $\mathrm{I}_{\tilde{\rho}}(\cdot \mid \cdot)$?

2. Can we find a suitable operational definition for $\mathrm{C}_{\alpha}(X;Y)$ when $\alpha \notin [0,1]$?

3. How to systematically agree on a Rényi extension of $\mathrm{C}_{\mathrm{W}}(X;Y)$?
   Proposition: via axiomization of $\mathrm{C}_{\mathrm{W}}(X;Y)$

Onto Relevant Common Information

Recall the example: $X = (U, V)$, $Y = (V, W)$ with $U \perp\!\!\!\perp V \perp\!\!\!\perp W$

$$\mathrm{C_W}(X;Y) = \mathrm{H}(V) \text{ and } \mathrm{C}(X;Y \to Z) = \mathrm{I}(V;Z)$$

Onto Relevant Common Information

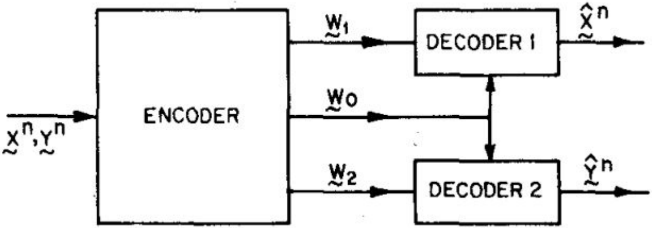
Recall the example: $X = (U, V)$, $Y = (V, W)$ with $U \perp\!\!\!\perp V \perp\!\!\!\perp W$

$$\mathrm{C_W}(X; Y) = \mathrm{H}(V) \text{ and } \mathrm{C}(X; Y \to Z) = \mathrm{I}(V; Z)$$

Task: Generalize this notion of relevant common information
following a suitable operational approach

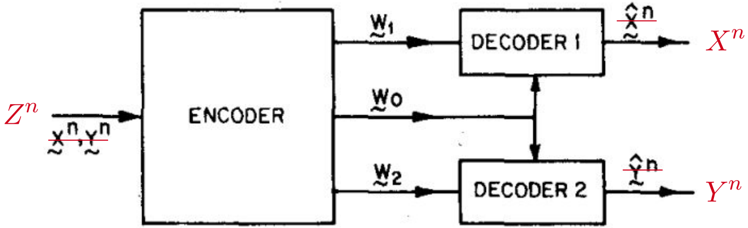
An Operational Approach Towards $\mathrm{C}(X; Y \to Z)$

Instead of ...



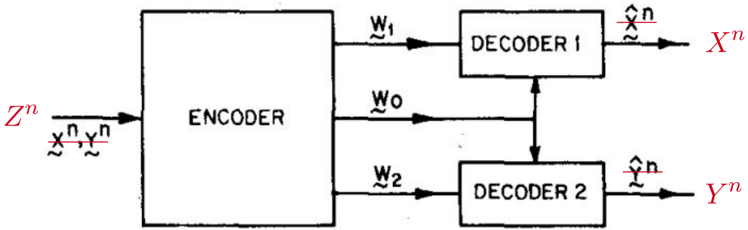
An Operational Approach Towards $\mathrm{C}(X;Y \to Z)$

... consider the setup



An Operational Approach Towards $\mathrm{C}(X;Y \to Z)$

. . . consider the setup



For a given PMF P_{XYZ} we require that Decoder 1 and 2 produce X^n and Y^n that, together with $Z^n \sim$ IID P_Z, coordinate P_{XYZ} in the weak sense that

$$\mathscr{D}(\Pi_{X^nY^nZ^n} \| P_{XYZ}) \to 0$$

where $\Pi_{X^nY^nZ^n}$ denotes the empirical distribution of (X^n, Y^n, Z^n)

An Operational Approach Towards $\mathrm{C}(X;Y \to Z)$

Define the common information $\mathrm{C}(X;Y \to Z)$ between X and Y relevant to Z as the least common rate R_0 that allows for weak coordination of P_{XYZ} subject to the no-excess-rate constraint

$$R_0 + R_1 + R_2 = \mathrm{I}(X,Y;Z)$$

Weak Coordination on the Gray-Wyner Network

Theorem (Graczyk, Lapidoth, and Wigger, 2022)

The set of all rate triples (R_0, R_1, R_2) that allow for weak coordination of P_{XYZ} subject to the no-excess-rate constraint $R_0 + R_1 + R_2 = \mathrm{I}(X, Y; Z)$ is given by

$$\bigcup_{\substack{P_{W|XYZ}: \\ X \to W \to Y \\ W \to (X,Y) \to Z}} \left\{ (R_0, R_1, R_2) \colon \begin{array}{r} R_0 \geq \mathrm{I}(Z; W) \\ R_0 + R_1 \geq \mathrm{I}(Z; X, W) \\ R_0 + R_2 \geq \mathrm{I}(Z; Y, W) \end{array} \right\}$$

An Operational Approach Towards $\mathrm{C}(X;Y \to Z)$

Corollary (Graczyk, Lapidoth, and Wigger, 2022)

$$\mathrm{C}(X;Y \to Z) = \min_{\substack{P_{W|XYZ}: \\ X \to W \to Y \\ W \to (X,Y) \to Z}} \mathrm{I}(Z;W)$$

Proof

1. By the theorem $R_0 \geq \mathrm{C}(X;Y \to Z)$

Proof

1. By the theorem $R_0 \geq \mathrm{C}(X;Y \to Z)$

2. Fix some $P_{W|XYZ}$ that achieves the minimum in $\mathrm{C}(X;Y \to Z)$ and let

$$R_0 = \mathrm{I}(Z;W)$$
$$R_1 = \mathrm{I}(Z;X,W) - \mathrm{I}(Z;W)$$

Proof

1. By the theorem $R_0 \geq \mathrm{C}(X;Y \to Z)$

2. Fix some $P_{W|XYZ}$ that achieves the minimum in $\mathrm{C}(X;Y \to Z)$ and let

$$R_0 = \mathrm{I}(Z;W)$$
$$R_1 = \mathrm{I}(Z;X,W) - \mathrm{I}(Z;W)$$

3. Show that we can set

$$R_2 = \mathrm{I}(Z;Y,W) - \mathrm{I}(Z;W) + \Delta$$

for some $\Delta \geq 0$ so that $R_0 + R_1 + R_2 = \mathrm{I}(X,Y;Z)$

Proof

This is equivalent to showing that

$$\mathrm{I}(Z;X,W)+\mathrm{I}(Z;Y,W)-\mathrm{I}(Z;W)\leq \mathrm{I}(X,Y;Z)$$

Proof

This is equivalent to showing that

$$\mathrm{I}(Z;X,W)+\mathrm{I}(Z;Y,W)-\mathrm{I}(Z;W)\leq \mathrm{I}(X,Y;Z)$$

Indeed, because $X\to W\to Y$ and $W\to (X,Y)\to Z$,

$$\begin{aligned}
&\mathrm{I}(Z;X,W)+\mathrm{I}(Z;Y,W)-\mathrm{I}(Z;W)\\
&\quad=\mathrm{I}(Z;X\mid W)+\mathrm{I}(Z;Y,W)\\
&\quad=\mathsf{H}(X\mid W)-\mathsf{H}(X\mid W,Z)+\mathrm{I}(Z;Y,W)
\end{aligned}$$

Proof

This is equivalent to showing that

$$\mathrm{I}(Z;X,W)+\mathrm{I}(Z;Y,W)-\mathrm{I}(Z;W)\le\mathrm{I}(X,Y;Z)$$

Indeed, because $X\to W\to Y$ and $W\to(X,Y)\to Z$,

$$\begin{aligned}&\mathrm{I}(Z;X,W)+\mathrm{I}(Z;Y,W)-\mathrm{I}(Z;W)\\&\quad=\mathrm{I}(Z;X\mid W)+\mathrm{I}(Z;Y,W)\\&\quad=\mathsf{H}(X\mid W)-\mathsf{H}(X\mid W,Z)+\mathrm{I}(Z;Y,W)\\&\quad\le\mathsf{H}(X\mid W)-\mathsf{H}(X\mid W,Y,Z)+\mathrm{I}(Z;Y,W)\end{aligned}$$

Proof

This is equivalent to showing that

$$\mathrm{I}(Z;X,W)+\mathrm{I}(Z;Y,W)-\mathrm{I}(Z;W)\le \mathrm{I}(X,Y;Z)$$

Indeed, because $X\to W\to Y$ and $W\to(X,Y)\to Z$,

$$\begin{aligned}
&\mathrm{I}(Z;X,W)+\mathrm{I}(Z;Y,W)-\mathrm{I}(Z;W)\\
&\quad=\mathrm{I}(Z;X\mid W)+\mathrm{I}(Z;Y,W)\\
&\quad=\mathsf{H}(X\mid W)-\mathsf{H}(X\mid W,Z)+\mathrm{I}(Z;Y,W)\\
&\quad\le \mathsf{H}(X\mid W)-\mathsf{H}(X\mid W,Y,Z)+\mathrm{I}(Z;Y,W)\\
&\quad=\mathsf{H}(X\mid W,Y)-\mathsf{H}(X\mid W,Y,Z)+\mathrm{I}(Z;Y,W)
\end{aligned}$$

Proof

This is equivalent to showing that

$$\mathrm{I}(Z;X,W) + \mathrm{I}(Z;Y,W) - \mathrm{I}(Z;W) \le \mathrm{I}(X,Y;Z)$$

Indeed, because $X \to W \to Y$ and $W \to (X,Y) \to Z$,

$$\begin{aligned}
&\mathrm{I}(Z;X,W) + \mathrm{I}(Z;Y,W) - \mathrm{I}(Z;W) \\
&\quad = \mathrm{I}(Z;X \mid W) + \mathrm{I}(Z;Y,W) \\
&\quad = \mathsf{H}(X \mid W) - \mathsf{H}(X \mid W,Z) + \mathrm{I}(Z;Y,W) \\
&\quad \le \mathsf{H}(X \mid W) - \mathsf{H}(X \mid W,Y,Z) + \mathrm{I}(Z;Y,W) \\
&\quad = \mathsf{H}(X \mid W,Y) - \mathsf{H}(X \mid W,Y,Z) + \mathrm{I}(Z;Y,W) \\
&\quad = \mathrm{I}(Z;X \mid W,Y) + \mathrm{I}(Z;Y,W) \\
&\quad = \mathrm{I}(Z;X,Y,W)
\end{aligned}$$

Proof

This is equivalent to showing that

$$\mathrm{I}(Z;X,W) + \mathrm{I}(Z;Y,W) - \mathrm{I}(Z;W) \le \mathrm{I}(X,Y;Z)$$

Indeed, because $X \to W \to Y$ and $W \to (X,Y) \to Z$,

$$\begin{aligned}
&\mathrm{I}(Z;X,W) + \mathrm{I}(Z;Y,W) - \mathrm{I}(Z;W) \\
&\quad = \mathrm{I}(Z;X \mid W) + \mathrm{I}(Z;Y,W) \\
&\quad = \mathrm{H}(X \mid W) - \mathrm{H}(X \mid W,Z) + \mathrm{I}(Z;Y,W) \\
&\quad \le \mathrm{H}(X \mid W) - \mathrm{H}(X \mid W,Y,Z) + \mathrm{I}(Z;Y,W) \\
&\quad = \mathrm{H}(X \mid W,Y) - \mathrm{H}(X \mid W,Y,Z) + \mathrm{I}(Z;Y,W) \\
&\quad = \mathrm{I}(Z;X \mid W,Y) + \mathrm{I}(Z;Y,W) \\
&\quad = \mathrm{I}(Z;X,Y,W) \\
&\quad = \mathrm{I}(Z;X,Y)
\end{aligned}$$

Two Follow-Up Questions

1. What result do we obtain in the strong coordination problem? (instead of $\Pi_{X^nY^nZ^n} \to P_{XYZ}$ require that $P_{X^nY^nZ^n} \to P_{XYZ}^n$)

2. How to systematically agree on a definition for relevant common information?

Thank you for your attention!